RECOMMENDATION OF CROP AND PESTICIDES USING MACHINE LEARNING

J. Ignashya preetha[1], N. Priyadharshini[2], P. Mageshwari[3], S. Rakshana[4], Dr. S. Jeyalakshmi[5]

1,2,3,4,5 Department of IT, SRM Valliammai Engineering College, Tamil Nadu 603203

--***---

Abstract – *The major resource for improving the economy of India is agriculture. From past farmers followed ancestral faming pattern and regularities within it. A single farmer cannot take action upon improving the crop yield of a nation and does not have enough potential to maximize the crop yield by adopting technical norms within plant growth and improving the yield in a large quantity. Severe change in climatic condition and several other pesticides attack cause shorting of crop yield and also led to food shortage. A simple misguided decision in farming can affect a farmer severe. In recent, there is lot of techniques applied by researchers and those techniques are available to raise the quantity of yield. This in turn changed traditional farming approach and introduced precision farming. Recently data mining performs vital role in identifying plant disease and providing solution prescribing pesticides to plant disease. But this study extends the application of data mining in agriculture to a greater extent. The cultivation of precious crop at right time is the major issues faced by farmer. This study proposes machine learning (ML) approach to resolve it and makes the farmer to choose right crop based on the nutrition content and quality of soil. The machine learning algorithms chosen for this study are Random forest, decision tree and K-nearest neighboring. Some of the factors mainly considered for recommendation of plant are humidity, rainfall, pH value, soil moisture. The recommended technique makes farmer to take decision on improving the crop yield; recommending crops as per climatic condition and quality of land.*

Key Words: Agriculture, Crop Recommendation, Machine Learning (ML), Random Forest, K-nearest Neighboring (KNN), Decision Tree

1. INTRODUCTION

Agriculture is said to be the backbone of Indian economy and it utilizing 60% of nation land to fulfil the food needs around 1.2 million people. Farmer doesn't have conquered knowledge about severe climatic changes and the soil moisture content. Mostly famers are difficult to understand those two factors. This in turn led to decrease in expected level of productivity. The selection of pesticide, usage of water and maintaining of it will make the crop growth even stronger. Every crop has special climatic factors. By precision farming technique, those factors are handled as per the crop planted. Precision farming not only focuses on productivity but also raising the yield rate of crop. To make

agriculture as a profitable business for farmers and satisfies the need of a nation, different kind of agricultural practices are carried out. In developing nation like India, sustainable agriculture is practised to manage the necessity of food. A lot of techniques were carried out to minimize the shortening of crop yield; but traditional agriculture having its own demerits. The demerits are further limited through precision farming. Other than that, some other factors affecting the yield of a plant are bacterial, fungal and viral diseases. The detailed explanation of various plant diseases occurring repeatedly in farms are given below:

Anthracnose: Mostly fungus is observed in genus collectotrichum and other regions; lesions occur on stem. The major reason for this disease is rotted waste and certain other wastages around it. During winter, the plants are affected by this disease and it is transmitted to nearest plant through watering and pollination. The dead tissues appear as anthracnose.

Bacterial blight: Lesions are converted into dead spots in this; later elongated lesions are appeared as like linear streaks and it is turned into milky green colour. As like anthracnose, this disease will affect in winter season and transmission through insects and water.

Alternaria alternate: A fungal disease found in different kinds of plants and the symptoms are observed as blights, leaf spot and rots. The spores in a leaf are created by conidia. Rainfall and humidity are the comfort zone for this disease.

Cercospora leaf spot: This does not have sexual stage and its genus is mycospharella.

Later data mining and ML techniques are used by researchers to bring revolution in the field of traditional agriculture to maximize the productivity by considering the necessity. ML can gain expertise without doing additional programming in a machine, so it maximizes the accomplishment of machine by differentiating and depict the consistency and format of drive data. In this research, combination of three different algorithms such as Random forest, KNN, decision tree algorithms were used to suggest crop, fertilizer and pesticides. As per the land condition, the proposed study will recommend crops and several other essentialities. This type of recommendation is carried out with the consideration of water level, moisture content, pH, temperature.

From a selected region, the data are collected from soil testing laboratory as well as from data.World. A wide variety of data is chosen for this analysis and has been processed through suggested ML approach. Following this crops are suggested and required fertilizer is chosen; disease is identified and pesticides are selected. The aim of the proposed study is to plant crops at right time and maximize the yield.

2. RELATED WORKS

K.Venkataramana et.al [1], proposed a Ml approach such as id3 algorithm to improving yield of tomato. The factors such as csv, moisture level, temperature are considered in selected dataset and the examination is done in Php platform.

R. Sujatha et.al, developed a data mining approach to improvise the yield of crop in a large cultivable land. The parameters mainly considered for this study are name of crop, type of soil, weather condition, seedling selection, level of water, pH value. The commonly affecting disease and the disease which are ease to affect are considered secondary during this process [2].

A. K. Tripathy et.al in [3], recommends Support Vector Mechanism (SVM) classifier algorithm to find out suitable crop for selected cultivable land by considering location, air moisture level, seed varieties. The dataset is prepared through Weka tool which makes pair of rules on present dataset. Through python, the entire process is done.

S. Veenadhari, B. Misra and C. Singh, created a website to know climatic condition of an area and yield of crop by c4.5 algorithm known as Crop Advisor. Dependent on c4.5 algorithm, decision tree and rule have been built. This explains how the crop is affected by the change in climatic condition [4].

Jun Wu et.al, suggested varieties of crops which is able to accept and grow under variable climatic condition. For this approach, decision tree classifier algorithm is utilized and it utilizes new factors which was helpful to improvise the yield of crops. 10-fold cross validation method is used to check dataset, horse-colic and soyabean dataset [5].

Murali Krishna et.al, defines interfacing of data mining technique with humidity and pesticide attack on plants. This explains the difficulties faced by farmers and problems in interfacing data mining technique with agriculture. Pesticide attack prevention is done by recommending pesticides [6].

Verheyen et.al, in [7], described about statistical mining approach to review the characteristics of soil. The K-means clustering classifier approach is utilized to classifying soil type and this system combines with GPS to select a region and does categorization.

Radhike et.al in [8] suggested smart farming approach by utilizing sensing devices. To calculate the intensity of light, yhe proposed technique utilizes pH, temperature and moisture sensor. The sensors are placed to sense the environmental condition and it collects the information of environment which is delivered towards processing without any change within it. This is suited for all kind of crop recommendation as per the environmental factors.

Kumar et.al, proposed an IoT model to guide farmers to select suitable crop and fertilizers. This system uses pH sensor and soil moisture sensor which is connected with radio frequency system to transmit data. The received information is processed by a ML approach so called decision tree. The necessary information such as selection of fertilizer, time of providing fertilizers to crops and cultivable stage is send to farmer. Through android application, the farmer will receive guidelines.

Suhas Athani et.al proposed neural network to process the data delivered by IoT based system. The neural network helps to find the condition of cultivable land in association with sensing devices and the farmer is guided through android application [10].

3. METHODOLOGY

The purpose of this proposed system is to go over the best practices for predicting crop area and yield in mixed and continuous cropping systems. List and area frames are both taken into consideration. The sampling frame for crop yield estimation will be determined by the sample selected for area estimation. It is briefly covered the sampling procedure used to select the sample for crop area and yield calculation. To merge the subjective and objective techniques, a double sampling regression estimator is used. The domain estimation technique and a double sampling regression estimator are also used to estimate crop area and yield. The notion of domain estimation enables the estimation of crop area and yield for a variety of crops and blends from a single sample. The criterion for estimating sample size is also included.
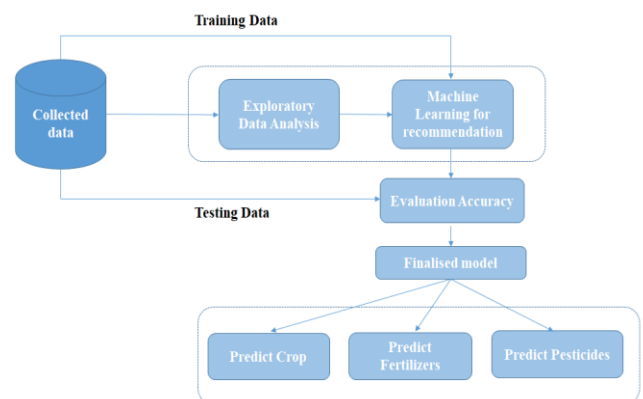

Fig -1: Architecture diagram

The detailed explanation of each module is explained bellow.

Collection of dataset: The dataset is made up of soil-specific properties that are gathered and evaluated in a soil testing lab. Moreover, similar web sources of general crop data were utilised. Rice, maize, chickpea, kidneybeans, pigeonpeas, mothbeans, mungbean, blackgram, lentil, and other crops are included in our model. In the training dataset, the number of instances of each crop is shown. Ph, temperature, rainfall, and humidity were all factors taken into account.

Data cleaning: Cleaning data is an essential part of every machine learning research. Data cleaning is performed in this module to prepare data for analysis by removing or changing data that is erroneous, incomplete, redundant, or badly formatted. In tabular data, you can investigate your data using a variety of statistical analysis and data visualisation approaches to find data cleaning activities you might wish to do.
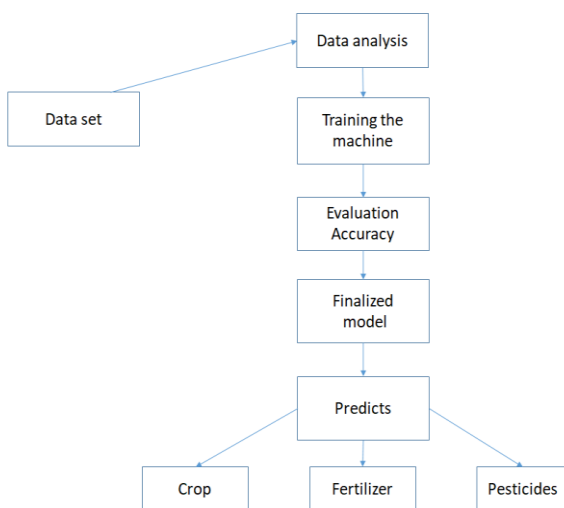

Fig -2: Flow analysis of proposed study

Training and testing of datasets: Separating the attributes and assigning the variables as X and Y as train and test data based on the dependent and independent variables.

Feature extraction: This is done to decrease the count of attributes in the dataset, resulting in benefits such as faster training and improved accuracy.

The feature is the major source for computational method of solving problems and tasks. Some image structure like as points, edges are characteristic. The extraction of features dataset is utilized for classification. Every character in the feature extraction is denoted by an identification feature vector. The main objective of this stage is to extract a collection of features that improvise prediction rate with few element count and create same feature for various instance of the same symbol.

Classifier Algorithms:

The most popular ML approach used here is Random forest also called as ensemble learning which a supervised learning technique is. It is mainly used as classifier and solving complex problems. It utilizes multiple numbers of classifiers to perform classification and maximize the performance in an effective manner. This technique have decision tree in a large count. To raise the accuracy of the model this method is suited. The prediction process is done with the usage of each tree without depending on a decision tree. Depending on the majority votes of prediction, this algorithm finalizes the output. With high dimensionality this method process large volume of dataset. Also it avoids over fitting issue to maximize the accuracy as much as possible.

KNN is a data mining technique. It takes every characteristic in training set as various dimensions in some space, and take the value an observation has for this characteristic to be its coordinate in that dimension, so getting a set of points in space. We can then consider the similarity of two points to be the distance among them in this space under some appropriate metric. The way in which the algorithm decides which of the points from the training set are similar enough to be considered when choosing the class to predict for a new observation is to pick the k closest data points to the new observation, and to take the most common class among these.

Decision Tree (DT) is a foreboding representation which functions by testing states at each stage of tree and tends to reach end tree in between that multiple decisions are recorded. The state depends on the application and the output is form in the form of decision. This algorithm calculates information gain of essential attributes such as area, vapor pressure, yield and cloud cover. The two different much needed attribute such as area and yield is in connection to make a decision and then extended.

4. RESULT AND DISCUSSION

The Random forest technique attained greater efficiency and it is visualized in chart 1. The decision tree attained 78% accuracy, KNN have 83% and finally Random forest attained 97%. The experimental analysis is more effective and can suggest right crop at right time to raise the yield.
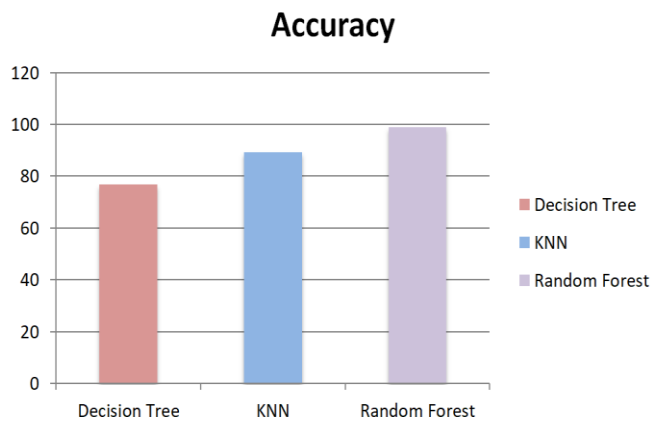
Chart -1: Accuracy Comparison

To recommend fertilizer the attributes considered are humidity, moisture level, soil type and finally crop to be grown. Fertilizer recommendation follows loading of external fertilizer datasets. The values of attributes are chosen either by sensors or human records.

5. CONCLUSIONS

In developing countries like India agriculture plays a vital role. Farmer and the country economy will rise when production of agricultural product is at larger proportion. The proposed system examines the quality of land, recommend pesticides and fertilizers. By this proposed study, farmer can planting suitable crops and can get more yields. This in turn led to improving the economy and there is no doubt within it. The experimental analysis evident that Random forest based crop selection would be a right choice.

In future, the proposed study led to a greater extends by considering additional features from dataset and double the productivity.

REFERENCES

[1] CH. Vishnu Vardhanchowdary, Dr.K.Venkataramana, "Tomato Crop Yield Prediction using ID3", March 2018,IJIRT Volume 4 Issue 10 pp,663-62.

[2] R. Sujatha and P. Isakki, "A study on crop yield forecasting using classification techniques" 2016 International Conference on Computing Technologies and Intelligent Data Engineering (ICCTIDE'16), Kovilpatti, 2016, pp. 1-4.

[3] N. Gandhi, L. J. Armstrong, O. Petkar and A. K. Tripathy, "Rice crop yield prediction in India using support vector machines" 2016 13th International Joint Conference on Computer Science and Software Engineering (JCSSE), KhonKaen, 2016, pp. 1-5.

[4] S. Veenadhari, B. Misra and C. Singh, "Machine learning approach for forecasting crop yield based on climatic parameters" 2014 International Conference on Computer Communication and Informatics, Coimbatore, 2014, pp. 1-5..

[5] Jun Wu, AnastasiyaOlesnikova, Chi- Hwa Song, Won Don Lee (2009), "The Development and Application of Decision Tree for Agriculture Data" IITSI, pp 16-20

[6] KiranMai,C., Murali Krishna, I.V, an A.VenugopalReddy, "Data Mining of Geospatial Database for Agriculture Related Application", Proceedings of Map India,New Delhi, 2006,pp 83-96.

[7] Verheyen, K., Adrianens, M. Hermy and S.Deckers (2001),"High resolution continuous soil classification using morphological soil profile descriptions"Geoderma, 101:31-48.

[8] Radhika, Y., & Shashi, M. (2009). Atmospheric Temperature Prediction using Support Vector Machines. International Journal of Computer Theory and Engineering, 55–58.

[9] Kumar, R., Singh, M. P., Kumar, P., & Singh, J. P. (2015). Crop Selection Method to maximize crop yield rate using machine learning technique. IEEE Xplore.

[10] Athani, S., Tejeshwar, C. H., Patil, M. M., Patil, P., & Kulkarni, R. (2017, February 1). Soil moisture monitoring using IoT enabled arduino sensors with neural networks for improving soil management for farmers and predict seasonal rainfall for planning future harvest in North Karnataka — India