Agricultural Optimization System Using Machine Learning

Vitthal Vikash
2018-B-29012000A
B.Tech Computer Science

Raushan Kumar
2018-B-20052001A
B.Tech Computer Science

KaushaL Kumar
2018-B-20052001
B.Tech Computer Science

AJEENKYA
D Y PATIL UNIVERSITY
THE INNOVATION UNIVERSITY

School of Engineering

Agricultural Optimization System Using Machine Learning

Thesis submitted in partial fulfillment

of the requirements of the degree of

Bachelor of Technology

in

Computer Engineering

by

Vitthal Vikash

2018-B-29012000A

Raushan Kumar

2018-B-20052001A

Kaushal Kumar

2018-B-20052001

Under the Supervision of

Dr. Raj Gaurav Mishra

May 2022

School of Engineering

Ajeenkya D Y Patil University, Pune

AJEENKYA D Y PATIL UNIVERSITY
THE INNOVATION UNIVERSITY
School of Engineering

May , 2022

CERTIFICATE

This is to certify that the dissertation entitled **"Agriculture Optimization System Using Machine Learning"** is a bonafide work of "**Vitthal Vikash" (2018-B-29012000A),** "**Kaushal Kumar" (2018-B-20052001),** and "**Raushan Kumar" (2018-B-20052001A)** submitted to the School of Engineering, Ajeenkya D Y Patil University, Pune in partial fulfillment of the requirement for the award of the degree of **"Bachelor of Technology in Computer Engineering"**.

Dr. Raj Gaurav Mishra
Supervisor

Internal-Examiner/s

External-Examiner

Dr. Biswajeet Champaty
Head-School of Engineering

Dr. Raj Gaurav Mishra
Associate Professor

May , 2022

Supervisor's Certificate

This is to certify that the dissertation entitled **"Agriculture Optimization System Using Machine Learning"** submitted by **Vitthal Vikash, 2018-B-29012000A, Kaushal Kumar, 2018-B-20052001** and **Raushan Kumar 2018-B-20052001A** is a record of original work carried out by him/her under my supervision and guidance in partial fulfillment of the requirements of the degree of **Bachelor of Technology in Computer Engineering** at **School of Engineering**, **Ajeenkya DY Patil University, Pune, Maharashtra-412105**. Neither this dissertation nor any part of it has been submitted earlier for any degree or diploma to any institute or university in India or abroad.

Dr. Raj Gaurav Mishra
Supervisor

AJEENKYA D Y PATIL UNIVERSITY
THE INNOVATION UNIVERSITY
School of Engineering

Declaration of Originality

We, ***Vitthal Vikash***, URN ***2018-B-29012000A***, ***Raushan Kumar***, URN ***2018-B-20052001A*** *and* ***Kaushal Kumar***, URN ***2018-B-20052001*** hereby declare that this dissertation entitled "***Agriculture Optimization System Using Machine Learning"*** presents my original work carried out as a bachelor student of School of Engineering, Ajeenkya D Y Patil University, Pune, Maharashtra. To the best of my knowledge, this dissertation contains no material previously published or written by another person, nor any material presented by me for the award of any degree or diploma of Ajeenkya D Y Patil University, Pune or any other institution. Any contribution made to this research by others, with whom I have worked at Ajeenkya D Y Patil University, Pune or elsewhere, is explicitly acknowledged in the dissertation. Works of other authors cited in this dissertation have been duly acknowledged under the sections "Reference" or "Bibliography". I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission.

I am fully aware that in case of any non-compliance detected in future, the Academic Council of Ajeenkya D Y Patil University, Pune may withdraw the degree awarded to me on the basis of the present dissertation.

Date: -05-2022

Place: Lohegaon, Pune

Vitthal Vikash

Raushan Kumar

Kaushal Kumar

Acknowledgement

We remain immensely obliged to **Dr. Raj Gaurav Mishra**, for providing me with the idea of this topic, and for his invaluable support in garnering resources for me either by way of information or computers also his guidance and supervision which made this Internship/Project happen.

We would like to say that it has indeed been a fulfilling experience for working out this Internship/Project.

Vitthal Vikash **Raushan Kumar** **Kaushal Kumar**

Abstract

Agriculture is the spine of a developing economic system such as India, and there's a big want to preserve agriculture sustainability. therefore, it is a huge contribution to the monetary and agricultural properly-being of states across the world. efficient use of agricultural land is critical to ensure meal protection and safety in an extremely poor u. s . a. This record is meant to advocate an ML-based totally agricultural system, to be able to help farmers or farmers in crop prediction supported with the aid of metrological agriculture principle through obtaining actual-time meteorological statistics from the crop subject to the usage of ML for prediction., Which it will allow smart farming and increase your ordinary yield. and consequently the great of the products.

food is visible as a primary human requirement that may be met via agriculture. Agriculture is not only a supply of meals for humans, but it is also a source of jobs across the world. Agriculture appeared as the spine of the Indian financial system and a supply of jobs. India's agricultural sector money owed for 15.4% of its GDP. Pre-harvesting, harvesting, and submit-harvesting are the 3 predominant categories of agricultural activity. Agriculture has benefited from advances in machine gaining knowledge. gadget studying is a current generation that enables farmers to reduce farming losses by way of offering detailed crop advice and insights. this article offers an in-intensity examination of the most recent system gaining knowledge of applications in agriculture, with the purpose of resolving demanding situations within the domains of pre-harvesting, harvesting, and submit-harvesting. machine gaining knowledge of in agriculture permits extra green and accurate farming with lower human hard work even as maintaining output. For extra powerful output, electronic agriculture integrates era and numerical procedures into agricultural activities.

This studies developed a mobile device studying device for agricultural land optimization making use of a spread of inputs which include area, crop type, and soil type. the pH of the soil, in addition to the distance among plant life when the device became carried out on the farmlands, it turned into additionally located that the users' information changed into optimized.

Index Terms - Machine learning and application, Data Science, Agriculture

Contents

List of Figures

List of Tables

CHAPTER 1

INTRODUCTION

Agriculture and farming are two of the world's most ancient and important vocations. Within the economic sector, it is critical. Agriculture is estimated to be worth $5 trillion worldwide. By 2050, the world's population is expected to reach over nine billion people, necessitating a 70 percent increase in agricultural production to fulfil demand. Because the world's population is growing, land, water, and resources are running out, making the demand-supply cycle unsustainable. As a result, we'd like to take a more strategic approach to farming and become more efficient. demanding situations faced by farmers through the use of conventional methods of farming and the manner computing is making a revolution in agriculture by means of replacing traditional methods via the usage of extra green techniques and helping the planet to turn out to be a better place. Al and ML in agriculture no longer only help farmers to automate their farming however additionally shifts to explicit cultivation for better crop yield and higher quality whilst the use of fewer resources. agencies worried in improving device mastering or artificial Intelligence-primarily based products or services like schooling facts for agriculture, drone, and automated gadget making gets technological improvements in the future and could offer greater useful applications to the cutting-edge area helping the planet cater to food manufacturing problems for the developing population.

As we all know that agriculture depends largely on the character of soil and therefore the climate and lots of a times, we face unpredictable changes in climate like, non-seasonal rainfall or heat waves or fluctuations in humidity levels, etc. and every one such events cause an excellent loss to our farmers and farming, due to which they're unable to utilize their agricultural land to it's fullest.So to unravel all such problems, i've got build a Machine Learning Model by the virtue of which we are able to help farmers, optimize the agricultural production, because this predictive model will help them understand that for a selected soil & given climatic condition, which crop are going to be best suitable for the harvest.According to the adage "Information is

Power,"retaining music of statistics on vegetation, the surroundings, and the market can assist farmers to make higher selections and clear up challenges related to agriculture. Blockchain, IoT, gadget studying, deep studying, cloud computing, and side computing are examples of technology that may be used to gather and manner information.computer vision, machine mastering, and IoT applications will help to enhance assembly, enhance standards, and ultimately grow the profitability of farmers and related regions. Precision learning is vital within the zone of agriculture to boom harvest output.

The latest growing tendencies in the laptop discipline are blockchain generation, cloud computing, the internet of factors (IoT), machine gaining knowledge of (ML), and deep getting to know (DL). It has formerly been used by academics to resolve complex problems in a spread of fields inclusive of healthcare, cybercrime, biochemistry, robotics, metrology, banking, medicinal drug, and meals. device gaining knowledge of and IoT packages in a selection of fields are discussed. gadget getting to know is turning into more effective and correct due to deep getting to know algorithms. the usage of automated machine mastering (AutoML), you can reduce the requirement for gadget gaining knowledge of know-how even as additionally automating the ML pipeline with greater precision.

We are using Multiple algorithms to coach our model and finding the simplest accuracy from all of the used machine learning algorithms to administer the most effective result like we are using five machine learning algorithms and therefore the first one is Support Vector Machine Algorithm. Both linear and non-linear data can be classified using the Support Vector Machine (SVM) technique. Each data item is initially mapped onto an n-dimensional feature space, where n is the number of features. The hyperplane that splits the information items into two groups is then identified, with the marginal distance for both classes maximised and classification errors minimised. The second Algorithm we are using in KNN (K Nearest Neighbor Algorithm). Both linear and non-linear data can be classified using the Support Vector Machine (SVM) technique. Each data item is initially mapped onto an n-dimensional feature space, where n is the number of features. The hyperplane that splits the information items into two groups is then identified, with the marginal distance for both classes maximised and classification errors minimised. The third algorithm we are using is Decision tree Classification Algorithm.One of the first and most well-known machine learning algorithms is the decision tree (DT). A call tree

represents the choice logic, or tests and results, for classifying data objects into a tree-like structure. The principal or top-most node of a DT tree is referred to as the base node, and the nodes of a DT tree usually have numerous layers. All internal nodes (those with at least one child) are input variable or attribute tests. The fourth algorithm we are using is the Random Forest Algorithm. A random forest (RF) is a type of ensemble classifier made up of multiple DTs, similar to how a forest is made up of many trees. Overfitting of the training data is common with deep DTs, causing a large variance in classification results for a small change in the computer file, We're also using Logistic Regression as the final algorithm. This type of statistical analysis (also known as a logit model) is extensively utilized in predictive analytics and modelling, as well as machine learning applications. The variable is either finite or categorical in this analytics approach: A or B (binary regression) or a variety of finite possibilities A, B, C, or D (multiple regression) (multinomial regression).

Things which are available in our data set for the prediction. We analyze a data set and there is a total (2200 rows and 8 column).Where 8 parameters that contain Nitrogen, Phosphorus, Potassium, Temperature, Humidity, Soil-ph value, Rainfall and lots of Crop which are given below.

Banana, Cotton, Watermelon, Lentil, Orange, Apple, Kidneybeans, Muskmelon, Mothbeans, Mungbean, Pomegranate, Coffee, Coconut, Jute, Pigeonpeas, Mango, Grapes, Papaya, Maize, Rice, Blackgram, Chickpea. we are using all these parameters to predict the crop.

Tools we have used in this project to predict and analyse the data set is anaconda, jupyter note. Anaconda is an open-source data science distribution for the Python and R programming languages that attempts to make package management and deployment easier. Anaconda is the most convenient way to combine Python/R data science and machine learning on a single computer. JupyterLab is the most recent interactive development environment for notebooks, code, and data that is available on the web. Users can create and arrange workflows in data science, scientific computing, computational journalism, and machine learning using its versatile interface. Extensions to enhance and enrich functionality are encouraged by a modular architecture.

