2021 5th International Conference on Information Systems and Computer Networks (ISCON)
GLA University, Mathura, India. Oct 22-23, 2021

Crop Recommendation Using Machine Learning Algorithm

*Daneshwari Modi
**Department of ECE,*
Basaveshwar Engineering College,
Bagalkot-587103, Karnataka, India
*daneshwarimodi@gmail.com

*Ashok V. Sutagundar
**Department of ECE,*
Basaveshwar Engineering College,
Bagalkot-587103, Karnataka, India
*sutagundar@gmail.com

**Vijayalaxmi Yalavigi
***Department of Information Science and Engineering,*
Cambridge Institute of Technology,
Bengaluru - 560036
**vijayalaxmiyalavigi@gmail.com

*Anupama Aravatagimath
**Department of ECE,*
Basaveshwar Engineering College,
Bagalkot-587103, Karnataka, India
*anupamaaravatagimath@gmail.com

Abstract - **Agriculture is extremely important to India's economy and employment. The most common issue faced by Indian farmers is that farmers do not select the appropriate crop for their soil. As a result, productivity is harmed. Agriculture is the main source of income and the backbone of our economy. The poor crop selection has reduced crop production and food shortages across the country which resulted in an increase in farmer suicide. Farmers' problems have been handled by recommendation of suitable crop before sowing. To overcome these issues it is necessary to analyze the soil parameters. This proposed work presents the SVM algorithm based crop recommendation system for the formers. In this work, it is necessary to analyze the profit of the particular crop, which eliminates the loss for the farmers and increase the productivity. SVM algorithm is used for classification to classify the different parameters of the soil and predict the most suitable crop.The proposed algorithm is simulated in anaconda navigator to analyze the soil parameters and recommend a suitable crop. The SVM algorithm is considered for classification. To test the effectiveness of the proposed algorithm accuracy and confusion matrix are computed.**

Keywords: Agriculture, Soil parameters, SVM algorithm, classification, Confusion Matrix.

I. INTRODUCTION

Agriculture is extremely important to India's economy and employment. Agriculture has changed dramatically in recent years as a result of globalization. India is one of the world's largest agricultural producers, but farm productivity remains low [1]. Many farmers are depending on agriculture for their livelihood. The most common problem faced by Indian farmers is that they do not select the appropriate crop for their soil. As a result, productivity is harmed. Farmers need to increase productivity so that they may obtain more money from the same amount of land with less labor. To reclaim one's health, many innovative technologies have been developed [2]. If there is increase in any crop price then most of the farmers are interested in cultivating that crop in their land because of increase in price so that they could also make more profits from the land [3]. When it comes to deciding which crop to sow in a given season, the majority of Indian farmers rely on their instincts. They seek peace in just keeping to traditional farming procedures and regulations, despite the reality that crop yields are strongly influenced by current weather and soil conditions [7].The main objective of the proposed work is to design a recommendation system for accurate crop selection, and to improve crop productivity which reduces the wrong choice on a crop. Our contributions in this work are classification, training the model, crop prediction and evaluation. Some of the related works on crop recommendation using algorithms are as follows. In paper [1], authors have considered many crops that are evaluated in the model for prediction. Depth, pH, erosion, permeability, texture, drainage, dater holding, and soil color were all factors taken into consideration for predicting the crop. Ensemble was the strategy used, which merged the power of two or more different models for better prediction. The Majority Voting Technique was utilized to assemble the group. The crops are examined and graded based on the results of the examination in order to estimate crop yields. Different data mining techniques have discovered this category. The several grouping rules are considered including K-Nearest Neighbor and Naive Bayes [2]. The dataset is collected from the lab.Ensemble model consisting of learner models (Random tree, SVM, ANN) are used. The crop is predicted by majority of the model and is voted by MVT. This crop is compared with the testing data and decides the crop to be recommended [3]. The system considers soil N, P, K, and pH values to determine the most productive crops that can be cultivated under such conditions. The approach assists the farmer in determining which crop to cultivate in their area by listing all available crops. This technique thus assists the farmer in determining the most profitable crop as well as identifying new crops that can be cultivated that the farmer has not before cultivated [4].The system is developed for smart farming. The task is inexpensive and simple to develop, resulting in better productivity. Farmers benefit from the crop's productivity. This mechanism works as follows: Random Forest is used to implement classification. Bayesian Network for statistical analysis and for pattern comparison, an artificial neural network is used [5]. The geographical location of a farm, soil qualities, and weather factors are all taken into consideration. Two subsystems are considered: Crop suitability predictor and the rainfall predictor.

