International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 05 Issue: 04 | April - 2021 ISSN: 2582-3930

Crop and Fertilizer Recommendation System Using Machine Learning

[1] Rushikesh Kanekar,[2] Kulsum Shaikh, [3]Ashish Golamade,[4]ShirkantSanas

[1, 2, 3] Department of Information Technology (IT), VasantdadaPatilPratishthan's College of Engineering and Visual Arts, Mumbai, Maharashtra, India.

[4] Professor, department of Information Technology (IT), VasantdadaPatilPratishthan's College of Engineering and Visual Arts, Mumbai, Maharashtra, India.

Abstract - We know that agriculture is the major factor on which Indian economy heavily relies but in many past years it has been observed that farmers are not getting high yields and thus resulting in them having lesser profit. Due to this the economy and gape of our country are getting badly affected. There are many factors affecting this outcome such as bad weather, bad soil study, poor knowledge, repetitive crops etc. but one of the vital reasons is not using the latest technology to improve the results. Recommending the crop in advance of sowing would help the farmers and policy makers for taking appropriate measures for agricultural processes and marketing. This system will not only recommend crops to farmers but it will also show them the best fertilizers to use according to their region and crops. We attempt to solve this problem by building a simple and interactive recommendation system for farmers. Implementation of such a system with an easy-to-use web based graphic user interface will be carried out. The final results after processing will be shown to the farmer in simplified format. Thus, for such a system for crop recommendation, there are different machine learning algorithms, and with the help of those algorithms we can achieve desired results. KNN algorithm is used in this system as it is one of the simplest yet effective supervised machine learning algorithms which is capable of performing both classification and regression. KNN algorithms use data and classify new data points based on similarity measures (e.g. distance function). Classification is done by a majority vote to its neighbors.

Keywords - ***Agriculture, Machine Learning, crop-recommendation, Supervised Algorithms, Crop yield, fertilizer-recommendation.***

1. INTRODUCTION

Agriculture is the most significant and important occupation in India. It is a big sector and plays an important role in the overall development of the country. The type of soil and the type of crops becomes very significant from an agriculture income point of view. Agriculture is very important and thus will boost our Indian economy significantly if farmers get profit. Data analytic (DA) is the process of examining data sets in order to draw conclusions about the information they contain, increasingly with the aid of specialized systems and software. Earlier crop selection was performed by considering the farmer's experience on a particular field and crop. As the conditions change every day very rapidly, farmers are forced to cultivate more and more crops. Being this as the current situation, many of the Farmers don't have enough knowledge about the new Crops and are not completely aware of the benefits they get while farming them. The farm productivity can be improved by checking and forecasting crop performance in different environmental conditions. Thus, the proposed system takes the location of the user as an input. Then it applies machine learning and prediction algorithms to identify the pattern among data and then process it as per input location given. This in turn will propose the best crops to invest in and also the best alternative crops according to available conditions. Thus, this system will only require the location of the user and it will suggest the number of profitable crops providing a choice directly to the farmer about which crop to cultivate and which fertilizers to use.

1.1Objectives

a. To use machine learning techniques to recommend the best crop.

b.To provide a user-friendly GUI to farmers for handling.

c. To increase the accuracy of crop yield prediction.

d. To analyze different climatic parameters

(Ph, rainfall, temperature)

e. To suggest best fertilizers to use based on

conditions.

2.LITERATURE REVIEW

"Predicting yield of the crop using machine learning algorithms. International Journal of Engineering Science Research Technology. This paper focuses on predicting the yield of the crop based on the existing data by using the Random Forest algorithm. Real data of Tamil Nadu state was used for building the models and the models were tested with samples.

Random Forest Algorithm can be used for accurate crop yield prediction." [1].

"Random forests for global and regional crop yield prediction. PLoS ONE Journal. Our generated outputs show that RF is an effective and adaptable 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. Random Forest is the most efficient strategy and it outperforms multiple linear regression (MLR)." [2].