1.1 Lifecycle of Agriculture

We can divide the Process of Agriculture into different parts:

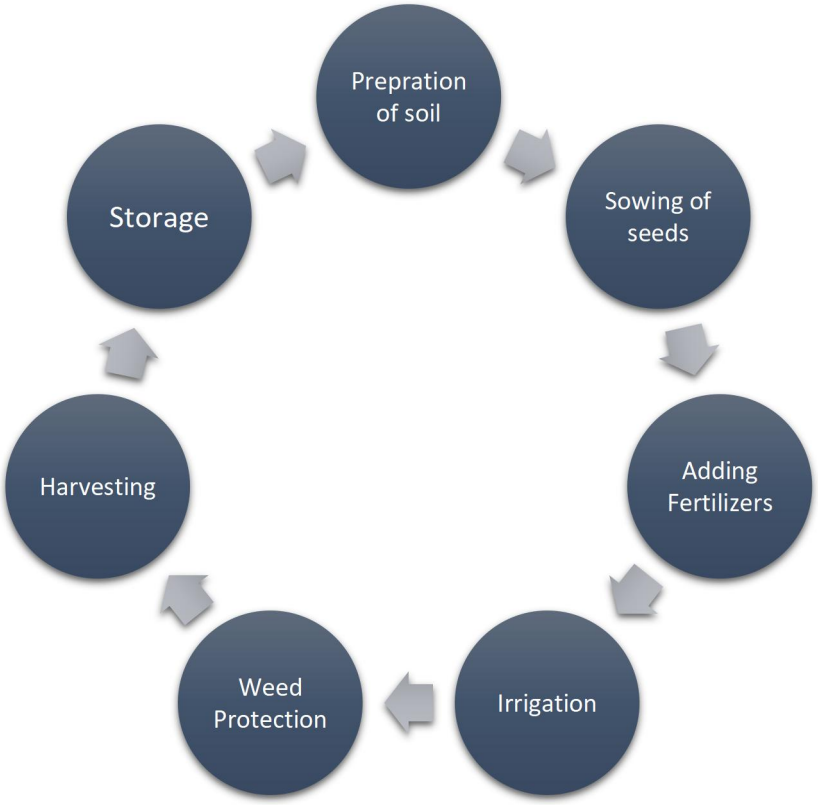

Fig. 1.1 Lifecycle of Agriculture

1.1.1 Soil preparation

"It is the initial stage of agriculture where farmers prepare the land for planting. This process involves breaking up large clods of soil and removing debris, such as sticks, rocks and roots. Also, add fertilizers and organic matter depending on the type of crop to create an ideal situation for crops." [1]

1.1.2 Sowing of seeds

"plant seeds This stage requires taking care of the distance between two seeds, the depth for planting. At this stage, climatic conditions such as temperature, humidity and precipitation play an important role." [1]

1.1.3 Adding Fertilizers

"Adding fertilizers Maintaining soil fertility is an important factor for the farmer to continue to produce nutritious and healthy crops. Farmers are turning to fertilizers because these substances contain plant nutrients such as nitrogen, phosphorus and potassium. Fertilizers are simply plant nutrients that are applied to agricultural fields to supplement the necessary elements found naturally in the soil. This stage also determines the quality of the crop." [1]

1.1.4 Irrigation

"Irrigation This stage helps keep the soil moist and maintain moisture. Insufficient or excessive watering can stop crop growth and, if not done correctly, can damage crops." [1]

1.1.5 Weed protections

"protection from weeds Weeds are unwanted plants that grow near crops or farm borders. Weed protection is important to consider as weeds reduce yields, increase production costs, interfere with harvesting, and reduce crop quality." [1]

1.1.6 Harvesting

"Harvesting refers to the act of gathering ripe harvests from the fields. This task necessitates a significant number of workers, making it a labor-intensive activity. Cleaning, grading, packing, and refrigeration are all part of the post-harvest process." [1]

1.1.7 Storage

storage This step of the post-harvest system involves storing products outside of agricultural seasons to ensure food safety. Packaging and transportation of crops are also included.

1.2 Farmers' challenges when adopting traditional farming methods

List the general challenges that exist in the agricultural sector.

Climate elements including rainfall, temperature, and humidity play a big effect in agriculture's life cycle. Climate change is caused by increased deforestation and pollution, making decisions about soil preparation, planting, and harvesting more challenging for farmers. "Each crop necessitates its own type of soil nutrition. Three nutrients are required in the soil: nitrogen (N), phosphorus (P), and potassium (K) (K). Crops of poor quality can be caused by nutrient deficiencies. Weed control is critical in the agricultural life cycle." [1] If left unchecked, it can raise production costs and take nutrients from the soil, perhaps leading to nutrient deficiencies.

1.3 Proposed Algorithm

1. Support Vector Machine Algorithm
2. KNN (K Nearest Neighbor Algorithm)
3. Decision Tree Classification Algorithm
4. Random Forest Algorithm
5. Logistic Regression

1.3.1 Support Vector Machine Algorithm

“Support Vector Machine (SVM) is a supervised machine learning algorithm that can be used for both classification or regression challenges. However, it is mostly used in classification problems. In the SVM algorithm, we plot each data item as a point in n-dimensional space (where n is a number of features you have) with the value of each feature being the value of a particular coordinate. Then, we perform classification by finding the hyper-plane that differentiates the two classes very well (look at the below snapshot).Support Vectors are simply the coordinates of individual observation. The SVM classifier is a frontier that best segregates the two classes (hyper-plane/ line)” [2]

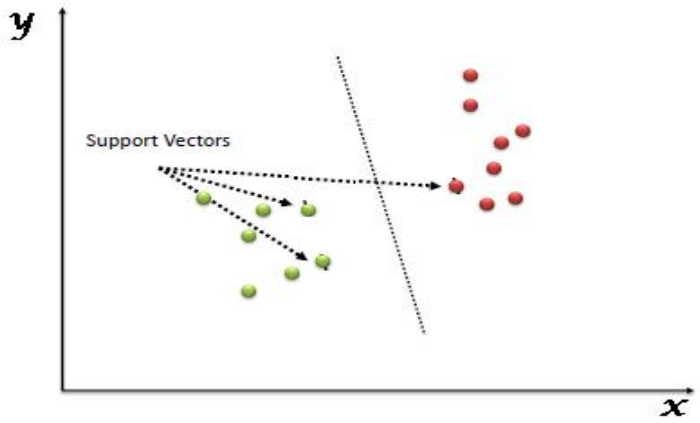

Fig. 1.2 Support Vector Machine

1.3.2 KNN (K Nearest Neighbor algorithm)

"K Nearest Neighbor algorithm falls under the Supervised Learning category and is used for classification (most commonly) and regression. It is a versatile algorithm also used for imputing missing values and resampling datasets." [3] As the name (K Nearest Neighbor) suggests it considers K Nearest Neighbors (Data points) to predict the class or continuous value for the new Datapoint.

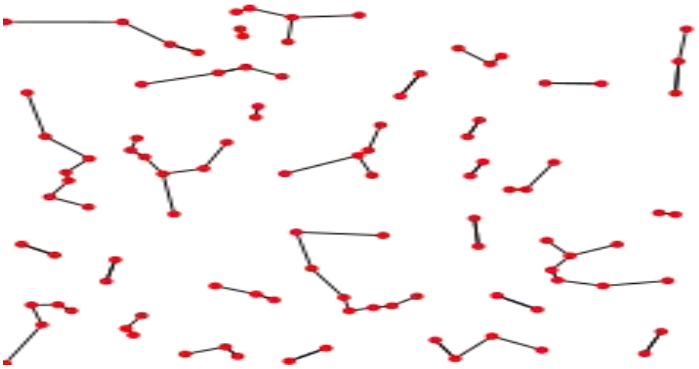

Fig. 1.3 "A nearest neighbor graph of 100 points in the Euclidean plane" [4]

1.3.3 Decision Tree Classification Algorithm

"Decision Tree is a Supervised learning technique that can be used for both classification and Regression problems, but mostly it is preferred for solving Classification problems. It is a tree-structured classifier, where internal nodes represent the features of a dataset, branches represent the decision rules and each leaf node represents the outcome." [6]

- "In a Decision tree, there are two nodes, which are the Decision Node and Leaf Node. Decision nodes are used to make any decision and have multiple branches, whereas Leaf nodes are the output of those decisions and do not contain any further branches.
- The decisions or the test are performed on the basis of features of the given dataset.
- It is a graphical representation for getting all the possible solutions to a problem/decision based on given conditions.
- It is called a decision tree because, similar to a tree, it starts with the root node, which expands on further branches and constructs a tree-like structure.
- In order to build a tree, we use the CART algorithm, which stands for Classification and Regression Tree algorithm.
- A decision tree simply asks a question, and based on the answer (Yes/No), it further split the tree into subtrees.
- Below diagram explains the general structure of a decision tree:" [6]

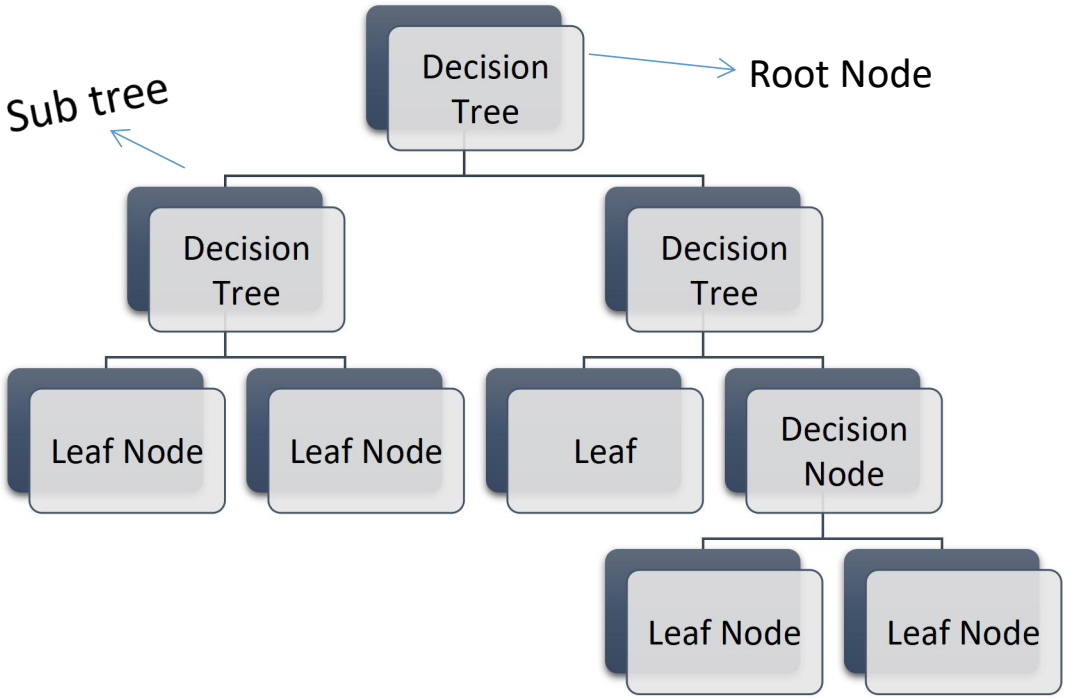

Fig. 1.4 Decision Tree

1.3.4 Random Forest Algorithm

Random forest is a supervised machine learning algorithm commonly used in classification and regression issues. " It constructs decision trees from various samples and uses the majority vote for classification and the average for regression. One of the most essential properties of the Random Forest Algorithm is that it can handle data sets with both continuous and categorical variables, as in regression and classification. It produces superior categorization results." [5]

