International Conference - Innovation-2021-Innovation-2021
International Journal of Scientific Research in Computer Science, Engineering and Information Technology ISSN : 2456-3307 (www.ijsrcseit.com)
Volume 8, Issue 3, May-June-2021

Crop Suggestion based on Regional Soil Quality using Machine Learning Techniques

Mayuresh Kulkarni[1], Rutuja Jade[1], Apekshita Bhosale[1], Bhagyashree Ramteke[1], Sunil Rathod[2]

[1]Students, Department Computer Engineering, Dr. D. Y. Patil School of Engineering, Lohegaon, Pune, Maharashtra, India

[2]Professor, Department Computer Engineering, Dr. D. Y. Patil School of Engineering, Lohegaon, Pune, Maharashtra, India

ABSTRACT

Agriculture in India plays a major role in economy and employment. The common difficulties present among the Indian farmers are they don't opt for the proper crop based on their soil necessities. Because of this productivity is affected. This problem of the farmers has been solved through precision agriculture. This method is characterized by a soil database collected from the farm, crop provided by agricultural experts, achievement of parameters such as soil through soil testing lab dataset. Agribusiness assumes a prevailing job in the development of the nation's economy. Atmosphere and natural changes have become a genuine danger inside the agri-field. Machine Learning ML is a significant methodology for accomplishing reasonable and compelling answers for this disadvantage. Harvest Yield Prediction technique includes foreseeing yield of the harvest from reachable historical and possible data like climate parameter, soil parameter and yield prediction. Real information of the state was utilized for building this model and furthermore the models were tried with tests acquired from the information. The expectation can make the farmer to foresee the yield of harvest before developing into the agribusiness zone. To anticipate the harvest yield in future precisely Random Forest, the most remarkable and popular administered machine learning rule is utilized.

With the impact of climate change in India, the majority of the agricultural crops are being badly affected in terms of their performance over a period of last two decades. Predicting the crop yield well ahead of its harvest would help the policy makers and farmers for taking appropriate measures for marketing and storage. Such predictions will also help the associated industries for planning the logistics of their business. Several methods of predicting and modeling crop yields have been developed in the past with varying rates of success, as these don't take into account characteristics of the weather, and are mostly empirical.

This software provides proper information to farmers and for that Data mining and machine learning is still an emerging technique in the field of agriculture and horticulture. In this paper we have proposed a method for classifying the soil according to the macro nutrients and micro nutrients and predicting the type of crop that can be cultivated in that particular soil type. Several types of machine learning algorithms are used such as K-Nearest Neighbor (K-NN), Support vector machine (SVM) and logistic regression.

Keywords: Machine learning, Agriculture, Soil, Classification, KNN Algorithm.

I. INTRODUCTION

Machine learning is a field of computer science where new developments evolve at recent times, and also helps in automating the evaluation and processing done by mankind, thus by reducing the burden on human power. In simple terms, machine learning provides basic algorithms that can provide information about a dataset without writing code to solve a program manually. Instead of writing code you provide data or the basic algorithm and it forms its own conclusions based on this data. In machine learning, the methods are derived from the learning process. Those methodologies need to learn through experiences to perform a particular task.

Classification is a data mining technique [1] based on machine learning which is used to categorize the data item in a dataset into a set of predefined classes. It helps in finding the diversity between the objects and concepts. Among these various machine learning techniques that are being used in this field; this system builds an efficient and accurate model to classify the soil type according to nutrients of soil and predict the type of crop that can be yielded in that particular soil. Also this information can further be used by the Minister of Agriculture, Food & Marketing Industrialist, etc. to grow the business and economy of the country.

India is one of the biggest producers of agricultural products and still has very little farm productivity. Productivity needs to be increased so that farmers can get more income from the same piece of land with less labor.

II. LITERATURE REVIEW

The requirements and planning are needed for developing software models. The author's [2] start from the basics of precision farming and move towards developing a model that would support it. It deeply studies the basics of precision farming. 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. 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 positioned elsewhere in India only with some modifications. 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.

The algorithms [3] used for yield prediction in this system are Support Vector Machine and K-nearest neighbor algorithm. The importance of crop selection and the factors deciding the crop selection like production rate, market price and government policies are discussed. This system will propose a Crop Selection Method (CSM) [4] which solves the crop selection problem and improves net yield rate of the crop. It suggests a series of crops to be selected over a season considering factors like weather, soil type, water density, crop type. The predicted value of influential parameters determines the accuracy of CSM. Hence there is a need to include a prediction method with improved accuracy and performance.

The system [5] aims to solve the crucial problem of selecting the classifiers for ensemble learning. The proposal aims to achieve higher accuracy and performance. This project emphasizes the need for accuracy because it depends on the dataset and the learning algorithm. The performance of the models was compared with accuracy and computational time. It proposes a framework which would predict the production, and import for that particular year. 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 has to be further imported or not. The soil dataset is analyzed and category predicted.

Table 1: Literature Survey

Sr. No.	Title	Author	Methodology Used	Result
1.	Crop Recommendation System for Precision Agriculture	S.Pudumalar, E.Ramanujam, 2016	1. Random tree 2. CHAID 3. KNN 4. Naïve Bayes 5. WEKA Tool	1. Pre-processing of Data 2. Handling missing and out of range values 3. Feature extraction 4. Ensemble model to get higher accuracy 5. Rule generation
2.	Agriculture decision support system using data mining	Prof. Rakesh Shirsath, 2017	1. Subscription based system 2. ANN 3. Android application	1. Android app with login module 2. Previously planted crops known to system 3. User feedback mechanism 4. Maintenance of crops
3.	A Study on Various Data Mining Techniques for Crop Yield Prediction	Yogesh Gandge, Sandhya, 2017	1.Attribute selection 2. Multiple Linear Regression 3. Decision Tree using ID3 4. SVM 5. Neural Networks 6. C4.5 K-Means and KNN	1. Selection of agricultural field 2. Selection of crop previously planted 3. Input from user 4. Pre-process 5. Attribute Selection
4.	RSF: A Recommendation System for Farmers	Miftahul Jannat Mukarrama, 2017	1. Location Detection 2. Data analysis and storage 3. Similar location detection 3. Recommendation generation module	1. Physiographic, thermal, crop growing period, crop production rate 2. Seasonal crop database 3. Similar location detection

III. TAXONOMY CHART

The Taxonomy chart represents the comparison of some existing systems with the proposed system. The chart clearly depicts that drawbacks of existing systems are overcome in the proposed system as shown in the following table.

Table 2: Taxonomy Chart

Attributes → Approaches ↓	K-Nearest Algorithm	Precision Agriculture	Agro-consultant	Recommendation generation module
Crop Prediction System using ML	Yes	No	No	No
Crop Recommendation System for Precision Agriculture	Yes	Yes	No	Yes
Proposed System	**Yes**	**Yes**	**Yes**	**Yes**

IV. PROPOSED SYSTEM

A. Problem Statement: To develop the system that classifies the soil type according to temperature and moisture content of soil and predicting the type of crop that can be yielded in that particular soil. The main target of the system are farmers and as most farmers are not experienced internet users to find the relevant information and services on the system, the proposed system should be easy for them. In the service requests with respect to these applications initiated using different delivery channels would be processed by the system.

B. Proposed System Overview: This project is made in consideration of farmers facing so many difficulties in which crops are suitable for production. Crop cultivation depends on the nature and the nutrients of the soil increasing the cultivation of land which brings a loss of supplements present in the soil. In crop cultivation, soil plays an important role. In this project, Machine learning is implemented using Python and KNN algorithms. Data mining is a challenging technology in the field of agriculture. Nowadays data mining has been used in the field of agriculture for soil classification, wasteland management, and crop and pest management. A soil test is carried out to identify the nutrients content, composition and other components contained in the soil. Soil tests are mainly conducted to measure the fertility and other decencies present in the soil so that suitable measures can be taken to resolve it. In simple terms, the meaning of machine learning is that basic algorithms can provide information about a dataset without writing code to solve this program manually.

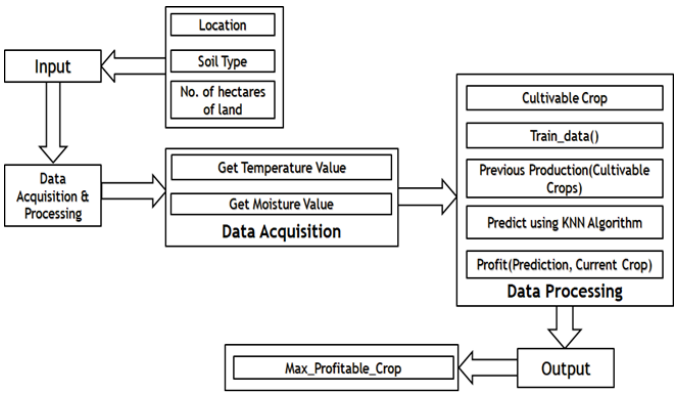

Figure 1: Proposed System Architecture

C. Algorithm: K-Nearest Neighbour (KNN)

Working:

Step 1 -For implementing any algorithm, we need a dataset. So during the first step of KNN, we must load the training as well as test data.

Step 2 -Next, we need to choose the value of K i.e. the nearest data points. K can be any integer.

Step 3 -For each point in the test data do the following –

- 3.1 – Calculate the distance between test data and each row of training data with the help of any of the methods, namely: Euclidean, Manhattan or Hamming distance. The most commonly used method to calculate distance is Euclidean.
- 3.2 – Now, based on the distance value, sort them in ascending order.
- 3.3 – Next, it will choose the top K rows from the sorted array.
- 3.4 – Now, it will assign a class to the test point based on the most frequent class of these rows.

Step 4 -End

Pseudo Code:

1. Calculate "d(x, xi)" i =1, 2, ….., n; where d denotes the Euclidean Distance between the points.
2. Arrange the calculated n Euclidean distances in non-decreasing order.
3. Let k be a +ve integer, take the first k distances from this sorted list.
4. Find those k-points corresponding to these k-distances.
5. Let ki denotes the number of points belonging to the ith class among k points i.e. $k \geq 0$
6. If ki >kj $\forall$ i ≠ j then put x in class i.

Mathematical Model:

Euclidean Distance is calculated using the following formula:$d=\sqrt{((x2-x1)^2+(y2-y1)^2)}$

V. RESULTS AND DISCUSSION

A. Experimental Setup: The system is built using the Flask framework on Windows platform. The PyCharm IDE is used as a development tool. The system doesn't require any specific hardware to run; any standard machine is capable of running this application.

B. Experimental Result: The following Figure 2 shows the accuracy of KNN Algorithm.

```
Confusion Matrix:
 [[ 16   0   0 ...   0   0   0]
 [  0  51   0 ...   0   0   0]
 [  0   1 132 ...   0   0   0]
 ...
 [  0   0   0 ...  50   0   0]
 [  0   0   0 ...   0  53   0]
 [  0   0   0 ...   0   0  70]]
Accuracy:  0.9921520033044197
```

Figure 2: Accuracy of KNN Algorithm

Figure 3 shows that for the given values the system predicts the accurate crop.

```
Confusion Matrix:
 [[ 16   0   0 ...   0   0   0]
 [  0  51   0 ...   0   0   0]
 [  0   1 132 ...   0   0   0]
 ...
 [  0   0   0 ...  50   0   0]
 [  0   0   0 ...   0  53   0]
 [  0   0   0 ...   0   0  70]]
Accuracy:  0.9921520033044197
Prediction on 36,625:
 ['Rice,Sugar beet ,Sunflower']
```

Figure 3: Prediction of Crop

Figure 4 shows the plotting of data points in which the testing module classifies the dataset in two sections.

When the value of K=1, it is represented by the red dots and when the value of K>1, it is represented by the green dots.

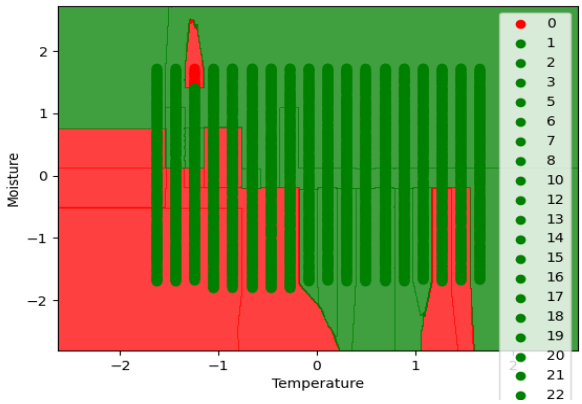

Figure 4: Plotting of Data Points

VI. FUTURE WORK

1. Crop disease detection using Image Processing where users can upload pictures of diseased crops and get pesticide recommendations.
2. Implementation of Smart Irrigation System to monitor weather and soil conditions, plant water usage etc. to automatically alter watering schedule.

VII. CONCLUSION

The application would definitely help in bridging the gap between the farmers and technology, and would prove beneficial to all sectors associated with farming. The proposed tool helps the farmer in determining crop based on contents of the soil like temperature and moisture. It also makes sure that the crops suggested follow crop rotation patterns so as to make sure that the land remains fertile for long. Our work would help farmers to increase productivity in agriculture, prevent soil degradation in cultivated land, and reduce chemical use in crop production and efficient use of water resources. Our future work is aimed at an improved data set with a large number of attributes and also implements yield prediction.

VIII. ACKNOWLEDGEMENT

It gives us great pleasure in presenting the paper on "Crop Suggestion based on Regional Soil Quality using Machine Learning Techniques". We would like to thank Dr. Pankaj Agarkar, Head of Computer Engineering Department, DYPSOE, Pune for giving us all the help and support we need during the course of the Paper writing work. We are really grateful to him for giving us an opportunity to work with the R&D cell of our department.

Our special thanks to Dr. Ashok Kasnale, Principal DYPSOE who motivated us and created a healthy environment for us to learn in the best possible way. We also thank all the staff members of our college for their support and guidance.

IX. REFERENCES

[1]. V. Rajeshwari and K. Arunesh, " Analyzing Soil Data using Data Mining Classification techniques," Vol 9(19),May 2016.

[2]. Satish Babu (2013), 'A Software Model for Precision Agriculture for Small and Marginal Farmers', at the International Centre forFree and Open Source Software (ICFOSS) Trivandrum, India.

[3]. Anshal Savla, Parul Dhawan, Himtanaya Bhadada, Nivedita Israni, Alisha Mandholia , Sanya Bhardwaj (2015), 'Survey of classification algorithms for formulating yield prediction accuracy in precision agriculture', Innovations in Information,Embedded and Communication systems (ICIIECS).

[4]. Rakesh Kumar, M.P. Singh, Prabhat Kumar and J.P. Singh (2015), '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).

[5]. Liying Yang (2011), 'Classifiers selection for ensemble learning based on accuracy and diversity' Published by Elsevier Ltd. Selection and/or peer-review under responsibility of [CEIS].