The crop suitability predictor estimates the crop, whereas the rainfall forecaster estimates the next 12 months' rainfall. A number of environmental and geographical factors will influence decisions is presented [6]. The system is based on the creation of an application. The location is taken from Google by the system. If the user's location is identified, he or she would enter all of the parameters' details. The algorithm is run by the app, which further includes a summary of crops. If the user's location is not identified, he or she might manually enter it. The data is obtained from a csv file recorded in the app. The agricultural productivity is estimated based on the location factor is shown in [7]. In this paper [8], authors suggested a system that considers soil N, P, K, and pH values to determine the most productive crops that can be cultivated under such conditions. The approach assists the farmer in determining which crop to cultivate in their area by listing all available crops. This technique thus assists the farmer in determining the most profitable crop as well as identifying new crops that can be cultivated that the farmer has not before cultivated. The Random Forest technique for forecasting agricultural yields with high accuracy is used. The Random Forest method produces the most crop yield models with the least amount of effort. It's ideal for predicting large amounts of crop yield in agricultural planning. As a result, farmers are more likely to make the best decisions is presented in [9]. The proposed system combines data from archives and climate offices to produce an expectation of the most plausible yields based on current natural conditions using machine learning calculations. This provides an agriculturist with a variety of harvest options from which to choose is shown in [10]. The primary goal of this research is to identify various types of leaf diseases. To improve classification accuracy, various feature extraction techniques were applied. SVM (Support Vector Machine), Random Forest, and Logistic Regression were used [11]. The objective of this study is to process the soil images to generate a digital soil classification system for rural farmers at low cost. Soil texture is the main factor to be considered before doing cultivation and the proposed method gives an average of 91.37% accuracy for all the soil samples [12]. The organization of this paper is as follows. Section 2 discusses the proposed model. In section 3 results are discussed. Section 4 concludes the paper.

II. PROPOSED SYSTEM

The designed system includes different steps such as Data collection, Data visualization, Data preprocessing, classification and evaluation as shown in figure.1.

A. Data Collection

This dataset contains soil metrics such as N, P, K, humidity, rainfall, PH and temperature for several types of soil as well as the crops that can be cultivated in those soils as labels. Let us consider $C_0, C_1, C_2 \ldots\ldots\ldots C_n$, as the 'n' number of crops which depends different soil parameters. The fig 2 shows the sample dataset. Dataset consists of 2200 rows and 8 columns. The last column is label that defines crop.

