

# Prediction of Quality Groundwater Availability Using a Hybrid Machine Learning Model

**\*Ihenetu Victoria Ndidiamaka<sup>1</sup>, Aniyom Ebenezer<sup>2</sup>, Wabilla Jean Claude<sup>4</sup>, Dr. Ugochukwu Ewuzie<sup>3</sup>, and Dr. Okoli Emeka Austin<sup>4</sup>**

<sup>1</sup>Department of Geology, Federal University of Technology, Owerri

<sup>2</sup>Department of Petroleum Engineering, Federal University of Technology, Owerri

<sup>3</sup>Department of Pure and Industrial Chemistry, Abia State University

<sup>4</sup>Software and Network Engineer, Integrated Oasis ICT Services.

## ABSTRACT

The prediction of groundwater availability is of keen interest and essential to water sustainability and drought mitigation. This strategy is essentially critical in areas with pore spaces of consolidated and unconsolidated sedimentary rocks, weathering layers, joint and fission of land rocks, fault zones, and karst areas. In recent times, the comprehensive approaches of remote sensing (RS) and geographic information system (GIS) have been applied to the exploration, assessment, and management of crucial groundwater resources. Therefore, this study employed a voting classifier machine learning model to train an ensemble of models consisting of a Support Vector Machine, Random Forest, Extreme Gradient Boosting, and Multilayer Perceptron for quality groundwater availability prediction. The comparative analysis results show that applying the voting classifier model increases groundwater availability prediction performance by 15% compared to using the individual models. Consequently, this technique is promising and will potentially enable geologists, geophysicists, and planners to (1) obtain water quality data from unsampled areas; (2) ascertain areas with high aquifer depth; (3) assay the sustainability of groundwater sources; (4) evaluate the quality of the water for human use.

**Keywords:** Groundwater, Remote sensing, Machine learning, Model, Karst, Sedimentary, Aquifer

## INTRODUCTION

Groundwater is the water present beneath the Earth's surface in rock and soil pore spaces, and the features of rock formations of the world. Water is a major component of living things, a facet that man cannot do without. Due to this situation, surface water cannot be dependable throughout the year; hence another alternative is needed in order to supplement surface water. The groundwater is the water that lies underground and it is the best quality freshwater that humans depend on as a major source. Groundwater occurrence is a subsurface phenomenon, and its identification and location are based on some directly observable terrain features. Water makes up about 71% of the earth's surface, while the other 29% consists of continents and Islands. To break this down, 96.5% of all

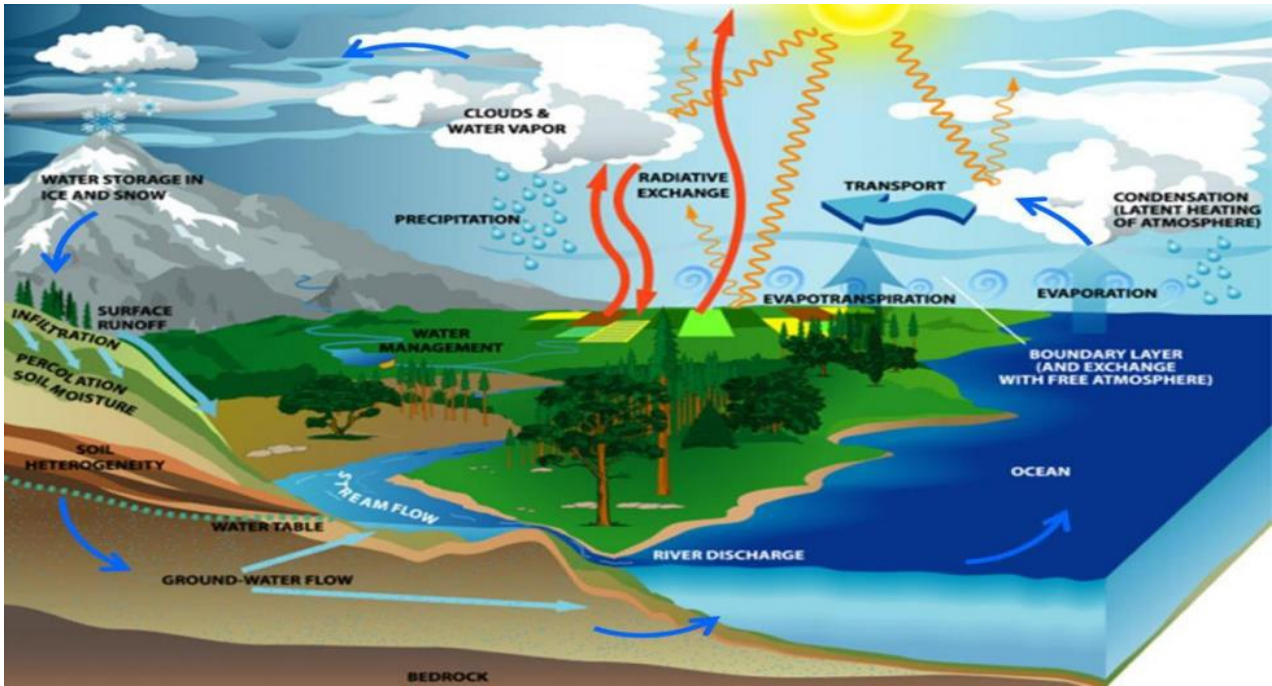
the Earth's water is contained within the ocean as salt water, while the remaining is fresh water, lakes, and frozen water locked up in glaciers, polar ice caps, and then groundwater which can be termed "Precious nature of freshwater".

The prediction of quality groundwater availability is of optimum importance due to the advantages when compared to surface water, it is generally of higher quality, better protected from chemical and microbial pollutants, less subjected to seasonal and perennial fluctuations, and more uniformly spread over long regions than surface water. The groundwater can be in the sedimentary terrain where it is less difficult to exploit or in the basement complex terrain in which it can be a bit difficult to locate especially in areas underlined by crystalline rocks (Fadele *et al.*, 2013) and the need for explorative measures comes to play.

In the past, due to the demand for available quality groundwater resources without the advances of modeling technologies, the resource was overexploited either by artisans which led to advert consequences resulting in to decline of water level, water quality, closure of spore

© Copyright 2023. Nigerian Association of Petroleum Explorationists.  
All rights reserved.

The authors wish to thank the Department of Geology, Federal University of Technology, Owerri, Department of Petroleum Engineering, Federal University of Technology, Owerri, Department of Pure and Industrial Chemistry, Abia State University, Software and Network Engineer, Integrated Oasis ICT Services and NAPE for providing the platform to present the paper during the Annual Conference.



**Figure 1:** An overview of Groundwater.

spaces of consolidated and unconsolidated sedimentary rocks, unforeseen areas of weathering layers, joint and fission of land rocks, unrealistic areas of fault zones as well as karst areas.

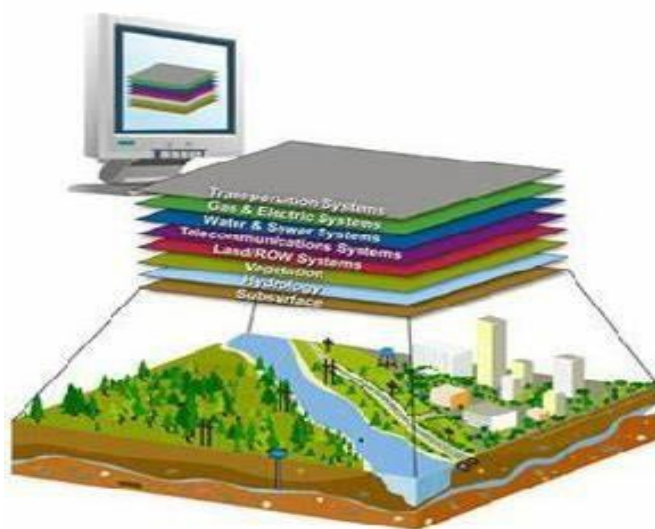
In recent times, the comprehensive approaches of remote sensing and geographic information system (GIS) have been applied to the exploration, assessment, and management of crucial groundwater resources. GIS is a computerized information system like any other database, all information in GIS must be linked to geographic (Spatial) reference (Latitude/Longitude or other spatial coordinates). The recent application of remote sensing and

a combination of GIS has helped identify water fluctuations and patterns, detect aquifer recharge zones, look at historical data to classify periods of hydrological drought, identification of drought-vulnerable aquifers, and mapping of groundwater potential zone, measuring, monitoring, evaluating and managing water availability to find a suitable approach for water management, which can provide sufficient fresh water for lives on earth.

The identification and location of groundwater resources using remote sensing data are based on an indirect analysis of some directly observable terrain features like geomorphology, geology, slope, land use/land cover, and hydrologic characteristics using satellite, and the use of a software program which enables users to store and manipulate large amounts of data from GPS and other sources. For this reason, researchers have proven that groundwater should be an alternative source of water which can be mapped using remote sensing and GIS technique through the application of multi-criteria analysis of certain hydrological and geological factors before the actual Electrical Resistivity observations for the siting of boreholes and wells Epuh *et al.*, (2018).

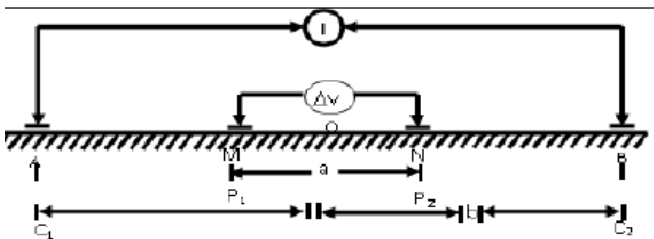
This direct analysis like the geophysical techniques which include Electrical Resistivity technique. This technique has been used in a wide range of geophysical investigations such as groundwater exploration, mineral exploration, archaeological investigation, engineering studies, and geothermal exploration (Fadele *et al.*, 2013).

The use of Electrical Resistivity technique has become very popular with groundwater prospecting due to the



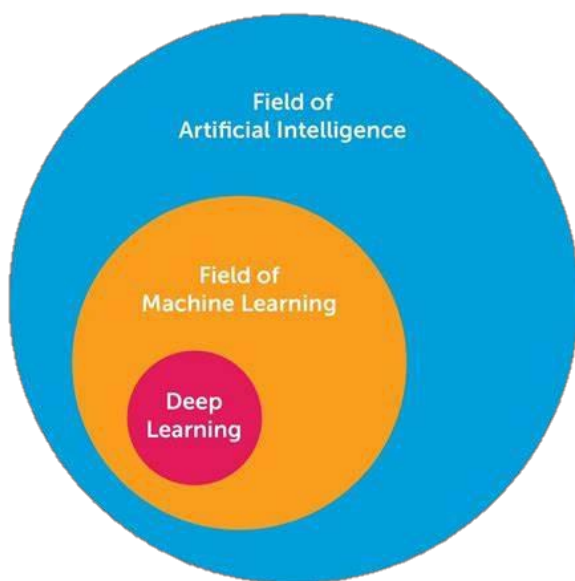
**Figure 1.1:** GIS and Remote Sensing Image.

simplicity of the technique and data provision. The purpose of this geophysical survey method/technique is to detect the sub-surface effects that is produced by the flow of electric current inside the earth. The geophysical techniques for groundwater exploration and water quality detection have increased and are widely accepted due to rapid advances in computer software and other numerical modeling technique, thereby a good data-driven approach to the application in machine learning for future predictions.



**Figure 3:** Diagrammatic Representation of a Werner Configuration.

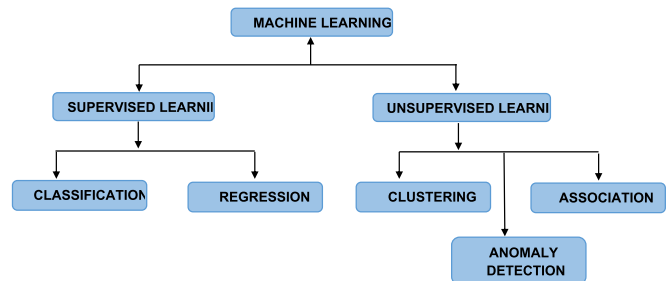
Therefore, this research aim at the use of Hybrid Machine learning model to analyze the performance and predict future availabilities and quality of groundwater from previous electrical resistivity technique, and data recorded. This is done by developing a voting classification algorithm embedding different machine learning algorithms, and as well compare the performance of the hybridized model with the other single models. Also, this won't be completed without referring to Artificial intelligence (AI). AI is a huge set of tools for making computers behave intelligently, and covers subfields that are Robotics & Machine learning.



**Figure 1.2:** Fields to Machine Learning.

## Machine Learning (ML)

Machine learning is a set of tools for making inferences and predictions from data. Where it predicts events, infers the causes of events and behaviour, and then infers patterns. ML is majorly divided into Supervised and Unsupervised learning and further divided.



**Figure 1.3:** Tree Diagram of Machine Learning Division

## How Machine Learning Works

- \* Interdisciplinary mix of statistics and computer science.
- \* Ability to learn without being explicitly programmed.
- \* Learn patterns from exciting data and applies them to new data.
- \* Relies on high-quality data.

## Data Science

Data science is about making discoveries and creating insights from data. machine learning is often an important tool for data science.

## Machine Learning Model

A statistical representation of a real-world process based on;  
Data new input → model → output.

## Training Data

Supervised learning and unsupervised learning is distinguished by their pattern. The existing pattern is known as the training data.

## Features

Different pieces of information that might help predict the data. We can use many features to train the model, when training is done, we now give new input.

Machine Learning tools and techniques have the best potential to drive groundwater knowledge and discovery and management by assisting in the prediction due to their easy availability, improved performance in comparison with geophysical techniques, the advancement of new machine learning tools like a voting classifier.

Therefore, this study employed a voting classifier machine learning model to train ensemble models consisting of a Support Vector Machine, Random Forest, Extreme Gradient Boosting, and Multilayer perception for quality



## Aim and Objective

**Aim:**

To produce a model that will predict groundwater availability and interpretation of datasets from different geophysical techniques/resources, which will perform further predictions of groundwater availability using the voting classifier ML model in different regions of Africa.

### Objectives:

- \* Detailed Exploratory Data Analysis (EDA)
- \* Make use of a hybrid model to obtain good predictions
- \* The Objective of this study is to develop a voting classification algorithm embedding different machine learning algorithms, and as well compare the performance of the hybridized model with the other single models.

## Background of the study

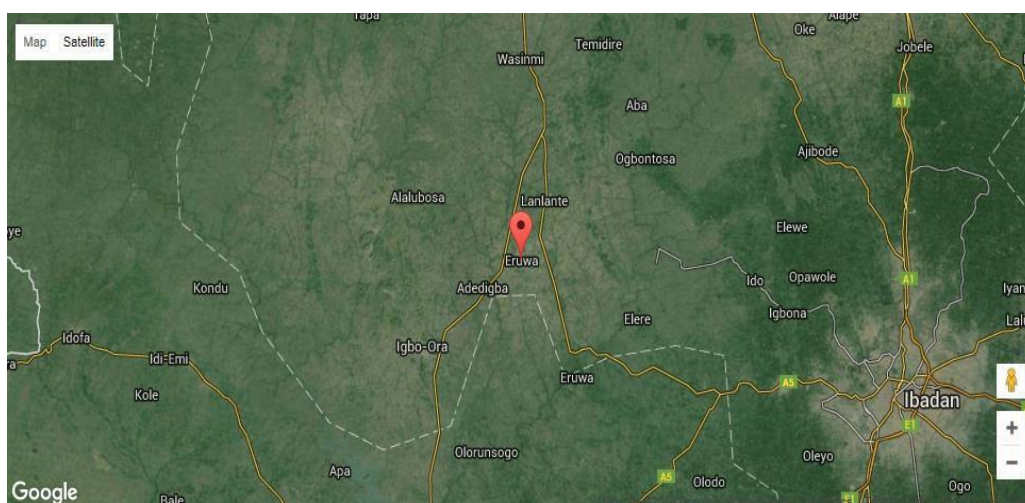
Eruwa is quite significant in so many aspects especially because it is the headquarters of the Ibarapa which is made up of seven towns namely Eruwa, Lanlate, Igboora, Idere, Ayete, Tapa and Igangan. It is about 72 km southwest of [Ibadan](#) and 60 km north east of [Abeokuta](#). [Eruwa derived its name from the way in which hawkers displayed and traded roasted yams to the northern caravans who constantly congregated in the town.](#) Eruwa town is situated in the heart (headquater) of Ibarapa east Local Govt area of Oyo state, Southwestern, Nigeria, of the [Greenwich meridian in Oyo State in the South Western part of Nigeria.](#) The latitude of Eruwa, Oyo, Nigeria is 7.536318, and the longitude is 3.418143. Eruwa, Oyo, Nigeria is located at Nigeria country in the Cities place category with Latitude: [7.536318, DMS Lat: 7° 32' 10.7448" and Longitude: 3.418143, DMS Long: 3° 25' 5.3148" E.](#)

## Hydrogeology of the Study Area

These rocks can be grouped into major and minor rock types. The major types are quartzite of the meta-sedimentary series and the migmatite complex comprising banded gneiss, augen gneiss, and magnetite, where the minor rock types include pegmatite, quartz, aplite, diorites, amphibolite, and xenoliths (Akintola 1994). The area has a low gentle undulating top. The type of rock in an area is an important factor governing the characteristics of its groundwater. Basement complex rocks, composed mainly of metamorphic and igneous rock types are relatively low in groundwater production in comparison with sedimentary rock areas to the south. The basement complex nature of the rocks in Ibadan does not however completely rule out the possibility of the presence of isolated good and productive aquifers if proper searching is carried out. The factors that account for the presence of good aquifers in particular locations over the basement complex rocks are the thickness of the regolith (weathered layer), the size and density of fractures, fissures and other cracks, and the permeability and porosity of the rocks. Ibadan is located near the forest grass-land boundary of Southwestern Nigeria (Amanam-bu, Ojo-Kolawole 2013).

### Geology Map of Oyo State, Showing the Geology of Ibadan

The area lies within the Nigeria Precambrian southwestern basement complex. The geology survey carried out in this area suggests that the basement complex is made up of granite and gneiss. There are also occurrences of pegmatite intrusion rich in feldspar and quartz. The dominant rock in the survey area is the granite followed by gneiss, both of which have been exposed to severe weathering. The presence of structures such as folds, faults, joints is an attestation that the area has been exposed to series of deformation.



**Figure 2.1: Google Location of Ewura**

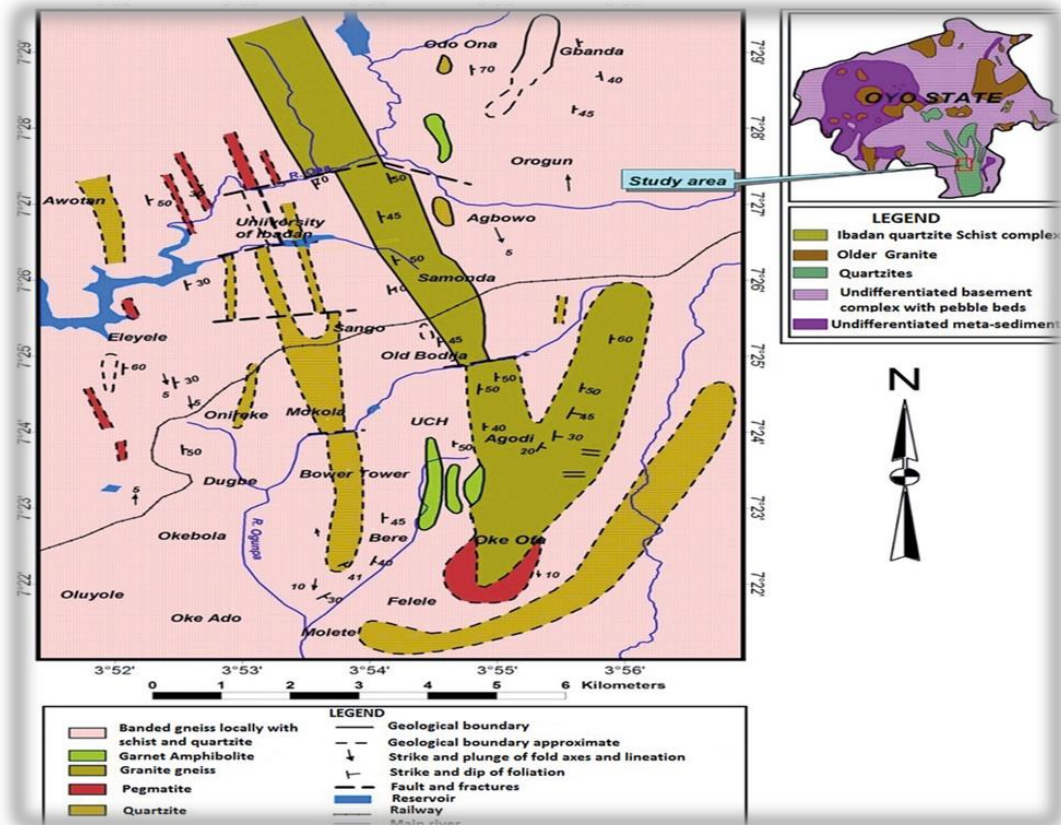


Figure 2.2: Geology Map of Oyo State, Showing Geology of Ibadan.

### Methodology (Data Collection, Description and Modelling)

Machine-learning tools technologies have the potential to drive groundwater knowledge discovery and management by assisting in the prediction of groundwater availability. This can be done by enabling the collection of massive water datasets, storing these datasets into databases and

processing these datasets to get useful insights which can be used by water resource managers to: (1) design targeted monitoring programs; (2) inform groundwater protection strategies; and (3) evaluate the sustainability of groundwater sources of drinking water (4) anticipate water quality in unsampled areas or depth zones;

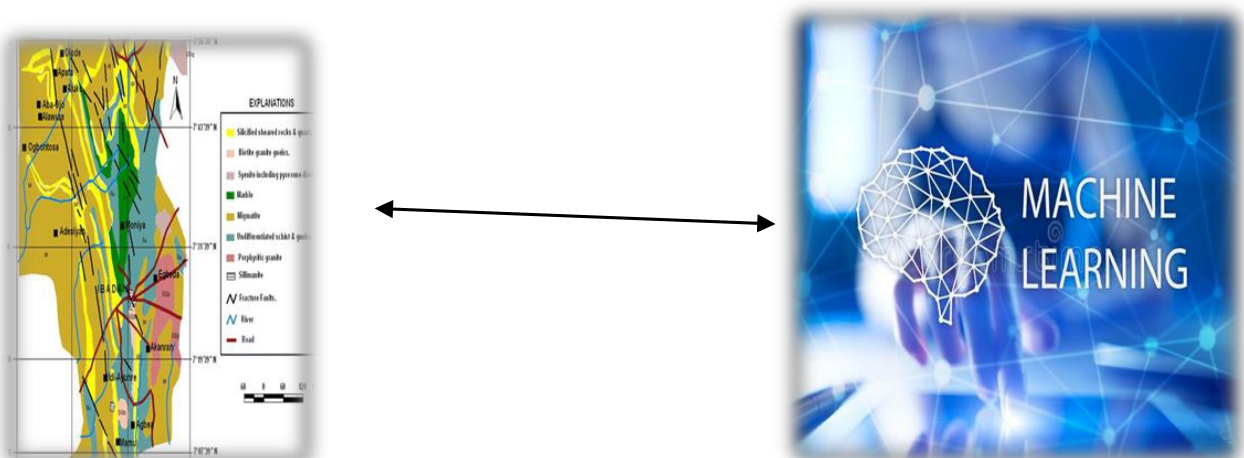


Figure 3: Image of Raw Data down to the Machine Modelling.

## Methodology

The method applied in solving the problem of groundwater exploration is summarized in the flowchart below;

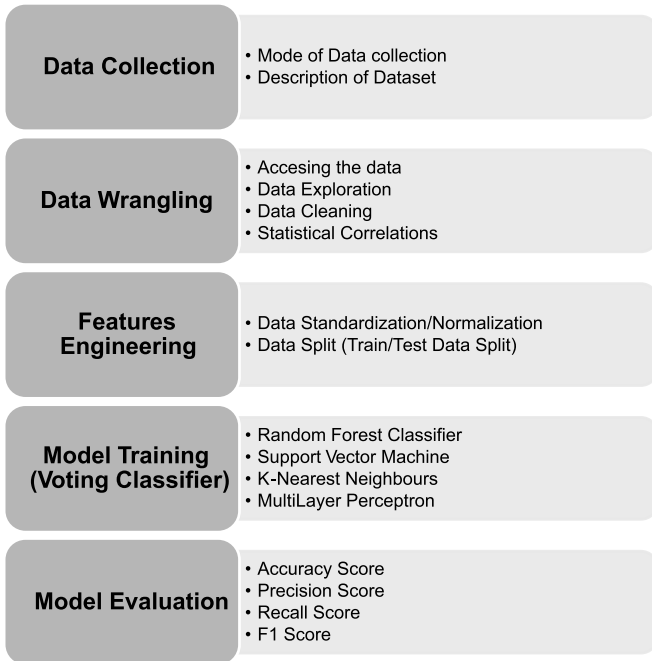


Figure 3.1: Methodology Flowchart.

## Data Collection

Data collection is a process of gathering information from relevant sources to provide an answer to a research problem (Dudovskiy, 2021). It involves the systematic process of gathering observations which could be done first-hand (i.e. primary collection) or secondary/Tertiary means of collection (Bhandari, 2022).

The area of data acquisition was from an open field opposite Plural hotel, Eruwa, Oyo state, south-western Nigeria. Located between latitude  $7^{\circ}35'57''N$  and longitude  $3^{\circ}25'17''E$ . Area is covered by moderate vegetation and is generally accessible. Data was acquired through the use of the electrical resistivity method, Constant Spacing Traversing (CST)

## Equipment used

Equipment used: resistivity metre, 4 electrodes, connecting wires.



Figure 3.2: Image of resistivity metre, 4 electrodes, connecting wires.

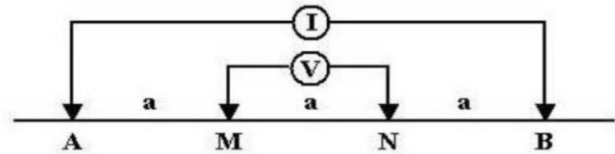
The configuration used for this project is the Werner array configuration and measurement was acquired for spacing of 5m, 10m, 15m. The Werner array consists of four collinear, equally spaced electrodes. The outer two electrodes are typically the current electrodes and the inner two electrodes are the potential electrodes. The array spacing expands about the array midpoint while maintaining an equivalent spacing between each electrode.

## Data Description

CST is also called electrical profiling; it is used to determine the lateral variations of resistivity in the shallow subsurface at a more or fixed depth of investigation. The current and potential electrodes are maintained at a fixed separation and moved along profiles. Common configuration for this method is the Werner array configuration. The current and potential electrodes are moved along a profile with constant spacing between the electrodes. The Werner configuration shown in the figure 10 is the best adapted for lateral line profiling to obtain precise location of the resistivity anomalies (Batte Muwanga and Sigrist, 2008).

The geometric factor (K-factor) used for the Werner array is given as:

$K = 2\pi a$ , where  $a$  is the spacing of the electrode  
**Resistivity = K x RESISTANCE**



$$\rho_A = 2\pi a \frac{V}{I}$$

Figure 3.3: Werner Array Configuration.

## Resistivity

As the water content and its distribution change the electrical properties of the rocks. The resistivity is inversely proportional to the rock porosity and its water saturation. Since it was in a basement terrain, we had a keen interest in the resistivity for our modelling and prediction, The presence of a low resistivity value together with a high thickness result (from the electrical Resistivity data) could be inferred as a possible zone for drilling for the geotechnical purpose, close attention should be made to areas with low resistivity value as they could be as a result of depression in the subsurface which is not good areas for building activities.

The resistivity of the fracture will be lower than that of



hard rock because in hard rock the moisture will collect in the fissures of the fracture zone. So why is it lower resistivity there? Because it contains moisture, and that moisture is more conductive than hard rock.

### Data Wrangling

Data wrangling involves the process of accessing, cleaning, structuring, and transforming raw data into the desired format for better decision-making or futuristic prediction in less time. Data wrangling done on the formation cut across accessing the data, data understanding and uncovering of patterns embedded in the dataset, this makes up for the data preparation section of this project. Let's work on the steps applied during the data wrangling stage of the project.

### Accessing the Data

Data accessing entails the process of visually or manually scanning of the dataset to spot the irregularities present in the dataset. Data accessing could also be done programmatically as applied in this project. Where the general overview of the dataset was displayed to understand the dataset properly with the display of each features data type and the value counts for each of the columns. The dataset used contained 8 features with over 159 observations, with its features having two unique datatypes; the integers and the floats.

### Data Exploration

Data exploration is used to explore and visualize data to uncover insights from the data and to understand patterns inherent in the dataset. During this project data exploration was applied to detect the missing values, the outliers present in the dataset. Visualizations used during the process of the data exploration include; bar plots, boxplots, pair plots, heat maps, etc. some of the visualizations are displayed below.

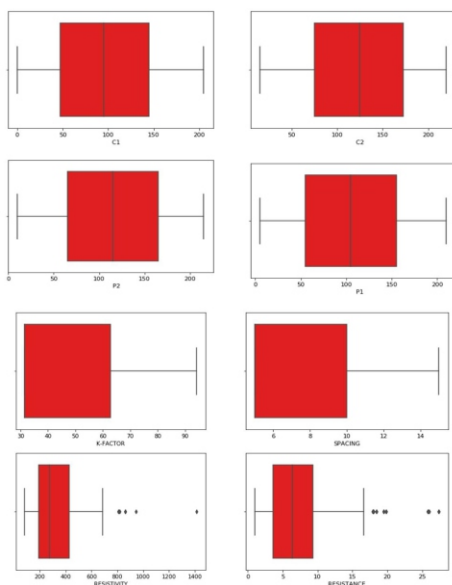


Figure 3.4: Boxplots

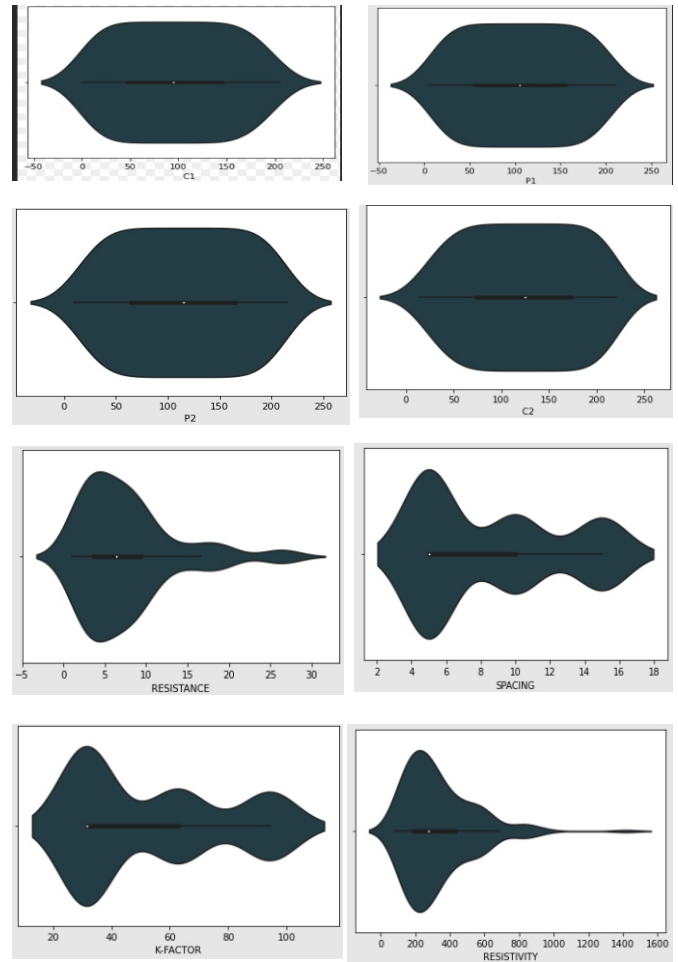


Figure 3.4: Violin Plots.

The boxplots and Violin plots above are used to check for outliers across the entire dataset with each feature being fully represented. Outliers are basically extreme data points that are very far from the range of the entire dataset. There exist minimal outliers in the resistance feature and the resistivity label. Other visualizations will be displayed in subsequent sessions of this project report.

### Data Cleaning

Data Cleaning involves the process of fixing the irregularities and bad data experienced in the dataset. The combined effort of detecting and correcting (or removing) corrupt or inaccurate records from a dataset is referred to as Data Cleaning. During the course of this project there exist a few irregularities in the datasets, it was experienced that there exist about 48 duplicated data points from the observation, which could lead to multi-collinearity during training of the datasets. These duplicated data points were dropped and completely removed from the dataset. Also, missing values in the observations were handled properly.

### Statistical Correlation

Statistical Correlation shows the extent to which variables

are related to each other either linearly or otherwise. This concept was implemented to understand the relationship there coexist between the features with itself and with the labels as well, as this provides a good fit to obtaining a good performance during the training of models. Although, the goodness of a model cannot only be based on this fact. There are other metrics to be considered as well. Below are a few plots to show how features are correlated to each other.



Figure 3.5: Heat map.

The heat map was applied to display the correlation that exist between the features and the labels. From fig. 3.5, it can be observed that there exists strong correlation between the signals sending devices and the receptors with a Pearson ratio score range of about 0.98 to 1.0. Considering the correlation of the features to the label (resistivity), we observe there exist a good correlation between the resistance feature to the resistivity, this indicates that the increase in resistance in that geological environment will result to an increase in the resistivity of the same formation with a Pearson score of 0.71 which is a good fit for the model training.

### Feature Engineering

Feature engineering as applied in this project cut across the selection of features that are paramount for the training of the models used for the prediction of ground water.

Data standardization being an integral part of feature engineering was applied to rescale the data points in the dataset to have a uniform number of observations. In this project the StandardScaler algorithm was applied in the transformation of the data points to have uniformity in the number of observations. This was applied due to the size of the dataset as it takes less time to transform.

After the standardization of the dataset, there was a need

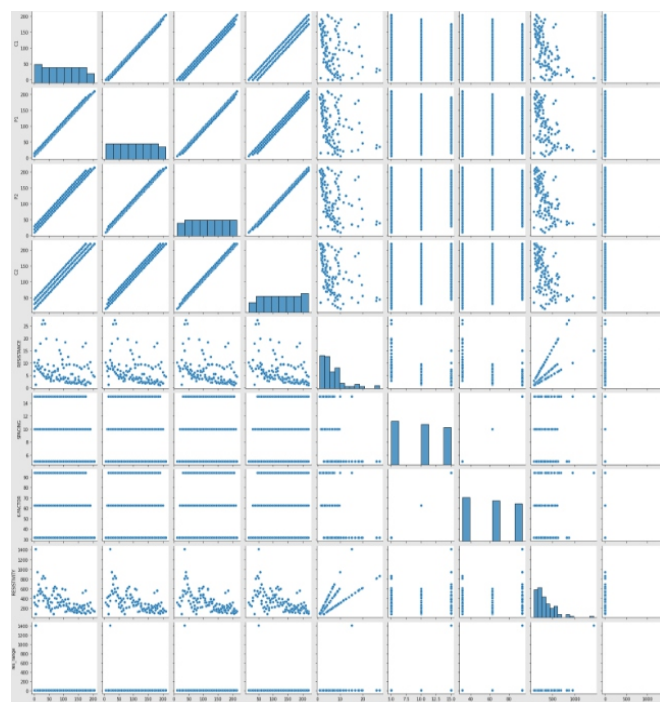


Figure 3.6: Pairplot.

for splitting of the dataset into train and test datasets with a test size of 0.1. the `train_test_split` algorithm was used from scikit learn. The split of the dataset into the train and test sets for both the features and the labels was necessary as the training dataset was used for training the model just as the name applies, while the test was used for prediction and to test the model in order to estimate the degree to which the model will perform.

### Model Training (Voting Classifier)

At this point the model (voting classifier) was trained with the train dataset as stated previously. This project proposes the use of voting classifier model to predict the availability of ground water.

#### Voting classifier

Voting classifier is a machine learning model which is trained by ensemble of several single machine learning models to predict an output based on the frequency of the highest probability of a chosen class as an output (Aniyom, Chikwe, & Odo, 2022). There exist two types of voting classifiers; Hard voting & Soft voting.

The hard voting classifier predicts its output based on the class with the highest number of votes, while the soft voting considers the average weight of each of the classes as a contribution to the prediction of the best outputs.

In this project, a voting classifier model was trained which ensembles the following single models; Random Forest Classifier, Support Vector Machine, K-Nearest Neighbors, MultiLayer Perceptron. This algorithm (voting classifier) was used to ensemble the mentioned four single models



with a better performance index. The voting classifier is a scikit learn algorithm under the ensemble module. The idea of the improvement of the performance index is made possible as a result of those single models in the ensemble accounting for the limitations of each other and bearing each other bias, while increasing the performance of the prediction.

### Model Evaluation

Result from the classification model (voting classifier) was evaluated using the following metrics built-in on scikit learn library. They include: accuracy score, precision score, recall score, and the F1\_score.

#### Accuracy score

Accuracy score is one of the metrics used to measure the model performance by evaluating the fraction of the prediction which the model got correctly to the total predictions. The values of accuracy score ranges from 0 to 1. The higher the accuracy score the better the performance of the model. Mathematically, it could be expressed as;

$$Accuracy = \frac{TN+TP}{TN+FP+TP+FN} \quad (i)$$

#### Precision Score

Precision score is the ratio of correctly predicted positive observations to the total predicted observations. It also varies from 0 to 1, with the higher score always showcasing a better performance. Mathematically, we have that;

$$Precision = \frac{TP}{TP+FP} \quad (ii)$$

#### Recall Score

Recall is the ratio of correctly predicted positive data points to all the observations in the actual class. This metric as well ranges from 0 to 1, with higher score showing a good fitness of the model as used for prediction. Mathematically, it is expressed as;

$$Recall = \frac{TP}{TP+FN} \quad (iii)$$

#### F1 Score

This is the weighted average of the precision and recall, which takes account of both the false positives and false negatives into account. It is expressed as;

$$F1\ Score = \frac{2 \cdot (Recall \cdot Precision)}{(Recall + Precision)} \quad (iv)$$

Where;

TN= True Negative Values

TP= True Positive Values

FP= False Positive Values

### Results and Discussion

#### Results

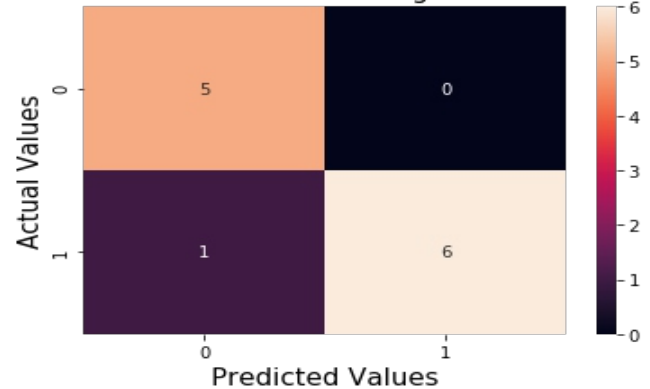
In the geological formation, our classification model was

able to predict the existence of two bodies in its environs, that is the ground water which is our major interest and the alluvium rocks. The result from the prediction by the model is displayed below;

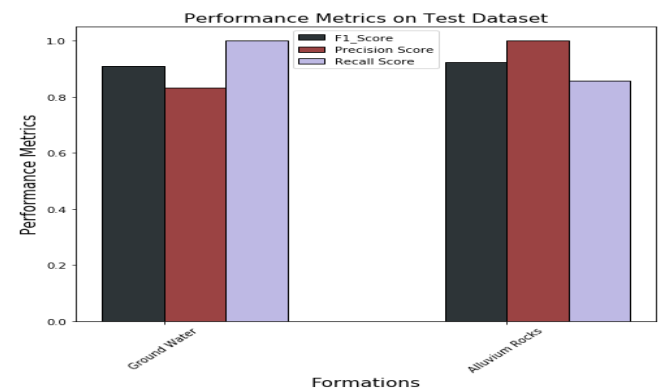
**Table 1:** Performance Score Table.

Geological Body	Precision	Recall	F1 Score
Ground Water	0.83	1.00	0.91
Alluvium Rocks	1.00	0.86	0.92
<b>Accuracy Score</b>	<b>0.91</b>		

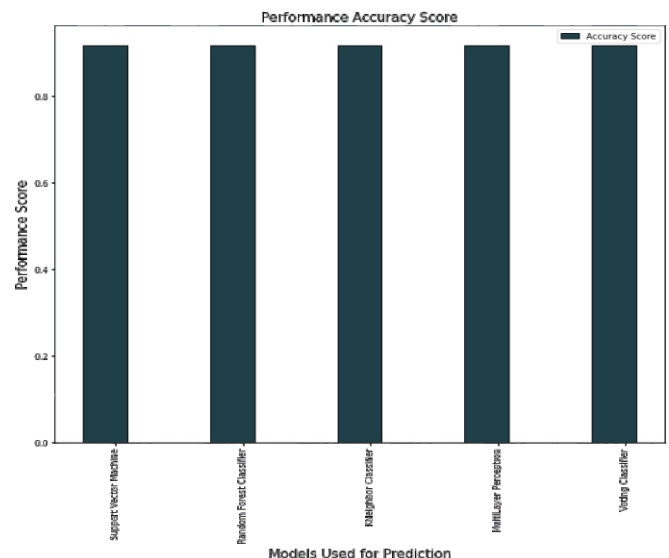
**Confusion Matrix Plot for Voting Classifier Model**



**Figure 4.1:** Confusion Matrix.



**Figure 4.2:** Performance Score Plot.



**Figure 4.3:** Models used for Prediction.

## Discussions

From the plots above, we can see that the model performed well. The performance table shows that the accuracy of the model is 0.91, which is a good fit for deployment.

Our model predicted the presence of groundwater and alluvium in large quantities which will be profitable to society. The prediction of groundwater was done with respect to a target variable which is resistivity. From the resistivity table put together by Keller & Frischknecht, 1966, as a standard resistivity to geological materials representations. They stated that, for geological materials with a resistivity range of 10 – 100 ohms' meters, their material is certainly bound to have groundwater in the dwelling. Thus, the prediction from our model classified those materials with groundwater resistivity range to be groundwater and the other as alluvium rock.

Material	Resistivity, $\Omega\cdot m$	Conductivity, milliSiemens/m (mS/m)
<b>Igneous and Metamorphic Rocks</b>		
Granite	5x10 <sup>3</sup> - 10 <sup>6</sup>	0.001 - 0.2
Basalt	10 <sup>3</sup> - 10 <sup>6</sup>	0.001 - 1
Slate	6x10 <sup>2</sup> - 4x10 <sup>7</sup>	2.5x10 <sup>-5</sup> - 1.7
Marble	10 <sup>2</sup> - 2.5x10 <sup>8</sup>	4x10 <sup>-6</sup> - 10
Quartzite	10 <sup>2</sup> - 2x10 <sup>8</sup>	5x10 <sup>-6</sup> - 10
<b>Sedimentary Rocks</b>		
Sandstone	8 - 4x10 <sup>3</sup>	0.25 - 125
Shale	20 - 2x10 <sup>3</sup>	0.5 - 50
Limestone	50 - 4x10 <sup>2</sup>	2.5 - 20
<b>Soils and Water</b>		
Clay	1 - 1000	1 - 1000
Alluvium	10 - 800	1.25 - 100
Groundwater (fresh)	10 - 100	10 - 100
Sea water	0.2	5000

Source: Keller and Frischknecht 1966.

## Relevance of Study

This study proposes a solution that will aid in the identification of ground water with a simple voting classifier models. it is an easy to used model as dataset fed into the model will predict groundwater availability with a good performance score as recorded here. The model can handle thousands of rows of data points with above-average performance metrics upon deployment.

## CONCLUSION

In this study, a voting classifier algorithm was applied in the prediction of groundwater. Being a classification problem, the ensemble of four different single models was assembled inside a voting classifier with the aim of improving the performance of the model's prediction.

At the end of the study, it was recorded that the prediction accuracy score is 91% which indicates a good fit for deployment. A voting classifier is proposed to perform better upon deployment because it accounts for the bias of each of the singles while maintaining a high-performance score. It can thrive in any terrain and whatever data is fed

on it will always almost give a very good performance.

## Limitations/Challenges and Recommendations

1. This particular solution has not in any way been deployed. Thus, there might be a slight reduction in the performance score of the model due to certain unconformities upon deployment. Further research can be tailored toward this area.
2. Accessing enough datasets in this area of study was indeed a setback. If enough data would be made available it will go a long way in making accurate results for this model.
3. ML can help in the ideal of automation, thereby saving time in the routine task, and cost.

## REFERENCES CITED

- Alteryx. (2020, May 20). Trifacta. Retrieved from trifacta.com: <https://www.trifacta.com/data-wrangling/>
- Akanbi, O. A., 2017. Lithologic Characterisation of the Basement Aquifers of Awe and Akinmorin Areas, Southwestern Nigeria Global Journal Of Geological Sciences Vol. 16, 2018: 1-11
- Aniyom, E., Chikwe, D. A., & Odo, E. J. (2022). Hybridization of Optimized Supervised Machine Learning Algorithms for Effective Lithology Identification at Global Scale. Onepetro.
- Batte, A.G., Muwanga, A. and Sigrist, W.P. (2008) Evaluating the Use of Vertical Electrical Sounding as a Groundwater Exploration Technique to Improve on the Certainty of Borehole Yield in Kamuli District (Eastern Uganda). African Journal Science Technology, 9, 72-85.
- Bhandari, P. (2022, September 19). Scribber. Retrieved from scribber.com: <https://www.scribber.com/methodology/data-collection>
- Dudovskiy, J. (2021, June 09). Business Research Methodology . Retrieved from research-methodology.net: <https://research-methodology.net/research-methods/data-collection/>
- Eslam A. Hussein et el Christopher Thron. Groundwater Prediction Using Machine-Learning Tools. Article. 2017, pg.2
- Epuh, E. E, Sanni, K. A., & Orji, M. J. (2018). Application of Remote Sensing, GIS and Hydrogeophysics to Groundwater exploration in parts of Lagos Metropolis: A case study of Oshodi/Isolo LGA. Nigerian Journal of Environmental Sciences and Technology (NIJEST). 2(1), 41- 55
- Fadele, S.I., *et al.* (2013) The Use of Vertical Electrical Sounding (VES) for Groundwater Exploration around Nigerian College of Aviation Technology (NCAT), Zaria, Kaduna State, Nigeria. The Pacific Journal of Science and Technology, 13, 549-555.