iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** III **Month of publication:** March 2022

DOI: https://doi.org/10.22214/ijraset.2022.40864

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue III Mar 2022- Available at www.ijraset.com

Soil Testing Prediction System

Rachit Sharma[1], Tushar Mittal[2], Ritik Chauhan[3], Dr. Ranjeet Kumar[4]
[1, 2, 3, 4]*Meerut Institute Of Engineering and Technology*

Abstract: ***Soil Testing Prediction is aimed to predict the soil functional properties (Calcium, Phosphorus, pH, Sand and Soil Organic Carbon) of a soil sample. Soil Testing Prediction finds its application in the field of agriculture, farming and research. It can help in economic crop management and better yield of crops. We have worked to find out how traditional soil testing methods can be replaced with modern Machine Learning techniques that can result in more economic , time efficient methods with no or little to no adverse effects on the environment. It trains to reduce the technical expertise required at users end, and aims to bring the labs to the user instead of taking the user to the lab.***

Keywords:

1) ***Linear Regression: Linear Regression is one of the many statistical techniques that has been adopted by the Machine Learning Society. It is a supervised learning technique i.e It works on labeled data. It assumes that there exists a linear relationship between the dependents and predictor. Although preemptive, the technique proves out to be at par with many ML techniques.***
2) ***Feature Selection: Feature Selection is the method of extracting out the useful features from the set of all available features, that can help us to predict the required targets . It aims to reduce the error in the prediction made by the model.***
3) ***Soil Functional Properties: Functional properties are the properties that define the behavior of the soil and its response to the environment and surroundings.***
4) ***Mehlich-3 Extraction Techniques: It is a chemical method to predict the value over a range of elements. It is a weak acid soil extraction procedure. The extract is composed of 0.2 M glacial acetic acid, 0.25 M ammonium nitrate, 0.015 M ammonium fluoride, 0.013 M nitric acid, and 0.001 M ethylene diamine tetraacetic acid.***

I. INTRODUCTION

In these modern days agricultural land is shrinking day after day so it is the need of the hour to test the soil before beginning with the agricultural activity to achieve best possible yield. The preemptive methods known have proved to be time consuming, costly and not so accurate. Hence, in this fast paced world we need a better testing system that can help us to overcome the mentioned issues.

In the near future, we need to explore new areas for farming due to the enormous development in the country area and we need to check the quality of our soil to make the best products. So, here comes the role of soil testing prediction which will help in:

1) Determining whether a particular type of soil would be good enough to use.
2) Predicting the best possible assessment of the soil's fertility to make fertilizer recommendations.
3) The diagnosis of plant problems and in the quality plant production etc.
4) The measurement can be typically performed in seconds, instead of the preemptive methods that are being used , they are slow and cost a hefty amount of money.

This soil testing could be done with the help of the traditional methods. These methods are generally known as WET Tests, that makes the use of some chemicals and are quite time consuming and leads to make these tests quite expensive ones. The reason for these tests to be that expensive is, there high cost of test kit as well as the requirement of professional expertise. So, here comes the role of machine learning. In order to train our machine learning models , we will use Africa soil Infrared Spectroscopy data with dimension 3600 features and 1157 samples. This data will solely help us to train and cross validate the data , Apart from that the data will be tested upon on a new unseen data to avoid overfitting .

II. BACKGROUND

A. Traditional WET Test

WET tests are the traditional testing methods that use chemical based techniques they are time consuming and expensive. Moreover they require high technical expertise and a slight deviation from ideal conditions can result in high error. This problem can be solved using automation and helps us to remove humans out of equation which results in more effective ways.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue III Mar 2022- Available at www.ijraset.com

B. Possibility of Kit

The existing possible method till now to test the soil is the WET Test, that could be done with the help of WET Testing kit. The major problem with this kit is that it requires a high level of expertise in order to use this kit as well as, this kit is quite expensive so this would not be available to be used for its direct consumers.

So the idea to use the kit directly by the consumers needed to be replaced with our proposed Machine Learning Model that would not require that much technical expertise as well as that much large amount of money by the direct consumers. They could easily use our friendly User Interface for the prediction of required content. This model will enable hassle free procedures in order to get the required values and these values will help them to make better decisions quickly and generate the best out of it.