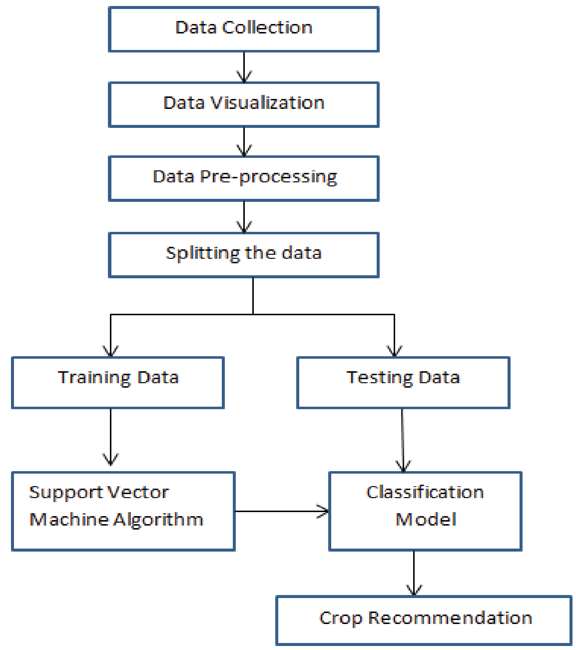

Fig. 1. Proposed System

	N	P	K	temperature	humidity	ph	rainfall	label
0	90	42	43	20.879744	82.002744	6.502985	202.935536	rice
1	85	58	41	21.770462	80.319644	7.038096	226.655537	rice
2	60	55	44	23.004459	82.320763	7.840207	263.964248	rice
3	74	35	40	26.491096	80.158363	6.980401	242.864034	rice
4	78	42	42	20.130175	81.604873	7.628473	262.717340	rice
...	...	...	...	...	...	...	...	...
2195	107	34	32	26.774637	66.413269	6.780064	177.774507	coffee
2196	99	15	27	27.417112	56.636362	6.086922	127.924610	coffee
2197	118	33	30	24.131797	67.225123	6.362608	173.322839	coffee
2198	117	32	34	26.272418	52.127394	6.758793	127.175293	coffee
2199	104	18	30	23.603016	60.396475	6.779833	140.937041	coffee

2200 rows × 8 columns

Fig. 2. Dataset

B. Data Visualization

One of the very important steps in machine learning is data visualization. In this process, data is checked for missing values, outliers, null values etc. It also helps to understand the data by using the various graphical tools such as bar charts, scatter plots, line plots etc. we used scatter plots to know the relationship between the two variables.

Dots in the plot represent the data points. The fig 3 shows the sample plot between K value and N values of the data. The graph clearly shows that there is clear boundary between the two columns, since there is no overlapping of the points. Similarly, we can use to visualize the other columns.

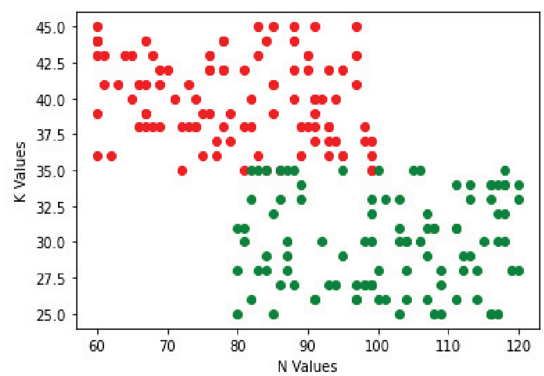

Fig. 3. Scatter Plot

C. Data Preprocessing

In this process, the data is prepared to feed to the machine learning algorithm. The columns that do not contribute for prediction are deleted such as row ids. All the machine learning algorithms operate on the numeric data. Data contains non-numeric values such as label that has to be converted to numeric values. One of process is called Label encoding. In label encoding, each class is assigned a numeric values so that algorithm can operate upon. In our dataset label is a non-numeric column, fit transform method of the pandas library is used to convert column with their respective encoded value. For example crop C_0 encoded value is 0 and crop C_1 encoded value is 1 and similarly for other crops. As there are 18 crops C_n encoded value are 17 and these respective crops encoded values are returned to the respective column in the data set. After encoding data is shuffled so that training data should have all the crops.

D. Split the Dataset

In this process, the pre-processed data is split into training set and test set. 80:20 ratios is used. The training set consist of 1726 rows x 8 columns and test dataset consists of 440 rows x 7 columns.

E. Support Vector Machine

Support vector machines are supervised learning algorithm that examines data for classification and regression analysis.

It is good for small datasets and used for multi-class classification problems. The idea behind the SVM is the optimization of the hyper planes that separates the different classes in n-dimensional space as shown in fig 4.

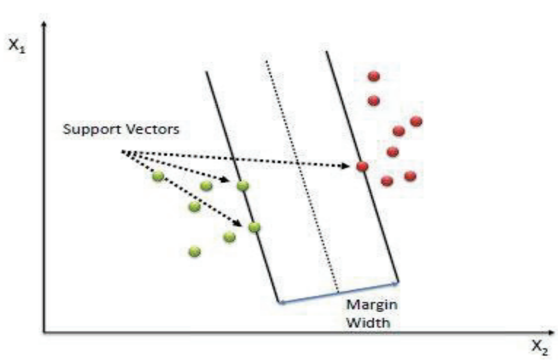

Fig. 4. Support Vector Machine

The points closest to the hyper plane are called as the ***support vector points*** and the distance of the vectors from the hyper plane are called the ***margins,*** *as shown in fig 4.* In 2-D space, hyper plane is a line that separates the categories. In 3-D space, it is a hyper plane. There can be many hyper planes that separate the data points, but the plane that maximizes the margin is chosen by the algorithm.

F. Algorithm of Proposed Work

The stages involved in SVM algorithm are as follows:

- **Step 1**: Initially training dataset is labeled
- **Step 2:** Classify the labels based on input parameters
- **Step 3:** The crops are predicted based on classification
- **Step 4:** The model is tested using testing data
- **Step 5:** Model is evaluated using the performance metrics such as accuracy, precision, recall, F1 score.

These steps are followed to recommend a suitable crop based on soil parameters that are given to the trained model. **The classification accuracy is the ratio of positive predictions to all predictions. The accuracy obtained for this work is 98.00%. The evaluation is done using confusion matrix diagonal matrix.**

III. SIMULATION

The simulation of the proposed system is done using Anaconda Navigator (Jupyter Notebook).**This section describes the simulation process, performance parameters and result analysis.**

A. Simulation Process

This section describes the process of simulation of the work.

Begin

- Importing dataset using pandas library
- Encoding the dataset using the scikit learn library
- Separate the dataset into training and testing data
- Apply SVM algorithm to training data for prediction
- Evaluate the model.

End

B. Performance Parameters

The performance parameters are given below:

- **Accuracy:** It is the ratio of number of positive predictions to all predictions.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

- **True Positive (TP) Rate**: It is also called as recall. It is used to measure the percentage of actual positives which are correctly identified.

$$TPR = \frac{TP}{TP+FN} \quad (2)$$

- **Precision:** It's the proportion of relevant cases identified among the retrieved ones. It's also known as a positive predictive model.

$$Precision\ (P) = \frac{TP}{TP+FP} \qquad (3)$$

- **F1-Score**: It's a metric for how accurate a model is on a given dataset. It's a way to integrate the model's precision and recall

$$F1 - Score = \frac{2*TPR*P}{TPR+P} \qquad (4)$$

C. *Result Analysis*

The Confusion Matrix is used for evaluation. The results are shown in the diagonal manner. The trained model and the testing data are considered for the confusion matrix. The Confusion Matrix is used for evaluation. The results are shown in the diagonal manner. The trained model and the testing data are considered for the confusion matrix.

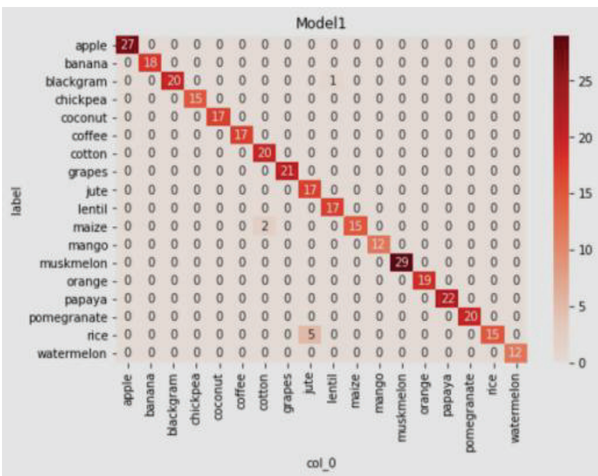

Fig. 5. Confusion Matrix

The precision parameter is used for result analysis. This mainly depends on the positive predictive values. The positive predicted values are the diagonal values in the recommended table. Here, the precision drop from 100 to 87 represents that the precision calculated for the label 7 is less. This variation is based on different soil parameters value.

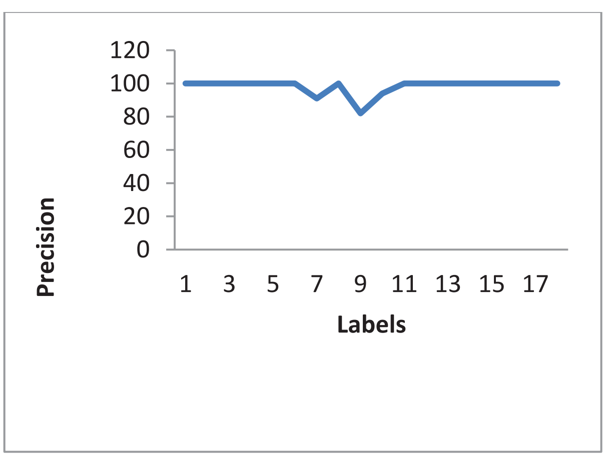

Fig. 6. Precision Analysis

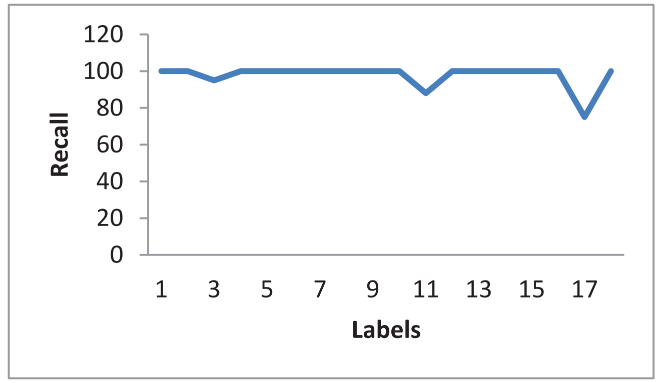

Fig. 7. Recall Analysis

This parameter is used to analyze the results. The recall value varies more because it counts the percentage of actual positives that are accurately identified. This parameter is used for result analysis which has very less variation in the accuracy.

The overall accuracy of the work is 97.65% hence this parameter is considered for evaluating the model. The accuracy rate is calculated by adding the number of correctly classified data and dividing by the total number of values. The clearly identified values can be found on the topmost left to extremely low. The confusion matrix accuracy is calculated by considering the total correct predictions to total predictions made. The correct predictions made are 333 and the total predictions are 341. Hence, the accuracy of evaluation model is **98.00%,** which is same as the SVM accuracy. Hence, the algorithm can be considered for the recommendation of crops.

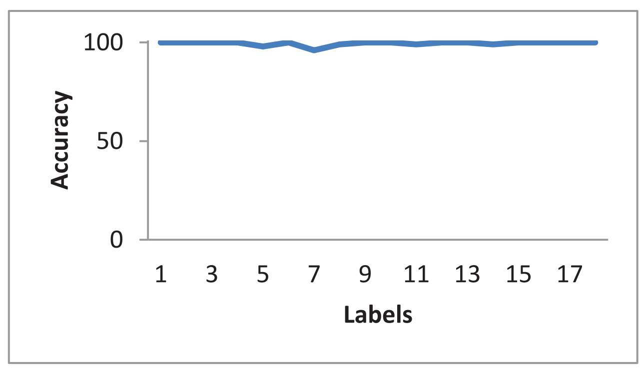

Fig. 8. Accuracy Analysis

IV. CONCLUSION

The designed methodology considers soil N, P, K, Ph, humidity, rainfall, and temperature for determining the most productive crops that can be cultivated in the given soil conditions. This methodology assists the farmer in selecting the most profitable crop. The crop selected will be precise and accurate. The wrong choice on a crop selection is reduced which increases the crop productivity. The accuracy of the algorithm obtained for this proposed work is 97%. This technique can be improved in the future by including IoT to obtain real-time soil values. Sensors are deployed on the farm to gather data about current soil conditions, providing the systems to increase the precision and accuracy of the outputs. As a result, farming may be done efficiently and productivity can be increased.

REFERENCES

[1] S.Pudumalar, E.Ramanujam, R.HarineRajashreen, C.Kavyan,T.Kiruthikan, J.Nishan, “Crop Recommendation System for Precision Agriculture”, IEEE Eighth International Conference on Advanced Computing (ICoAC), pp. 32-36, 2016.

[2] Rakesh Kumar, M.P. Singh, Prabhat Kumar and J.P. Singh, “Crop Selection Method to Maximize Crop Yield Rate using Machine Learning Technique”, International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM), pp. 138-145 , 2015.

[3] Rohit Kumar Rajak, AnkitPawar, MitaleePendke,PoojaShinde, Suresh Rathod, AvinashDevare, “Crop Recommendation System to Maximize Crop Yield using Machine Learning Technique”, International Research Journal of Engineering and Technology (IRJET), vol. 4 Issue. 12,pp. 950-953, 2017.

[4] Kevin Tom Thomas, VarshaS ,Merin Mary Saji , Lisha Varghese , Er. Jinu Thomas, “Crop Prediction Using Machine Learning”, International Journal of Future Generation Communication andnetworking,vol.13 Issue.3 pp.1896-1901, 2020.

[5] S.R Rajeswari, ParthKhunteta, Subham Kumar, Amrit Raj Singh, VaibhavPandey , “ Smart Farming Prediction Using Machine Learning”, International Journal of Innovative Technology and Exploring Engineering (IJITEE), vol. 8 Issue.7, pp. 45-51, 2019.

[6] Z. Doshi, S. Nadkarni, R. Agrawal and N. Shah, "Agro Consultant: Intelligent Crop Recommendation System Using Machine Learning Algorithms", Fourth International Conference on Computing Communication Control and Automation (ICCUBEA),pp. 1-6,2018.

[7] Meeradevi and H. Salpekar, "Design and Implementation of Mobile Application for Crop Yield Prediction using Machine Learning" Global Conference for Advancement in Technology (GCAT), pp. 1-6, 2019.

[8] R. Medar, V. S. Rajpurohit and S. Shweta, "Crop Yield Prediction using Machine Learning Techniques", IEEE 5th International Conference for Convergence in Technology (I2CT), pp. 1-5, 2019.

[9] P.Priya, U.Muthaiah and M.Balamurugan, “Predicting Yield Of The Crop Using Machine Learning Algorithm”, International Journal Of Engineering Sciences & Research Technology(IJESRT), vol.7 Issue.4, pp. 1-7, 2018.

[10] M.V.R. Vivek, D.V.V.S.S. Sri Harsha, P. SardarMaran, “A Survey on Crop Recommendation Using Machine Learning”, International Journal of Recent Technology and Engineering (IJRTE), vol. 7 Issue.5, pp. 120-125, 2019.

[11] Debasish Das, Mahinderpal Singh, Sarthak Swaroop Mohanty, S. Chakravarty, “ Leaf Disease Detection using Support Vector Machine”, 2020 International Conference on Communication and Signal Processing (ICCSP) Chennai, India pp.1036 – 1040, 2020.

[12] Utpal Barman a, Ridip Dev Choudhury “Soil texture classification using multi class support vector machine”, Journal on Information processing in Agriculture, vol. 7, Issue.2, pp. 1 – 13, 2020.