International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 11 Nov. 2022, pp: 85-90 www.ijaem.net ISSN: 2395-5252

Crop and fertilization recommendation using machine learning

Pulicherla HariKrishna Reddy
Student, Master of computer applications, Sree Vidyanikethan institute of management, Andhra Pradesh, India

Submitted: 25-10-2022 Accepted: 04-11-2022

ABSTRACT: A Crop and Fertilizer recommendation system can predict suitable crops and Fertilizer based on weather of a particular region and display them on a Graphical user Interface (GUI), which will help the farmers to improve the production of their crops. Due to sudden changes in weather conditions, farmers suffer throughout the country as they fail to produce enough crops. This system overcomes this issue by recommending suitable crops analyzing various parameters like Nitrogen, Potassium, Phosphorous, Temperature, Rainfall, Humidity, (Soil moisture, Soil Type and Crop for Fertilizer Recommendation). The large volume of data is pre-processed in python environment, K-Nearest Neighbor (KNN), Naive Bayes (NB), Random Forest (RF), Decision Tree (DT) classifiers are employed and the best classification is chosen for the crop prediction. It will also suggest fertilizer if the farmer wants the specific crop for the field which contains less Soil Nutrients. It has been a major problem to identify what to grow, any man has adequate space in the owner's land. Not only domestic lands but also for farming lands. Why it has become a problem is that environmental factors such as temperature, water levels, and soil conditions are uncertain as they change from time to time. Due to these problems, this solution of crop recommendation system predicts the user, what crop type would be the most suitable for the selected area by collecting the environmental factors for plant growth and processing them with the trained sub-models of the main of the system.

Keywords: Machine Learning, Agriculture, K-Nearest Neighbor, Naïve Bayes, Random Trees and Decision Tree, Recommendation System

I. INTRODUCTION

In general, agriculture is the backbone of India and plays a major role in Indian economy. More than 70% of the population are involved in the crop production and related activities. But crop production and prediction are getting depleted due to sudden changes in weather conditions which will affect the economy of farmers by producing a poor yield. This system helps the farmers by recommending reasonable crops based on weather conditions using agricultural and meteorological data collected from open dataset sources. This system collects and analyses temperature, rainfall, humidity, nitrogen, phosphorous, potassium, pH to predict the crops and displays them on aGraphical User Interface. Large volume of data is pre-processed in Python environment, a wide range of classifiers, namely K-Nearest Neighbor (KNN), Naïve Bayes (NB), Random Forest (RF), Decision Tree (DT) are employed and the best classifier is chosen for the crop prediction. The recommended crop to sow is displayed on the Graphical User Interface (GUI). If the farmer wants specific crop for the field which contains very less Soil Nutrients and need to use fertilizer, this system will recommend most suitable fertilizer for the crop.

The more inputs and statistical data collected, and higher the algorithmic rule is at predicting the outcomes. And the aim is thatfarmers willuse thesetechnologies to attain their goal of improved harvest by creating better selections within the field. By implementing the system of temperature, soil hydrogen ion concentration and soil wetness detection, the information captured are processed with an explicit algorithmic rule and passed to a centralized database that is connected to different modules of the research, so the main system will predict the most effective crop kind that the farmer should grow to require the most outcome of the crop kind that is farmed in a home garden or the respectable land area.

II. LITERATURE SURVEY

In this research, it is proposing a system where the major factors are taken into consideration at the same time and come up with a

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 11 Nov. 2022, pp: 85-90 www.ijaem.net ISSN: 2395-5252

solution so that the system will not be complicated for the user. As mentioned above in the sentence, the major factors taken into consideration at once is unlike other models proposed in previous researchers, this system considers all the major factors which are essential for plant growth, are processed together using various algorithms whereas the other models consider only parameters at once keeping the other factors constant.