"Crop production Ensemble Machine Learning model for prediction. In this paper, AdaNaive and AdaSVM are the proposed ensemble model used to project the crop production over a time period. Implementation done using AdaSVM and AdaNaive. AdaBoost increases efficiency of SVM and Naive Bayes algorithms." [3].

"Machine learning approach for forecasting crop yield based on parameters of climate. The paper was provided in the International Conference on Computer Communication and Informatics (ICCCI). In the current research a software tool named Crop Advisor has been developed as a user friendly web page for predicting the influence of climatic parameters on the crop yields.C4.5 algorithm is used to produce the most influencing climatic parameter on the crop yields of selected crops in selected districts of Madhya Pradesh. The paper is implemented using the Decision Tree algorithm." [4].

"Prediction on Crop Cultivation. International Journal of Advanced Research in Computer Science and Electronics Engineering (IJARCSEE) Volume 5, Issue 10, October 2016.Presently, soil analysis and interpretation of soil test results is paper based. This in one way or another has contributed to poor interpretation of soil test results which has resulted in poor recommendation of crops, soil amendments and fertilizers to farmers thus leading to poor crop yields, micro-nutrient deficiencies in soil and excessive or less application of fertilizers. Formulae to Match Crops with Soil, Fertilizer Recommendation." [5].

"Analysis of Crop Yield Prediction by Making Use Data Mining Methods. IJRET: The paper provided in the International Journal of Research in Engineering and Technology. In this paper the main aim is to create a user-friendly interface for farmers, which gives the analysis of rice production based on the available data. For maximizing the crop productivity various Data mining techniques were used to predict the crop yield. Such as the K-Means algorithm to forecast the pollution factor in the atmosphere." [6].

"Applications of Machine Learning Techniques in Agricultural Crop Production. Indian Journal of Science and Technology, Vol 9(38), DOI:10.17485/ijst/2016/v9i38/95032, October 2016.
From GPS based color images is provided as an intensified indistinct cluster analysis for classifying plants, soil and residue regions of interest. The paper includes various parameters which can help the crop yield for better enhancement and ratio of the yield can be increased during cultivation." [7].

"In this paper, we present a comprehensive review of research dedicated to the application of machine learning in agricultural production systems. Machine learning (ML) has emerged together with big data technologies, techniques, methods and high-performance computing to generate new opportunities to unravel, quantify, and analyze data intensive processes in agricultural operational sectors. By using Support Vector Machines (SVP) the Paper is implemented." [8].

"A Study to Determine Yield for Crop Insurance using Precision Agriculture on an Aerial Platform. Symbiosis Institute of Geoinformatics Symbiosis International University 5th & 6th Floor, Artur Centre, Gokhale Cross Road, Model Colony, Pune – 411016. Precision agriculture (PA) is the application of geospatial methodologies and remote sensors to identify variations in the field and to deal with them using different strategies. The causes of variability of crop growth in an agricultural field might be due to crop stress, irrigation practices, incidence of pest and disease etc. The Paper is implemented using Ensemble Learning (EL)." [9].

"Random Forests for Global and Regional Crop Yield Predictions. Institute on the Environment, University of Minnesota, St. Paul, MN 55108, United States of America. The generated outputs show that RF is an effective and different machine-learning method for crop yield predictions at regional and global scales for its high accuracy. The Paper is Implemented using k-nearest neighbor, Support Vector Regression (SVG)." [10].

3. METHODOLOGY

Dataset Collection: Data is a very important part of any Machine Learning System. To implement the system, we have decided to focus on Maharashtra state in India. As the climate changes from place to place, it is necessary to get data at district level. Historical data about the crop and the climate of a particular region is needed to implement the system. This data will be gathered from different government websites. The data about the crops of each district of

Maharashtra is gathered from Kaggle and the data about the climate is gathered from IBM weather forecast. The climatic parameters which affect crops the most are acidity, temperature, solidity, wateravailability, rainfall. So, the data about these climatic parameters is updated at a monthly level on above websites and hence we will be using these datasets.

There is no such complete and exact dataset available on any of the websites and hence we will make our own dataset from available existing datasets. We will take in consideration only important parameters which are basic but vital for precise output. Data cleaning is an important step before preparing a finalized dataset as in this step we remove all unwanted data. Next step would be to fill in missing values in dataset cells as this will help the algorithm and overall systems efficiency.

The dataset contains 5 columns which includes following parameters -Temperature, acidity, Solidity, Water availability, and crop label. For at least ten years the yearly abstracts of a crop will be used. These dataset usually accepts behavior of anarchic time series.

Data Partitioning: The Entire dataset is partitioned into 2 parts: for example, say 75% of the dataset is used for training the model and 25% of the data is set aside to test the model. There are two types of machine learning algorithms which are supervised and unsupervised machine learning algorithms. In Supervised learning the system learns from past experiences and applies that knowledge to get future predictions and outcomes. In unsupervised machine learning algorithms the system learns from current experiences and adapts to the changes as quickly as possible. We will be using supervised approach as supervised machine learning algorithms can apply what has been learned in the past to new data using labelled examples. After Sufficient training the system can provide targets for any new input. In order to change the model accordingly the learning algorithm can also differentiate its results with the correct, intended output and find errors.

Unsupervised learning: IN comparison, unsupervised machine learning algorithms are used when the information used to train is neither labelled nor classified. Unsupervised learning does analysis of how systems can infer a function to describe a hidden structure from unlabeled data. In order to describe hidden structures from unlabeled data the system doesn't figure out the right output, but it examines the data and can draw inferences from datasets.
Because of above requirements the system becomes more complex and less accurate thus supervised algorithms are better.

We will be using the KNN algorithm for this system as it has both classifier and regressor.It is the most popular and simple supervised machine learning algorithm, K-NN algorithm assumes the similarity between the new case/data and available cases and put the new case into the category that is most similar to the available categories. Itstores all the available data and classifies a new data point based on the similarity. This means when new data appears then it can be easily classified into a well suite category by using K-NN algorithm.

KNN algorithm fairs across all parameters (Ease of output, calculationtime, and predictive power) of considerations. It is commonly used for its ease of interpretation and low calculation time. We can implement a KNN model by following the below steps:

1. Load the data
2. Initialize the value of k
3. For getting the predicted class, iterate from 1 to total number of training data points: calculate the distance between test data and each row of training data. Here we will use Euclidean distance as our distance metric since it's the most popular method. The other metrics that can be used are Chebyshev, cosine, etc.
    1. Sort the calculated distances in ascending order based on distance values
    2. Get top k rows from the sorted array
    3. Get the most frequent class of these rows
    4. Return the predicted class

Fig 1:

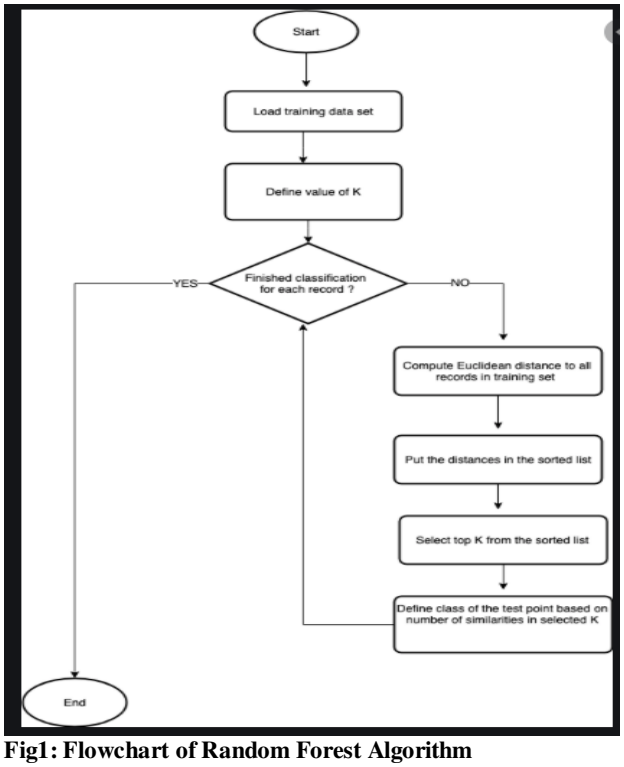

Fig1: Flowchart of Random Forest Algorithm

Fig 2:

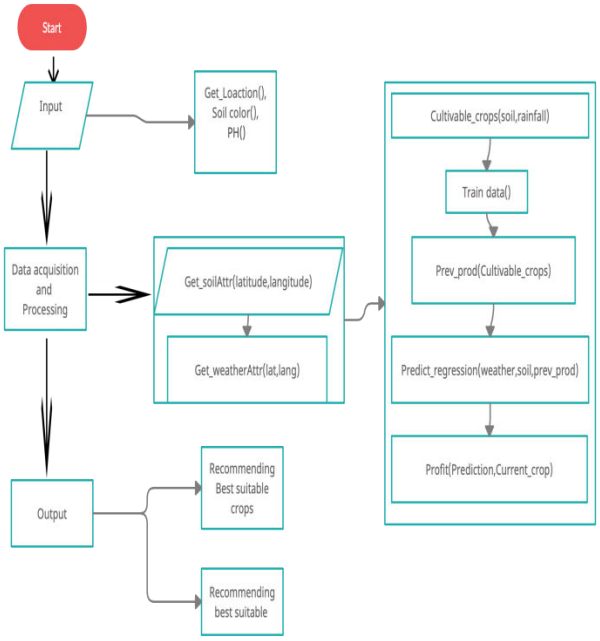

Fig 2: Block Diagram

The working of the system will be as follows. First of all the farmer has to enter his location coordinates or select his farm from the map shown. The he will be asked for the Ph. value of that soil, solidity, humidity and water availability. This is done as there is no exact dataset available. After taking inputs, the data acquisition stage starts. In this stage for that particular location coordinates the weather reports are fetched for next crop season and they are also taken in consideration. Also for that location soil attributes for that overall region are looked after in the dataset. The next stage is data processing. In this stage all the available data is combined and converted into needed format. Previousyear's crops are found out and depending on weather and all other parameters further algorithm is looked after. After applying final knn classification and regression and depending on weather and soil quality and previous year's crops as main factors the result is shown to the farmer in the form of graphical and easily understandable format. This includes best suitable crop which will give best production and maximum profit and along with that alternate 3 to 4 crops will also be shown so that farmers have multiple choices to choose from.

Best fertilizers to use according to soil PH value and crop will also be shown so that best results could be obtained.

Taking in consideration that the main users are farmers, the system will be made as simple as it can be so that naive users can use it easily.

Dataset Snapshots:

temperature	humidity	ph	rainfall	label
20.87974371	82.00274423	6.502985292	202.9355362	rice
21.77046169	80.31964408	7.038096361	226.6555374	rice
23.00445915	82.3207629	7.840207144	263.9642476	rice
26.49109635	80.15836264	6.980400905	242.8640342	rice
20.13017482	81.60487287	7.628472891	262.7173405	rice
23.05804872	83.37011772	7.073453503	251.0549998	rice
22.70883798	82.63941394	5.70080568	271.3248604	rice
20.27774362	82.89408619	5.718627178	241.9741949	rice
24.51588066	83.5352163	6.685346424	230.4462359	rice
23.22397386	83.03322691	6.336253525	221.2091958	rice
26.52723513	81.41753846	5.386167788	264.6148697	rice
23.97898217	81.45061596	7.50283396	250.0832336	rice
26.80079604	80.88684822	5.108681786	284.4364567	rice

