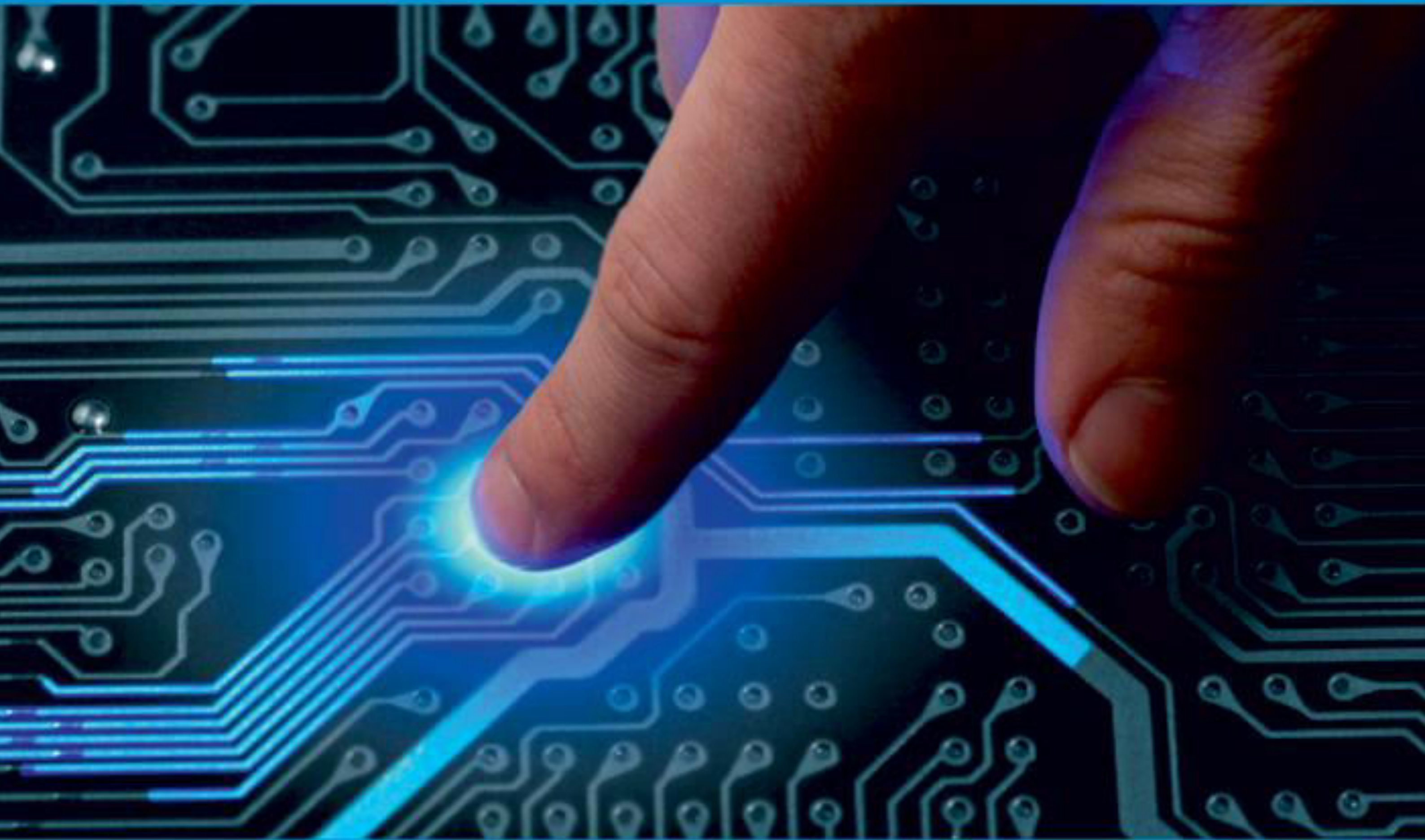
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 10, Issue 5, May 2022