The requirements and planning needed for developing a software model for precision farming is discussed. It deeply analyses the basics of precision farming. The author's start from the basics of precisionfarming and move towards developing a model that would support it. This paper describes a model that applies Precision Agriculture (PA) principles to small, open farms at the individual farmer and crop level, to affect a degree of control over variability. The comprehensive objective of the model is to deliver direct advisory services to even the smallest farmer at the level of his/her smallest plot of crop, using the most accessible technologies such as SMS and email. This model has been designed for the scenario in Kerala State where the average holding size is much lower than most of India. Hence this model can be deployed elsewhere in India only with minor modifications. The papermakes a comparative study of classification algorithms and their performance in yield prediction in precision agriculture. These algorithms are implemented in a data set collected for several years in yield prediction on soya bean crop. The algorithms used for yield prediction in this paper are KNN, Random Forest, Naïve bayes,decision Trees.

As anexample, some tests are carried outto findthe rate evaporation and how the plant growth is affected when there isno sufficient water present. And a derived equation is presented as a result

ETo = K pan × E pan

ETo: reference crop evapotranspiration

Kpan: pan coefficient E pan: pan evaporation

Even though an equation isproposed, there aresome limitations. Mainly this could be done only for a land with a less area. This is notsuitable forcommercialization as the profit will be low when you use less area for cultivation. The second limitation is the average rainfall of Sri Lanka mostly suitable for many crop types to grow without any deficiency. Therefore, the water level alone itself is not a big issuewhereas other factors are. If there isenough waterfor plantsbut no temperature that would be a problem as the major environmental factorshave amutual relationshipwith each other in plant growth. The goal of the previous researches is also to predict the best crop type. But once the farmer or user has cultivated the predicted crop type, the system's job is done. But in the system that proposed from this paper, has a feedback system as well. Even after suggesting the best crop type, the system can track the plant growth and it provides feedback if the farm is malnourished. So that the usercan take necessary precautions prior

The paper [9] tries to solve the problem of food insecurity in Egypt. It proposes a framework which would predict the production, and import for that particular year. It uses Artificial Neural Networks along with multi-layer perceptron in WEKA to build the prediction. At the end of the process, we would be able to visualize the amount of production import, need and availability. Therefore, it would help to make decisions on whether food must be further imported or not. The soil datasets in paper [10] are analysed and a category is predicted. From the predicted soil category, the crop yield is identified as a Classification rule. Naïve Bayes and k-Nearest Neighbor algorithms are used for crop yield prediction. The future work stated is to create efficient models using various classification techniques such as support vector machine, principal component analysis.

III. METHODOLOGY

3.1 Dataset Collection

The open-source dataset (crop.xlsx) containing attributes such as Nitrogen, Phosphorous, Potassium, Temperature, Rainfall, Humidity, and pH of the soil is collected from Kaggle. [16] (https://www.kaggle.com/datasets)

The open-source dataset (fertilizer.xlsx) containing attributes such as Nitrogen, Phosphorous, Potassium, Temperature, Rainfall, moisture, Crop, Soil Type is collected from Kaggle. [16] (https://www.kaggle.com/datasets)

The dataset comprising the soil specific attributes which are collected for Madurai district tested at soil testing lab, Madurai, Tamil Nadu, India. In addition, similar online sources of general crop data were also used. The crops considered in our model include millet, groundnut, pulses, cotton, vegetables, banana, paddy, sorghum, sugarcane, coriander. Figure 1 gives an analysis of the dataset. The number of instances of each crop available in the training dataset is depicted. The attributes considered where Depth, Texture, Ph, Soil Colour, Permeability, Drainage, Water holding and Erosion.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 11 Nov. 2022, pp: 85-90 www.ijaem.net ISSN: 2395-5252

Soil is the anchor of the roots. The water holding capacity determines the crop's ability to absorb nutrients and other nutrients that are changed into ions, which is the form that the plant can use. Texture determines how porous the soil is and the comfort of air and water movement which is essential to prevent the plants from becoming waterlogged. Soil texture which affects the soil's ability to hold onto nutrients. The level of acidity or alkalinity (Ph) is a master variable which affects the availability of soil nutrients. The activity of microorganisms presents in the soil and the level of exchangeable aluminium can be affected by PH. The water holding and drainage determine the penetration of roots. Hence for the following reasons the above stated parameters are considered for choosing a crop.

3.2 Machine learning algorithms:

3.2.1 RANDOM TREE:

Random tree [11] is like that of a decision tree. But it differs from random tree in a way that for each split only a random subset of attributes is available. Random trees can be built for both nominal and numerical data. The Random Tree is like C4.5 or CART but it varies in the fact that before it is applied for training it selects only a random subset of attributes. At each node it considers K randomly chosen attributes. The subset ratio parameter specifies the size of the subset.

3.2.2 K-NEAREST NEIGHBOR:

K-Nearest Neighbor [15] can be used for both classification and regression. K-Nearest Neighbors is a non-complex algorithm which stores all the available cases and classifies new cases based on some similarity measure. The sample set is classified based upon the “closeness” that is the distance measure such as Euclidean distance or Manhattan distance.

3.2.3 NAÏVE BAYES:

Naive Bayes [14] classifier is a simple probabilistic classifier which works based on applying Bayes' theorem (from Bayesian statistics) with strong naive independence assumptions. Naive Bayes is a technique for constructing classifier models which assign class labels to problem instances which are represented as vectors of feature values, where the class labels are drawn from some finite set. It is not just a single algorithm for training such classifiers, but a family of algorithms based on a common principle. All naive Bayes classifiers assumes that the value of a particular feature is independent of the value of any other feature, given the class variable.

3.2.4 DECISION TREES

A decision tree is drawn upside down with its root at the top. In the image on the left, the bold text in black represents a condition/internal node, based on which the tree splits into branches/ edges. The end of the branch that doesn’t split anymore is the decision/leaf, in this case, whether the passenger died or survived, represented as red and green text respectively.

Although, a real dataset will have a lot more features and this will just be a branch in a much bigger tree, but you cannot ignore the simplicity of this algorithm. The feature importance is clear and relations can be viewed easily. This methodology is more commonly known as learning decision tree from data and above tree is called Classification tree as the target is to classify passenger as survived or died. Regression trees are represented in the same manner, just they predict continuous values like price of a house. In general, Decision Tree algorithms are referred to as CART or Classification and Regression Trees.

3.3 CROP PREDICTION

Since the environmental conditions differ from region to region, a machine learning model is used to predict the best crop type for the selected land. To train the crop recommending model with the data collected from the Arduino sensors, machine learning algorithms are used toidentify the best crop to cultivate with the highest probability of growing. Naïve Bayes & Support vector machine algorithms are used to select the best crop type. From this model, it decided what type of crops that the farmer should grow. This is done by analyzing factors of humidity, temperature, soil moisture, pH level, and sunlight. Mainly the system suggests crop types by analyzing the above-mentioned factors using two machine learning algorithms. Naïve Bayes - Naive Bayes is a technique for constructing classifier models that assign class labels to problem instances which are represented as vectors of feature values, where the class labels are drawn from some finite set.

3.4 MONITORING AND FEEDBACK

This proposed system product would mainly identify four types of crops according to the environmental factors of the selected plot of land. But the reason to obtain a probability of more than 90% for the above-identified crops would be the soil condition or any other changes in the selected land. But to avoid these factors affecting the crop prediction, the farmer’s feedback system is included in the system. Once the crop type is

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 11 Nov. 2022, pp: 85-90 www.ijaem.net ISSN: 2395-5252

recommended, the farmer is asked for details and feedbacks regularly through the mobile application to guide the farmer with necessary precautions. To provide the necessary feedback, the feedback system is used in the mobile application by selecting the crop type. By this, the overall accuracy and the reliability of the product are increased with time.

Fig Proposed Methodology

This paper presents spatial information mining techniques particularly decision tree algorithm applying to agriculture arrive reviewing. The clue is to pool spatial information mining/decision tree strategies with master framework strategies and apply them to shape savvy agriculture arrive reviewing data framework.

Data mining is the practice of monitoring and amplifying purposeful information from the data. Data mining explores its applicability in diverse fields like banking, retail, medicine, farming, etc. In agriculture, it is used for analysing the numerous biotic and abiotic constituents. Horticulture in India represents a significant role in the market and employment. The common dilemma breathing among the Indian farmers is they do not select the right crop based on their soil specifications. Due to this, they suffer a grave hindrance to fertility. This difficulty of the producers has been notified about through precision agriculture. Precision agriculture is a modernized farming practice that utilizes research data of soil properties, soil varieties, crop yield data gathering and introduces the farmers to the best crop based on their site-specific parameters. This diminishes the wrong judgment on a crop and enhanced fertility.

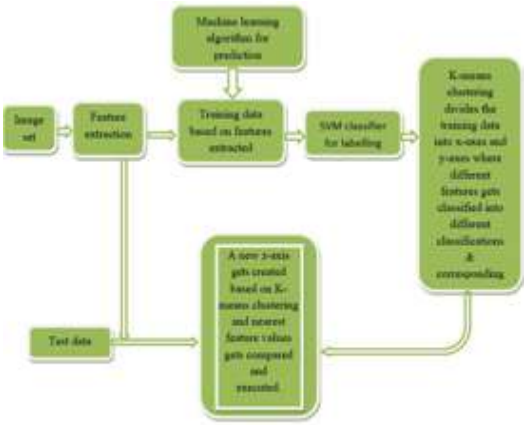

Fig: Image of overall system working

IV. RESULT

Extensive work has been done, and many ML algorithms have been applied in the agriculture sector. The biggest challenge in agriculture is to increase farm production and offer it to the end-user with the best possible price and quality. It is also observed that at least 50% of the farm produce gets wasted, and it never reaches the end-user. The proposed model suggests the methods for minimizing farm produce wastage

The proposed system was successfully implemented in a selected land. The data sampled from the sensors at an interval of one hour on a normal sunny day is presented in Table I. Based on the tests conducted and the data collected, it can suggest the best crop to cultivate for maximum harvest. Using the feedbacks collected from the farmer the accuracy of the predictions is sharpened by neglecting the invalid data. As an example, if the farmers are continuously providing the negative feedback for cultivating strawberry in Galle, the system itself learned, and the accuracy of the final output is increased which means in future it will not suggest strawberry to grow in Galle. The farmer can enter his own feedback using his native language.

According to the prevailing environmental factorsin the selected land, the best suitable four crop types are suggested to the farmer.

Fig: sample output given to the farmers

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 11 Nov. 2022, pp: 85-90 www.ijaem.net ISSN: 2395-5252

Figure shows a sample result where the farmer gets, once the environmental factors are entered. PLU code is an attribute that is used to identify the crop uniquely. The overall accuracy of the proposed system is more than 92%. As the farmer continue to use this, the more data system feeds, the more accurate it will be. Therefore. by the time of long usage, the farmer can obtain more the 95% of accuracy from the whole system.
Based on the result analysis of the algorithm vice accuracy result in percentage. Where KNN algorithm has the highest accuracy [99.31%] and the Decision trees has the lowest accuracy [90%].

Table 1
Algorithm wise accuracy result in percentage

Algorithm	Accuracy
Decision trees	90%
Naïve bayes	99%
KNN	99.31%
Random forest	99

V. CONCLUSION

India is a nation in which agriculture plays a major role. The Crop and Fertilizer recommendation system would help farmers in sowing the right crop based on the weather conditions of the region. It will also suggest the most suitable fertilizer for a specific crop if the field contains fewer soil nutrients. The crop recommendation system aims at analysing temperature, rainfall, pH, Nitrogen, Phosphorus, Potassium content in the soil, humidity, soil moisture, and moisture (for fertilizer) and recommends a reasonable crop or fertilizer. Large volumes of data are pre-processed in a Python environment, and a wide range of classifiers, namely K-Nearest Neighbour (KNN), Naïve Bayes, Random Forest (RF), and Decision Tree are employed on the data and the results are displayed on a Graphical user Interface. The system design is scalable and can be used to predict crops for different regions. Thus, the farmers can sow the right crop based on weather conditions, increasing their yield and the overall productivity of the nation. It also suggests fertilizer if the farmer wants a specific crop for the field which contains fewer soil nutrients.

This application focused on how Raw Data in the field and past data set are used to foretell the pattern in plant diseases by implementing the Naive Bayes algorithm. Climate plays an important role in the development or production of various troubles. Pests cause major harm in damaging the crops or by transporting the bacteria or virus to the crops, because of which there is a fall in the production of crops at the time of harvest. This produces huge damage to the farmer's community. Each season, pests spoil the crops.
To overcome this obstacle, a weather-based forewarning pest prediction model is proposed. The recommended model uses the Multiple Regression model plus Generalized Linear model procedures.

VI. FUTURE SCOPE

We are thinking to modify the project from a prototype to a comprehensive end to the end-user product. This can be achieved using the TensorFlow library within Python IDE with high processors. The finished product would be perfectly predicting disease/pest attacks simultaneously with recognizing them. A larger set of data would be provided for the training network.

The entire algorithm would be created using the TensorFlow library for better processing. OpenCV is utilized for Image analytics like Image Processing Toolbox in MATLAB.
Accordingly, the farmer must just take a snap of the leaf, upload it to the cloud where the back-end processing will do analysis, and provide remedial measures for restricting and excluding external hosts.

REFERENCES

- Priyadharshini A, Swapneel Chakraborty, Aayush Kumar and Omen Rajendra Pooniwala, “Intelligent crop recommendation system using machine learning”, 2021 5th International Conference on Computing Methodologies and Communication (ICCMC) (2021)
- Rahul Katarya, Ashutosh Raturi, Abhinav Mehndiratta and Abhinav Thapper, “Impact of Machine Learning Techniques in Precision Agriculture”, 2020 3rd International Conference on Emerging Technologies in Computer Engineering: Machine Learning and Internet of Things (ICETCE) (2020)
- Avinash Kumar, Sobhangi Sarkar and Chittaranjan Pradhan, “Recommendation System for Crop Identification and Pest Control Technique in Agriculture”, 2019 International Conference on Communication and Signal Processing (ICCSP) (2019)
- Rashmi Priya, Dharavath Ramesh and Ekaansh Khosla, “Crop Prediction on the Region Belts of India: A Naïve Bayes MapReduce Precision Agricultural Model”, 2018 International Conference on Advances in Computing, Communications, and Informatics (ICACCI)

- Niketa Gandhi, OwaizPetkar, and Leisa J. Armstrong, "Rice Crop Yield Prediction Using Artificial Neural Networks", 2016 IEEE Technological Innovations in ICT for Agriculture and Rural Development (TIAR) (2016)
- S. Fan, and C. Hanrong, "BP network prediction model for occurrence trend of rice gall midge," in Computer Simulation, vol. 17, no. 3, pp. 61- 63, 2000.
- C. Guanhao, Z. Yiu-chung, and Y. Guofeng, "Prediction model of occurrence degree of main damage generation of late rice gall midge," in Journal of Biosafety, vol. 15, no. 2, pp. 137-139, 2006.
- W. Xuan, L. Jiake, and X. Deti, "Application of BP neural network integrated with GASA hybrid learning strategy in rice pest prediction," in China Agronomy Bulletin, vol. 23, no. 4, pp. 404-409, 2007.
- X. Zhenxiang, "Improvement of genetic algorithm and BP Model and its Application in Water Resources Engineering," Northeast Agricultural University, 2004.
- Z. Lei, H. Zhiguo, and W. Li, "Effects of Climate change on crop pests in China," in Journal of Ecology.
- W. Congcong, "Application Research of pest prediction based on rough set and artificial neural network," Jilin University, 2011.
- G. Ning, "Prediction and prediction of crop insect condition based on BP neural network and its MATLAB realization," Anhui Agricultural University, 2003.
- X. Xiao, "Research on Railway Passenger Volume Prediction based on genetic algorithm and BP Neural Network," Southwest Jiaotong University, 2010.
- S. Wenqi. "Research on Regional Highway Traffic forecasting based on genetic algorithm and BP Neural Network," Southwest Jiaotong University, 2015.
- C. Yugui, Dawn, and L. Mingyu, "Middle and long-term power load forecasting and analysis based on genetic algorithm and BP neural network," in Computer applications, vol. 30, no. 1, pp. 224-226, 2010.