Original Article

Physical Analysis Of Soil By Using Machine Learning Algorithm In Kanpur Nagar (U.P.)

[1]Mr. Avinash Bajpai*,[2] Dr. Jitendra Nath Shrivastava, [3]Dr. Swati Saxena, [4]Dr. Pratibha Mishra

[1]Research Scholar, INVERTIS UNIVERSITY, Bareilly (UP), India
[2] Professor, INVERTIS UNIVERSITY, Bareilly (UP), India
[3] Assistant Professor, Maharana Pratap Group of Institutions,Kanpur,India
[4] Assistant Professor, DAV PG COLLEGE,Kanpur,India
*Corresponding Author's Email ID:avinashbajpai151@gmail.com
Subject Area: Computer Science and Engineering
DOI: 10.47750/pnr.2023.14.04.12

Abstract

In India, Agriculture is playing important role in growth of the economy. It is a non technical sector but today we can also apply technology that improve the agriculture growth .We fulfill the requirement of agriculture and fast implementation using Artificial Intelligence. The agriculturalist usually follows a way that called crop transformation successive to produce to good result. Using of technology, we increase the growth and achieving good result in precision agriculture. We have designed a model that achieves for soil parameter. Analyze the quality of soil using machine learning algorithm we can predict the quality of soil using various parameters. Regularly testing of soil is very tedious task. So requirement of automate the process, we are using Machine leaning approach to analyze the soil dataset and visualize of various parameter that would help to take decisions. Using machine learning algorithm, we have Design a model that is monitoring the physical component of soil in a proper way for different area of Kanpur. In our model, we collect the soil data sample from different area of the Kanpur that use for machine learning algorithm to analyze the physical properties. In physical property ,we analyze the pH and humidity data of soil using machine learning algorithm .here we use 30%testing data and 70% training data for analysis in random forest algorithm .this algorithm is find the model accuracy and error values.

Keywords: Machine learning, soil data, physical properties, transformation, Artificial Intelligence

INTRODUCTION

Backbone of Indian economy is based on the agriculture. But farmer of India is suffering lot of difficulty .their uncomfortable situation of our agriculturalist are increased dept and poor result. Lack of information our former have not know about soil quality and composition. Which is very useful to predict the crop for the particular soil. proper management of soil that increase the agriculture productivity and ecological balance. Soil word has been derived from Latin word solum which means earthly material in which all the living things survives on the inside and outside of the periphery. Soil is the most vital natural resource, can be termed as 'Soul of in (Naphade et al., 2021) [5]. finite life' and it is the exclusive source of infinite living organisms which supports the life of crop plants by acting as a medium for growth along with providing nutrients, air, and water (Rajshri et al., 2021) [6]. In order to make accurate decisions based on the type of crops to be planted and to achieve a good harvest, data such as use of pesticides, fertilizers, meteorological, and soil Hindawi Journal of Nanomaterials Volume 2022, Article ID 5343965, 7 pages https://doi.org/10.1155/2022/5343965 information must be made available to the farmers in an accurate and timely manner [1].

Proper knowledge and information about soil nutrition and components can improve the agriculture productivity. It provides support to manage the crop. Analysis of soil is very important to decide which crop is better for productivity as per requirement of farmers. Analysis of soil reduces the damage and loss of crops that occur due to unfavorable conditions. Soil physical components analysis is very difficult task in lab regular bases which is traditional process.

Here we use machine learning algorithm in computer to analyze the unstructured data of soil .and predict the accuracy of model and data on the bases of analytical analysis. It trains the machine on the bases of testing data and improves the performance of the system also reduce the chances of error. High accuracy provide by the system, so error chances very less in this analysis. The soil physical components accuracy is measure through the system that help to farmer for cultivation of the crop. We find the accuracy of ph and humidity for soil using machine learning algorithm. To appropriate information about soil physical components that helps the loss of crop or wastage of money on the view of farmer.

About 16-20% of total crops produced in India suffer wastage annually according to survey [3].Analyze the physical components of soil are depends on different factor pH and humidity. Soils can be classified according to their pH value: 6.5 to 7.5 –neutral,Over 7.5 alkaline. A lot of soil moisture sensors make available measurement in the unit "water fraction by volume". Different machine algorithms are available to analyze the physical property of soil. Here we are using Random Forest machine algorithm, which belongs to the supervised learning. Supervised learning is a subcategory of AI and machine learning. In this algorithm we use labeled dataset for training purpose and 25% to 30% data use for testing purpose and classify the data or predict outcomes accurately. . In terms of soil classification, the highest accuracy is observed with SVM classification in this work. SVM, Naïve Bayes, decision trees, neural networks, and several other ML algorithms are used by Pramudyana et al. [8]. Organic agricultural crop protection has been the major goal in the research proposed by Patil and Umarji [9]. Various scientific features and parameters are used for grading and classification of soil samples using the approach proposed by Ashwini et al. [10]. A CNN algorithm is used for prediction of various crop yields in a precision agriculture model by Alex and Kanavalli [11].

PROPOSED SYSTEM WITH REQUIREMENT SPECIFICATION

The proposed system using machine learning algorithm which is the subset of Artificial Intelligence. It is uses statistical technique to analyze of data .different machine learning algorithm are available to improve the model by using number of iterative process that gradually increase efficiency of the system and we calculate machine accuracy and predict soil physical component with less human intervention.

Using machine learning algorithm we can handle multi dimensional and different variety of information very quickly and fast. It gives result in real time mode. Complicated task can be easily handled.

Machine Learning: Machine learning provide concept that system is learn and improve by training and experience .

Supervised Learning: In this machine is trained by data and already provide question and answer to train the machine and similar type question is raise and machine give output that on the basis store answer.

Example of one such type is Chabot, [2] where the questions and answers are given while training and if the similar queries are asked the given answers would be the output in the form of text or audio.

Unsupervised Learning: in this learning we have the input data to machine but output of corresponding input data is not known. This is learning process without guidance. Output help to decide some specific conclusion. Example visual recognition and text recognition where text data is given and conclusion on related content search.

Reinforcement Learning: In this type of learning after giving the data input ,process and output from that received from the machine ,some feedback is taken from the output ,this data again stored as input. The question we raise is the input and the feedback/credit system (+,-) is the output to learn. Example : automated car system where car learn driving by their own experience and reward or feedback on training model.

The system was use following module for evaluation of data:

a. Dataset Module: The dataset consist of raw data that is stored in excel sheet in tabular format. Raw data converted into comma separator value(CSV file). That file imported in python.

b. Training Module: The training module is essential part of the system it train the system using machine learning algorithm and to perform specific task without human interference

c**. Testing Module:** The testing module is used to test the training model is verify or correct as per requirement of task, if some modification are required to in training module for accurate output,it can also be check by this module. Several graphs and matrix forms are used to make the user understand [4].

System predicts the accuracy of pH and humidity of soil by using Random forest algorithm, thus we analyze the accuracy of pH and humidity, it would help the formers to add fertilizer in accurate range and to make the production better.

Object-oriented evaluation and design methods become the most widely used computer system design techniques. [7].

METHODOLOGY

Here we are using Random Forest machine algorithm, which belongs to the supervised learning. Supervised learning is a subcategory of AI and machine learning. In this algorithm we use labeled dataset for training purpose and 25% to 30% data use for testing purpose and classify the data or predict outcomes accurately. These parameters play a vital role in growth and development of crop and its production. The following parameters are used as variables. Here we apply data that is collected from lab, that data apply for learning algorithm i.e. (Decision Tree, Random Forest) after processing that we find the model accuracy and error. We choose the training data is 70% and testing data is 30%, and apply number of iteration that provide best model. The output quality is given as a percentage based on the inputs.

Step 1: A soil data collect from lab is stored in excel file and converted into comma separated file using pandas in python .

Step 2: csv file further process,it is use for creating dataset ,further dataset is divided into two part training data and testing data. About 65% of dataset is use as training data and 35% of dataset is use as testing data. Further using RandomForestRegressor algorithm to find accuracy of the model and mean absolute error.

SYSTEM IMPLEMENTATION

The project has executed and show the accuracy of the system in this stage and system fulfills the requirement of the user. The stage starts after the user has tested [4] the scheme and approved it.

Algorithm: Random forest algorithm is related to machine learning algorithm .it is used in supervised learning concept .The classification and regression problem can be easily solved by random forest. It is based on ensemble learning. In this concept, combining multiple classifiers to solve complex problem and to improve the performance of the model. it contains a number of decision tree from the subset of the input dataset and take average that improve the projecting accuracy of the dataset. it take prediction from each tree ,instead of one tree, and based on the majority votes of prediction, it forecast the final output. The number of tree is greater in the random forest that generate more accuracy and high performance and reduce the over fitting problem.

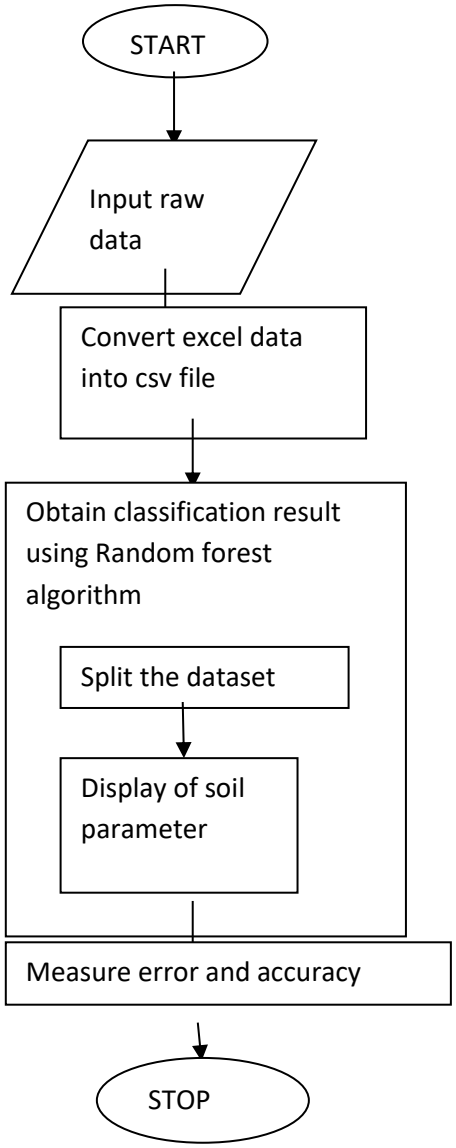

Fig 1: Flowchart of the proposed system

Step by step follow the process

1. Finding the problem application
2. Collecting dataset
3. Load and summarize dataset
4. Segregating Dataset into X & Y
5. Splitting Dataset to Train & Test
6. Algorithm Random Forest
7. Training our Model for preprocessed Dataset
8. Evaluation/Validation
9. Prediction(Observing how our model is classifying our new data)

Table 1:pH value of Soil in selected sites of Kanpur

Year	Month	IIPR	TIKRA	CSA	BITHOOR
2021	9	8.6	8.5	8.7	8.6
2021	10	8.6	8.3	8.5	8.7
2021	11	8.4	8.5	8.5	8.4
2021	12	8.1	7.7	8.2	7.9
2022	1	8.2	7.9	8.1	8.2
2022	2	8.2	8.1	8.4	8
2022	3	8.4	8.3	8.3	8.6
2022	4	8.8	8.4	8.6	8.7
2022	5	8.9	8.5	8.8	8.8
2022	6	9	8.6	8.9	8.9
2022	7	8.8	8.4	8.7	8.6
2022	8	8.4	8.3	8.6	8.5

Table 1A:Standard deviation of pH sep 21 to Aug 22

	Year	Month	IIPR	TIKRA	CSA	BITHOOR
count	12.000000	12.000000	12.000000	12.000000	12.00000	12.000000
mean	2021.666667	6.500000	8.533333	8.291667	8.52500	8.491667
std	0.492366	3.605551	0.296444	0.267848	0.24168	0.311764
min	2021.000000	1.000000	8.100000	7.700000	8.10000	7.900000
25%	2021.000000	3.750000	8.350000	8.250000	8.37500	8.350000
50%	2022.000000	6.500000	8.500000	8.350000	8.55000	8.600000
75%	2022.000000	9.250000	8.800000	8.500000	8.70000	8.700000
max	2022.000000	12.000000	9.000000	8.600000	8.90000	8.900000

Training Features Shape: (8, 6)
Training Labels Shape: (8,)
Testing Features Shape: (4, 6)
Testing Labels Shape: (4,)
Mean Absolute Error: 0.18 degrees.
Accuracy: 98.0 %.

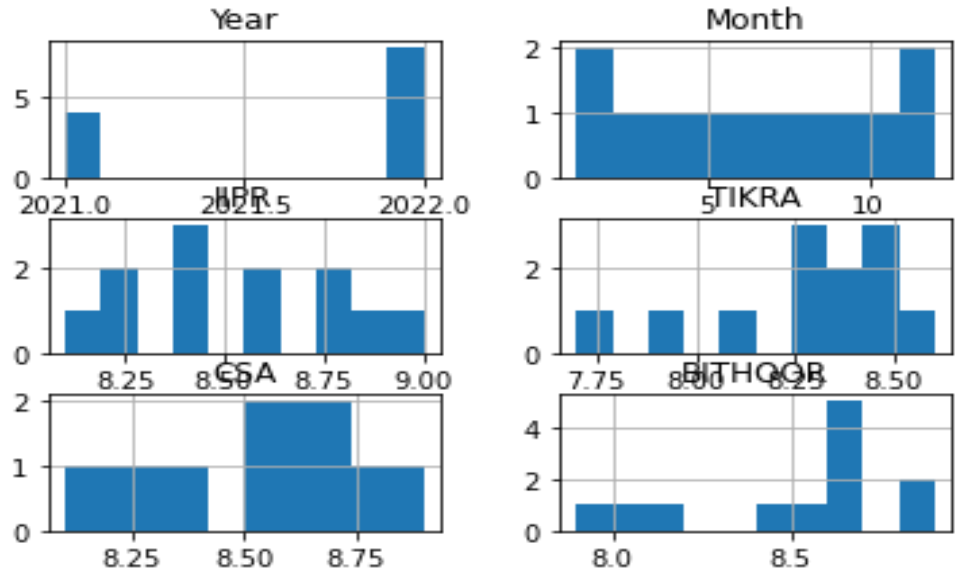

Fig 2: Histograph for pH value of Soil in selected sites of Kanpur

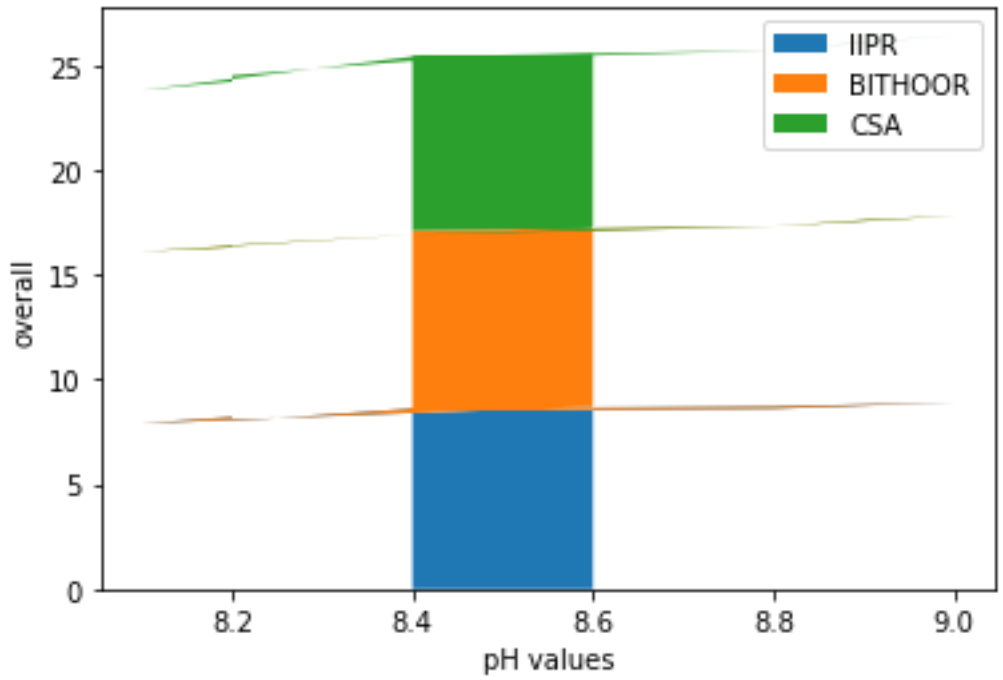

Fig 3: Stackplot for pH value of Soil in selected sites of Kanpur

Table 2:Humidity value of Soil in selected sites of Kanpur

Year	Month	Bithoor	Tikra	IIPR	CSA
2021	9	59	57	58	57
2021	10	55	53	52	54
2021	11	55	54	53	52
2021	12	50	51	50	49
2022	1	48	47	47	46
2022	2	47	46	47	45
2022	3	46	45	45	44
2022	4	42	44	43	41
2022	5	40	41	40	39
2022	6	35	36	34	33
2022	7	65	62	63	61
2022	8	64	63	62	60

Table 2 A:Standard deviation of Humidity sep 21 to Aug 22

	Year	Month	Bithoor	Tikra	IIPR	CSA
count	12.000000	12.000000	12.000000	12.000000	12.000000	12.000000
mean	2021.666667	6.500000	50.500000	49.916667	49.500000	48.416667
std	0.492366	3.605551	9.375985	8.262273	8.712582	8.680787
min	2021.000000	1.000000	35.000000	36.000000	34.000000	33.000000
25%	2021.000000	3.750000	45.000000	44.750000	44.500000	43.250000
50%	2022.000000	6.500000	49.000000	49.000000	48.500000	47.500000
75%	2022.000000	9.250000	56.000000	54.750000	54.250000	54.750000
max	2022.000000	12.000000	65.000000	63.000000	63.000000	61.000000

Training Features Shape: (7, 6)
Training Labels Shape: (7,)
Testing Features Shape: (5, 6)
Testing Labels Shape: (5,)
Mean Absolute Error: 1.64 degrees.
Accuracy: 96.96 %.

Fig 4:Histograph for Humidity value of Soil in selected sites of Kanpur

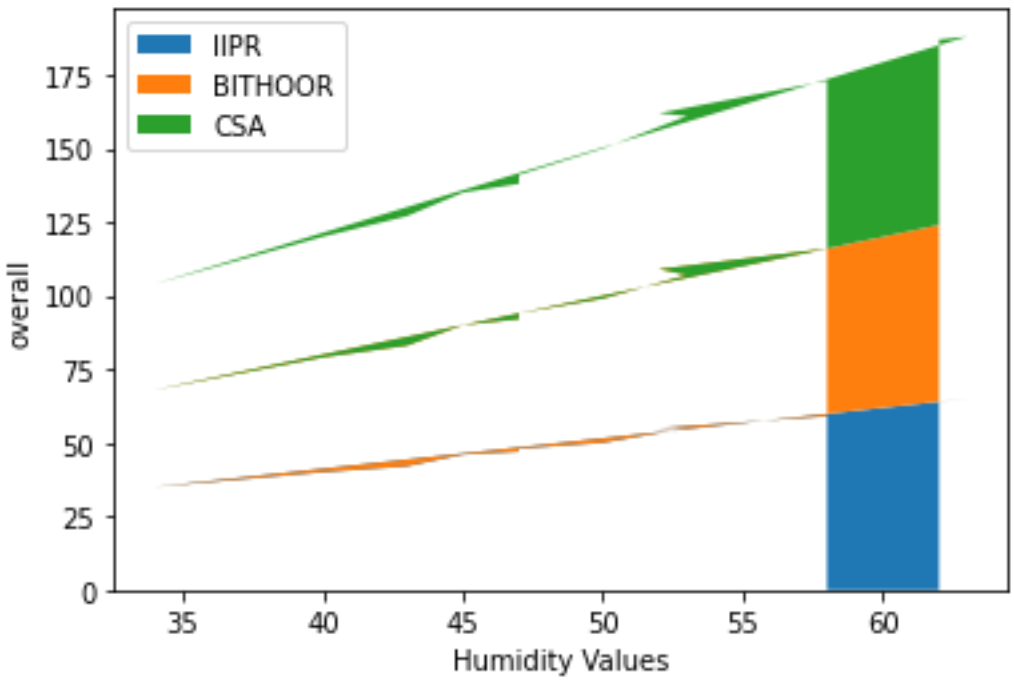

Fig 5: Stackplot for Humidity value of Soil in selected sites of Kanpur

RESULT AND DISCUSSION

Data is used as an input after applying algorithm, we analyzed the quality of soil is based on the input data, that is given an accuracy of the pH of soil is 98% and accuracy of humidity of soil 96.96% which is good accuracy for accepting the model .Mean absolute error is 0.18 degrees for pH of soil .similarly mean absolute error is 1.64 degree for humidity of soil. In statistics, mean absolute error (MAE) is a measure of errors between paired observations expressing the same phenomenon. Example of Y versus X include comparison of predicted verses observed ,subsequent time verses initial time, and one technique of measurement verses an alternative technique of measurement. Hence we conclude that physical analysis of soil using machine learning algorithm predict more accurate result.

REFERENCES:

[1] G. Pan, F.-m. Li, and G.-j. Sun,"Digital camera based measurement of crop cover for wheat yield prediction," in 2007 IEEE International Geoscience and Remote Sensing Symposium, pp. 797–800, Barcelona, 2007.

[2] P. Vinciya, Dr. A. Valarmathi, ―Agriculture Analysis for Next Generation High Tech Farming in Data Mining‖ IJARCSSE,vol. 6, Issue 5, 2016.

[3] T. Islam, T. A. Chisty, and A. Chakrabarty, "A deep neural network approach for crop selection and yield prediction in Bangladesh," in IEEE region 10 humanitarian technology conference (R10-HTC), pp. 1–6, Malambe, Sri Lanka, 2018

[4] Sabri Arik, Tingwen Huang, Weng Kin Lai, Qingshan Liu , ―Soil Property Prediction: An Extreme Learning Machine Approach‖ Springer, vol. 3, Issue 4,666-680,2015.

[5]Naphade M, Sidhu GS, Patil VD, Shinde R. Assessment of Physico-chemical properties and micro nutrient status of Jalgaon district, Maharasthra states. International Journal of Current Microbiology and Applied Sciences 2021;10(03):52-59.

[6]. Rajshri S, Sidhu GS, Patil VD, Naphade M. Evaluation of Physico-Chemical properties in Hot Semi-arid Ecoregion of Beed district, Maharashtra. International Journal of Current Microbiology and Applied Sciences 2021;10(03):45-51.

[7] http://agricoop.nic.in/sites/default/files/Annual_rpt_ 2 016 17_E.pdf

[8] P. A. Harlianto, T. B. Adji, and N. A. Setiawan, “Comparison of machine learning algorithm for soil type classification,” in 2017 3rd International Conference on Science and Technology-Computer (ICST), pp. 7–10, Indonesia, 2017.

[9] M. Patil and I. R. Umarji, “Identification of crop diseases using deep learning,” International Journal of Research in Engineering, Science and Management, vol. 2, no. 6, 2019.

[10] A. Rao, A. Gowda, and R. Beham, “Machine learning in soil classification and crop detection,” IJSRD-International Journal for Scientific Research and Development, vol. 4, no. 1, pp. 792– 794, 2016.

[11] S. A. Alex and A. Kanavalli, “Intelligent computational techniques for crops yield prediction and fertilizer management over big data environment,” International Journal of Innovative Technology and Exploring Engineering (IJITEE), vol. 8, no. 12, 2019