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.165

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 5, May 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1005318 |

Prediction of Soil Moisture Using Random Forest Classifier

Elika Ajay[1], Chilaka Karuna Kumari[2], Bollimera Grace Bhavana[3], Chandan Jaladi[4], Sk. Wasim Akram [5]

U. G. Student, Department of Computer Science and Engineering, Vasireddy Venkatadri Institute of Technology, Nambur, Andhra Pradesh, India. [1, 2, 3, 4]

Assistant professor, Department of Computer Science and Engineering, Vasireddy Venkatadri Institute of Technology, Nambur, Andhra Pradesh, India. [5]

ABSTRACT: India is a country in which a vast number of people depended on agriculture for their livelihood. So, soil and its nutrients play a vital role in cultivating different kinds of crops. Soil is very fertile in India and supports a high number of crops for cultivation. Our motto is to predict the moisture content in the soil which is very important in Agricultural sector activities. In agricultural lands, soil moisture affects important farming activities from crop selection to time of tilling, planting, usage of fertilizer, and harvesting due to infiltration, evaporation, runoff and heat etc. By predicting the moisture content in soil will help farmers know when to irrigate the crop that means groundwater also play a key role in soil moisture and by considering each and every factor in our sight we have different machine learning models like Decision Tree and Random Forest Algorithm which are trained using the database that contain state wide collection of different sources of how the groundwater is being recharged that might be due to the rainfall, household activities, lakes, rivers, considering monsoon and non-monsoon seasons which was surveyed by central groundwater resource department and classified moisture in soil by taking situation feature which contains four labeled values defined as excess, moderated, semi-critical, critical. So, by applying this data as a training dataset to the machine learning model Random Forest Classifier to get the model to predict the moisture content in soil and the performance of this model is evaluated on basis of Mean Squared Error (MSE) which is recorded as 0.0038, Root Mean Square Error (RMSE) which is 0.062 and test scores. After applying the Random Forest Classifier model to the dataset which shows 0.937500 of test accuracy which is the relatively best method to predict the moisture compared to the other machine learning models.

KEYWORDS: Decision Tree, Random Forest Classifier, Mean Square Error (MSE), Root Mean Square Error (RMSE), Support Vector Machine (SVM), Decision Tree.

I. INTRODUCTION

The water that's flowing down the soil is referred as the groundwater and that groundwater is classified into two categories based on the depth of the water from the surface which are named saturated and unsaturated. Basically the water on the surface starts sinking down to the small pore spaces in the soil and gets accumulated and starts sinking down to the down and the region just down the surface is the saturated zone which contains a mixture of air with water and the water present at certain depth is considered as the unsaturated zone which has dusty water which is mixture of rocks and water.

Basically the moisture in the soil is predicted mainly on the reference of the hydrological cycle and assessing the climatic conditions. By observing the seasons, hydrology, climatic changes and the runoff generation during rainfalls the farmers used to predict the moisture and start the different farming activities like irrigation, cultivating, seeding, fertilizing etc. There are some remote sensor tools which are used to predict soil moisture but the efficiency and usage is still under study for any better usage. People predict the moisture based on different factors that influence the soil that might be any of the following factors.

1. Rainfall
2. Recharge of groundwater
3. Humidity
4. Organic matter (different microorganisms)
5. Climate

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 5, May 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1005318 |

6. Evapotranspiration

On considering more than one factor from above the soil moisture is predicted approximately and the decision of performing agricultural activities or any other necessary tasks are performed. But we can make this prediction more accurate by using the machine learning models like Random Forest and Decision Tree Classifiers and other classification models.