Another possible aspect could be to use IOT based infrastructure that can generate onsite real time data. That can be piped into real time processing systems and resulting in real time predictions.

C. Limitations

1) Current methods used for the soil testing can't be directly used by most of the small scale end consumers as well as some of the large consumers too.
2) This current method for the soil testing prediction can't be easily used by the novice.
3) This current method used for the soil testing can't exist without making use of harmful chemicals, that would anyhow cause environmental disruption.
4) Conventional Methods these days require ample amount of time to generate the test results , Considering the size of indian consumer market this time constraint proves out to be crunching.

III. PROPOSED METHODOLOGY

The Project aims to predict the following :

1) SOC: Soil organic carbon
2) pH: pH values
3) Ca: Mehlich-3 extractable Calcium
4) P: Mehlich-3 extractable Phosphorus
5) Sand : It is the percentage of sand present within soil.

IV. DATA WRANGLING

Wrangling refers to cleaning data i.e making it more normal and removing any outlier or unscaled features.

A. Outlier Removal

Outliers are the data points that do not follow the general trend within the data i.e They will impact the result in the wrong manner and fluctuate the result from the actual one and can increase the error i.e deviation from the predicted behavior..

- Scaling
- Normalization

$$x_{scaled} = \frac{x - x_{min}}{x_{max} - x_{min}}$$

Where
x is the sample point
$x_{min}$is the minimum value over the domain
$x_{max}$is the maximum value over the domain

B. Dimensionality Reduction

Since the dimension of the data set were high hence in order to reduce it to a feasible number PCA is used.

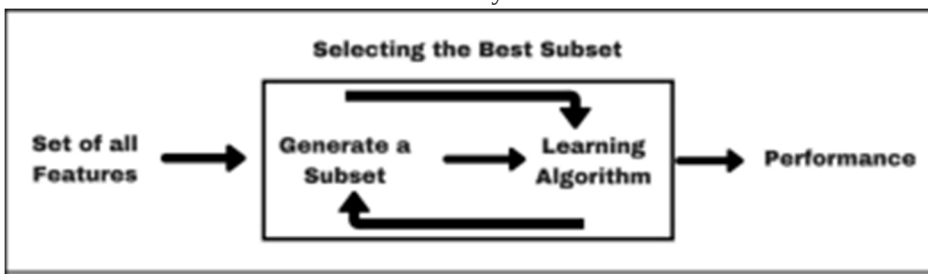
Principal component analysis is used for reducing the dimensionality(Feature set) of a dataset , it creates a projection of a higher dimension feature set to a lower dimension it generates a set of new features with higher variance resulting in better predictive power as expected it makes us loose some amount of knowledge as compared to original dataset.

C. Feature Selection

For extracting the required number of features (50 Features to be precise)we've used recursive feature elimination and Random forest Regressor.

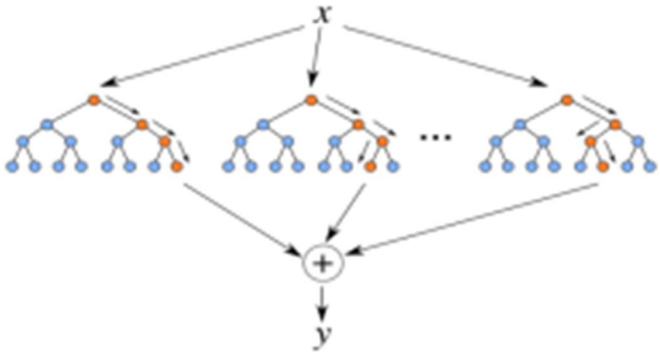
D. Recursive Feature Elimination

Recursive feature elimination (RFE) is a feature selection method it helps us to get the best subset of features from the feature set. To do so , it fits the model on the given data and tries to remove the least significant feature one at a time. The process is repeated until the desired number of features is reached or the desired accuracy is achieved.



E. Random Forest Regressor

A random forest regressor as the name suggest it uses a number of Random trees in order to generate a random forest based on the random forests regression scores it tries to fetch out the best possible features for the model in order to predict the targets, it is one of the most popular feature selection technique it is available for bout classification task and regression task . It trains each tree on the subset of a given dataset and increasing accuracy by sampling.