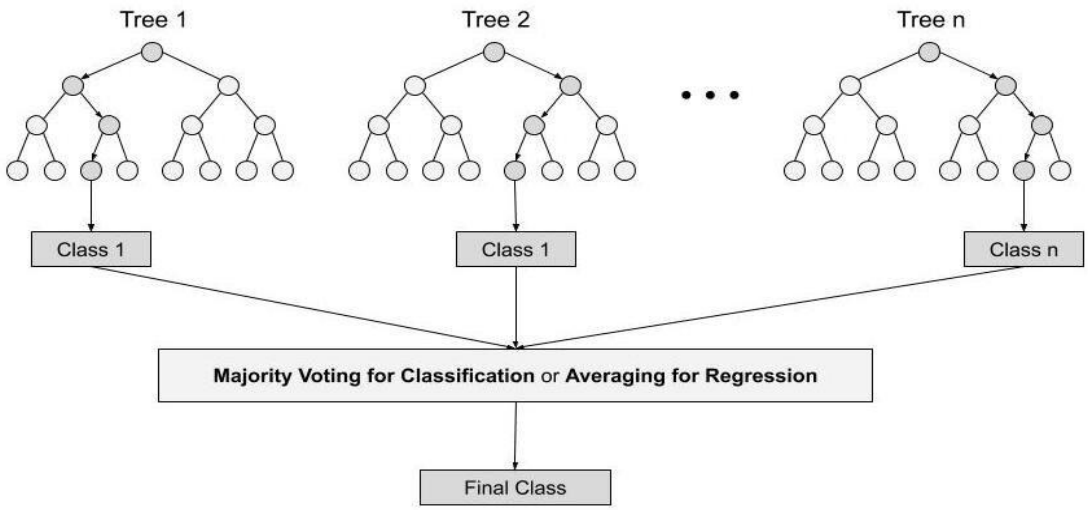

Fig. 1.5 Random Forest

1.3.5 Logistic Regression Algorithm

"This type of statistical analysis (also known as *logit model*) is often used for predictive analytics and modeling, and extends to applications in machine learning. In this analytics approach, the dependent variable is finite or categorical: either A or B (binary regression) or a range of finite options A, B, C or D (multinomial regression). It is used in statistical software to understand the relationship between the dependent variable and one or more independent variables by estimating probabilities using a logistic regression equation. " [7]

This sort of statistical analysis (also known as a logit model) is commonly used for predictive analytics and modelling, as well as machine learning applications. The

dependent variable in this analytics technique is finite or categorical. “Your analysis can look at known characteristics of visitors, such as sites they came from, repeat visits to your site, behavior on your site (independent variables). Logistic regression models help you determine a probability of what type of visitors are likely to accept the offer — or not. As a result, you can make better decisions about promoting your offer or make decisions about the offer itself.” [7]

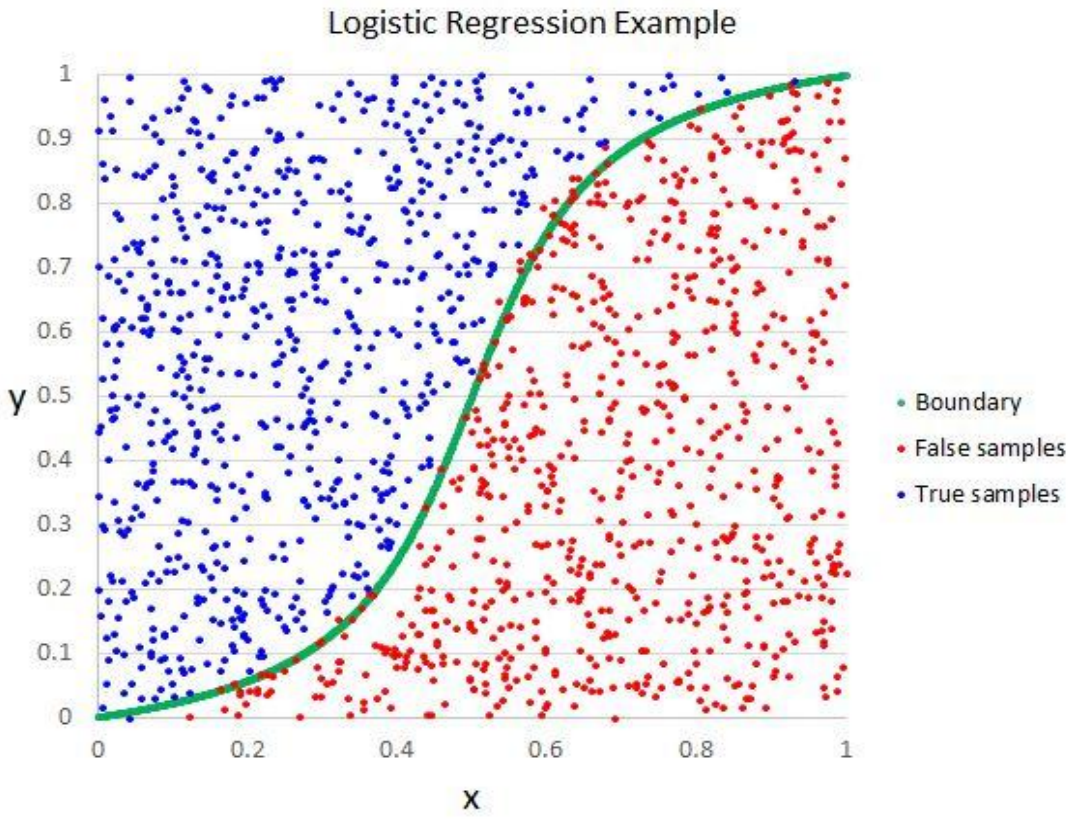

Fig. 1.6 Logistic Regression

OBJECTIVE

1. The model can produce the best agricultural production plan by merging multiple competing criteria into a single utility function.
2. Machine learning is used to analyze data.
3. These data set suggesting crops according to the given data set of soil and using the prediction of weather.
4. In given dataset we have different value of Nitrogen, Phosphorus, Potassium, Temperature, Humidity, Ph, Rainfall.
5. Machine learning process is used for prediction of best soil,weather and climatic condition and get the best result for our crops grow.
6. Through all these process we are getting the best predicition to grow the crops in depending soil ,weather and climatic condition.

CHAPTER 2

LITERATURE REVIEW

Liakos et. al. (2018)

"The work on this paper involved a research into the use of machine learning agricultural production systems. The methods in this paper include work applied artificial neural networks. The result showed that machine learning models have been used in several agriculture-related areas. Mainly in crop production and aiding management decision making processes." [8]

Priya et al. (2018)

"The work on this paper was concerned with the use of the random forest algorithm to generate predictions for crop yield and improvement. The method include in this paper is the random forest algorithm was used for yield production using a data-set with four features or parameters. A training set as used to train the algorithm rules which were then applied to the remaining datasets. In this paper The results showed that we can attain an accurate crop yield prediction using the random forest algorithm. Random forest algorithm achieves a largest number of crop yield models with lowest models. It is suitable for massive crop yield prediction in agricultural planning." [9]

Jeong et. al. (2016)

"The work on this paper aimed at examining the performance efficiency of the random forest algorithm in crop yield prediction for the wheat crop, potato crop, and maize crop. The methods in this paper include The random forest algorithm was used to train the datasets, and the same datasets were applied to an MLR model as a benchmark for the random forest algorithm. The result showed that the random forest algorithm is far more effective in crop yield prediction." [10]

Ming et. al. (2016)

"The work on this paper involved classification of land cover based on image and remote sensing. The methods in this paper include Random forest machine learning

algorithm was used in the classification of image data. The result showed Random forest is an efficient classification algorithm and performs effectively without using special selected features." [11]

Nitze et al. (2012)

"The work on this paper compared the effectiveness of several machine learning algorithms: support vector machine, artificial neural networks, and random forest. The method in this paper include several classifiers, Naïve Bayes for ML, random forest (RF), multilayer perceptron in case of ANN, and Lib SVM for support vector machine, were used in this work for the classification of crops. The result showed Even though classification results depended strongly on the number of images used, the SVM classifiers performed much better than the RF and ANN in most of the cases." [12]

Chen and Cournede (2017)

"The work on this paper focused on finding the most efficient way to predict the yield of corn based on meteorological records. The method in this paper studied a new methodology named multiple scenarios parameter estimation and used the CORNFLO model. The result showed Random forest regression was shown to be the most efficient for crop yield prediction." [13]

Mitra et al. (2017)

"The work on this paper focused on simulating and predicting crop yield for effective crop management and adequate results. The method in this paper include A three-layered artificial neural network (ANN) and R language were used in this work for prediction and simulation of crop yield. The result showed the artificial neural network was effective for simulation and prediction." [14]

Marion Olubunmi Adebiyi et al. (2020)

"This work focused on classification and the aim of data classification here was to divide the crops into classes based on their respective data. This work machine learning applied what had been used to set parameters and embedded it into a dataset on a mobile application. classification was carried out using random forests to allow all inputs to be considered multiple times for better accuracy since the algorithm comprises multiple decision trees." [15]

Shi, L., Duan, Q., Ma, X., Weng, M. (2012)

"The work on this paper involved a research into the use of machine learning algorithm. the work on this paper was concerned with the use of support vector machine algorithm to generate a prediction. this paper uses a agriculture data set and classify a data from dataset using classification.this paper shows experimental results which shows that the svm algorithm outperforms two popular algoritghm that is naive bayes and artifical neural network in the term of f2 measure." [16]

Satyanarayanan M., Bahl P., Caceres R., Davies N. (2012)

"the work on this paper involved a research into rice yield production using machine learning algorithm. the algorithm using in this paper is support vector machine.Accurate forecasting may help stakeholders in taking appropriate and timely decision.methodology using in this paper on svm and their application basic concept involved in this paper related svm.the data set encompasses the rice yield in india from the year 1950 to 2014." [17]

Xiao Z., Song W., Chen Q (2018)

"Proposed an approach using Markov Chain Theory, this model focused on corn and cotton yield and the prediction of diseases and pests related to these crops. This method gives better results than the regression model. Predicting diseases and pests is done using the sensor based approach. The Naive Bayes Kernel algorithm is used to compare the patterns from the crop data. A sample dataset of crops is given as a training set to Naive Bayes Kernel algorithm and the raw dataset of soil sample and temperature. The model gives the pattern comparison of both datasets. If the pattern is consistent then there is no disease and crop growth will be good, If it is inconsistent then the disease is predicted." [18]

Thomas van Klompenburg, Ayalew Kassahun, Cagatay Catal (2020)

"The work on this paper involved a review into the use of machine learning algorithm. in this paper machine learning based crop yield prediction have been synthesized. this paper uses fifty ml based papers and thirty deep learning based paper. this paper focus on temperature, rainfall and soil types. in this paper used most widely used machine

learning algoritm neural networks and also used deep learning algorithm cnn.the methodology in this paper are assigned in different stages like first define research question, define search protocol, select database, define search string.this paper performed an additional search in electronic databases to identify deep learning-based studies, reached 30 deep learning-based papers, and extracted the applied deep learning algorithms. According to this additional analysis, Convolutional Neural Networks (CNN) is the most widely used deep learning algorithm in these studies, and the other widely used deep learning algorithms are Long-Short Term Memory (LSTM) and Deep Neural Networks (DNN).” [19]

Santosh T. Jagtap (2022)

“The work on this paper involved a research into the uses of application of various machine learning techniques in agriculture. this paper used various machine learning techniques such as knn algoritm for classification a data this paper also focus on svm.Although k-Nearest Neighbor classification is a simple and effective technique, it slows down the classification of each object. Furthermore, the classification’s effectiveness suffers as a result of the uneven distribution of training data. The purpose of this paper is to look into the applicability of various machine learning techniques in agriculture.” [20]

