

Model Approach of Crop Classification Using Logistic Regression

¹Dr. K. A. Waghmare, ²Sheetal A Zare

^{1,2}Department of Computer Science, GOVT Amravati(MH), India.

¹waghmare.kamlesh@gmail.com, ²Sheetal.jhare@gmail.com

Abstract—Relation between agriculture and the human development is very old. From the beginning era all participant of food chain in second stage depends on agriculture. At the beginning state life was natural and moving. With the stability of humans use of specific land increased and now stage is , where , humans are useable to chemical products for increasing the quantity of crop production in the land. Though the use of external chemicals result in quantitative growth of crop, but internally soil health get suffer from it and one –day it might be loss her fertility. Soil testing tools has a vital role in testing the soil for nutrient in soil and test its productivity. Easy classification of soil on the basis of its different features and also from testing the quality of soil to suggest the additional supplement to improve the health and nutrient in the soil. Key objective of this paper is to capture soil health in concern of nutrient. In this paper we have shown the classification approach of soil nutrient and detecting the soil health. We have built model using machine leaning algorithm (Logistic Regression) in Python. Results are compared with standard chart of soil health contains from the agriculture laboratory. Our detection accuracy lies between 95 to 99%.

Keywords— *classification, soil health, logistic refression; feature extraction; Machine Learning*

I. INTRODUCTION

Classification techniques are essential part of machine learning and data mining application. Learning Classification tools, detection tools, suggestion tools are being developed and modifying using artificial intelligence, Machine learning, Neural network and different classifiers as support vector machine, rule base system classifiers. Brute force classifiers and many more but logistic regression is common and useful regression method for solving binary classification problems.

Primary factor of soil are potassium content, sulphate contains, Hydrogen, Oxygen and nitrogen, carbon, contains. Secondary factors are related to minerals like calcium, magnesium or phosphorus. And other are iron, manganese, copper, zinc, molybdenum, boron and chlorine. If the soil losing her nutrient health it shows the symptoms on crop growth. And hence if crop get affected without external interference then we can assume that soil has nutrient deficiency.

In this project we focus on soil health and suitability of crop in specific soil. Crop management practice during the cultivation may arise heavy loss in soil quality. We use the features of soil to detect the soil health and her nutrient deficiency. As so our research base is building a framework for crop suggestion and soil classification we have built up this model using python tool. Logistic regression is one of the most popular method for two-class

classification problems. Its basic fundamental concept are also constructive in deep learning. Logistic Regression describes and estimates relationship between one dependent variable and independent variable. The outcome or target variable is dichotomous in nature. That is there are only two cases. Productive soil or unproductive soil.

II. LITRATURE SURVEY

Machine learning approach is effective method for application in the field of analytics of the data to predict the results of the system using specific models or algorithms. Ample amount of applications are available for machine learning (ML). The most significant among them is classification and regression approach. In the field of engineering associated with agriculture motive especially for soil science, classification and regression is most interested area for research using various models and methods [1].

M Martins and others from Africa investigated the performance of eight machine learning algorithms for soil mapping a tropical mountainous area of official rural settlement in Zona da mata region. They also mention that satellite images and climatic maps were the part of recursive feature elimination algorithm to predict soil types using machine learning approach. The final performance measured was determine by confusion matrix and Z-test.[2].

In the paper [3], various application of machine learning techniques in agriculture have been listed such as crop selection and crop yielding prediction, weather forecasting, smart irrigation system, Crop disease prediction, deciding the minimum support prize. The research was based on enhance the productivity of field among with a reduction in input efforts of farmers. Along with the advance in machine and technologies used in farming. Useful and accurate information about different matter also play a significant role .Machine Learning providing many effective algorithm which may be useful to determine the relation between crop selection and yielding prediction.

Accuracy of classification is depend on the training performance. The prediction is computed for the accuracy and if the accuracy acceptable then machine learning algorithm deployed and if accuracy is not acceptable then machine learning algorithm is again set for training repeatedly with the prior dataset [4].

In [5] M.S Sucheta and Maya L Pai , researchers in this field has been given a classification model for the area-wise soil fertility . In their work comparison among the soil samples from the different villages has been compared. Their research study was limited to Kerala and parker’s nutrient index was considered for measurement. System was design to measure the PH value and FI (fertility index) using ELM learning classifier .System result focus on the fertility of soil is decreasing using the chemical and with the key objective best accuracy obtain by the researchers is 90%.

In [6] researchers had discussion about implementation of geo-referenced soil fertility map showing distribution of soil fertility and there special variability. Laboratory analysis soil sample was used in work.

III. PROPOSED SYSTEM METHODOLOGY

The main purpose of the proposed work is to create a suitable model for classifying various kinds of soil series data along with suitable suggestion for improving the fertility of soil using the logistic algorithm due to its advantage or power as- It require low computational power, Low complexity, High rate of analysis in scientific data, and does not require scaling of features. Basic idea for use of machine learning is to form a surprised and unsupervised algorithm that can accept training data as input and use statistical data on that training data to forecast an output while updating output as new testing data available.

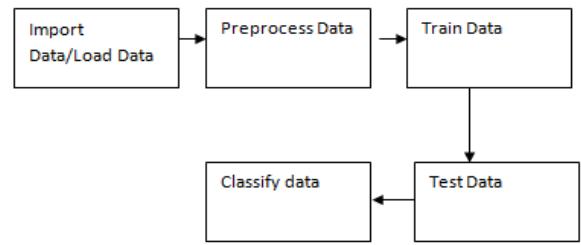


Figure 1: Proposed Model

In the first stage Import dataset with the function Import Dataset has been store in .CVS format and contain numerous values under the proper head. Load and preprocess data: preprocessing is one of the most important and initial stage while solving classification and ML problems.

Train dataset: It is another step where labeled or unlabeled dataset used to train and produce a model. The system accept the trained data for comparison. function used while building the system is-

```

Train = n=pd.read_csv('Soil_train.csv')train.head()
train['soil'].hist(color='brown',bins=40,figsize=(8,4))
  
```

IV. MATHEMATICAL FOUNDATION

In the research work linear regression used, where target variable is categorical in nature. It use log of odds as dependent variables. Logistic regression predicts the probability of occurrence of binary event.

$$y = B_0 + B_1 X_1 + B_2 X_2 + \dots + B_n X_n \tag{1}$$

i.e

$$\log\left[\frac{y}{1-y}\right] \Rightarrow y = c + B_0 + B_1 X_1 + B_2 X_2 + \dots + B_n X_n$$

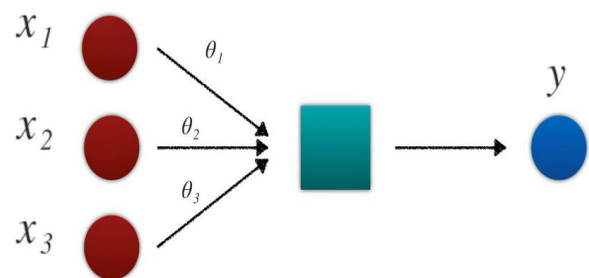


Figure 2:Flow chart for logistic regression

There are two event depend on the conditions. The event ‘y’ given the ‘1’ if the event happens and wise versa if the event does not happen then the ‘y’ gives ‘0’. The value ‘1’ represents the healthy soil and ‘0’ represent the unhealthy soil.

Our main objective is to classify the soil and suggest the respective crop. With this aim we have taken the laboratory tested soil samples and its contents like PH value, Micro nutrients and primary factors to create a

database. Different classes have been created using the factors associated with soil. In this stage of feature selection we need to divide the data with two types of variables as dependent and independent those are features variable. Few samples are trained and some of them are store for testing.

TN (true negative) FN (False negative), TP (true positive) and (False positive) values are used for computing the accuracy .

FPR- False positive rate is the Ratio of samples classified incorrectly ‘FP’ to the actual number of normal records ‘TN’. It is calculated as-

$$FPR = \frac{FP}{TN + FP} \dots\dots\dots eq(2)$$

TPR-True positive rate is the ratio of sample classified correctly for a specified class to the actual total samples of that class. It can be formulated as-

$$TPR = \frac{TP}{TP + FN} \dots\dots\dots eq(3)$$

Detection Rate: It represent the rate of correctly identified considering all instance predicted. In practical setting it may represent as-

$$DR = \frac{TP}{TP + FP} \dots\dots\dots eq(4)$$

Accuracy: Accuracy Value is the ratio of number of sample correctly classified by the system to the total number of total samples . It can be calculated as-

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \dots\dots\dots eq(5)$$

V. RESULT AND CONCLUSION

There are many factors that determine the soil fertility. This depend on the soil classification, Generally, evaluated soil fertility depend on several parameter such as texture , organic matter, soil PH, electrical conductivity, total calcium carbonate, total Nitrogen, C to N ration, available content of phosphorus, potassium , calcium, magnesium and also on the micro nutrients

Logistic regression predicts whether something is True (1)or False(0)instead, predicting something that is continuous like size.

It has an S-shaped line.

Logistic Regression’s ability to provide probabilities and classify new samples using continuous and discrete measurements makes it a popular machine learning method.

Table 1: Performance Analysis

| Researchers | Average Accuracy | Method |
|----------------------------|------------------|-----------------------|
| M.S Sucheta and Maya L Pai | 90% | ELM classifier and NN |
| Proposed Method | 95-98% | Logistic Algorithm |

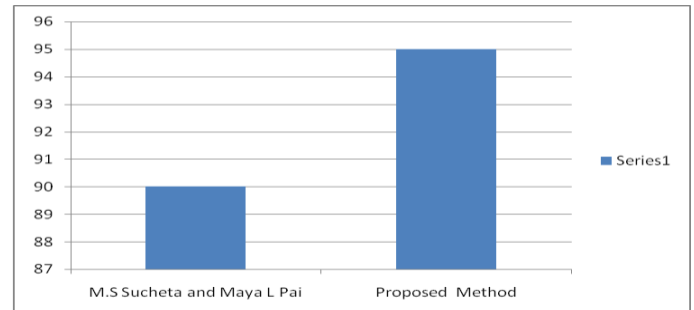


Figure 3: Performance Chart

From the table 1 and graph as shown in figure 3 we can observer that logistic algorithm better perform about 95 to 80 % accuracy for classification of soil. This model is not build as web application . and hence in future we can design it as web application .

REFERENCES

- [1] Pant, Himanshu & C., Manoj & Bhatt, Ashutosh. (2019). Impact of Physico-Chemical Properties for Soils Type Classification of OAK using different Machine Learning Techniques. International Journal of Computer Applications. 177. 38-44. 10.5120/ijca2019919617.
- [2] Meier M, Souza E, Francelino MR, Fernandes Filho EI, Schaefer CEGR. Digital soil mapping using machine learning algorithms in a tropical mountainous area. Rev Bras Cienc Solo. 2018;42:e0170421. <https://doi.org/10.1590/18069657rbcs20170421>
- [3] Jain, N., Kumar, A., Garud, S., Pradhan, V., and Kulkarni, P. (2017). Crop selection method based on various environmental factors using machine learning.
- [4] Bishop, C. M. (2006), Pattern Recognition and Machine Learning, Springer, ISBN 978-0-387-31073-2
- [5] M.S Sucheta and Maya L Pai ‘Improving the prediction accuracy of soil nutrient classification by optimizing the extreme machine parameter’ , information processing in agriculture <https://doi.org/10.1016/j.inpa.2019.05.003> . China agricultural university. Production and hosting by Elsevier B.V on behalf of KeAi. Pp 2214-3197 (2019)
- [6] K.A.Nketia ‘Using soil fertility index to evaluate two different sampaling scheme in soil fertility mapping: a case study of Hvanneyri,Iceland, UNU land Restoration training program (2011).