25.93730186	78.89864446	5.915568968	98.21747528	banana
25.13686519	84.88394407	6.195152442	91.46442491	banana
27.50527651	80.79783998	6.156373499	105.0776992	banana
26.21009246	82.34429458	6.313197204	112.0700033	banana
29.10403455	79.19588629	6.324270089	92.07835761	banana
29.2440638	77.32017166	5.707488987	90.66727868	banana
25.56202173	77.38229006	6.119216009	93.10247183	banana
26.3985515	81.36028902	5.571401169	98.16752001	banana
28.09577643	77.79586769	5.63127166	109.5408614	banana
27.84767901	83.31110751	6.101241579	117.2878912	banana
27.39341554	81.4654833	6.438137279	94.31102057	banana
25.14517635	81.38204104	6.098369122	119.218154	banana
28.65456263	79.28693687	5.695267822	102.4633775	banana
29.16093424	76.67484233	5.618094446	109.575944	banana
27.57278064	82.0638878	6.435785799	91.34276507	banana

Future Scope & improvements:

For improving predictions there has to be more clarity in data and hence hardware microprocessor modules could be used to collect data from that exact location. These live soil samples can produce more than 95% accurate predictions.

4. CONCLUSION

Based on the climatic parameters the present study provided the demonstration of the potential use of machine learning techniques in recommending the crop based on certain factors. The developed website will be user friendly and the accuracy of predictions is expected to be 75-80 percent in all the crops and districts selected in the study indicating higher accuracy of prediction. The website will be having such an easy user interface that any naive user can use it. The best crop to invest in will be prompted and also fertilizer to use willalso be suggested.

ACKNOWLEDGEMENT

This paper on ‘Crop and fertilizer recommendation system using ML’ has been possible only because of the kind cooperation lent by our project guide Professor ShrikantSanaswithout which this would not have been possible.

At last, we would like to thank Dr. SeemaLadhe (HOD-IT) for his guidance and support.

REFERENCES

[1] P.Priya, U.MuthaiahM.Balamurugan.Predicting yield of the crop using machine learning algorithm. International Journal of Engineering Science Research Technology.

[2]. J.Jeong, J.Resop, N.Mueller and team.Random forests for global and regional crop yield prediction.PLoS ONE Journal.

[3].Narayanan Balkrishnan and Dr. GovindarajanMuthukumarasamy.Crop production Ensemble Machine Learning model for prediction. International Journal of Computer Science and Software Engineering (IJCSSE).

[4]. S.Veenadhari, Dr. Bharat Misra, Dr. CD Singh.Machine learning approach for forecasting crop yield based on climatic parameters. International Conference on Computer Communication and Informatics
(ICCCI).

[5]. Shweta K Shahane ,Prajakta V Tawale.Prediction On Crop Cultivation. IInternational Journal of Advanced Research in Computer Science and Electronics Engineering (IJARCSEE) Volume 5, Issue 10, October 2016.

[6]D Ramesh B Vishnu Vardhan. Analysis Of Crop Yield Prediction Using Data Mining Techniques. IJRET: International Journal of Research in Engineering and Technology.

[7]SubhadraMishra,Debahuti Mishra, GourHariSantra. Applications of Machine Learning Techniques in Agricultural Crop Production. Indian Journal of Science and Technology, Vol 9(38), DOI:10.17485/ijst/2016/v9i38/95032, October 2016.

[8].Konstantinos G. Liakos,PatriziaBusato,DimitriosMoshou, Simon Pearson ID,DionysisBochtis. Machine Learning in Agriculture. Lincoln Institute for Agri-food Technology (LIAT), University of Lincoln, Brayford Way, BrayfordPool,LincolnLN67TS, UK,spearson@lincoln.ac.uk.

[9]. Baisali Ghosh. A Study to Determine Yield for Crop Insurance using Precision Agriculture on an Aerial Platform. Symbiosis Institute of Geoinformatics Symbiosis International University 5th & 6th Floor, Atur Centre, Gokhale Cross Road, Model Colony, Pune – 411016.

[10]. Jig Han Jeong, Jonathan P. Resop, Nathaniel D. Mueller, David H. Fleisher ,Kyungdahm Yun, Ethan E. Butler,Soo-Hyung Kim. Random Forests for Global and Regional Crop Yield Predictions. Institute on the Environment, University of Minnesota, St. Paul, MN 55108, United States of America.