H. K. Karthikeya, K. Sudarshan, Disha S. Shetty (2020)

“The work on this paper involved a research into the prediction of agriculture crops. in this paper they have using a k-nearest neighbor algorithm to predict a crop. the problem statement of this paper specially focus on agriculture of indian farmers.the method in this paper include a collection of data set which contains various of food items and also they have uses the data collect related to Rainfall. Humidity, Soil type, Irrigation type, Previous Yields,Location, Price, Year, type of crop, Crop diseases and its symptoms. this paper uses knn to predicting of crop yield and the accuracy they get 56% in kodagu region and they have also try in different region like mangalore and they get 63% accuracy.the result shows knn is not much more accurate.” [21]

Dr. V. Latha Jothi, Neelambigai A, Nithish Sabari S, Santhosh K (2020)

"The work on this paper involved a research into the data mining and the problem of predicting the yielding of crops an exciting challenge. they have discussed about various related works already done in data mining techniques using previous years dataset. the method involved in this paper is to classifies the groundwater level dataset to predict the future test data record dataset using knn. the result shows that the knn could be useful in analysing rainfall prediction, temperature prediction and the ground water levels in the past and which predict the future levels." [22]

A. Suresh, P. Ganesh Kumar and M. Ramalatha (2018)

"The work on this paper involved a research on prediction of major crops of tamilnadu using k-means and modify knn. in this paper they have proposed a prediction method for the major crops of Tamilnadu using K-means and Modified K Nearest Neighbor (KNN). the method include in this paper uses matlab and weka as the tool for clustring and classification respectively.The number result shows that in this paper knnand modified knn method is better than traditional data mining approach." [23]

K. PORCHILAMBI,Dr. P. SUMITRA (2019)

"The work on this paper involved a research on machine learning is used in different fields like machine, banking, web based appliaction and agriculture for prediction, classifiaction and clustring. in this paper machine learning application in the field of agriculture has been reviewed and discussed. machine learning algorithm is suitable for predicting large data set and certain unpredicatable situtation prevailing in crop prediction can be solved using these algorithms." [24]

Nebeesath Sunaina (2020)

"The work on this paper involved a analysis of Crop Yield Prediction by using Machine Learning Algorithms. in this paper two machine learning algorithm have been used knn and svm. the method include in this paper initially it gathers basic user information like the Name, Email id, Password and phone number from the user and store in the databaes. the result shows that the user can then select the option to

predict the crop which is suitable for the particular location and the soil type and He will also get the estimated yield of the predicted crop according to the number of acres of land he has given as the input." [25]

Gao Yi-yang, Ren Nan-ping (2009)

"The work on this paper involved data mining and analysis of our Agriculture Based On The Decision Tree.in this paper introduces the concept of classification and the method of the decision tree. Then, this paper analyses the data of rural labor, arable land area and the gross output value of agriculture about 30 cities of China based on the decision tree, and adopts clustering analysis method to discretize continuous data during the process of data mining in order to subjectivity comparing to the traditional classification methods. Finally, generating the decision tree of our agriculture, thereby gaining the spatial classification rules and analyzing the rules." [26]

B. Ganga Bhavani, G. L. N. V. S. Kumar, Moram Lakshim Rekha, K N V P S Brahma Ramesh (2021)

"The work on this paper involved machine learning algortihm which predict the Various Crops in Agricultural Field. the method include in this paper For better crop yield the artificial Neural Network have demonstrated to be an effective tool for modeling and prediction by using Decision tree and Navi Bayes Algorithm. This paper mainly focused on the techniques and measures taken to improve farming by in calculating the technical knowledge and developments in order to make the agricultural sector more reliable and easier for the farmers by predicting the suitable crop by using Machine Learning techniques. different types of crops are using in this paper like wheat, jowar, mustard, sugarcane and two types of soil taken balck soil and alluvial soil for prediction and making results.Classification and Regression Trees (CART) are one implementation of Decision Trees." [27]

N. Selvam , M. Yuvaraj (2018)

"The work on this paper involved machine learning and its algorithm. the study of this paper proposes the decision tree algorithm to investigate the different elements which affects and demanding to establish the root component of production control.in this

paper they describes the fundamental theory of decision tree technique which is used to forecast or predict and classify the agriculture data. For all class of agriculture data this decision tree method is utilized for prediction and classification.the working methodology using in this paper about decision tree is First choose the training experience or data, then Choose target fuction that is to be learned, then Choose how to represent the target function.Finally choose the learning algorithm to inter the target function. the result shows that The experimental data are from the total output value of an annual agriculture production,Use decision tree to classify the attributes contains city,labours in rural area, land area and yield of agriculture and the results are listed in Table." [28]

Qiao Jie (2022)

"The work on this paper involved Precision and intelligent agricultural decision support system based on big data analysis.this paper combines big data technology to carry out precision mining of agricultural data, and combines decision tree algorithm to carry out data classification processing.this paper obtains the most effective agricultural decision reference data through data mining, combines the agricultural decision support process to set the functional modules of the decision system, and analyzes the implementation process of each functional module.this paper verifies the structure of the model in this paper combined with experimental analysis. From the comparison of experiments, it can be seen that the precision and intelligent agricultural decision support system constructed in this paper has significant effects." [29]

Kalesanwo Olamide,Awodele Oludele,Eze Monday,Kuyoro 'Shade, Ajaegbu Chigozirim (2020)

"The work on this paper involved autonomousirrigation system,Various AISs have been proposed to ensure effective management of water resources, soil water content optimization and increasing crop yield. these systems require some level of decisionmaking algorithm in order to make the appropriate irrigation decisions that are critical. Decision tree algorithms were evaluated and results are presented in this paper.the outcome of this paper is to Choosing the right features from your data sets

can make all the difference between medium performance which has recorded a long training time and great efficiency with short training time." [30]

Kiran Moraye, Aruna Pavate, Suyog Nikam, Smit Thakkar (2021)

"The work on this paper involved machine learning based Crop Yield Prediction Using Random Forest Algorithm. this paper focus on major cities of maharashtra to analyse the data and predict the crop. the goal of this research is to build a user-friendly web application which will help the farmers, user and more policy planners to predict the crop yield based on the factor of climate change. the method use in this paper is to analyse and predict the data from dataset.Data plays an important role in Machine Learning. To design and perform crop yield prediction system data is taken from various cities of Maharastra state. the dataset involve three things which is very important for this paper Data about the crop, climate of a particular district, and region needed to perform the system. random forest is used to predict and analyse the data set.the result shows that the model used in this paper used 10-fold cross-validation technique which is indicates which is gives high accuracy and correlation between the climate and the crop yield and accuracy of the model is found 87%." [31]

L. Karthigadevi, G. Srinivasagan (2020)

"The work on this paper involved Agricultural Data Analysis in Tirunelveli District using Random Forest Classification Algorithm. The formers to identify the deficiencies in the soil, PH value, Soil type, EC and soil texture to choose the correct crops to increase the production. In this paper soil fertility level is identified and crop selection by the use of a random forest classification algorithm in Tirunelveli district.The prosed work, using random forest classification method produced efficient results comparing with the other feature selection method. The experimental results shows that the random forest classification method gives 98% of high accuracy with the less time than the other feature selection methods. The proposed method also used to identify the user location. Compared the existing feature selection methods, the proposed experimental results shows that the random forest classification algorithm for agricultural data analysis produce a high accuracy and less processing time." [32]

M. Suresh et al. (2021)

"The work on this paper involved machine learning algorithm such as randm forest. this paper allow farmers to capture the yield of their crops before cultivation in the field of agriculture and thus help them make the necessary decisions.The results obtained are granted access to the farmer. And yet there are various methods or protocols for such very data analytics in crop yield prediction, and we are able to predict agricultural productivity with guidance of all those algorithms. It utilizes a Random Forest Algorithm." [33]

V. Geetha, A. Punitha, M. Abarna, M. Akshaya, S. Illakiya and A. P. Janani (2020)

"The work on this paper involved predicting crop using random forest.In this system, a machine-learning method, Random Forest algorithm has an ability to analyze crop growth related to the current climatic conditions and biophysical change.they have collected crop growth datasets from various sources.the data set used both training and testing process.Random Forest classifier was found huge ability to predict crop yield.From different outputs, it shows that Random Forest is an efficient learning algorithm to analyze crop at current climatic condition and has a huge exactness in data investigation." [34]

Jeong JH, Resop JP, Mueller ND, et al. (2016)

"The work on this paper involved global and regional crop yield predictions using random forest. this paper evaluated a machine-learning method, Random Forests (RF), for its ability to predict crop yield responses to climate and biophysical variables at global and regional scales in wheat, maize, and potato in comparison with multiple linear regressions (MLR) serving as a benchmark. the paper used crop yield data from various sources and regions for model training and testing. the resullt shows in this paper that RF is an effective and versatile machine-learning method for crop yield predictions at regional and global scales for its high accuracy and precision, ease of use, and utility in data analysis. RF may result in a loss of accuracy when predicting the extreme ends or responses beyond the boundaries of the training data." [35]

CHAPTER 3

RESULTS AND DISCUSSION

3.1 Applications of AI and ML in Agriculture

Artificial intelligence is being used by the food sector to help produce healthier crops, control pests, monitor soil and growing conditions, organise data for farmers, reduce effort, and improve a variety of agriculture-related operations across the food supply chain.

3.1.1 Use of weather forecasting

Farmers are having a difficult time determining the best time to plant seeds as weather conditions change and pollution rises. Farmers may study weather conditions using weather forecasts with the use of artificial intelligence, which helps them plan the type of cultivation that can be done. when the seeds should be sowed, and when they should be grown

3.1.2 Monitoring system for soil and crop health

Artificial intelligence is being used in the soil and crop health monitoring system business to assist farmers grow healthier crops, control pests, monitor soil and growing conditions, organise data, reduce effort, and improve a variety of agriculture-related activities. throughout the food web The sort of crop planted and its quality are heavily influenced by soil type and nutrition. The condition of the soil is diminishing as a result of growing deforestation, making it impossible to identify its quality. PEAT, a German technology company, has created Plantix, an AI-based software that can detect nutrient deficits in the soil, as well as pests and plant diseases, so farmers can decide whether or not to use fertilisers to increase crop quality. This application employs picture recognition technologies. Smartphones can be used by the farmer to photograph plants. This app also includes short films that demonstrate floor repair processes, as well as recommendations and other solutions. Trace Genomics is another machine learning-based startup that assists farmers with soil testing. Farmers can use this type of application to monitor soil and crop health and generate healthier, more productive harvests.

3.1.3 Predictive Analytics and Precision Farming

"AI and ML applications in agriculture have developed applications and tools which help farmers inaccurate and controlled farming by providing them proper guidance to farmers about water management, crop rotation, timely harvesting, type of crop to be grown,Farmers can satisfy the world's growing food demand by using IoT and AI-driven technologies to enhance productivity and revenue while preserving irreplaceable natural resources. AI will assist farmers in becoming agricultural technologists in the future, allowing them to use data to improve yields down to individual plant rows.

"The model can achieve the optimum production plan of an agricultural region combining in one utility function different conflicting criteria. Machine learning is used to analyze data."[6] These data set suggesting crops according to the given data set of soil and using prediction of weather. In given data-set we have different value of Nitrogen, Phosphorus, Potassium, Temperature, Humidity, soil-Ph value and Rainfall. Machine learning process is used for prediction of best soil, weather and climatic condition and get the best result for our crops grow.

3.2 Process Diagram

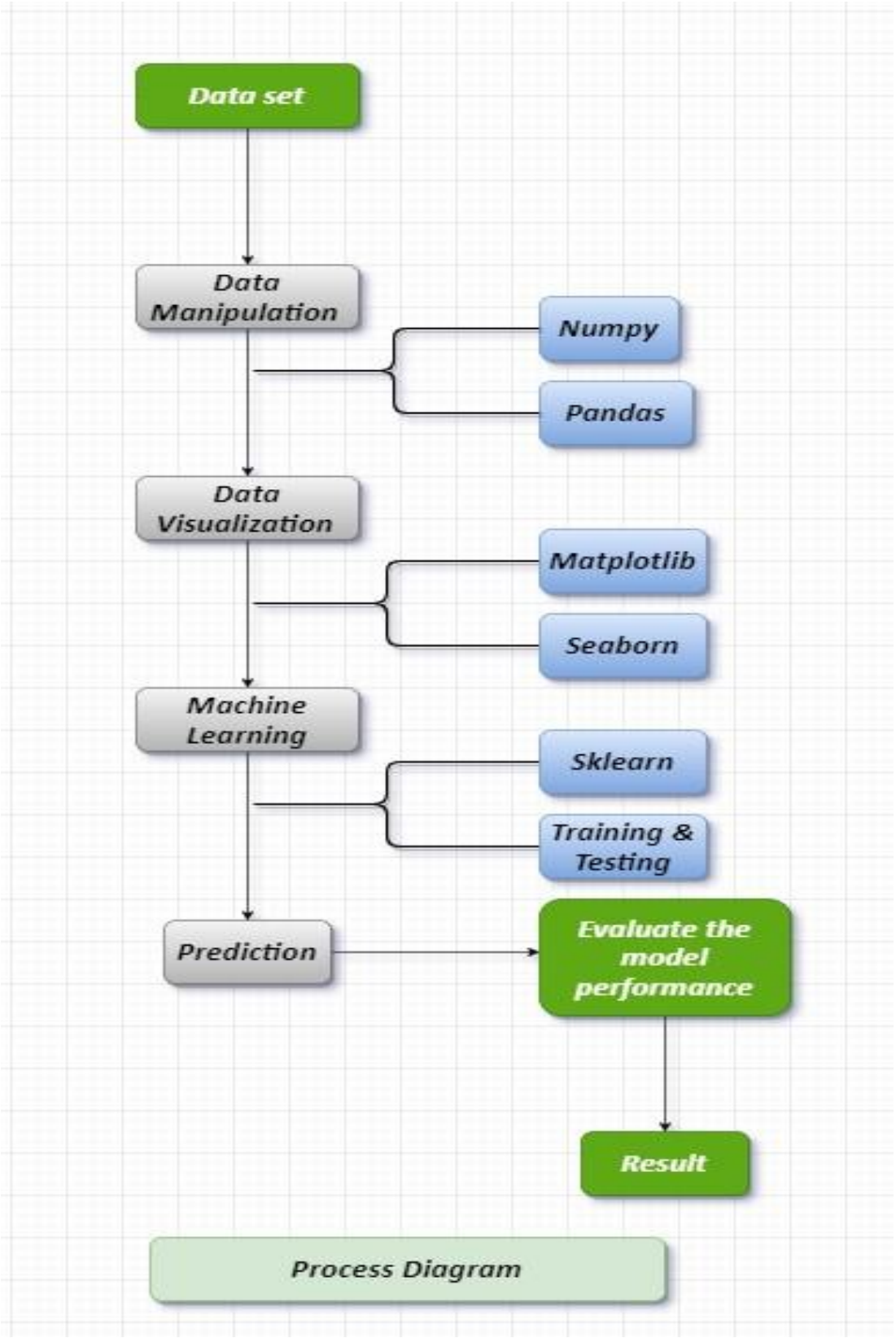

Fig. 3.1 Process Diagram

In the process diagram, we explain how our machine learning model processes data. First, we analyse the data set, and then we manipulate the data, which involves organising or arranging the type of structured data that computer programmes read to make it easier to interpret. Then we import the numpy and pandas libraries to read the dataset and visualise the results. The graphic display of data and information is the subject of visualisation, a multidisciplinary science. Then we import matplotlib and seaborn. Matplotlib is a charting library for the Python programming language and its NumPy numerical mathematics extension, while seaborn is a data visualisation library for the Python programming language. Seaborn is a matplotlib-based Python data visualisation package. It offers a high-level interface for creating visually appealing and useful statistical graphs. After that, we use the sklearn machine learning model. Simple and effective predictive data analysis tools Everyone can use it, and it can be reused in different situations. · We use a machine learning approach to train and test the data set, which is built on NumPy, SciPy, and Matplotlib. The model's performance is then evaluated using a multiple machine learning technique, and the final result is obtained.

3.3 COMPARISON OF ALGORITHM

Comparison Of Algorithm In Our Model For Best Result

In first step we train all the machine learning algorithm individually using the size of training data 50% and testing data 50% and try to find the accuracy of all the algorithm

- ***Support Vector Machine Algorithm***

In first algorithm we check the accuracy of Support Vector Machine Algorithm and check what is the accuracy of Support Vector Machine Algorithm are

After training a model we find the accuracy of Support Vector Machine Algorithm is 0.960

Now lets check another algorithm and try to find the accuracy of the algorithm in our model

- ***KNN (K Nearest Neighbor algorithm)***

In second algorithm we check the accuracy of KNN (K Nearest Neighbor algorithm) and check what is the accuracy of K Nearest Neighbor algorithm are

After training a model we find the accuracy of KNN (K Nearest Neighbor algorithm) is 0.966

Now lets check another algorithm and try to find the accuracy of the algorithm in our model

- ***Decision Tree Classifier***

In third algorithm we check the accuracy of Decision Tree Classifier and check what is the accuracy of Decision Tree Classifier are

After training a model we find the accuracy of Decision Tree Classifier is 0.961

Now lets check another algorithm and try to find the accuracy of the algorithm in our model

- ***Random Forest Algorithm***

In fourth algorithm we check the accuracy of Random Forest Algorithm and check what is the accuracy of Random Forest are

After training a model we find the accuracy of Random Forest is 0.991

Now lets check another algorithm and try to find the accuracy of the algorithm in our model

- ***Logistic Regression Algorithm***

In fifth algorithm we check the accuracy of Logistic Regression Algorithm and check what is the accuracy of Logistic Regression Algorithm are

After training a model we find the accuracy of Logistic Regression Algorithm is 0.950

Now in second step we combine all the five algorithm in one frame and check which algorithm is best for our model

After combining all the five algorithm in one image we can easily check which algorithm is best for our model.

Out[70]:

	name	test_size score 5.0
0	LogisticRegression()	0.950000
1	DecisionTreeClassifier(random_state=24)	0.961818
2	RandomForestClassifier(random_state=24)	0.991818
3	KNeighborsClassifier()	0.966364
4	SVC(random_state=24)	0.960000

Fig. 3.2 Comparison of algorithm using 50% of Training and 50% of Testing size

According to the above graph as we can see in image 3.2 we find that the best accuracy is 0.991.

Now this time in first step we train all the machine learning algorithm individually but we are going to change the size of training data from 50% to 70 % and testing data from 50% to 30% and then we will try to find the accuracy of all the algorithm.

- ***Support Vector Machine Algorithm***

In first algorithm we check the accuracy of Support Vector Machine Algorithm and check what is the accuracy of Support Vector Machine Algorithm are

After training a model we find the accuracy of Support Vector Machine Algorithm is 0.968

Now lets check another algorithm and try to find the accuracy of the algorithm in our model

- ***KNN (K Nearest Neighbor algorithm)***

In second algorithm we check the accuracy of KNN (K Nearest Neighbor algorithm) and check what is the accuracy of K Nearest Neighbor algorithm are

After training a model we find the accuracy of KNN (K Nearest Neighbor algorithm) is 0.981

Now lets check another algorithm and try to find the accuracy of the algorithm in our model

- ***Decision Tree Classifier***

In third algorithm we check the accuracy of Decision Tree Classifier and check what is the accuracy of Decision Tree Classifier are

After training a model we find the accuracy of Decision Tree Classifier is 0.983

Now lets check another algorithm and try to find the accuracy of the algorithm in our model

- ***Random Forest Algorithm***

In fourth algorithm we check the accuracy of Random Forest Algorithm and check what is the accuracy of Random Forest are

After training a model we find the accuracy of Random Forest is 0.996

Now lets check another algorithm and try to find the accuracy of the algorithm in our model

- ***Logistic Regression Algorithm***

In fifth algorithm we check the accuracy of Logistic Regression Algorithm and check what is the accuracy of Logistic Regression Algorithm are

After training a model we find the accuracy of Logistic Regression Algorithm is 0.965

Now in second step we combine all the five algorithm in one frame and check which algorithm is best for our model

After combining all the five algorithm in one image we can easily check which algorithm is best for our model

Out[44]:

	name	test_size score 3.0
0	LogisticRegression()	0.965152
1	DecisionTreeClassifier(random_state=24)	0.983333
2	RandomForestClassifier(random_state=24)	0.996970
3	KNeighborsClassifier()	0.981818
4	SVC(random_state=24)	0.968182

Fig. 3.3 Comparison of algorithm using 70% of Training and 30% of Testing size

According to the above graph as we can see in image we find that the best accuracy is 0.996.

Now this time in first step we train all the machine learning algorithm individually but we are going to change the size of training data from 70% to 80 % and testing data from 30% to 20% and then we will try to find the accuracy of all the algorithm.

- ***Support Vector Machine Algorithm***

In first algorithm we check the accuracy of Support Vector Machine Algorithm and check what is the accuracy of Support Vector Machine Algorithm are

After training a model we find the accuracy of Support Vector Machine Algorithm is 0.986

Now lets check another algorithm and try to find the accuracy of the algorithm in our model

- ***KNN (K Nearest Neighbor algorithm)***

In second algorithm we check the accuracy of KNN (K Nearest Neighbor algorithm) and check what is the accuracy of K Nearest Neighbor algorithm are

After training a model we find the accuracy of KNN (K Nearest Neighbor algorithm) is 0.981

Now lets check another algorithm and try to find the accuracy of the algorithm in our model

- ***Decision Tree Classifier***

In third algorithm we check the accuracy of Decision Tree Classifier and check what is the accuracy of Decision Tree Classifier are

After training a model we find the accuracy of Decision Tree Classifier is 0.993

Now lets check another algorithm and try to find the accuracy of the algorithm in our model

- ***Random Forest Algorithm***

In fourth algorithm we check the accuracy of Random Forest Algorithm and check what is the accuracy of Random Forest are

After training a model we find the accuracy of Random Forest is 0.997

Now lets check another algorithm and try to find the accuracy of the algorithm in our model

- ***Logistic Regression Algorithm***

In fifth algorithm we check the accuracy of Logistic Regression Algorithm and check what is the accuracy of Logistic Regression Algorithm are

After training a model we find the accuracy of Logistic Regression Algorithm is 0.968

Now in second step we combine all the five algorithm in one frame and check which algorithm is best for our model

After combining all the five algorithm in one image we can easily check which algorithm is best for our model

	Classifier	Predicted Value	F1-Score	Precision	Recall	Roc-Auc-Score
Agriculture Optimization System Using Machine Learning	SVM	0.98	0.97	0.97	0.97	0.996
	KNN	0.98	0.98	0.98	0.98	0.995
	Decision Tree	0.99	0.98	0.98	0.98	1
	Random Forest	0.99	0.99	0.99	0.99	1
	Logestic Regression	0.97	0.97	0.97	0.97	0.9983

Table 3.1 Comparison of Algorithm

According to the above table as we can see in table 4.1 we find that the best accuracy is 0.997, so for our model we find the best Machine Learning Algorithm is Random Forest Classifier. The accuracy of this algorithm is higher than other algorithm, so we can say that from all five algorithm, the best algorithm for our model is the second last from our table which is Random Forest Classifier due to its their accuracy.

3.4 USER INTERFACE

This is our final stage in this stage we are ready to make UI of our Project.

The user interface (UI) is the point at which a computer, website, or application interacts with humans. The purpose of good UI is to make the user's experience simple and intuitive, requiring the least amount of work from the user to achieve the maximum desired result.

We are using flask and Python to create a user Interface (UI)

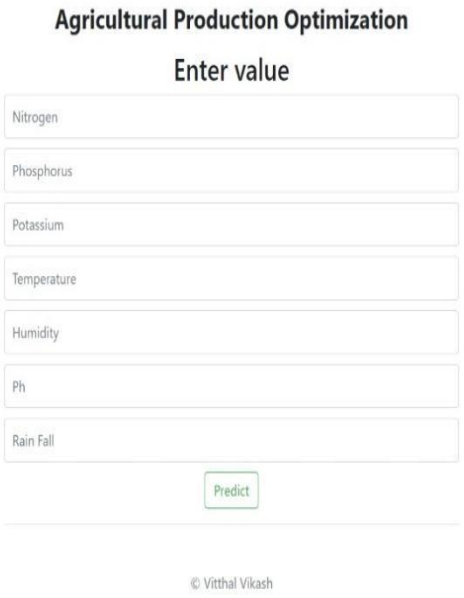

Fig. 3.4 Web Application Front Page

Fig. 3.5 Web Application Optimizer Output Area

Fig. 3.6 Web Application Optimizer With Output

These are the Screen-shot of our Project UI which we have done in our project.

1. First image we are showing the starting page of UI without data and
2. In second image we are giving inputs according to the parameter and
3. In third and last image we are showing the predicted output of our given data and getting a result

CHAPTER 4

CONCLUSION AND FUTURE WORK

As we all know that agriculture depends largely on the nature of soil and the climatic conditions and many a times, we face unpredictable changes in climate like, non-seasonal rainfall or heat waves or fluctuations in humidity levels, etc. and all such events cause a great loss to our farmers and farming, because of which they are not able to utilize their agricultural land to it's fullest.So to solve all such problems, I have build a Machine Learning Model by the virtue of which we can help farmers, optimize the agricultural production, because this predictive model will help them understand that for a particular soil & given climatic condition, which crop will be best suitable for the harvest.

There are 7 key factors that I've taken into account which will help us in determining, exactly which crop should be grown and at what period of time, viz. Amount of Nitrogen, Phosphorus and Potassium in soil, Temperature in degree celcius, Humidity, pH and Rainfall in mm.

In agriculture, artificial intelligence and machine learning are not only assisting farmers in automating their operations, but they are also assisting farmers in shifting to express cultivation for increased crop output and quality while using fewer resources. Companies that improve machine learning or Artificial Intelligence-based products or services, such as training data for agriculture, drones, and automatic machine manufacturing, will benefit from technological advancements in the future, allowing them to provide more useful applications to the current sector, assisting the planet in addressing food production issues for a growing population. i've got build a Machine Learning Model by the virtue of which we are able to help farmers, optimize the agricultural production, because this predictive model will help them understand that for a selected soil & given climatic condition, which crop are going to be best suitable for the harvest. According to the adage "Information is Power," keeping track of data on crops, the environment, and the market can help farmers make better decisions and solve challenges related to agriculture.

Outcome of this model is as we can see in table 4.1 we find that the best accuracy is 0.997, so for our model we find the best Machine Learning Algorithm is Random

Forest Classifier. The accuracy of this algorithm is higher than other algorithm, so we can say that from all five algorithm, the best algorithm for our model is the second last from our table which is Random Forest Classifier due to its their accuracy.

Future Work

Machine learning and deep learning have numerous uses in agriculture, each with its own set of obstacles. As a result of this research, we will endeavour to improve the implementation process faster, more precise, smoother, and more effective in the future.

1) Currently our model is accessible through the website only, but in future we will try to make a complete application.

2) In order to train the model, we will attempt to develop our own dataset and make it available to other academics using open platforms such as Kaggel, Meandly, IEEE Dataport, and so on.

3) Use a publicly available dataset for model testing and validation.

4) AutoML is a cutting-edge method for creating more accurate and high-quality machine learning models in less time.

5) Deployment of the model in real-time application is recommended to help the intended users in their mundane work.

6) It is suggested that the model be deployed in a real-time application to assist the intended users in their daily tasks.

7) In future when any algorithm comes in ml for prediction then we will also apply those algorithm in our model.

REFERENCES

[1] Pravar Jain, "Artificial Intelligence in Agriculture : Using Modern Day AI to Solve Traditional Farming Problems",Data Science Blogathon,November 4, 2020.

[2] sunil,"Understanding Support Vector Machine(SVM) algorithm from examples (along with code)",Data Science Blogathon,September 13, 2017

[3] Sai Patwardhan,"Simple understanding and implementation of KNN algorithm!",Data Science Blogathon, April 21, 2021

[4] Anton, Howard (1994), Elementary Linear Algebra (7th ed.), John Wiley & Sons, ISBN 0-471-58742-7

[5] Sruthi E R,"Understanding Random Forest",Data Science Blogathon, June 17, 2021

[6] javatpoint,"machine learning decision tree classification algorithm", www.javatpoint.com machine learning decision tree classification algorithm

[7] Ranjit Kumar Paul,"LOGISTIC REGRESSION",Indian Agricultural Statistics Research Institute Library Avenue,cabin publication smfa Module,2011

[8] K. Liakos, P. Busato, D. Moshou, S. Pearson, and D. Bochtis, "Machine learning in agriculture: a review," Sensors, vol. 18, no. 8, p. 2674, 2018.

[9] P. Priya, U. Muthaiah, and M. Balamurugan, "Predicting yield of the crop using machine learning algorithm," International Journal of Engineering Sciences & Research Technology, vol. 7, pp. 1–7, 2018.

[10] J. H. Jeong, J. P. Resop, N. D. Mueller et al., "Random forests for global and regional crop yield predictions," PLoS One, vol. 11, no. 6, Article ID e0156571, 2016.

[11] D. Ming, T. Zhou, M. Wang, and T. Tan, "Land cover classification using random forest with genetic algorithm-based parameter optimization," Journal of Applied Remote Sensing, vol. 10, Article ID 35021, 2016, http://remotesensing.spiedigitallibrary.org/.

[12] I. Nitze, U. Schulthess, and H. Asche, "Comparison of machine learning algorithms random forest artificial neural network and support vector machine to maximum likelihood for supervised crop yield classification," in Proceedings of the 4th GEOBIA, pp. 35–40, Salzburg Austria, May 2012.

[13] X. Chen and P. Cournede, “Model-driven and data-driven approaches for crop yield prediction: analysis and comparison,” International Scholarly and Scientific Research & Innovation, vol. 11, no. 7, pp. 334–342, 2017.

[14] P. Mitra, A. Mukherjee, and A. Chandra, “Machine learning for prediction of crop yield: a case study on rain-fed area rice yield from West Bengal, India,” in Proceedings of the DSANRM2017: Workshop on Data Science for Agriculture and Natural Resource Management, International Institute of Information Technology (IIIT, Hyderabad), December 2017.

[15] Marion Olubunmi Adebiyi, Roseline Oluwaseun Ogundokun, Aneoghena Amarachi Abokhai, "Machine Learning–Based Predictive Farmland Optimization and Crop Monitoring System", Scientifica, vol. 2020, Article ID 9428281, 2020.

[16] Shi, L., Duan, Q., Ma, X., Weng, M. (2012). The Research of Support Vector Machine in Agricultural Data Classification. In: Li, D., Chen, Y. (eds) Computer and Computing Technologies in Agriculture V. CCTA 2011. IFIP Advances in Information and Communication Technology, vol 370. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-27275-2_29.

[17] Satyanarayanan M., Bahl P., Caceres R., Davies N., 2009.”Machine Learning Methodologies for Paddy Yield estimation in India: a Case Study”, IEEE transaction.

[18] Xiao Z., Song W., Chen Q, 2018,”Predicting the Yield of Crop Using Machine Learning Algorithms”, International Journal of Engineering Sciences & Research Technology.

[19] Thomas van Klompenburg, Ayalew Kassahun, Cagatay Catal,Crop yield prediction using machine learning: A systematic literature review, Computers and Electronics in Agriculture,Volume 177,2020,105709,ISSN 0168-1699.

[20] Santosh T. Jagtap, Khongdet Phasinam, Thanwamas Kassanuk, Subhesh Saurabh Jha, Tanmay Ghosh, Chetan M. Thakar,Towards application of various machine learning techniques in agriculture,Materials Today: Proceedings,Volume 51, Part 1,2022,Pages 793-797,ISSN 2214-7853

[21] H. K. Karthikeya, K. Sudarshan, Disha S. Shetty, "Prediction of Agricultural Crops using KNN Algorithm",International Journal of Innovative Science and Research Technology,Volume 5, Issue 5, May – 2020,ISSN No:-2456-2165

[22] Dr. V. Latha Jothi, Neelambigai A, Nithish Sabari S, Santhosh K, 2020, Crop Yield Prediction using KNN Model, INTERNATIONAL JOURNAL OF

ENGINEERING RESEARCH & TECHNOLOGY (IJERT) RTICCT – 2020 (Volume 8 – Issue 12)

[23] A. Suresh, P. Ganesh Kumar and M. Ramalatha, "Prediction of major crop yields of Tamilnadu using K-means and Modified KNN," 2018 3rd International Conference on Communication and Electronics Systems (ICCES), 2018, pp. 88-93, doi: 10.1109/CESYS.2018.8723956.

[24] K. PORCHILAMBI,Dr. P. SUMITRA,"MACHINE LEARNING ALGORITHMS FOR CROP YIELD PREDICTION: A SURVEY", International Journal of Emerging Technologies and Innovative Research, ISSN:2349-5162, Vol.6, Issue 3, page no.112-116, March-2019

[25] Nebeesath Sunaina,"Analysis of Crop Yield Prediction by using Machine Learning Algorithms",International Research Journal of Engineering and Technology (IRJET),e-ISSN: 2395-0056,Volume: 07 Issue: 03 | Mar 2020

[26] Gao Yi-yang,Ren Nan-ping,"Data Mining And Analysis Of Our Agriculture Based On The Decision Tree",September 2009,IEEE Xplore Computing, Communication, Control, and Management, 2009. CCCM 2009. ISECS International Colloquium on, Volume: 2

[27] B. Ganga Bhavani, G. L. N. V. S. Kumar, Moram Lakshim Rekha, K N V P S Brahma Ramesh, 2021, Prediction of Various Crops in Agricultural Field Using Decision Tree and Naviebayes Algorithm in Machine Learning, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) ICRADL – 2021 (Volume 09 – Issue 05)

[28] N. Selvam , M. Yuvaraj, "DEVELOPMENT OF AGRICULTURE BASED ON DECISION TREE ALGORITHM", International Journal of Pure and Applied Mathematics,Volume 118 No. 20 2018, 2427-2435,ISSN: 1314-3395

[29] Qiao Jie (2022) Precision and intelligent agricultural decision support system based on big data analysis, Acta Agriculturae Scandinavica, Section B — Soil & Plant Science, 72:1, 401-414, DOI: 10.1080/09064710.2021.2008477

[30] Kalesanwo Olamide,Awodele Oludele,Eze Monday,Kuyoro ‘Shade, Ajaegbu Chigozirim,"Evaluation of Decision Tree Algorithms in Precision Agriculture",IJCAT - International Journal of Computing and Technology, Volume 7, Issue 3, March 2020.

[31] Kiran Moraye, Aruna Pavate,Suyog Nikam,Smit Thakkar,"Crop Yield Prediction Using Random Forest Algorithm for Major Cities in Maharashtra

State",International Journal of Innovative Research in Computer Science & Technology (IJIRCST) ISSN: 2347-5552, Volume-9, Issue-2, March 2021

[32] K. Karthigadevi,G. Srinivasagan,"Random Forest Classification Algorithm for Agricultural Data Analysis in Tirunelveli District",Journal of Xi'an University of Architecture & Technology,ISSN No : 1006-7930,Volume XII, Issue VIII, 2020

[33] N. Suresh et al., "Crop Yield Prediction Using Random Forest Algorithm," 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS), 2021, pp. 279-282.

[34] V. Geetha, A. Punitha, M. Abarna, M. Akshaya, S. Illakiya and A. P. Janani, "An Effective Crop Prediction Using Random Forest Algorithm," 2020 International Conference on System, Computation, Automation and Networking (ICSCAN), 2020, pp. 1-5.

[35] Jeong JH, Resop JP, Mueller ND, et al. Random Forests for Global and Regional Crop Yield Predictions. PLoS One. 2016;11(6):e0156571. Published 2016 Jun 3.

ANNEXURES

```
# data manipulate
import numpy as np
import pandas as pd
# data visualization
import matplotlib.pyplot as plt
import seaborn as sas
import klib #==> new model for visualization
# for interactivity
from ipywidgets import interact
# pandas (all lowercase) is a popular Python-based data analysis toolkit which can be imported using import pandas as pd . It presents a diverse range of utilities, ranging from parsing multiple file formats to converting an entire data table into a NumPy matrix array
#lets read the dataset
data = pd.read_csv("data.csv")
# lets cheak teh shape of the dataset
print("shape of the datset:" ,  data.shape)
data.head() #head 5
data.tail() #last 5
# #Pandas isnull() and notnull() methods are used to check and manage NULL values in a data frame.
# let find missing value
data.isnull().sum()
# - klib.describe - functions for visualizing datasets
# - klib.cat_plot(df) # returns a visualization of the number and frequency of categorical features
# - klib.corr_mat(df) # returns a color-encoded correlation matrix
# - klib.corr_plot(df) # returns a color-encoded heatmap, ideal for correlations
# - klib.dist_plot(df) # returns a distribution plot for every numeric feature
# - klib.missingval_plot(df) # returns a figure containing information about missing values
```

```
klib.cat_plot(data) # returns a visualization of the number and frequency of categorical features

klib.corr_mat(data) # returns a color-encoded correlation matrix

klib.corr_plot(data) # returns a color-encoded heatmap, ideal for correlations

klib.dist_plot(data) # returns a distribution plot for every numeric feature

klib.missingval_plot(data)# returns a figure containing information about missing values
#let cheak the crops present in this dataset
data["label"].value_counts()

# The format() method formats the specified value(s) and insert them inside the string's placeholder.
# The placeholder is defined using curly brackets: {}. Read more about the placeholders in the Placeholder section below.
#
# The format() method returns the formatted string

# lets check the summary for all the crops
#       Nitrogen phosphorous ---potassium temperature humidity ph rainfall
print('==> Avg Ratio of Nitrogen in the soil :{0:.2f}'.format(data["N"].mean()))
print('==> Avg Ratio of phosphorous in the soil :{0:.2f}'.format(data["P"].mean()))
print('==> Avg Ratio of potassium in the soil :{0:.2f}'.format(data["K"].mean()))
print('==> Avg Tempature in celsius :{0:.2f}'.format(data["temperature"].mean()))
print('==> Avg Relative humidity :{0:.2f}'.format(data["humidity"].mean()))
print('==> Avg ph value of the soil :{0:.2f}'.format(data["ph"].mean()))
print('==> Avg RailFall in nm :{0:.2f}'.format(data["rainfall"].mean()))
```

@interact automatically creates user interface (UI) controls for exploring code and data interactively. It is the easiest way to get started using IPython's widgets.

Mean = in the data set in have numerical value(1,2,5,9) then replace the missing value helping through mean
Median = The mean value of numerical data is without a doubt the most commonly used statistical measure.
Outlier Analysis is a data mining task which is referred to as an “outlier mining”
mode = in the data set in have categorical value(age , time etc) then replace the missing value helping through mean

#let cheak the summary statistics of the crops
Nitrogen phosphorous ---potassium temperature humidity ph rainfall

```
@interact
def summary (crops = list(data["label"].value_counts().index)):
   x = data[data["label"] == crops ]
   print("------------------------------------------------")
   print("==============statistics for Nitrogen============")
   print("Minimun Nitrogen reqired:", x["N"].min())
   print("avg Nitrogen reqired:", x["N"].mean())
   print("Maximun Nitrogen reqired:", x["N"].max())
   print("------------------------------------------------")
   print("==============statistics for phosphorous  ============")
   print("Minimun phosphorous  reqired:", x["P"].min())
   print("avg phosphorous  reqired:", x["P"].mean())
   print("Maximun phosphorous  reqired:", x["P"].max())
   print("------------------------------------------------")
   print("==============statistics for potassium  ============")
   print("Minimun potassium  reqired:", x["K"].min())
   print("avg potassium  reqired:", x["K"].mean())
   print("Maximun potassium  reqired:", x["K"].max())
```

```
    print("-----------------------------------------------")
    print("==============statistics for temperature  ============")
    print('==> Minimun temperature reqired :{0:.2f}'.format(x["temperature"].min()))
    print('==> Avg temperature reqired :{0:.2f}'.format(x["temperature"].mean()))
    print('==> Maximun temperature reqired :{0:.2f}'.format(x["temperature"].max()))
    print("-----------------------------------------------")
    print("==============statistics for humidity  ============")
    print('==> Minimun  humidity reqired :{0:.2f}'.format(x["humidity"].min()))
    print('==> Avg humidity reqired :{0:.2f}'.format(x["humidity"].mean()))
    print('==> Maximun humidity reqired :{0:.2f}'.format(x["humidity"].max()))
    print("-----------------------------------------------")
    print("==============statistics for ph  ============")
    print('==> Minimun ph reqired :{0:.2f}'.format(x["ph"].min()))
    print('==> Avg ph reqired :{0:.2f}'.format(x["ph"].mean()))
    print('==> Maximun ph reqired :{0:.2f}'.format(x["ph"].max()))
    print("-----------------------------------------------")
    print("==============statistics for rainfall  ============")
    print('==> Minimun rainfall reqired :{0:.2f}'.format(x["rainfall"].min()))
    print('==> Avg rainfall reqired :{0:.2f}'.format(x["rainfall"].mean()))
    print('==> Maximun rainfull reqired :{0:.2f}'.format(x["rainfall"].max()))

#lets compere the avg requirement for each each avgrage conditions
#       Nitrogen        phosphorous ---potassium temperature humidity       ph
        rainfall

@interact
def compare (conditions = ["N" , "P" , "K" , "temperature" , "humidity" , "ph" ,
"rainfall"]):
    print('avg value for ' , conditions , 'is {0:2f}'.format(data[conditions].mean()))
    print("-----------------------------------------------------------------")
    print("grapes          :          {0:2f}".format(data[(data["label"]          ==
"grapes")][conditions].mean()))
    print("muskmelon          :          {0:2f}".format(data[(data["label"]          ==
"muskmelon")][conditions].mean()))
```

```
    print("kidneybeans : {0:2f}".format(data[(data["label"] == "kidneybeans")][conditions].mean()))
    print("apple : {0:2f}".format(data[(data["label"] == "apple")][conditions].mean()))
    print("lentil : {0:2f}".format(data[(data["label"] == "lentil")][conditions].mean()))
    print("maize : {0:2f}".format(data[(data["label"] == "maize")][conditions].mean()))
    print("jute : {0:2f}".format(data[(data["label"] == "jute")][conditions].mean()))
    print("Rice : {0:2f}".format(data[(data["label"] == "rice")][conditions].mean()))
    print("coffee : {0:2f}".format(data[(data["label"] == "coffee")][conditions].mean()))
    print("mothbeans : {0:2f}".format(data[(data["label"] == "mothbeans")][conditions].mean()))
    print("mango : {0:2f}".format(data[(data["label"] == "mango")][conditions].mean()))
    print("pomegranate : {0:2f}".format(data[(data["label"] == "pomegranate")][conditions].mean()))
    print("chickpea : {0:2f}".format(data[(data["label"] == "chickpea")][conditions].mean()))
    print("coconut : {0:2f}".format(data[(data["label"] == "coconut")][conditions].mean()))
    print("banana : {0:2f}".format(data[(data["label"] == "rice")][conditions].mean()))
    print("pigeonpeas : {0:2f}".format(data[(data["label"] == "pigeonpeas")][conditions].mean()))
    print("orange : {0:2f}".format(data[(data["label"] == "orange")][conditions].mean()))
    print("blackgram : {0:2f}".format(data[(data["label"] == "blackgram")][conditions].mean()))
    print("cotton : {0:2f}".format(data[(data["label"] == "cotton")][conditions].mean()))
    print("mungbean : {0:2f}".format(data[(data["label"] == "mungbean")][conditions].mean()))
    print("watermelon : {0:2f}".format(data[(data["label"] == "watermelon")][conditions].mean()))

#let make this function more intvitive :==>
# in below avg and above agv conditions :==>
@interact
```

```
def compare (conditions = ["N" , "P" , "K" , "temperature" , "humidity" , "ph" , "rainfall"]):
    print("Crops which require greater than avg", conditions,"\n")
    print(data[data[conditions]>data[conditions].mean()]["label"].unique())
    print("----------------------------------------------------------")
    print("Crops which require less than avg", conditions,"\n")
    print(data[data[conditions]<=data[conditions].mean()]["label"].unique())

#seabron  #displot distribution plot function ?
##      Nitrogen      phosphorous ---potassium temperature humidity      ph
        rainfall

plt.subplot(2, 4, 1)
sas.distplot(data["N"],color = "darkblue")
plt.xlabel("Ratio of Nitrogen" , fontsize = 12)
plt.grid()

plt.subplot(2, 4, 2)
sas.distplot(data["P"],color = "black")
plt.xlabel("Ratio of phosphorous" , fontsize = 12)
plt.grid()

plt.subplot(2, 4, 3)
sas.distplot(data["K"],color = "grey")
plt.xlabel("Ratio of potassium" , fontsize = 12)
plt.grid()

plt.subplot(2, 4, 4)
sas.distplot(data["temperature"],color = "lightgreen")
plt.xlabel("Ratio of temperature" , fontsize = 12)
plt.grid()

plt.subplot(2, 4, 5)
sas.distplot(data["humidity"],color = "darkgreen")
```

```
plt.xlabel("Ratio of humidity" , fontsize = 12)
plt.grid()
plt.subplot(2, 4, 6)
sas.distplot(data["ph"],color = "pink")
plt.xlabel("Ratio of ph" , fontsize = 12)
plt.grid()

plt.subplot(2, 4, 7)
sas.distplot(data["rainfall"],color = "lightgrey")
plt.xlabel("Ratio of rainfall" , fontsize = 12)
plt.grid()

plt.suptitle("distribution Agricultural conditions" , fontsize=20)
plt.show()

#lets find out some interestings facts
##      Nitrogen      phosphorous      ---potassium      temperature      humidity      ph
        rainfall

print("some Intersting patterns")
print("============================")
print("crops which requires very high Ratio of Nitrogen content is soil:",
data[data["N"]>120]["label"].unique())
print("crops which requires very high Ratio of phosphorous content is soil:",
data[data["P"]>100]["label"].unique())
print("crops which requires very high Ratio of potassium content is soil:",
data[data["K"]>200]["label"].unique())
print("crops which requires very high rainfall:",
data[data["rainfall"]>200]["label"].unique())
print("crops which requires very low temperature :",
data[data["temperature"]<10]["label"].unique())
print("crops which requires very high temperature: ",
data[data["temperature"]>40]["label"].unique())
```

```
print("crops which requires very low humidity :", data[data["humidity"]<20]["label"].unique())
print("crops which requires very low ph:", data[data["ph"]<4]["label"].unique())
print("crops which requires very high ph:", data[data["ph"]>9]["label"].unique())

##  lets understand which crops can only be grown in summer  , winter , Rainy
print("====================summer crops=====================")

print(data[(data["temperature"]>30) & (data["humidity"]>50)]["label"].unique())

print("====================winter crops=====================")
print(data[(data["temperature"]<20) & (data["humidity"]>30)]["label"].unique())
print("====================Rainy crops=====================")
print(data[(data["rainfall"]>200) & (data["humidity"]>30)]["label"].unique())

#clustering analysis +> used to classication      data point into realvite groups
#that mean we assing samely data aasing one group

from sklearn.cluster import KMeans

# in unsupervised learning we do not have need labels
#removing the labels colunm
x = data.drop(["label"],axis=1)

#selecting all the value of the data
x = x.values

print(x.shape)

# What is elbow method in K-means?
# Image result for elbow method k means
# The elbow method runs k-means clustering on the dataset for a range of values for k (say from 1-10) and then for each value of k computes an average score for all
```

clusters. By default, the distortion score is computed, the sum of square distances from each point to its assigned center.

```
#In clustering analysis 1th we knows how many cluster we have
#so knowing cluster help elbow method
# lets determine the optimun number of clusters within dataset
plt.rcParams["figure.figsize"]=(10,4)
wcss = []
for i in range(1,11):
   km = KMeans(n_clusters = i , init = "k-means++" , max_iter = 300 , n_init = 10 , random_state = 0)
   km.fit(x)
   wcss.append(km.inertia_)
#lets plot the results
plt.plot(range(1,11),wcss)
plt.title("The Elbow method" , fontsize = 20)
plt.xlabel("No of clusters")
plt.ylabel("wcss")
plt.show()

# lets implement the k means algorithm to perfron clustering analysis
km = KMeans(n_clusters = 4 , init = "k-means++" , max_iter = 300 , n_init = 10 , random_state = 0)
y_means = km.fit_predict(x)

#lets find out the results
a = data["label"]
y_means = pd.DataFrame(y_means)
z = pd.concat([y_means,a],axis= 1)
z = z.rename(columns = {0:"cluster"})

#lets cheak the clusters of each crops
print("lets check results the k means algorithm to perfron clustering analysis \n ")

print("Crops in frist cluster:", z[z["cluster"]==0]["label"].unique())
```

```
print("==========================================================
=====")
print("Crops in frist cluster:" ,z[z["cluster"]==1]["label"].unique())
print("==========================================================
=====")

print("Crops in frist cluster:" ,z[z["cluster"]==2]["label"].unique())
print("==========================================================
=====")
print("Crops in frist cluster:" ,z[z["cluster"]==3]["label"].unique())

# #prediction ==> uesd machine learning mobel
# #predictive modeline is part of data set used dataset and make a partten of model
# #finaly model train so used  model to predicition of unseen data
# #machine learning mobel used dataset and trend and partten to  make dision rules to
predicition finaly rejeltt

#machine learning mobel used  ==> logical reasoning

# lets split tge adtaset for predictive modeling
y = data["label"]
x = data.drop(["label"],axis = 1)

print("shape of x:",x.shape)
print("shape of y:",y.shape)

from sklearn.model_selection import train_test_split
x_train , x_test ,  y_train , y_test =  train_test_split(x,y, test_size = 0.2 , random_state
= 0)
print("the shape x train:",x_train.shape)
print("the shape x train:",x_test.shape)
print("the shape y train:",y_train.shape)
print("the shape y train:",y_test.shape)
```

```
# lets created a predicitive model
from sklearn.linear_model import LogisticRegression

model = LogisticRegression()

model.fit(x_train , y_train)

y_pred = model.predict(x_test)

# letts evalute tge model performace
from sklearn.metrics import confusion_matrix

#lets prit yhe confusion motrix frist
plt.rcParams["figure.figsize"]  =  (10,10)
cm = confusion_matrix(y_test ,y_pred)
sas.heatmap(cm , annot = True , cmap = "Wistia" )
plt.title("confusion maxrtix for logistic regression" , fontsize = 15)

# letts print the classifixcation report also
from sklearn.metrics import classification_report
cr = classification_report(y_test, y_pred)
print(cr)
data.head()
predicition = model.predict((np.array([[90,40,40,20,80,7,200]])))
print("the suggested crop for given climatic condition id :", predicition)
```

Annexures 1

Train and Compare Machine Learning Models

```
# the libraries we need
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeClassifier
from sklearn.ensemble import RandomForestClassifier
from sklearn.neighbors import KNeighborsClassifier
from sklearn.svm import SVC
from xgboost import XGBClassifier
from sklearn.metrics import classification_report
from sklearn.metrics import accuracy_score
from sklearn.tree import export_graphviz
import graphviz
# data manipulate
import numpy as np
import pandas as pd
# data visualization
import matplotlib.pyplot as plt
import seaborn as sas
#import klib #==> new model for visualization
# for interactivity
from ipywidgets import interact
import klib
from sklearn.metrics import roc_curve
from sklearn.metrics import roc_auc_score

data = pd.read_csv("data.csv")
y = data["label"]
x = data.drop(["label"],axis = 1)

#from sklearn.model_selection import train_test_split
```

```
x_train , x_test ,  y_train , y_test =  train_test_split(x,y, test_size = 0.4 , random_state = 0)
# first, initialize the classificators
tree= DecisionTreeClassifier(random_state=24) # using the random state for reproducibility
forest= RandomForestClassifier(random_state=24)
knn= KNeighborsClassifier()
svm= SVC(random_state=24)
xboost= XGBClassifier(random_state=24)

# now, create a list with the objects
models= [tree, forest, knn, svm ,xboost ]

for model in models:
    model.fit(x_train, y_train) # fit the model
    y_pred= model.predict(x_test) # then predict on the test set
    accuracy= accuracy_score(y_test, y_pred) # this gives us how often the algorithm predicted correctly
    clf_report= classification_report(y_test, y_pred) # with the report, we have a bigger picture, with precision and recall for each class
    print(f"The accuracy of model {type(model)} is {accuracy:.2f}")
    print(clf_report)
   print("\n")
```

Annexures 2

Deployment of Ml Algorithm Using Falsk

```
import pandas as pd
import numpy as np
import pickle

data = pd.read_csv("data.csv")

data.isnull().sum()

y = data["label"]
x = data.drop(["label"],axis = 1)

from sklearn.model_selection import train_test_split
x_train , x_test , y_train , y_test = train_test_split(x,y, test_size = 0.3 , random_state = 0)
#from sklearn.svm import SVC
from sklearn.ensemble import RandomForestClassifier

forest = RandomForestClassifier()
sv = forest.fit(x_train, y_train)
pickle.dump(sv, open('iri.pkl', 'wb'))
```

Annexures 3

Taking Input Data and Pass to Machine Learning Algorithm and Getting Output

```
  from flask import Flask, render_template, request
  import pickle
  import numpy as np

  model = pickle.load(open('iri.pkl', 'rb'))
app = Flask(__name__)

  @app.route('/')
  def man():
      return render_template('home.html')

  @app.route('/', methods=['POST'])
  def home():
      data1 = request.form['a']
      data2 = request.form['b']
      data3 = request.form['c']
      data4 = request.form['d']
      data5 = request.form['e']
      data6 = request.form['f']
      data7 = request.form['g']
      predicition = model.predict((np.array([[data1, data2, data3, data4 , data5 , data6 , data7]])))
      #print("the suggested crop for given climatic condition id of RandomForestClassifier:", predicition)
      output = predicition[0]

      return render_template ("home.html" , pr_text = output)
  if __name__ == "__main__":
      app.run(debug=True)
```

Report

ORIGINALITY REPORT

29%	25%	18%	%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.hindawi.com Internet Source	6%
2	www.analyticsvidhya.com Internet Source	2%
3	Vishal Meshram, Kailas Patil, Vidula Meshram, Dinesh Hanchate, S.D. Ramkteke. "Machine Learning in Horticulture Domain: A state-of-art Survey", Artificial Intelligence in the Life Sciences, 2021 Publication	2%
4	ejmcm.com Internet Source	2%
5	www.ijert.org Internet Source	1%
6	acadpubl.eu Internet Source	1%
7	ar.kalasalingam.ac.in Internet Source	1%
8	www.science.gov Internet Source	

		1 %
9	ijircst.org Internet Source	1 %
10	www.x-mol.com Internet Source	1 %
11	www.ibm.com Internet Source	1 %
12	www.cse.dmu.ac.uk Internet Source	1 %
13	Gao Yi-yang. "Data mining and analysis of our agriculture based on the decision tree", 2009 ISECS International Colloquium on Computing Communication Control and Management, 08/2009 Publication	1 %
14	V. Geetha, A. Punitha, M. Abarna, M. Akshaya, S. Illakiya, AP. Janani. "An Effective Crop Prediction Using Random Forest Algorithm", 2020 International Conference on System, Computation, Automation and Networking (ICSCAN), 2020 Publication	1 %
15	Namgiri Suresh, N.V.K. Ramesh, Syed Inthiyaz, P. Poorna Priya et al. "Crop Yield Prediction Using Random Forest Algorithm", 2021 7th	1 %

International Conference on Advanced Computing and Communication Systems (ICACCS), 2021
Publication

16	jaanoandseekho.com Internet Source	1 %
17	www.ijraset.com Internet Source	1 %
18	www.researchgate.net Internet Source	1 %
19	www.dqindia.com Internet Source	1 %
20	Abbas Saad Alatrany, Abir Hussain, Mustafina Jamila, Dhiya Al-Jumeiy. "Stacked Machine Learning Model for Predicting Alzheimer's Disease Based on Genetic Data", 2021 14th International Conference on Developments in eSystems Engineering (DeSE), 2021 Publication	1 %
21	Santosh T. Jagtap, Khongdet Phasinam, Thanwamas Kassanuk, Subhesh Saurabh Jha, Tanmay Ghosh, Chetan M. Thakar. "Towards application of various machine learning techniques in agriculture", Materials Today: Proceedings, 2022 Publication	1 %

publication.babcock.edu.ng

22 Internet Source

1%

Exclude quotes	On	Exclude matches	< 1%
Exclude bibliography	On		

Report

GRADEMARK REPORT

FINAL GRADE

/10

GENERAL COMMENTS

Instructor

PAGE 1

PAGE 2

PAGE 3

PAGE 4

PAGE 5

PAGE 6

PAGE 7

PAGE 8

PAGE 9

PAGE 10

PAGE 11

PAGE 12

PAGE 13

PAGE 14

PAGE 15

PAGE 16

PAGE 17

PAGE 18

PAGE 19

PAGE 20