II. LITERATURE SURVEY

Paper Name	Authors	Features	Algorithm Used	Proposed System	Drawbacks
Soil moisture content prediction model for tea plantations	1. Ying Huang 2. Hao Jiang 3. Wen-feng Wang 4. Weixing Wang 5.Daozhong Sun	1.Humidity 2.Rainfall 3.Light Intensity 4.Air Temperature 5.Soil Temperature	Support Vector Machine (Bald Eagle Algorithm, Genetic Algorithm)	Sensor nodes collect soil information from different areas, and the weather station node collects air environment information of the tea garden	The authors suggested an approach that is only suitable to calculate moisture in soil only in the view of planting tea. It is not suitable to assess moisture requirements for other crops.
Research on soil moisture prediction model based on deep learning	1.Yu Cai 2.Wengang Zheng 3.Xin Zhang 4.Lili Zhang Zhong 5. Xuzhang Xue	1.Average temperature 2.Average Pressure 3. Humidity 4.Average Wind Speed	Adaptive Gradient Algorithm	The DNNR model training involves supervised training, in that the training set and the test set features all needed labels, and the model parameter.	In Adaptive Gradient Algorithm, an approach of finding train loss and test loss, is used to measure performance. In this paper the average test and train losses are 1.42 and 0.56. Here the model is overfitted as test loss is much greater than the train loss.

Machine Learning for Predicting Field Soil Moisture Using Soil, Crop, and Nearby Weather Station Data in the Red River Valley of the North	1. Umesh Acharya, 2. Aaron L. M. Daigh 3. Peter G. Oduor	1.cumulative rainfall, 2. PET, 3. weather station VWC, 4. crop field's Ksat	1.classification and regression trees 2.Random Forest Regression 3.Support Vector Regression 4.Boosted Regression Tree	This study was conducted along the Red River Valley of the North (RRVN) in North Dakota and Minnesota using four machine learning regressions algorithms CART, RFR, SVR, and BRT.	The authors just validated the MSE and RMSE values of the 4 regression algorithms to check which one suits best to predict moisture in soil in north Dakota. But they have not accessed the conditions required to plant different crops at the region of Red River Valley.
Root zone soil moisture estimation with Random Forest	Coleen Carranza, Corjan Nolet, Michiel Pezij, Martine van der Ploeg	Precipitation, wind speed, Relative Humidity, Global Radiation, Sun Hours,	Random Forest Regression	Using the Random Forest Regression Algorithm the authors estimated the root zone soil moisture using limited information on soil hydraulic properties.	The results from the Random forest model do not elaborate on Controlling soil moisture state and may suffer from bad extrapolation outcomes. But it provides necessary variables influencing prediction accuracy.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 5, May 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1005318 |

Ⅲ. PROPOSED SYSTEM

We can predict moisture in soil more accurately using the machine learning models. There are different machine learning models each has its algorithm and can be trained using a database that contains a set of features which are correlatively used to classify the data. Here we have taken the data from a survey report on groundwater by the central government of the groundwater department. Based on that statewide report we have collected the data and we have extracted 13 different features indicated in billion cubic meters (bcm) which are state, recharge done by rainfall and other resources in both monsoon and non-monsoon seasons , total annual rainfall in each area, Natural discharge, net annual groundwater availability, industrial uses and other features.

Based on these features we can predict the moisture into 4 labels: excess, moderated, semi-critical, critical, like this we have a situation column in our dataset which specifies the amount of moisture present in soil. In the previous system we used to predict the moisture directly using necessary factors whereas here we find the moisture based on groundwater that perfectly classifies whether the moisture is high or low and suggests the necessary tasks to do. We are entirely using the groundwater recharging resources and using resources to know its contents and based on it the moisture is classified into 4 contents

3.1 Data Pre-Processing :

Based on the survey conducted by the government in order to learn the groundwater levels and the moisture content in soil we have constructed our dataset with total 13 features which are categorized in two ways one is recharge of groundwater by rainfall and other resources and other is usage of the ground water for different works in various sectors. This data is used to predict the situation of the soil that states whether the moisture is high or low based on the features we are using to train our model. In our dataset we have considered 13 features in order to classify the soil moisture in each state as categorical values excess, moderate, critical. The features are

1. Recharge from rainfallMonsoon season
2. Recharge from other sources
3. Recharge from rainfall Non-monsoon season
4. Recharge from other sources
5. Total_Rainfall
6. Natural discharge during non-monsoon season
7. Net annual groundwater availability
8. Irrigation
9. Domestic and industrial uses
10. Total_Usage
11. Projected demand for domestic and industrial uses upto 2025
12. Groundwater availability for future irrigation use
13. Excess

The different statistical properties like Mean, Mode, Standard deviation, minimum, maximum and others can be viewed using the describe() method in python. These numerical parameters can be useful in calculation of the error in data.

Feature extraction is the method of constructing a combination of variables to get around these problems while still describing the data with sufficient accuracy. In this dataset we have a total of 13 features and in those 13 we have features of two categories one acts as the resource of the water that is rainfall and other is utilization of water by different sectors of people may be household use, industrial use, Agricultural usage. Heat map can be used to see the features and its data weight either a higher or lower value can be found. Usually the Heat map represents each feature in a unique color shade. If the color shade is darker then it might be a higher value else if it is lighter shade then is probably a lower value relatively lesser.

In order to train the dataset using machine learning algorithms like Random tree and Decision tree classifier the features should be numerical. If any categorical values are observed in any feature then that particular feature values should be converted into numerical type. We can convert such categorical values into numerical using this feature extraction. We have a method get_dummies() in pandas module which can be used to make each categorical value into individual column and each column has two numerical values 0 (for false) and 1(true).1 represents that the output in training set resembles the column name or else it is not equal to the feature name. Here situation is the only feature that has categorical values excess, moderated, semi-critical and this is to be converted into numerical values using get_dummies () method that creates 3 new columns with the values of situation as feature name.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 5, May 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1005318 |

3.2 Random Forest Algorithm:

Random Forest is a popular machine learning algorithm that belongs to the supervised machine learning technique. It can be used for both Classification and Regression problems in ML. It is based on the concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model. Instead of depending on a single decision tree we can actually can improve our model prediction accuracy using random forest which can be a combination of the number of decision trees. Based on the features total_rainfall, Recharge of groundwater through rainfall and other sources and considering other industrial uses randomly and training model with random number of features can improve the model test score.

The more the number of trees in the forest leads to higher accuracy and prevents the problem of overfitting. The features in the dataset are randomly chosen

```
#Fitting Random Forest classifier to the training set
From sklearn.ensemble import RandomForestClassifier
Random = RandomForestClassifier ()
Random.fit (x_train, y_train)
```

Prediction using Random Tree Classifier:

- Select random samples from a given dataset.
- Construct a decision tree for each sample and get a prediction result from each decision tree.
- Perform a vote for each predicted result.
- Select the prediction result with the most votes as the final prediction

IV. RESULTS

In this study we are considering input data directly not from users directly but, we are considering statewide annual factors that affect the groundwater from a recently done survey by the central department of groundwater. Based on the level of groundwater we are assessing the moisture content. Soil moisture is previously predicted by many machine learning algorithms. By comparing each algorithm with every other algorithm we can say that using random forest classifier provides higher test accuracy. Based on users input we are predicting the amount moisture present in soil (either high or low).

