IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.918 ⁜ Vol. 11, Issue 12, December 2022

DOI: 10.17148/IJARCCE.2022.111247

Soil Properties Prediction Using Machine Learning Algorithm

A. Vineeta[1], Mr. Ramesh Ponnala[2]

MCA IV Sem, Dept. of MCA, Chaitanya Bharathi Institute of Technology (A), Gandipet, Hyd, India[1]

Asst Professor, Dept. of MCA, Chaitanya Bharathi Institute of Technology (A), Gandipet, Hyd, India[2]

Abstract: Information about soil properties help the farmers to do effective and efficient farming, and yield more crops with less usage of resources. An attempt has been made in this paper to predict the soil properties using machine learning approaches. The main properties of soil prediction are Calcium, Phosphorus, pH, Soil Organic Carbon, and Sand. These properties greatly affect the production of crops. Two well-known machine learning models, namely, support vector machine, and K-Nearest Neighbour, are used for prediction of these soil properties. The performance of these models is evaluated on Africa Soil Property Prediction dataset. Experimental results reveal that the Support Vector Machine outperforms the other models in terms of coefficient of determination. Support Vector Machine is able to predict all the soil properties accurately. It will be helpful for the farmers to know the properties of the soil in their particular terrain.

Keywords: Machine learning, soil properties are Calcium, Phosphorus, pH, Soil Organic Carbon, and Sand.

I. INTRODUCTION

Agriculture is one of the main occupation in India. Large population of India depends upon agriculture as their main source of income. With time, the demand for production has been increased exponentially. The availability of accurate and timely information such as soil, meteorological, usage of fertilizers, usage of pesticides can help farmers to make accurate decision as per their need for their cropping situations.

In this project, machine learning algorithms such as Support Vector Machine, and K-Nearest Neighbour with a different degree of accuracy are used for comparative analysis. The dataset is split into a training and testing dataset (80% training data and 20% testing data). The machine learning models are trained on the training data. After a model is trained, the testing data is used to check the accuracy of the trained model. Here, the coefficient of determination (COD) is calculated to check the working of the models after being trained. After training the models, the best working model is deployed to predict the properties of the soil (Calcium, Phosphorous, pH, Soil Organic Carbon, and Sand). These predicted values of the soil properties are going to be helpful in choosing the different suitable fertilizers.

II. LITERATURE SURVEY

Pranay Malik, Sushmita Sengupta and Jitendra Singh Jadon **[1]** performed a comparative analysis of ML algorithms i.e. K Nearest Neighbor algorithm, Naïve Bayes algorithm and Decision Trees classifier for crop yield prediction has been done on self –obtained dataset, for three crops - tomato, potato and chilli,. KNN and Decision Trees have higher accuracy than compared to Naïve Bayes by 14.453% and 18.935% respectively. The proposed work can help farmers to identify the yield of crops in different soil and atmospheric conditions.

Kaushik Chandra Mitra et al. [2] model is proposed for predicting soil series with land type and according to prediction it can suggest suitable crops,for this different machine learning models are used. Two datasets are used: Soil dataset and crop dataset.The research has been done on soil datasets of six upazillas of Khulna district, Bangladesh. Experimental results show that the proposed SVM based method performs better than many existing methods.

Vaishali Pandith and Haneet Kour et al. [3] various machine learning techniques have been implemented in order to predict Mustard Crop yield in advance from soil analysis .To assess the performance of each technique under study; five parameters namely accuracy, recall, precision, specificity and f-score have been evaluated. In this study, KNN and ANN are found to be most accurate techniques for mustard crop yield prediction.

T. Venkat Narayana Rao [4] proposed a model that helps in maintaining the soil fertility consistently. There are three soil parameters that come into consideration when we have to predict the soil quality they are Chemical Parameters ,Physical Parameters and Biological Parameters.From the result it has been observed that Gradient Boosted Decision Trees has high ability to perform even better owing to the heavy dependence on Random Forests.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.918 ⁜ Vol. 11, Issue 12, December 2022

DOI: 10.17148/IJARCCE.2022.111247

Keerthan Kumar was proposed [5] a machine learning based solution for the analysis of the important soil properties and based on that we are dealing with the Grading of the Soil and Prediction of Crops suitable to the land. The various soil nutrient EC, pH , OC etc. are the feature variables. Dataset is preprocessed and regression algorithm is applied and RMSE (Root Mean Square Error) is calculated for predicting rank of soil and we applied various Classification Algorithm for crop recommendation and found that Random Forest has the highest accuracy score.

III. METHODOLOGY

1.Removing irrelevant variables:

Removing the pointless/immaterial factors like PIDN are taken out during the preprocessing stage. These factors in the informational collection could prompt the model not working as expected because of a potential bogus relationship

2.Normalizing data:

Normalization is a method frequently applied as a component of information groundwork for machine learning. The objective of standardization is to change the upsides of numeric segments in the informational index to utilize a normal scale without mutilating contrasts in the scopes of values or losing data. Standardization is additionally expected for certain calculations to accurately display the information.

3.Dividing data into train and test sets:

The train-test part technique is utilized to gauge the exhibition of ML calculations. They are utilized to make forecasts on information. It is a quick and simple methodology to play out, the consequences of which permit you to think about the execution of ML calculations for your prescient displaying issues portraying the video content naturally has two phases. The principal stage is grasping the items. This spotlights on visual acknowledgment with profound learning models and concentrates the entertainer, activity, and the object of the activity (for example human and movement location) from the video cut. The video cut is taken care of as a progression of edges that are considered as pictures. Thus, we have a progression of edges in each clasp that are input pictures. Then, at that point, the separated data from the clasp is placed in a typical component vector. This vector is taken care of into the subsequent stage. The subsequent stage is subtitle age which portraying is removed in a linguistically right normal language sentence, in this manner planning the items distinguished in the main stage. Here, we bring a mix of profound learning designs for the encoding and disentangling stages.

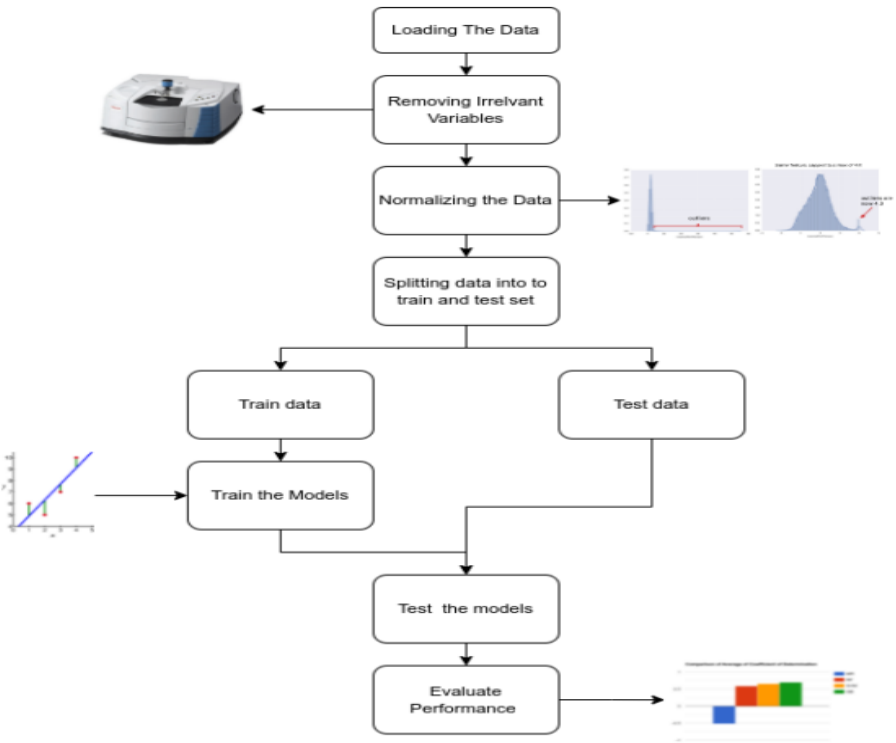

Figure 1: Methodology Used for Soil Prediction

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified ⁜ Impact Factor 7.918 ⁜ Vol. 11, Issue 12, December 2022
DOI: 10.17148/IJARCCE.2022.111247

IV. RESULTS AND DISCUSSION

In this part, the presentation assessment of ML models is assessed.

V. CONCLUSION

This paper concentrated on the AI methods to anticipate soil properties for accuracy farming. Four AI methods were utilized to assess the dirt properties like Calcium, Phosphorus, pH, Soil Organic Carbon, and Sand. These strategies were prepared and tried on the Africa Soil Property Prediction dataset. It is seen from the outcomes that Support Vector Machine performed better compared to different strategies. Support Vector Machine had the option to anticipate all properties better than K-Nearest Neighbor. It tends to be seen that there is a possibility to involve spectroscopy as an elective strategy for soil part examination. Profound learning and crossover models might be utilized for anticipating soil properties in a compelling and proficient way. The principal constraint of our review is the utilization few soil parts for expectation. This study can be stretched out by utilizing an enormous dataset and different models.

REFERENCES

[1] Pranay Malik, Sushmita Sengupta and Jitendra Singh Jadon, "Comparative Analysis of Soil Properties to Predict Fertility and Crop Yield using Machine Learning Algorithms ." IEEE,2021.
[2] Sk Al Zaminur Rahman, Kaushik Chandra Mitra, S.M. Mohidul Islam, "Sk Al Zaminur Rahman Kaushik Chandra Mitra S.M. Mohidul Islam." 2018 21st International Conference of Computer and Information Technology (ICCIT).
[3] Vaishali Pandith, Haneet Kour, Surjeet Singh, Jatinder Manhas and Vinod Sharma, "Performance Evaluation of Machine Learning Techniques for Mustard Crop Yield Prediction from Soil Analysis." Volume 64, Issue 2, 2020 Journal of Scientific Research Institute of Science.
[4] T. Venkat Narayana Rao, "Prediction Of Soil Quality Using Machine Leaning Techniques." International Journal Of Scientific & Technology Research Volume 8, Issue 11, November 2019.
[5] Keerthan Kumar T G, Shubha C, Sushma S A, "Random Forest Algorithm for Soil Fertility Prediction and Grading Using Machine Learning." International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-9 Issue-1, November 2019.