Predicting the Farmland for Agriculture from Soil Features Using Machine Learning

P.Poorna Chandu, T.Nagasheshu, P.B.S Sai Prasad, K.Venkatesh

Abstract— **Agriculture has been a crucial part of human society due to the fact that the civilization growth directly depends on agriculture. Advancements in agriculture are necessary to raise food production as the population is increasing day by day. We need to explore the use of innovative methods to improve crop selection for the soil depending on its texture, type, moisture, temperature, humidity, soil erosion, and slope. Machine Learning, can help us in learning the data, analyzing the data fields predicting the future. For predicting the farmland, supervised machine learning techniques are used. India's agriculture land over all its states has been used to predict whether it is farmland or not using Machine Learning Algorithms and making the resources available beforehand.**

Index Terms—**Agriculture, food production, Supervised Machine Learning, Machine Learning Algorithms.**

I. INTRODUCTION

From the past, agriculture is one of the leading sectors practiced in India. Ancient people farm the crops in their land. Therefore, natural crops are cultivated and had been used by human beings, animals and birds. The invention of new technologies in the field of agriculture is slowly disgracing. Due to, plenty of inventions in the agriculture sector people are more concentrated on cultivating hybrid products which ultimately leads to an unhealthy life, lack of nutrition. Nowadays, modern people don't have awareness about farming crops. The changes in the farming seasons, climatic conditions, soil features like water availability, and soil erosion, etc affects the cultivating techniques. By examining all these issues and problems like temperature, weather, and several factors, there is no true solution and technologies to overcome the situation faced by farmers. In India, the economical growth of agriculture can be increased in many ways. There are different ways to increase and improve the crop yield and the quality of the

.

P Poorna Chandu, CSE, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai, India,9866269790,
(e-mail:patichaandu7@gamil.com).
T NagaSheshu, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai, India, 9566088027,
(e-mail: tsheshu99@gamil.com).
P B S Sai Prasad, CSE, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai, India, 8428968134, (e-mail: saiprasadpadala6@gmail.com).
K Venkatesh, CSE, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai, India, 9894340294,
(e-mail: venkiur@gmail.com).

crop. [11] Machine Learning algorithms can be used for predicting crop yield production-based on the soil features. Supervised Learning Classification technique is used to predict the accuracy. In supervised learning, both input and output are given to the system by splitting the data into training and testing data. Training data is used to train the system and testing data is used for predicting the accuracy. For predicting the accuracy three algorithms are used and accuracy among them is compared.
1) DNN -Deep Neural Network

2) Random Forest

3) Linear Discriminant Analysis

Practical implementation of the models developed in decision support tools that would provide a snapshot of areas and classify it whether agriculture could be practiced there or not. Along with these proposed algorithms, the RANDOM FOREST will be implemented to predict the data and to improve accuracy.

A. Aim of the project

The aim of the project is to use Machine Learning to find whether a particular land is adaptable for agriculture, with the perfect conditions for agriculture, and to improve the accuracy of the algorithms and compare to find the best among the three.

B. Scope of the Project

The scope of the project is to identify the Farmlands based on the agriculture conditions of all types of soil features. It required many inputs such as temp, soil conditions, erosion conditions, humidity, etc and output are collected as a dataset which includes all information about agriculture lands that is needed for the project. The agriculture sector is necessary to balance the supply as the population and demand are increasing day by day. Agriculture will be formed as the management of land parcels and up productivity. It serves solely up to date society and additionally our future generations. The project is to give valuable information to users about profit yielding farmlands based on the soil features. So the need of exploring the use of innovative methods to improve the crop growing accuracy in that farmland from the soil depending on its texture, type, moisture is there. Machine Learning, can help in learning, analyzing and predicting the data for the future.

II. LITERATURE SURVEY

The author in [1] presents that Machine Learning deals with issues where the relationship between input and output variables. This characteristic is useful to model sophisticated non-linear behaviors, sort of operate for crop yield prediction. Machine learning techniques most successfully applied to Crop Yield Prediction (CYP). The work methodology continues until the model achieves a desired level of accuracy on the work info. This will facilitate farmers to grow the right amount of crops within the needed land and to understand the precipitation temporary worker, min temporary worker of that space. This paper concentrates on the prediction of the foremost profitable crop that may be adult within the agricultural land exploitation machine learning techniques. [12] This paper includes the utilization of the android system which will offer them 64000-time crop analysis victimization varied lookout reports and soil quality. Therefore, farmers would adult the foremost profitable crop within the best appropriate months.

The author in paper [2] tells about developing accurate models for crop yield estimation using Information and Communication Technologies that may help farmers and other stakeholders improve deciding in reference to national food import/exports and food security. And this will examine the performance of the RF and the MLC methods. The author in [3] describes, however, the recent agricultural info is often wont to describe the long run prediction of crops and yield. It conjointly suggests the farmers regarding what style of the crop is often big victimization the weather station info and provides the acceptable info to like the correct season for excellence farming. The data processing techniques area unit mentioned thoroughly.

The author in [4] describes the vital role performed by data processing ways in the agricultural field. They have conferred the various metric capacity unit algorithms like Random forest, SVM, ANN, etc. The crops were foreseen primarily supported climatically options which provide accuracy score of regarding ninety-fifth with the C4.5 formula.

The main objective of this paper [5] is to the analysis of main soil properties like organic matter, essential plant nutrients; micronutrient that affects the expansion of crops.BPN will find and suggest the proper correlation percentage among those properties. The machine learning system will be split into three steps. The first sampling second one is Back Propagation Algorithm and the third one is Weight updating. The performance of the rear Propagation Neural network model is going to be evaluated employing a test data set.

The innovation of this paper [6] lies in combining CNN-based learning methods for producing geo-objects and tree-based learning methods for mapping soil property. To improve the precision of predicting soil properties in a geographic space, the author developed a novel geo-object-based soil property mapping procedure using machine learning algorithms.

III. ALGORITHMS

This project has chosen the DNN – Deep Neural Network LDA and Random Forest algorithms for prediction. [7] Neural networks use randomness intentionally by choice to verify they effectively learn the function being approximated for the matter. Randomness is employed because this class of machine learning algorithm performs better with it than without. The most common kind of randomness utilized in neural networks is that the random initialization of the network weights. DNN can be implemented using a sequential classifier by adding layers. [8] In this add, dense methods are used to add layers. The compile method is used to run the layers conversion using adam optimizer, loss binary cross- entropy, and accuracy as metrics. The fit method is used to pass the input to the model and epoch can be set for the number of iterations. Similarly, LDA and Random Forest algorithms[9] [13] are implemented and the respective confusion matrix and the accuracy of each algorithm are predicted.

IV. OUTPUT DISCUSSION

The work has chosen Random Forest, LDA, and DNN. Neural networks use randomness by design to ensure that they effectively learn the function being used for the problem. Randomness is used because this class of machine learning algorithm performs better with it than without it. Experienced mathematical modeling is used to process data in complex ways by deep neural networks. Different classifiers are used for predicting the accuracy, the difference between them is clearly observed.

```
Original feature number: 1422
Reduced feature number: 2
Original feature number: 1422
Reduced feature number: 2
[[4807   43]
 [  24 1997]]
Accuracy for LDA 0.9902488720710232
[[4813   37]
 [  64 1957]]
Accuracy for Random Forest 0.9853005384951244
```

Fig: 4.1 Accuracy with confusion matrix

Evaluation of the performance of the machine learning model is done by predicting the accuracy of the algorithms under supervised learning. Another method is a confusion matrix by comparing the Actual data and Predicted data. It contains the predicted outcomes of the Y label compared with the actual output of the Y test data. There are various ways to check the performance of the machine learning model is

1) Confusion matrix – [10] For simplicity, it'll mostly discuss things in terms of a binary classification problem some common terms to be clear with are: True positives (TP): Predicted positive and are literally positive. False positives (FP): Predicted positive and are literally negative. True negatives (TN): Predicted negative and are literally negative. False negatives (FN): Predicted negative and are literally positive. A confusion matrix is just a representation

of the above parameters in a matrix format. Better visualization is usually good.

2) Accuracy - The most commonly used metric to judge a model and is actually not a clear indicator of the performance.

V. GRAPH

The graph for the accuracy of two algorithms is shown below in figure 5.1. There is slight difference between the accuracies of two algorithms that is both the algorithms are almost giving equal accuracy. The accuracy of LDA is slightly high compared to Random Forest.
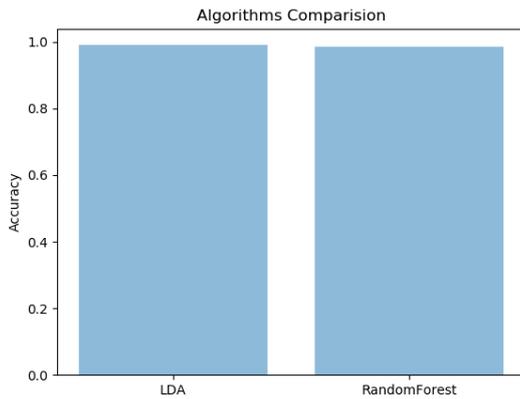

Fig: 5.1 Comparisons between LDA and Random Forest

VI. CONCLUSION

We could see that the dataset has been processed and trained. The output of the trained data has been checked against the output set of the train data. The complete portion has once again checked with the test train and test output data. The accuracy of the DNN has also been verified. It works fine for the dataset and predicts the outcome accurately.

VII. FUTURE ENHANCEMENTS

In the future, the system is developed in the form of a website where the user is able to get information about different crops and the fertilizers used for a particular crop and the information about the workshops will be provided on the webpage. And the user can maintain a separate profile and can able to upload and find the accuracy if the dataset is available with the user. The methodology will help to improve the net profit of the farmers having a better organic farming practice. The method will help to get rid of the poverty of farmers and agricultural runoff. Nowadays the information technology plays a key role in the agriculture industry. All the details will be stored in the cloud which will be used for easy access and used as a reference by other users in the form of notifications to mail and to mobile as text messages.

REFERENCES

[1] Priya, P., U. Muthaiah, and M. Balamurugan. "Predicting yield of the crop using machine learning algorithm." International Journal of Engineering Sciences Research Technology 7.1 (2018): 1-7.

[2] Narasimhamurthy, V. and Kumar, P., 2017. Rice Crop Yield Forecasting Using Random Forest Algorithm. Int. J. Res. Appl. Sci. Eng. Technol. IJRASET.

[3] R. Sujatha, Dr. P.Isakki,A Study on Crop Yield Forecasting Using Classification Techniques, 978-1-4673- 8437-7/16/31.00 c 2016 IEEE.

[4] Veenadhari, S., Bharat Misra, D Singh, Data mining Techniques for Predicting Crop Productivity – A review article IJCST International Journal of Computer Science and technology, march 2011.

[5] Shivnath Ghosh,Santanu Koley, "Machine Learning for Soil Fertility and Plant Nutrient Management using Back Propagation Neural Networks" IJRITCC, vol. 2, Issue 2,292-297,2014.

[6] Wu, Tianjun, et al. "Geo-Object-Based Soil Organic Matter Mapping Using Machine Learning Algorithms With Multi-Source Geo-Spatial Data." IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 12.4 (2019): 1091-1106.

[7] Maier, Daniel, et al. "Applying LDA topic in communication research: Toward a valid and reliable methodology." Communication Methods and Measures 12.2-3 (2018): 93-118.

[8] Ayaz, Muhammad, et al. "Internet-of-Things (IoT)-Based Smart Agriculture: Toward Making the Fields Talk." IEEE Access 7 (2019): 129551-129583.

[9] Zingade, D. S., et al. "Crop prediction system using machine learning." Int. J. Adv. Eng. Res. Dev. Spec. Issue Recent Trends Data Eng 4.5 (2017): 1-6.

[10] Rahman SA, Mitra KC, Islam SM. Soil classification using machine learning methods and crop suggestion based on soil series. In2018 21st International Conference of Computer and Information Technology (ICCIT) 2018 Dec 21 (pp. 1-4) IEEE.

[11] Emrullah, A. C. A. R., Mehmet Sirac OZERDEM, and Burak Berk USTUNDAG. "Machine Learning based Regression Model for Prediction of Soil Surface Humidity over Moderately Vegetated Fields." 2019 8th International Conference on Agro-Geoinformatics (Agro-Geoinformatics). IEEE, 2019. 50

[12] S.R.Rajeswari , Parth Khunteta, Subham Kumar, Amrit Raj Singh, Vaibhav Pandey,2019. Smart Farming Prediction Using Machine Learning. International Journal of Innovative Technology and Exploring Engineering (IJITEE).

[13] R. Shrestha (2016), Regression based corn yield assessment using MODIS based daily NDVI in Iowa state, IEEE Fifth International Conference on AgroGeoinformatics (Agro- Geoinformatics), 24(12), 7-9.

Pati Poorna Chandu, Computer Science and Engineering, *CSE, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai.*

Tadipatri Nagasheshu, Computer Science and Engineering, *CSE, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai.*

Padala Bala Surya Sai Prasad, Computer Science and Engineering, *CSE, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai.*

K Venkatesh. M.Tech. M.B.A. P.hD*
Assistant professor. Dept of cse. Vel Tech
Rangarajan Dr.Sagunthala R&D institute of
science and Technology.