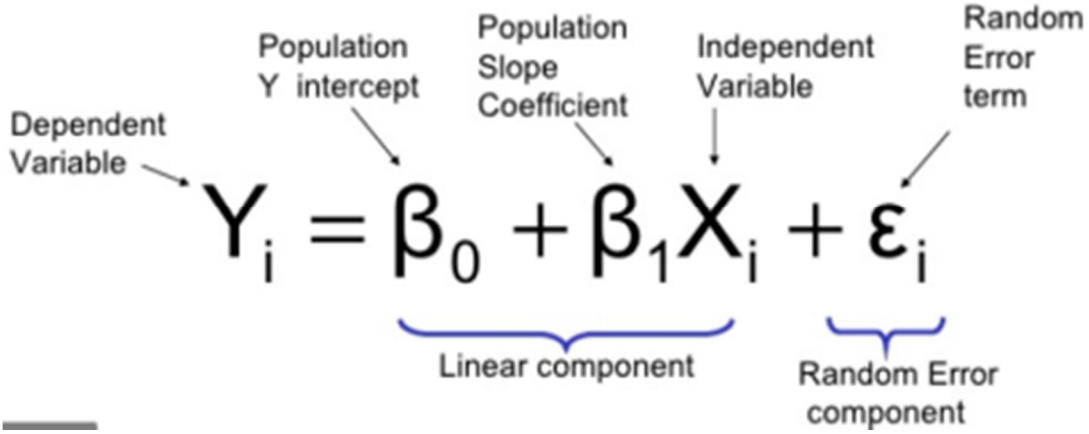
V. MODEL TRAINING

For making model we've used two algorithms namely ordinary least square regression and light gradient boosting machine.

A. Ordinary Least Square Regression

Ordinary Least Square Regression, also known as simple linear regression, is one of the many techniques that have been adopted from the field of statistics in the field of machine learning . It works on the assumption that there is linear relation between the target value and the features. It estimates the best fit line for the feature - target relation, although simple but this technique has proven to give exceptional accuracy.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue III Mar 2022- Available at www.ijraset.com



B. Light Gradient boosting Machine

LightGBM, short for Light Gradient Boosting Machine, is a free and open source distributed gradient boosting framework for machine learning originally developed by Microsoft. It is based on decision tree algorithms and used for ranking, classification and other machine learning tasks.

VI. ACCURACY MEASURES

The Predictions are scored upon Mean Column wise root mean squared error (MCRMSE).

$$\text{MCRMSE} = \frac{1}{5}\sum_{j=1}^{5}\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_{ij} - \hat{y}_{ij})^2},$$

VII. RESULT DISCUSSION

Accuracy for OLSR - 0.83577
Accuracy for LGBM - 0.46892

MCRMSE (mean columnwise root mean squared error)

```
[ ]  pans = 0.0
     for i in range(len(pres)):
       pans  = pans + (pres[i] - y_test_P.values[i])**2

     socans = 0.0
     for i in range(len(socres)):
       socans  = pans + (socres[i] - y_test_SOC.values[i])**2

     phans = 0.0
     for i in range(len(phres)):
       phans  = phans + (phres[i] - y_test_pH.values[i])**2

     sandans = 0.0
     for i in range(len(sandres)):
       sandans  = sandans + (sandres[i] - y_test_Sand.values[i])**2

     MCRMSE = (((ans+pans+socans+phans+sandans)/382)**0.5)/5

[ ]  MCRMSE

     0.835778088242345
```


VIII. CONCLUSION

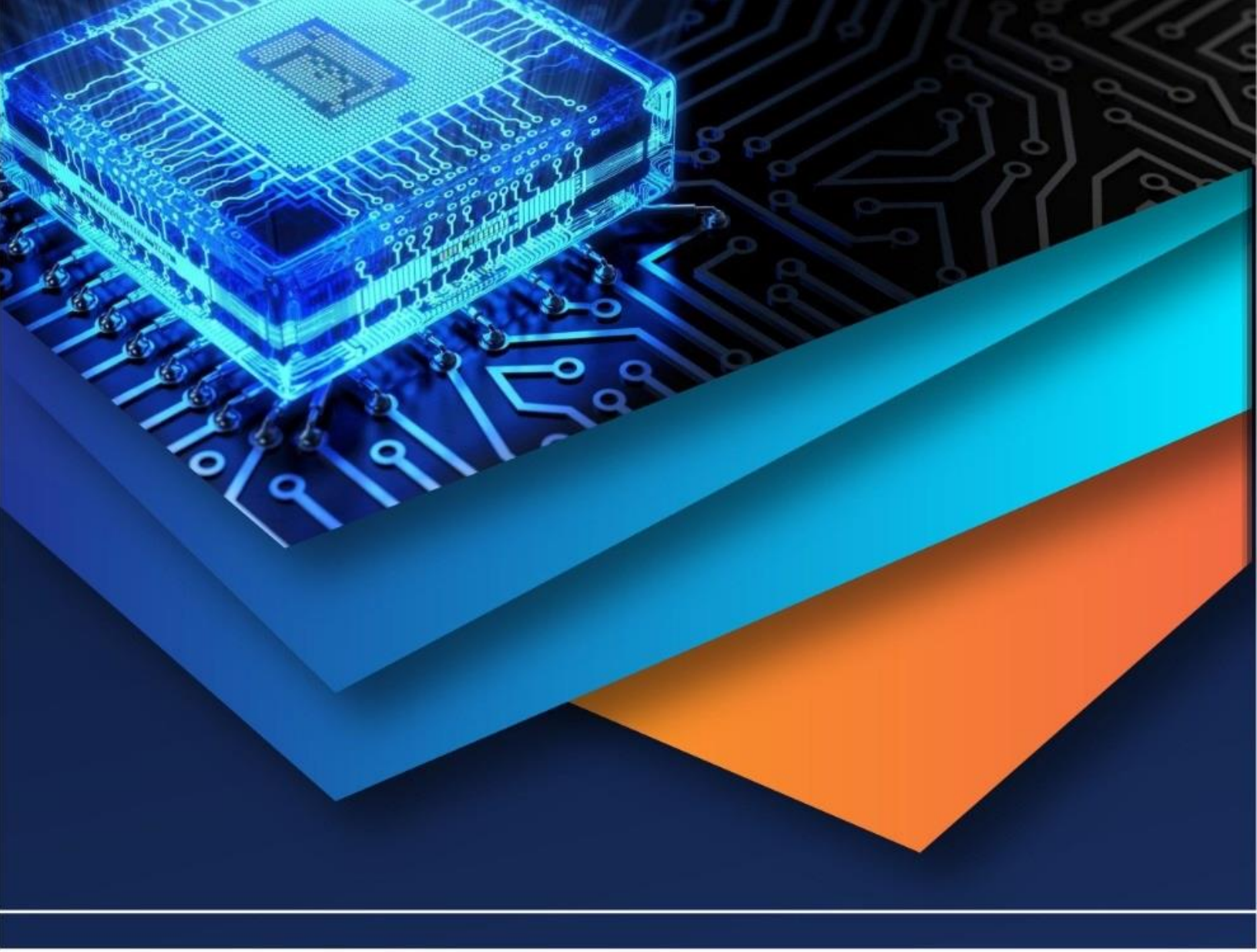
As we all know that soil testing has been an essential need for today's world. But doing that with the help of traditional methodology takes almost 5 to 6 business days while our model proposes to give the almost same result within seconds with a Mean column wise root mean square error of 0.83577 and 0.46892 for OLSR and LGBM respectively.

IX. ADVANTAGES

1) Cost Efficient
2) Fast processing and predictions in real time
3) Better Accuracy
4) Minimal or no use of Chemicals

REFERENCES

[1] "Carge`le Nduwamungu , Noura Ziadi , Le´on-E´tienne , Gae¨tan F. Tremblay , and Laurent Thurie`s (2009) Opportunities for, and limitations of, near infrared reflectance spectroscopy applications in soil analysis: A review"
[2] sklearn.linear_model.LinearRegression
[3] "Soil Analysis using Mehlich 3 Extractant Technique for Sample Preparation . ÚKZUZ (Ústřední kontrolní a zkušební ústav zemědělský) Central Institute for Supervising and Testing in Agriculture Hroznová 2, CZ-65606 Brno, Czech Republik L. Vlk, M. Horová, R. Krejča; R. Špejra Chromservis S.R.O., Jakobiho 327, CZ-10900 Praha-10, Petrovice, Czech Republik"
[4] https://flask-doc.readthedocs.io/en/latest
[5] https://docs.aws.amazon.com/ec2/index.html



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY