ISSN 2063-5346

CROP ENDORSEMENT SYSTEM OPTIMIZATION USING GAUSSIAN NAIVE BAYES CLASSIFICATION TECHNIQUE

Faheem Ahmed[1] Dr. Sunil Vijay Kumar Gaddam[2]

Article History: Received: 10.05.2023 Revised: 29.05.2023 Accepted: 09.06.2023

Abstract

Vegetables, fruits, cereals, natural textile fibres like cotton and jute, as well as many other goods, are heavily dependent on agriculture in India. Additionally, the agriculture industry is crucial to the nation's economic development. As a result, employment in India is significantly impacted by agricultural productivity. The soil in India has been used for thousands of years, which has caused nutrient and mineral depletion and fatigue, which lowers crop productivity. Additionally, the lack of contemporary applications creates a demand for precision agriculture. A number of techniques and instruments are used in precision agriculture, commonly referred to as satellite farming, to manage farms based on the observation, measurement, and reaction to crop variability both within and between fields. The recommendation of precise crops is one of the key uses of precision agriculture. It aids in boosting crop productivity and generating revenue. The purpose of this research is to review and evaluate the effectiveness of various techniques on crop recommendation systems. In this research, we analysed and fed a crop recommendation dataset from the internet to a number of machine learning algorithms. With an accuracy of roughly 99.5%, we see that the model developed using the Gaussian Naive Bayes technique exceeds all other models.

Index Terms **—** Corp, Support vector machines,Machine learning algorithms,Linear regression,Artificial neural networks, Soil,Mobile applications,Random forests.

[1]Research Scholar, Dept. of Computer Science and Engineering, Lords Institute of Engineering & Technology, Hyderabad, Telangana

[2] Associate Professor, Dept. of Computer Science and Engineering, Lords Institute of Engineering & Technology, Hyderabad, Telangana

DOI:10.48047/ecb/2023.12.9.31

I. INTRODUCTION

India's population relies on the agriculture industry for about 58% of their income. The world's second-largest producer of important food staples including wheat and rice, as well as a variety of textile raw materials based on agriculture, coconut, sugarcane, different types of dry fruits, and a huge variety of vegetables and fruits, is India. The growth of the nation's economy depends heavily on the agriculture industry. Site-specific farming, often known as precision agriculture, is one of the many methods used to enhance crop health and lower environmental concerns.

One of the main domains in precision agriculture is the recommendation of crops to increase crop yield. The suggestion of crops helps in increasing the productivity of crops without the consumption of too many resources in the process. The main objective of precision farming is profitability and sustainability. It also helps the farmers acquire knowledge to determine the diseases that the crops could get infected by in advance using different methods of precision farming. The inputs given to the models are characteristics of soil, rainfall, humidity, and temperature of the environment. Based on those parameters provided as the input, the farmer gets suggested the most suitable crop for their field to achieve the maximum crop yield.

The agriculture sector started using Machine Learning algorithms and the Internet of Things to solve problems including low crop yields, irrigation, and loss of nutrients and minerals in the soil. Using Arduino microcontrollers and sensors, such as sunlight intensity sensors, soil moisture sensors, soil pH sensors, and humidity and temperature sensors, IoT is used to collect environmental data. The gathered data is sent to the database by connecting the sensors to the Arduino Wifi Module. A supervised learning algorithm is a form of machine learning method in which machines are trained using the training data and then used to predict the output. The labeled data indicates that some of the input data have already been tagged with the appropriate result. Unsupervised learning algorithms are a subclass of machine learning algorithms that train models on unlabeled data before allowing them to operate on it without supervision. Unsupervised learning aims to uncover a dataset's underlying structure, categorize data based on similarities, and display it in a compact way.

Naive Bayes, Support Vector Machine (SVM), Decision Tree, KNN, Random Forest, are the different kinds of supervised learning algorithms and unsupervised learning algorithms used in the systems. The predictions generated by the classification models are summed up and
used.

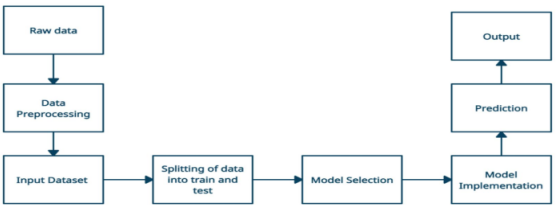


Figure 1. Block Diagram of Crop recommendation Model

The block diagram depicts the basic stages involved in creating the machine learning model above. The system's first stage involves cleaning and processing raw data. Crop data and fertilizer data are combined in this process to produce a final dataset. This dataset is then separated into training and testing data and utilized as input to the system. Model selection is the process of comparing various classification models in order to find the best appropriate one. The most suited crops are predicted, and the output is used for the following.

II. SYSTEM ANALYSIS

Problem Statement:

The soil in India has been in use for thousands of years, resulting in depletion and exhaustion of nutrients and minerals, which leads to a reduction of crop yield. Also, there is a lack of modern applications, which causes a need for precision agriculture. Precision Agriculture, also known as Satellite farming is a series of strategies and tools to manage farms based on observing, measuring, and responding to crop variability both within and between fields. One of the main applications of precision agriculture is the recommendation of accurate crops. It helps in increasing crop yield and gaining profits.

The three most basic amenities required for the survival of a human being are food, shelter and clothing. In today's tech-savvy generation, the latter two have witnessed a huge scientific boost. Unfortunately, even today, agriculture is considered as more of a man-power oriented field. Most of the farmers are untutored and have little to no scientific knowledge of farming. So, they have to rely on the hit and trial method to learn from experience which leads to wastage of time and resources.

Aim Of The Project

We have to take into consideration these factors which will be taken as an input from the farmer/authority that will be deciding which crop to sow. A detailed analysis has to be carried out on these factors and various inferences were generated from the results that we then took into consideration while predicting the optimum crop. The dataset has to be analyzed by plotting graphs and maps to check the effect and correlation of different attributes. The attributes taken into consideration are the ratio of nitrogen in soil, ratio of phosphorous in soil, ratio of potassium in soil, temperature, humidity, pH of soil and the amount of rainfall. This covers varied factors which affect the growth of crops and are the deciding factors in which crop to be chosen to sow. Need to design a model that shows the label as output which is the recommended crop. Need to apply different algorithms like - logistic regression, SVM and Random Forest Classifier and find which Classifier gaves the best results and so the model has been made using the same. The model has then been incorporated with GUI using Tkinter which makes it a complete application which can directly be put to use without further updates required.

Scope Of The Project

A. Crop recommendation using machine learning or deep learning

Crop Recommendation Systems use inputs of soil parameters like the ratio of Nitrogen, the ratio of Phosphorous, the ratio of Potassium and ph value of the soil, environmental factors such as humidity, rainfall, temperature, and many more.

The paper [3] displayed by Kiran Shinde, Jerrin Andrei, and Amey Oke proposes a crop recommendation system by comparing Naive Bayes, ID3, and Random Forest algorithms. The Random Forest algorithm is utilized to create the model, which is more accurate than the Naive Bayes and ID3 algorithms. The paper also proposes two more systems: fertilizer recommendation and crop rotation guidance. The crop rotation recommendation model is developed by using the Naïve Bayes algorithm. The fertilizer recommender calculates all of the fertilizer combinations that will suit the crop's needs at the lowest cost.

Proposed System:

We have to use a sample data set from Kaggle, which consisted of various factors needed for the proper growing of a crop.

We have to take into consideration these factors which will be taken as an input from the farmer/authority that will be deciding which crop to sow. A detailed analysis has to be carried out on these factors and various inferences were generated from the results that we then took into consideration while predicting the optimum crop. The dataset has to be analyzed by plotting graphs and maps to check the effect and correlation of different attributes. The attributes taken into consideration are the ratio of nitrogen in soil, ratio of phosphorous in soil, ratio of potassium in soil, temperature, humidity, pH of soil and the amount of rainfall. This covers varied factors which affect the growth of crops and are the deciding factors in which crop to be chosen to sow. The model then shows the label as output which is the recommended crop. After applying three different algorithms- logistic regression, SVM and Random Forest Classifier, we found out that Random Forest Classifier gave the best results and so the model has been made using the same. The model has then been incorporated with GUI using Tkinter which makes it a complete application which can directly be put to use without further updates required.

III. PROJECT IMPLIMENTATION

Algorithm/ Technique Used

The Technical Approach

Below is the technical approach to address the problem:

1. Identification of dataset
2. Explorative Data Analysis
3. Cleaning the dataset and applying NLP techniques
4. Feeding the dataset to multiple algorithms and finding the best algorithm that suits the scenario
5. Training the final classifier and creating a model for the final classifier
6. Testing the final classifier and saving the results.

Creating Model:

This project helps to understand the emotional well being of the users through their posts and detect and self destructive intentions from their tweets. Timely support can curb suicides and save lives. In this work, we consequently removed casual inactive subjects from tweeter which communicating self-destructive ideations. We first abstractly assessed the dormant subjects and afterward contrasted them with chance components. Proposed models are also used to predict the urgency of the posts.

This notebook includes the following:

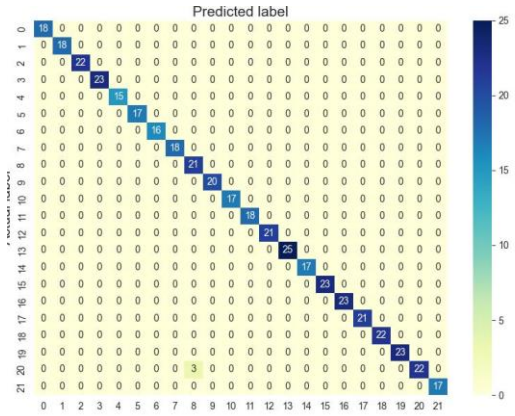
1. Dataset splitting
2. Text pre-processing and cleaning
3. Vectorizing with Tfidf
4. Random Forest classification
5. Hyperparameter tuning
6. Evaluation of Validation and Test Set

Most significant features for prediction

Final Classifier using Gaussian Naïve Bayes:

Training Accuracy Score: 99.5%

Validation Accuracy Score: 99.3%



	precision	recall	f1-score	support
apple	1	1	1	18
banana	1	1	1	18
blackgram	1	1	1	22
chickpea	1	1	1	23
coconut	1	1	1	15
coffee	1	1	1	17
cotton	1	1	1	16
grapes	1	1	1	18
jute	0.88	1	0.93	21
kidneybeans	1	1	1	20
lentil	1	1	1	17
maize	1	1	1	18
mango	1	1	1	21
mothbeans	1	1	1	25
mungbean	1	1	1	17
muskmelon	1	1	1	23
orange	1	1	1	23
papaya	1	1	1	21
pigeonpeas	1	1	1	22
pomegranate	1	1	1	23
rice	1	0.88	0.94	25
watermelon	1	1	1	17
accuracy			0.99	440
macro avg	0.99	0.99	0.99	440
weighted avg	0.99	0.99	0.99	440

A. Proposed Modular Implementation

Below is the proposed modular implementation of the project. It consists of modules:

Admin Module:

1. Login
2. Upload crop recommendation dataset that was downloaded from Kaggle
3. Exploratory Data Analysis
4. Data Preprocessing
    a. Check for duplicates in the dataset.
    b. Transform Categorical features using label encoding.
    c. Split the data into Training and Testing Datasets.
5. Feeding the dataset to multiple classification algorithms
    a. Multinomial Naive Bayes
    b. Random Forest
    c. Decision Trees
    d. Logistic regression
    e. Support Vector Machine
6. Creation of model using **Gaussian Naïve Bayes**

IV. PROJECT EXECUTION

Home page:

This is the starting page of the application when the application is executed on Pycharm, the application is hosted on a web server and URL is generated to access the application once the user clicks on the URL the below page is opened on the browser.

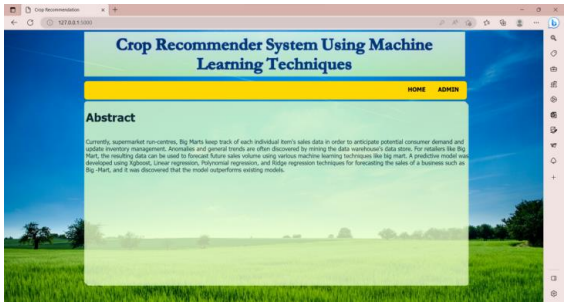


Fig: Home Page

Admin Login:

This is the login page for the admin module. The admin need to login into the system with his credentials in order to perform operations like uploading the dataset, Training the dataset, Exploratory data Analysis of the dataset, Feeding the dataset to different Machine learning Algorithms to find the Algorithm that can meet the best accuracy and Create a model that can be hosted on the Flask Application to be used by the users.

Fig: Admin Login

Upload Dataset:

On this page, the administrator of the system can upload datasets that are used for training the machine learning models. The admin has to select the file by clicking on the Choose file button and click on the upload button to upload the file to the server. Once the upload is complete, a success message would be displayed that the file is successfully uploaded. For this project we are using Crop_recommandation.csv as a dataset.

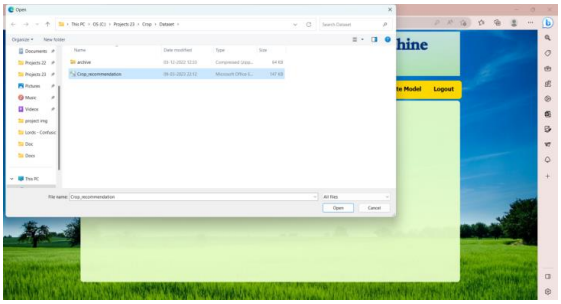


Fig: Upload Dataset & File Uploaded Successfully.

Data Analysis:

Exploratory Data Analysis is performed on the dataset in order to clean the dataset for any missing data, identify patterns, identify the relationships of various parameters of the outputs with the help of graphs, statistics etc.

Fig: Data Analysis

Crop vs Soil Analysis - Nitrogen:

The below graph shows the Crop vs Soil Analysis - Nitrogen over data present in the dataset.

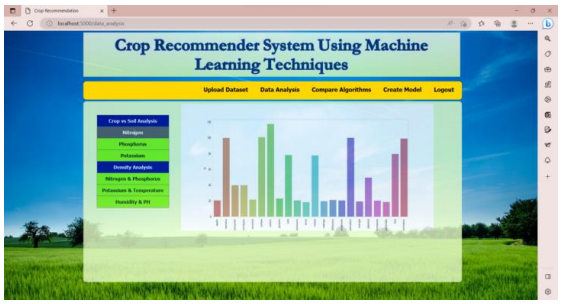


Fig: Crop vs Soil Analysis – Nitrogen

Crop vs Soil Analysis – Phosphorus Analysis:

The below graph shows the Crop vs Soil Analysis - Phosphorus Analysis over data present in the dataset.

Fig: Crop vs Soil Analysis – Phosphorus Analysis

Crop vs Soil Analysis - Potassium Analysis:

The below graph shows the Crop vs Soil Analysis - Potassium Analysis over data present in the dataset.

Fig: Crop vs Soil Analysis - Potassium Analysis

Density Analysis - Nitrogen & Phosphorus Analysis:

The below graph shows the Density Analysis - Nitrogen & Phosphorus Analysis over data present in the dataset.

Fig Density Analysis - Nitrogen & Phosphorus Analysis

Density Analysis - Potassium and Temperature Analysis:

The below graph shows the Density Analysis - Potassium and Temperature Analysis over data present in the dataset.

Fig Density Analysis - Potassium and Temperature Analysis

Density Analysis - Humidity and PH Analysis:

The below graph shows the Density Analysis - Humidity and PH Analysis over data present in the dataset.

Fig Density Analysis - Humidity and PH Analysis

Compare Algorithms:

On this page, the admin can feed the dataset to various Algorithms to train them and get the test accuracy for each algorithm. When the dataset is feed to various algorithms to evaluate the situation with some parameters like Accuracy, F1-Score , Recall…

Multinomial Naive Bayes Classifier:

When the dataset is feed to Multinomial Naive Bayes Classifier algorithm we observe that the test accuracy is 85.91%.

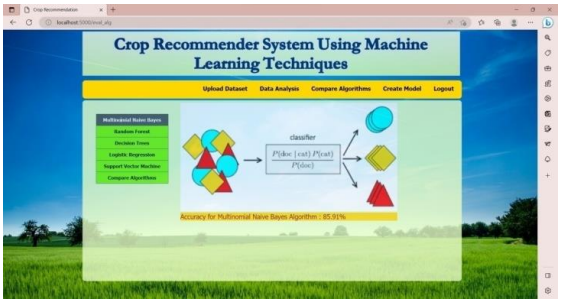


Fig: Multinomial Naive Bayes Classifier

Random Forest Classifier:

When the dataset is feed to Random Forest algorithm we observe that the test accuracy is 99.32%.

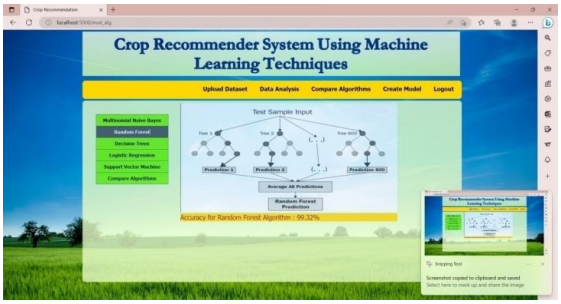


Fig: Random Forest Classifier

Decision Tree Classifier:

When the dataset is feed to Decision Tree Classifier algorithm we observe that the test accuracy is 98.18%.

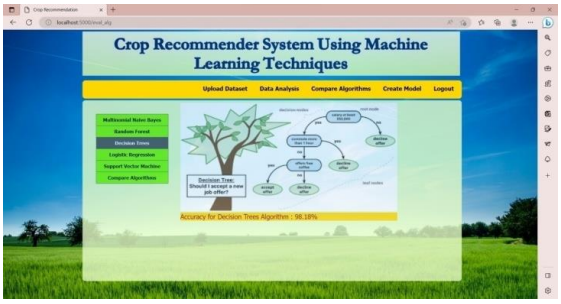


Fig: Decision Tree Classifier

Logistic Regression Classifier:

When the dataset is feed to Logistic Regression Classifier algorithm we observe that the test accuracy is 94.55%.

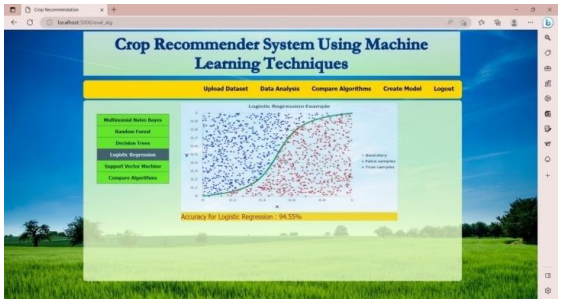


Fig: Logistic Regression Classifier

Support Vector Machine Classifier:

When the dataset is feed to Support Vector Machine Classifier algorithm we observe that the test accuracy is 96.14%.

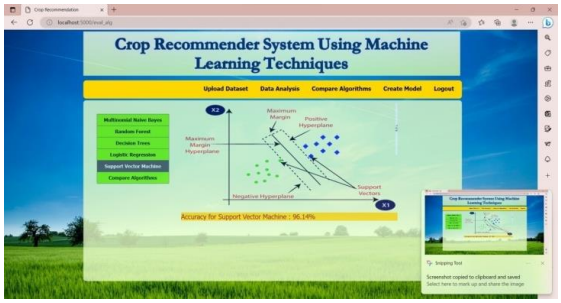


Fig: Support Vector Machine Classifier

Compare Algorithm Summary:

On this page, the admin can feed the dataset to various Algorithms to train them, get the test accuracy for each algorithm and their accuracies are summarized here.

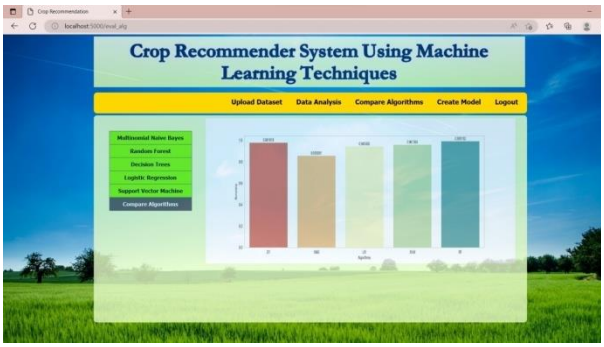


Fig: Compare Algorithm Summary

Create Model:

This screen shows the Training Accuracy of the Model is 99.55% and Test Accuracy of the Model is 99.32%.

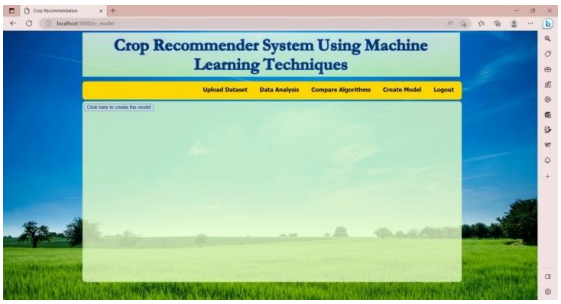
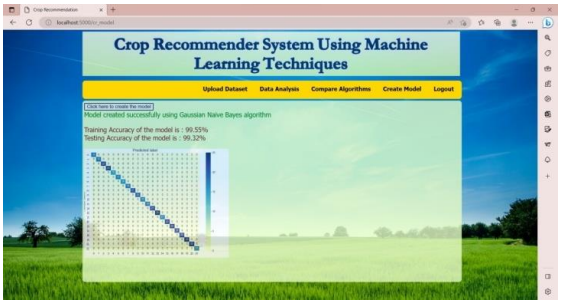


Fig: Create Model



CONCLUSION

The limits of current technologies and their usefulness for yield prediction were emphasised in this paper. The built-in prediction technology aids farmers in forecasting crop yields. The built-in recommender system enables the user to investigate potential crops and their yield in order to make more informed judgements. Several machine learning methods, including Random Forest, SVM, and KNN, were constructed and tested on the provided datasets for yield to accuracy. The accuracy of the various algorithms is contrasted. The results show that, with an accuracy of 99.5%, Gaussian Naive Bayes is the most accurate standard algorithm employed on the provided datasets. The suggested model looked into when fertilisers should be applied and suggested a suitable time frame. As we all know, a lot of agricultural research has been done and is still being done to increase

production, strengthen the Indian economy, and, most significantly, help farmers earn more money. To do this, a proposed model utilising the Gaussian Nave Bayes algorithm will provide farmers with guidance regarding the optimum crop to grow on their area. to benefit farmers through it.

REFERENCES

[1] Umamaheswari S, Sreeram S, Kritika N, Prasanth DJ, "BIoT: Blockchain-based IoT for Agriculture", 11th International Conference on Advanced Computing (ICoAC), (pp. 324-327). IEEE.

[2] Jain A. "Analysis of growth and instability in the area, production, yield, and price of rice in India", Journal of Social Change and Development.

[3] Manjula E, Djodiltachoumy S, "A model for prediction of crop yield" International Journal of Computational Intelligence and Informatics, 6(4):2349-6363.

[4] Sagar BM, Cauvery NK., "Agriculture Data Analytics in Crop Yield Estimation: A Critical Review", Indonesian Journal of Electrical Engineering and Computer Science, 1087-93.

[5] Wolfert S, Ge L, Verdouw C, Bogaardt MJ, "Big data in smart farming– a review. Agricultural Systems",;153:69-80.

[6] Jones JW, Antle JM, Basso B, Boote KJ, Conant RT, Foster I, Godfray HC, Herrero M, Howitt RE, Janssen S, Keating BA, "Toward a new generation of agricultural system data, models, and knowledge products: State of agricultural systems science. Agricultural systems",;155:269-88.

[7] Johnson LK, Bloom JD, Dunning RD, Gunter CC, Boyette MD, Creamer NG, "Farmer harvest decisions and vegetable loss in primary production. Agricultural Systems", 1;176:102672.

[8] Kumar R, Singh MP, Kumar P, Singh JP, "Crop Selection Method to maximize crop yield rate using a machine learning technique", International conference on smart technologies and management for computing, communication, controls, energy, and materials (ICSTM), (pp. 138-145). IEEE.

[9] Sriram Rakshith.K, Dr.Deepak.G, Rajesh M, Sudharshan K S, Vasanth S, Harish Kumar N, "A Survey on Crop Prediction using Machine Learning Approach", In International Journal for Research in Applied Science & Engineering Technology (IJRASET), pp(3231- 3234)

[10] Veenadhari S, Misra B, Singh CD, "Machine learning approach for forecasting crop yield based on climatic parameters", In International Conference on Computer Communication and Informatics, 2014 Jan 3 (pp. 1-5). IEEE.