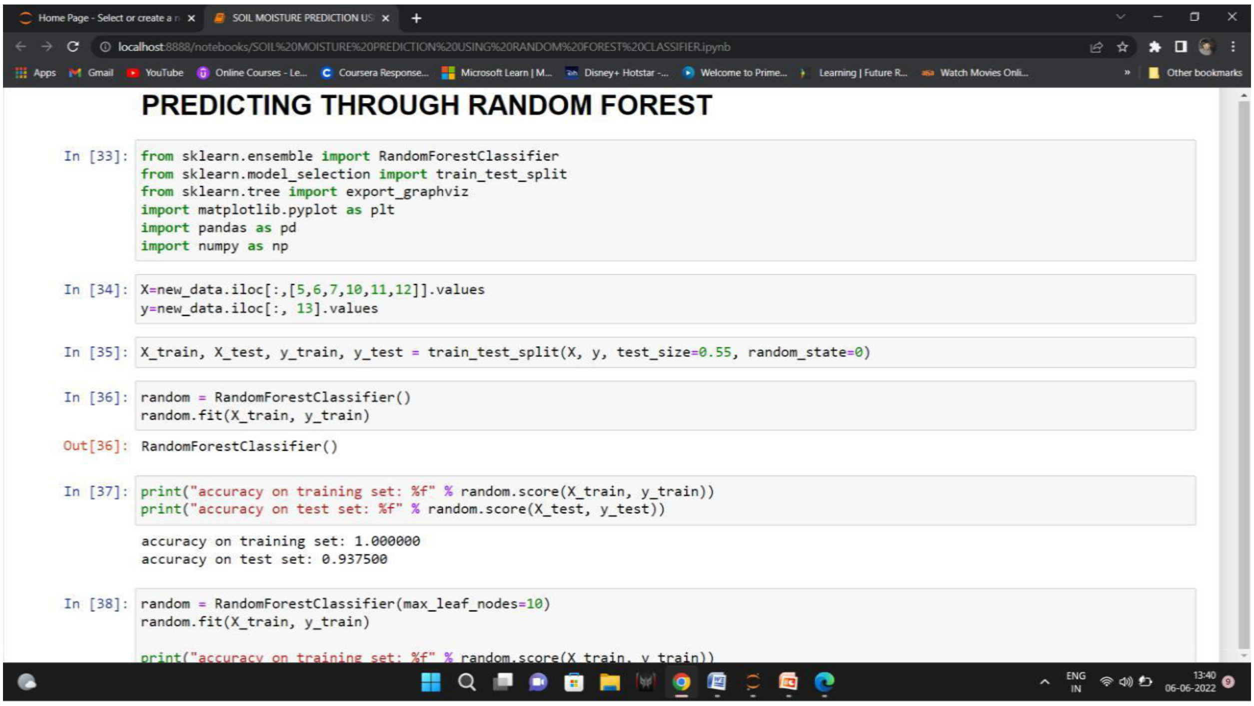


fig 2 – Predicting Through Random Forest

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 5, May 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1005318 |

Using the sklearn module we had imported the random forest classifier model which can be assigned using RandomForestClassifier (). Here we have first split the dataset into two one is for training the model and the other is to test the model accuracy.

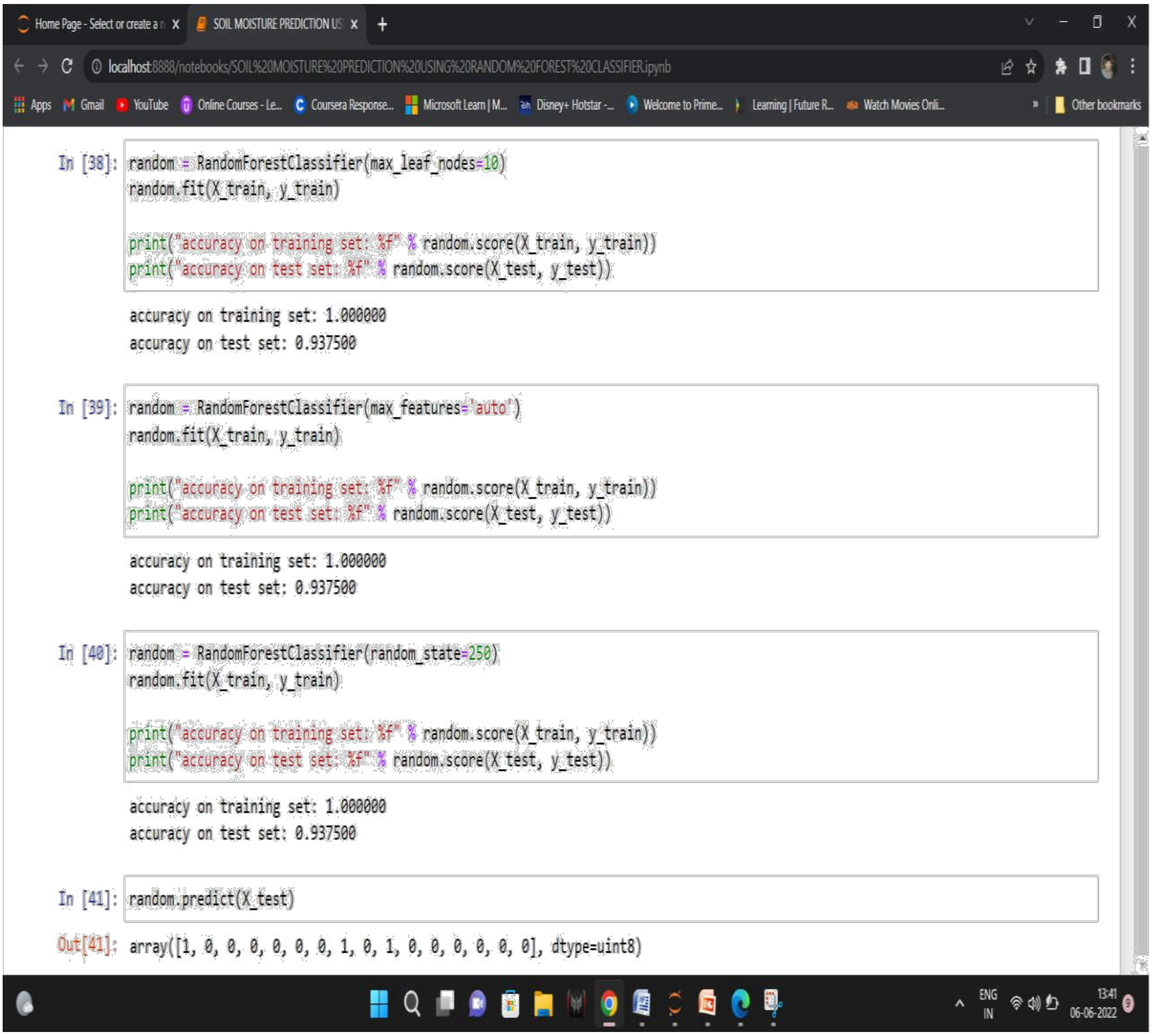


Fig 3 – Measuring of Test Accuracy Using Source Method

The accuracy can be found using the score() method to the model name. Here random is our machine learning model and we have trained our model using a training dataset and then we have applied our model to a testing set in order to find the test accuracy using random. Score(x-test, y-test).

V. CONCLUSION

Based on the latest assessment involving estimation of dynamic ground water resources that was conducted jointly by both respective State and Central Ground Water Board, we have taken that database and converted it to a data frame and with the help of different factors of how the groundwater is recharged and how the water comes in different seasons. Here features are the resources of water that causes the recharge of ground water. This data is used to train the different machine learning models like Random Forest Classifier and Decision Tree model and applying the Data

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 5, May 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1005318 |

Preprocessing methods and using the Anaconda Python Environment we have applied the data to the models using sklearn module and this is a classification model. The model predicts the amount of moisture present in soil using the data that is given as input to the model which assesses the features and predicts one of the four labels excess, moderated, critical, semi-critical which represents a range of quantity of moisture present in soil.

In order to find the best machine learning model to fit this data we have taken both a decision tree and random forest classifier to get the best out of it by calculating the test scores which predicts the test accuracy of the model which is trained with a certain number of examples. Here to predict the moisture in soil using that data frame we have got a test score higher using Random Forest Classifier Algorithm that is 0.937500 whereas Decision tree classifier got accuracy of 0.9000 and Support Vector Machine Algorithm got an accuracy of 0.875. So, Random Forest Classifier is the suitable model for higher accuracy and also MSE and RMSE values are relatively better than other algorithms which are 0.0038 and 0.062.

REFERENCES

[1] R. Yamini Nivetha and C. Dhaya, "Developing a Prediction Model for Stock Analysis ", *Technical Advancements in Computers and Communications (ICTACC) 2017 International Conference on IEEE*, 2017.

[2] D. R. Legates, R. Mahmood, D. F. Levia, T. L. DeLiberty, S. M. Quiring, C. Houser, and F. E. Nelson, "Soil moisture: A central and unifying theme in physical geography," Progress in Physical Geography, vol. 35, no. 1, pp. 65–86, 2011.

[3] E. G. Njoku, T. J. Jackson, V. Lakshmi, T. K. Chan, and S. V. Nghiem, "Soil moisture retrieval from amsr-e," IEEE transactions on Geoscience and remote sensing, vol. 41, no. 2, pp. 215–229, 2003.

[4] S. Lambot, E. C. Slob, I. van den Bosch, B. Stockbroekx, and M. Vanclooster, "Modeling of ground-penetrating radar for accurate characterization of subsurface electric properties," IEEE Transactions on Geoscience and Remote Sensing, vol. 42, no. 11, pp. 2555–2568, 2004.

[5] S. Keller, F. M. Riese, J. Stötzer, P. M. Maier and S. Hinz, "Is it possible to retrieve soil-moisture content from measured VNIR hyperspectral data?", *CoRR*, 2018, [online] Available: http://arxiv.org/abs/1804.09046.

[6] M. Lobato, M. Juston, W. R. Norris and R. Nagi, "Comparison of semantic segmentation deep learning models for land use mapping: Site characterization for construction site preparation in press", *2021 Institute of Industrial and Systems Engineers (IISE) Annual Conference*.

[7] F. M. Riese and S. Keller, "Introducing a Framework of Self-Organizing Maps for Regression of Soil Moisture with Hyperspectral Data", *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, pp. 6151-6154, July 2018.

[8] M. Aboutalebi, N. Allen, A. Torres-Rua, M. McKee and C. Coopmans, "Estimation of soil moisture at different soil levels using machine learning techniques and unmanned aerial vehicle (uav) multispectral imagery", *Autonomous Air and Ground Sensing Systems for Agricultural Optimization and Phenotyping IV*, pp. 26, 05 2019.

R. Sobayo, H. -H. Wu, R. Ray and L. Qian, "Integration of convolutional neural network and thermal images into soil moisture estimation", *2018 1st International Conference on Data Intelligence and Security (ICDIS)*, 04 2018.

[9] Wang, W., et al.: WSN monitoring system with adaptive transmitting power based on low-power-consumption in rice fields. Trans. Chin. Soc. Agric. Mach. 49(3), 150–157 (2018)

[11] Alsattar, H.A., Zaidan, A., Zaidan, B.: Novel meta-heuristic bald eagle search optimization algorithm. Artif. Intell. Rev. 53(8), 2237–2264 (2020)

[12] Wang, W., GAO, M., Wang, J.: Hyperspectral parameters and prediction model of soil moisture in apple orchards. IOP Conf. Ser. Earth Environ. Sci. 687(1), 012085 (2021)

[13] Zhang, W., et al.: Effect of different monitoring sampling intervals on soil moisture prediction model performance. J. Gansu Agric. Univ. 55(1), 221–228 (2020)

[14] Gao, P., et al.: Prediction models of soil moisture content and electrical conductivity in citrus orchard based on Internet of Things and LSTM. J. South China Agric. Univ. 41(6), 134– 144 (2020)

[15] Romashchenko, M.I., et al.: Numerical simulation of irrigation scheduling using fractional Richard's equation]. Irrig. Sci. (2021)

INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 8.165

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर
NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING