Received: 2022-09-12
Accepted: 2022-12-10
Published Online: 2022-12-30

International Journal of Multidisciplinary Research and Explorer (IJMRE)
E-ISSN: 2833-7298, P-ISSN: 2833-7301

Water Resources Mapping along the Jilango /Shabel-Dulla Areas of Laghdera Subcounty in Northern Kenya

Dr. Meshack Owira Amimo[1], Andrew Rage Eysimkele[2]

[1]Water Resources Authority Headquarters, Nairobi Kenya

[2]Ag CEO, Northern Water Works Development Agency, Garissa County Kenya

[1] bmoamimo@gmail.com

[2]eysimkele@yahoo.com

Abstract— Jilango and Shabel-Dulla areas in Laghdera Sub-County, Garissa County, Kenya, experience severe water scarcity that contributes significantly to poverty among nomadic pastoralist communities. This study assessed water resource availability, quality, and priority using field surveys, geophysical mapping, GIS, and decision tree algorithms. Groundwater potential and expected quality were evaluated, with results showing high prediction accuracy (over 90%). Feasibility of various water supply options—including wells, earth pans, springs, and piped water—was analyzed based on cost and community needs. Groundwater was ranked as the second priority source due to its saline nature and low yield, though it remains suitable for livestock use. The study recommends constructing a large storage dam to improve domestic water supply while complementing limited groundwater resources.

Keywords—Decision tree, Distal Merti aquifer, python, R, precision, accuracy.

I.Introduction

The Project targets a population of at least 500 persons and 400 animal stocks, on the average, respectively for domestic and livestock use. For water scarcity reasons, the centre appears literally deserted at the time of our visit this BARAKI/GURUFA subsets of the Merti aquifer, along the Laghdera river course, to fetch domestic water supply via donkey drawn carts, as well as water their livestock. Only women and a few children of school-going age are left at home, alongside

Received: 2022-09-12
Accepted: 2022-12-10
Published Online: 2022-12-30

International Journal of Multidisciplinary Research and Explorer (IJMRE)
E-ISSN: 2833-7298, P-ISSN: 2833-7301

the elderly. Some 50 cubic meters per day yield (MINIMUM) is feasible from this well, given the hydraulics inferred from the geophysical curves generated. This is a donor-funded water supply project meant to enhance the rapid settlement of the local population and enhance rapid growth of the township. It will also help having the local children easily absorbed into schools for vocational, academic and spiritual growth. The village has a daily requirement of approximately 30 cubic metres of water to address the needs forestated. The Direct Aid Organization, formerly African Muslim Agency (AMA) has funded the drilling and equipping of the facility, and will hand it over to the county government for further development of infrastructure systems. The borehole will be developed upon drilling completion and should be preferably encased with steel casings. Once the productivity of the borehole has been determined, a suitable submersible pump will be installed to pump water into the proposed storage tanks. The schematic design and the detailed itemization for the proposed borehole shall be the subject of phase two work for the planning and designing unit, but will be predicated on the borehole performancein terms of aquifer yields and recharge. In case the yield will be too low for a submersible motor powered pumpage, a hand pump or windmill driven pump system is suitable as well.

A. Project Ownership

The proposed site is a public facility owned by the local Shabel-Dula community

II. Hydrogeology

A. Geology and Stratigraphy

The topography is undulating dotted with several anthills which are clayey rich, and support vegetations that comprise mainly thorny shrubs, undergrowth and acacia family trees. Several acacia family units flourishing are favored and devoured by the camels. The thorny shrubs also act as building materials suited to putting up dwelling structures.

The geology is defined by red to light toned sandy clayey sediments, the Jurassic clays and sands, which overlies the carbonates – namely corallites, aragonitic sediments and calcite. The sandy clayey species are mainly the Miocene-Pliocene gravels and Sandstones.

The Jurassic limestone carbonates are fairly fractured and possess water at the shallow depths, though highly mineralized, via the fractures and karstification veins. Water also forms at the contact points between the carbonates and the Archaean metamorphic basement units.

Received: 2022-09-12
Accepted: 2022-12-10
Published Online: 2022-12-30

International Journal of Multidisciplinary Research and Explorer (IJMRE)
E-ISSN: 2833-7298, P-ISSN: 2833-7301

Groundwater in the upper sediments shall enjoy annual precipitation recharge through direct infiltration, while the deep-seated zones shall be recharged via regional flow aided by the karstification channels and plate tectonics in the Jurassic – cretaceous period. Evapo transpiration rates of up to 3,000mm per annum over shadow the annual rains of up to 400mm per annum.

B. Physiography

The area is endowed with a low/ unfavorable physiography. It stands at an average altitude of 265 metres above sea level within a gently dipping terrain punctuated with several ant-hills and flood plains both on the south eastern and north western flanks

. III. Project Location

A. Location

The project area lies in North-eastern Region within the Lagderasubcounty. It is located on the southwestern sides of the main catchments course way. The area is defined by longitudes and latitudes shown in the geophysical curves analyzed, and at an altitude of approximately 265m above sea level. Oblique dipping sediments litter the terrain alongside some zero degree dipping units of flood-prone Miocene Pliocene sediments.

B. Nature of the Project

This is a proposed institutional project meant to enhance the rapid settlement of the local population by having them practice small scale agriculture as well as enhanced rapid absorption of children into schools for academic and spiritual growth. The project shall have a daily requirement of approximately 50 cubic meters, computed from the population figures. This is a figure computed to include daily domestic requirements as well.The borehole/shallow well will be developed upon drilling completion and should be preferably encased with UPVC casings. Once the productivity of the borehole has beendetermined, a suitable submersible pump will be installed to pump water into the proposed storage tanks. The schematic design and the detailed itemization for the proposed borehole shall be the subject of phase two work for the planning and designing unit, but shall be predicated on the borehole performance in terms of aquifer yields and recharge.

Received: 2022-09-12
Accepted: 2022-12-10
Published Online: 2022-12-30

International Journal of Multidisciplinary Research and Explorer (IJMRE)
E-ISSN: 2833-7298, P-ISSN: 2833-7301

The proposed site is a public facility owned by the un registered Shabel Dulla- Community Water Project. At an appropriate stage, the community will be trained on management related issues pertaining to sanitation, as well as borehole operations and maintenance.

C. Site alternative and proposed

action- Any site located within a radius of 10m away from the pegged spot should have promising groundwater potential as the Stratigraphic units possess continuous permeability, both lateral and vertical.

D. Geology, Geomorphology and Stratigraphy

Geomorphologically speaking, the slopes, soil types, vegetation and drainage of the project area is not conducive for perennial flow from the annual rainfall runoff.

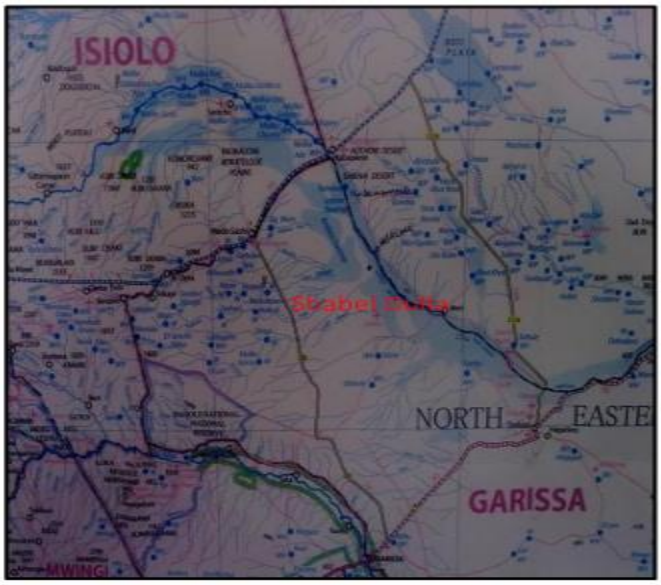
The topography is generally flat, and is clayey rich, supporting vegetations that comprise mainly thorny shrubs, undergrowth savannah grass, equatorial weeds typical of the coastal strips and acacia family trees. The area is apparently a buried Quaternary Lake, with evidence of fresh water systems fish populations. Quaternary Tectonics probably resulted into the landmass uplift with the results that some of the waters were emptied the surrounding Coastal plains. There is more evidence to suggest a coastal climate than a desert climate for the Project Area. The geology is defined by dark to light toned sandy clayey sediments, the Mansa Guda formation, which overlies the carbonates – namely corallites, aragonitic sediments and calcite. The sandy clayey species are mainly the Mariakani Sandstones.The Jurassic limestone carbonates are fairly fractured and possess water at the shallow depths, though highly mineralized, via the fractures and karstification veins. Water also forms at the contact points between the carbonates and the Archaean metamorphic basement units.

Groundwater in the upper sediments shall enjoy annual precipitation recharge through direct infiltration, while the deep-seated zones shall be recharged via regional flow aided by the karstification channels and plate tectonics in the Jurassic – cretaceous period. Evapo transpiration rates of up to 3,000mm per annum over shadow the annual rains of up to 400mm per annum.

Map 1-Map showing the Location of Shabel Dulla Study area in the Laghdera subcounty

Received: 2022-09-12
Accepted: 2022-12-10
Published Online: 2022-12-30

International Journal of Multidisciplinary Research and Explorer (IJMRE)
E-ISSN: 2833-7298, P-ISSN: 2833-7301



E. Hydrology, Hydrochemistry and Structural Geology

1. Recharge Mechanisms within the Lower Tana Aquifer Systems:Evidences abound of jointing and fracturing of the carbonate sediments on the surface, alluding to intense forces of fracturing, carbonation and quaternary tectonic faulting. Much of the south westerly – north easterly directed stress fields helped sculpture the terrain into its present geological state. Owing to the relatively high fractions of clays in the beds, there is no sufficient time available for maximum catchment input infiltrations into the sub surface zones lying on the adjacent aquifer units in the proposed well sites. This explains the anomalous salinity levels of the boreholes done to great depths in the area.

2. Drainage:Owing to the relative flat nature of the terrain, there is flood rampancy. The permanent civil structures on the ground to stand the risk of destruction added to the occasional loss of lives for both livestock and human persons. Most of the housing units are constructed through shrubs and dry acacia trees locally available, lightening the task of evacuation in the event of impending flood disasters.

3. Climate:The project area falls within zone 7 of the classification of climatic/ecological zones of Africa, that is to say arid to semi arid with temperatures averaging 30 to 34 degrees per day and occasioning evapo transpiration rates of up to 3000mm per annum. The rainfall average falls well below 500mm per year.

Received: 2022-09-12
Accepted: 2022-12-10
Published Online: 2022-12-30

International Journal of Multidisciplinary Research and Explorer (IJMRE)
E-ISSN: 2833-7298, P-ISSN: 2833-7301

F. Geophysics:

1. Introduction:In order to determine the Projects Area's hydrostragraphy and aquifer suitability, a total of 3No. Vertical Electrical Soundings were undertaken using the modern ABEM SAS 4000B Terrameter. Schlumberger arrays were used so that current electrode spreads of up to 320m against potential spreads of between 5m and 25m were employed to conduct the surveys. Copper electrodes were used for the potentials, while steel iron electrodes were used for the currents. The best site is analysed and tabulated hereunder.

Table 1: The Geonalysed Data

Resistivity Curve No	Schlumberger Probe Depth Interval(m)	Resistivity In OhmM	Expected Geological sediment/Formation
R-002/2021	0-1	80	Top Alluvial Sediments
The second site	1-3	32	Subsoils/Clayey Sediments
under acacia trees.	3-40	20	Sandstones/gypsites
The recommended	40-80	8	Quartzites/Clays
site as shown to the	80-250	12.5	Clays and sandstones
team by an elder	Over 250	7.5	Fine sandstones and clays

2. Background A theoretical treatise on how the ML algorithm of Decision trees work is of essence, to provide a background for understanding how the predictions may be made using this method in groundwater hydrology. This discussion hereunder will explain both the theoretical and practicalusage of the Decision tree (DT), using both python and R softwares. It is essential that one understands the technical terminologies and vital concepts, which relate to the DT algorithm.

a) The Decision Tree Algorithm:In simple terms, the decision tree algorithm is a graphical representation of all the feasible solutions to a decision, from the dataset array used, having made certain assumptions. The method is so name, since it is begins with a single variable. This variable subsequently branches off into a number of solutions, forming what looks like a tree that has been turned upside down.

b) Composition of a Decision Tree Model:The first important component is the Root Node. It lies at the apex of the DT model, forming an appearance akin to that of the tree that has been turned upside-down. It represents the best predictor out of all the variables appearing in the dataframe being used for predictive analytics. The next component is the Decision / Internal Node, on which predictors in the dataset, forming several branches, each one of which represents an outcome of the test so performed using the DT. The leaf node is the third component regarded as being of vital

Received: 2022-09-12
Accepted: 2022-12-10
Published Online: 2022-12-30

International Journal of Multidisciplinary Research and Explorer (IJMRE)
E-ISSN: 2833-7298, P-ISSN: 2833-7301

significance. It represents the final result of the exercise of classification. It is also known as the terminal node.

3. The strengths and weaknesses of DT Model

The DT algorithm has strength of being not sensitive to outliers in the dataset being modeled. The DT model also works perfectly, regardless of the non-linearity relationships in the datasets. The conventional simple or linear regression models will give rise to unintelligible results with variables that are non-linearly related to each other. Owing to the simple branching and splitting of variables, the DT model is easier to interpret. On the downside, the DT algorithm may tend to overfit easily, thereby giving false predictions. The model assumes that all variables are related to each other and this may not always be necessarily the case.

We can identify overfitting by looking at validation metrics, like loss or accuracy. Usually, the validation metric stops improving after a certain number of epochs and begins to decrease afterward. The training metric continues to improve because the model seeks to find the best fit for the training data.

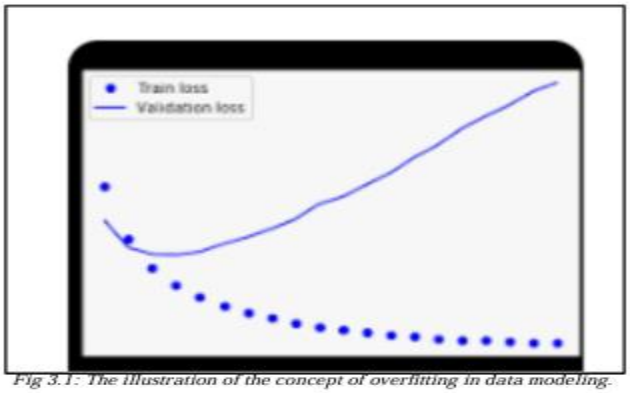


Fig 3.1: The illustration of the concept of overfitting in data modeling.

4. Basics terms relating to the DT Algorithm

1) Pruning: Correct Overfitting: Literally, in plant science or agriculture, pruning implies removing excess leaves from a plant so that the remaining ones will maximize their life-giving processes such as evaporation and photosynthesis. The excess leaves will let the plant occasion unnecessary loss of water, which means that pruning reduces the unnecessary water loss from then plant.

In Data science, it is a technique employed to help diminish the impacts of overfitting in a DT model. It logically and progressively reduces the size of decision trees, by eliminating the sections

Received: 2022-09-12
Accepted: 2022-12-10
Published Online: 2022-12-30

International Journal of Multidisciplinary Research and Explorer (IJMRE)
E-ISSN: 2833-7298, P-ISSN: 2833-7301

of the tree, which bestows or grants little power, during the process of DT classification problems. Pruning is employed to eliminate anomalies, in the training phase, for the data being modeled, on the account of noise or statistical outliers. The pruned trees generate a model which is easier to understand and use, as well as interpret.

a) **The Pre-Pruning phase:** This term defines the stage or phase at which the modeler stops growing the tree, when it appears that there is no (longer) a statistically significant association, existing between the variable being grown and the class anticipated, at a particular node. To determine this, a test has to be performed. One such test is the chi square. The chi-squared test is undertaken, to make a determination of the existence of statistically significant association between the two variables.

b) **The Post-pruning phase:**This term refers to the process of building the full tree, which as noted earlier, is a tree standing upside-down, as well as pruning the tree. The process involves determining cost complexity. The cost complexity is one of the most popular post-pruning methods. It is measured by the two parameters, namely, the Number of leaves in the tree and the error rate of the tree implyingthe misclassification rate or the SSE) Complexity Parameter of the tree is also abbreviated as the CP. The modeler has to determine value of the smallest tree that has the smallest cross-validation error value. This is different from the conventional linear regression modeling processes and error metrics. In MLR regression problems, this would imply that the overall R squared must increase by cp, at each step. This is of significance in arresting the nuisance of overfitting in DT models. It implies a trade-off between the size of a tree and the error rate, to help diminish or prevent (altogether) the problem of overfitting. Consequently, the large trees with a low error rate are penalized, whereas the small trees with low error rates are rewarded implying in favor of the latter category. This Cost Complexity is the tuning parameter in the Classification and Regression tree algorithms, which are abbreviated as CART.

c) **Splitting:** As explained earlier, the DT model turns the variables into a tree, complete with roots and leaves. For this to happen, lots of splitting shall be involved. The splitting is thus a process of dividing a node into two or more sub-nodes

Received: 2022-09-12
Accepted: 2022-12-10
Published Online: 2022-12-30

International Journal of Multidisciplinary Research and Explorer (IJMRE)
E-ISSN: 2833-7298, P-ISSN: 2833-7301

. d) Branch: Several branches will give the tree, as the branch comprises the leaves and nodes. The branch is therefore a sub-section of entire tree.

e) Parent Node: A super node splits into several others in the tree model. The node which splits into sub nodes is the parent node.

f) Child Node: The parent node branches out into sub-nodes. The sub-node is a child node.

g) Surrogate Split: In several instances during development of a DT model, data may be missing for certain variables.in the event that some data is missing, and the DT model will return output predictions that may include surrogate splits. As an example, in case the surrogate set value is 2, it means if the primary splitter is missing, the modeler may use the number one surrogate. They may be used inter-changeably, so that if the number one surrogate is missing, then we use the number two surrogate.

5. Classification Tree

When one wishes to estimate the class of attributes in a data set, the classification tree generated by the DT model is of essence.The outcome variable is a categorical variable (and as an instance, this may refer to saline water or fresh, in the case of two-class (binary DT problems). The predictors may be continuous variables (with discrete numerical values) or a mixture of both categorical and continuous variables.

In groundwater hydrology, the predicted variable (dependent) may be the TDS. The independent values being used to make these predictions are variables like the GPS coordinates, depths to aquifer and the radii away from the nearest recharge river-courses and or major hydrological drainage parameters.

6. The workings of a DT model

a) Using the gini index to pick the variable giving out the best split-the modeler is advised to take the variable that gives the best split usually the one giving the lowest gini index value.

b) Once this is done the splitting begins in earnest. The data shall then be subject to the necessary splits, based on the gini index value. Partition the data based on the value of this variable

c) The process a) and b0 are repeated , so that the splitting may only stop when the CARET model detects no further information gain can be made

Received: 2022-09-12
Accepted: 2022-12-10
Published Online: 2022-12-30

International Journal of Multidisciplinary Research and Explorer (IJMRE)
E-ISSN: 2833-7298, P-ISSN: 2833-7301

7. Algorithms of Classification

Tree The Gini Index for splitting: The most important parameter in DT models is the gini index value. This parameter measures impurity levels in a node. The value ranges between 0 and (1-1/n) where n is the number of categories in a dependent variable. The categories may refer to the number of classes being modeled. As an example, a three-class model in ground water hydrology may include parameters like hard water, fresh water and saline water.

$$Gini = 1 - \sum_{i=1}^{C} (p_i)^2$$

In the above equation, 'p' refers to probability of class. In layman's language, it can be read as thus: Gini=1 – (P(class1)^2 + P(class2)^2 + … + P(classN)^2) It should be noted that gini index favors larger partitions. It is of vital essence that the modeler takes several factors into consideration.

Number one, the modeler may get a Gini index value of zero. The zero Gini index implies perfect classification. Two, (1 - (1/ No. of classes) implies worst classification. The modeler always hunts for /wants a variable split having a low Gini Index. For binary dependent variable, max Gini index value can be 0.5. this is worked as shown hereunder:

$$\begin{aligned} &= 1 - (1/2)2 - (1/2)2 \\ &= 1 - [\,2*(1/2)2] \\ &= 1 - 2*(1/4) \\ &= 1 - (0.5) \\ &= 0.5 \end{aligned}$$

8. Entropy and Information Gain

A rather advanced method of splitting in DT modeling exists. It is known as entropy. It is a wonderful criterion for classification using the DT algorithm. The formula is illustrated hereunder:

$$Entropy = \sum_{i=1}^{C} -p_i * \log_2(p_i)$$

Received: 2022-09-12
Accepted: 2022-12-10
Published Online: 2022-12-30

International Journal of Multidisciplinary Research and Explorer (IJMRE)
E-ISSN: 2833-7298, P-ISSN: 2833-7301

The above equation may be read as follows: P(class1)*log(P(class1),2) + P(class2)*log(P(class2),2) + … + P(classN)*log(P(classN),2)

a) It favors partitions that have small counts but many distinct values:Just like the gini index method, low value of Entropy implies a good classification.

b) Information Gain can be evaluated, metrically, as shown hereunder:

= Entropy (Parent) - Weighted Sum of Entropy (Children) One may use either the gini index formula or the entropy formula as both have been shown to be such reliable. Both splitting procedures are approximately similar and will generate near- similar result in ninety-five percent of the known instances of their use. The gini index method is comparatively faster than Entropy as it does not requireCalculation of log. One may thus opt for whichever one works best as different dataset may favor the use of either criterion.

C) Split Method Least-Squared Deviation or Least Absolute Deviation:The impurity of a node is measured by the Least-Squared Deviation (LSD), which is simply the within variance for the node.

IV. Literature Review

Khader et al (2013) undertook a study on how to employ the DT algorithms to help monitor groundwater quality of aquifers. In the study, well-water contaminated with nitrate was deemed to stand out as a public health hazard, to the low age-brackets, especially the infant children in terms of raising the mortality rates. This happened when the water was used in domestic purposes. ML algorithms were employed to determine the contamination status of aquifers with nitrogen-based contaminants. This called for a deliberate effort at designing a working groundwater quality assessment template to aid gathering pertinent info on aquifer conditions that gave rise to the waters being studied. This is the info that was proposed for use in management decisions on whether or not to condemn the waters. The ML algorithm, Decision Tree or simply DT was used by the decision making study groups to help come up with the expected outcomes from the data so assembled.

In 2020, Mirhashemi et al mapped the Qazvin Plains in Iran. He employed the use of the DT algorithm. The study found that in recent years, more focus had been accorded issues of water resources, on the account of the development of agriculture and management strategies of the aquifers. The considered the impacts of various causalities on aquifer depth variations, in that study. The study subsequently concluded that human and environmental factors played the most

Received: 2022-09-12
Accepted: 2022-12-10
Published Online: 2022-12-30

International Journal of Multidisciplinary Research and Explorer (IJMRE)
E-ISSN: 2833-7298, P-ISSN: 2833-7301

significant roles in the Qazvin plain aquifer dynamics. The Classification and Regression Tree found use in the study and especially the DT classifier component, adopted for assessment and predictive analytics of aquifers. The studied fund out that the highest probability levels of the aquifer drop came up in the months of July, August and September, and was estimated at approximately eighty seven percent.The highest probability of rise in aquifer depths was observed to be in the Months of December and January, amongst others.

Yoo et al (2016) also undertook a study involving aquifer assessments using decision tree methods. The study deliberately targeted the development of a new methodology that would be effective at establishing estimated patterns of groundwater pollution sensitivity levels, using data mining procedures. The proposed algorithms utilized seven hydrogeological properties as input variables for the study, and the variables were:

i) depth to aquifer water tables

ii) net recharge of aquifers

iii) aquifer media

iv) aquifer soil media

v) topography of aquifer locality

vi) aquifer vadose zone media

vii) Aquifer hydraulic conductivity levels.

The study employed four data mining algorithms, namely-

(a) Artificial neural network

(b) Decision tree (DT),

(c) case-based reasoning

(d) Multinomial logistic regression (MLR) was tested. The study concluded that the Decision trees -based data analysis and the rule induction methods both displayed and yielded more reliable and accurate predictions accuracy and consistency.

Stumpp et al (2016) also mapped aquifers vulnerability index by employing the powerful predictive capability of the DT algorithm. The study established the specific aquifers deemed to be at risk were subject to more precise risk assessment where the concept of vulnerability was factored in course of the study design as thus:

Received: 2022-09-12
Accepted: 2022-12-10
Published Online: 2022-12-30

International Journal of Multidisciplinary Research and Explorer (IJMRE)
E-ISSN: 2833-7298, P-ISSN: 2833-7301

Source–Pathway–Receptor.The study vouched for a systematic operational approach, primarily based on a decision tree, which led the user through the stages of aquifer vulnerability assessment. This meant formulating the problem, before relating it to a threat, quantitatively or qualitatively, on an aquifer.

Carretero et al (2020)undertook a study in Argentina, again using DT in mapping aquifers. The decision tree formed the basis of defining the hydrogeologic parameters (aquifer depths, aquifer thickness, aquifer surface area and the radii existing between the wells). The DFT helped pick the most favorable groundwater abstraction procedure, deemed most viable, in the study, which involved mapping of un-confined coastal aquifers. The negative impacts occasioned by the use of the inappropriate abstraction procedure for groundwater were analysed. It was noted that as a result of excessive extraction, and un-controlled abstractions, there was dramatic decrease of the freshwater reserves. The decision tree was thus a useful tool, in unraveling decision making on matters abstraction in the sensitive fresh-saline water interfaces.

Dauji (2021) undertook a study on chlorideions, on its role as an important parameter defining of water quality status. The procedures adapted for field measurement of the ions was deemed involving. Moreover the lab procedures were deemed to be both time-consuming and chemical involving. The chemistry of chlorine and its high values of correlation coefficient with electrical conductivity suited the latter for use as proxy for estimating chloride ions.

Krhoda et al (2019) did research water which came up with a model to be used to predict expected water quality before one drill it for the Garissa County. This involved in putting the longitudes, latitudes, elevations, aquifer depths and the resistivity values of depths deemed as productive. The model proved effective with prediction accuracy of over 90 percent.

Another study was undertaken and published these years on DT modeling of aquifers, by Singha et al (2022). The study focused on the spatial variations inherent in the aquifers systems mapped, which explained the groundwater hydrology parameters in the aquifers surrounding the large coal mining fields of Central India. The CART models of the ML order were employed to undertake this kind of study, the aim being to assess the variables that played the most significant roles in lowering the quality of the aquifer waters.

The results generated indicate that the factors deemed pertinent, in their order of decreasing importance, in this respect were related in the order hereunder: Aquifer Slope (34%) > Distance to

Received: 2022-09-12
Accepted: 2022-12-10
Published Online: 2022-12-30

International Journal of Multidisciplinary Research and Explorer (IJMRE)
E-ISSN: 2833-7298, P-ISSN: 2833-7301

Active Mines Deemed Contaminant-Laden (23%) > Aquifer Water Table Depths (16%) > aquifer Drawdowns (15%) > Aquifer Locality Elevation Levels (12%). All these variables were used to predict the groundwater quality index

V. Methods Used In The Study

Part One: Mapping the Available Water Resources for Study Area

In the present study, existing hydrochemical data and water resources assessment reports in the NWWDA database were used to make an informed position on what structure suits the Shabel-Dulla populace. The hydrology and drainage of study area as well as soils chemistry and mechanics was subjected to insitu assessment and the information so generated summarized for the purpose of determining da suitability. This involved fieldwork and transects, as well as some surface geophysics employing the VES probes, which was done and helped generate a 2D model of soils in study area. Springs development potential was noted to exist but unreliable so, as the spring points may only be active in the wet season. Sand wells were thus more rated more favorable than them, along the riverbed.

a) Earthpan Potential-The soils were determined to be of favorable order for dam siting and excavations they have lots of favorable levels of clays , which may be used as both an impounding layer as well as backfilling material for the few places found to be leaking. The Earthpans can thus be developed to tap in the inflow from the tributaries of the seasonal flow course. One may do a separate pan for livestock as well.

Model 1-Model of Tomographic Image of sediments from point 0 meters to 30 meters.

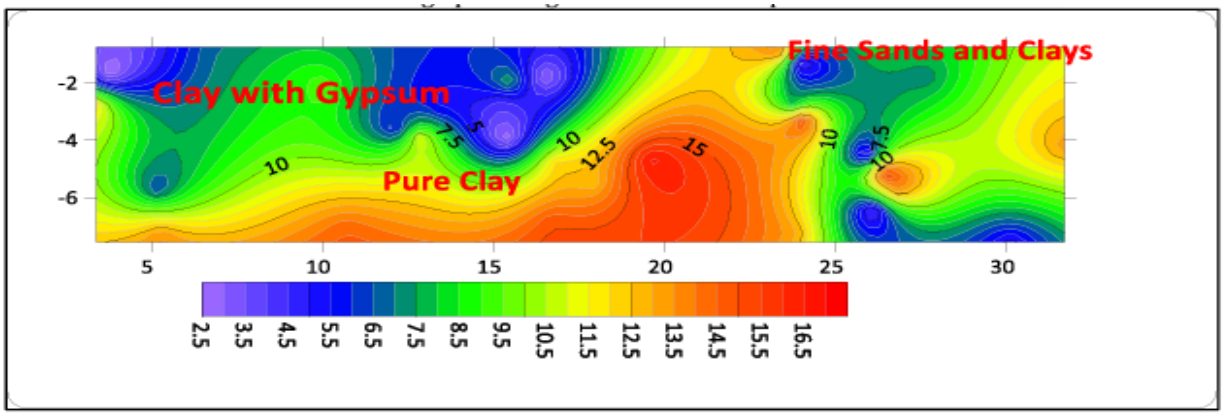


Fig 5.1: Tomographic Image model showing the sediment layers

Received: 2022-09-12
Accepted: 2022-12-10
Published Online: 2022-12-30

International Journal of Multidisciplinary Research and Explorer (IJMRE)
E-ISSN: 2833-7298, P-ISSN: 2833-7301

Explanation: The Model shows that in the depth brackets 0.1m to 7.0m bgl, resistivity of material oscillates between 2.5ohmM and 16. 5ohmM.The lowest resistivity values represent zones enriched in gypsum-based deposits, mainly the gypsites. These are zones that tend to be bluish in color. The zones that are greenish in color represent fine sands with clays, and this has acceptable soil mechanics required for average to maximum impoundment potential. The other zones that are orange are pure clays in mineralogy and bear the requisite plastic properties for the most favorable impermeability deemed conducive for the layers to act as good impoundment material for earthdams. On the basis of the foregoing geophysics, one may conclude that the proposed project area has acceptable levels

b) Groundwater Potential-This was undertaken using the VES probes and the results indicated that there exists a limited potential for abstracting saline water essential for livestock and sanitation purposes.

c) Spring Development Potential-This potential exists within such a far-fetched probability to the point of being deemed non-existent, altogether.

d) Sand wells-These are ranked as feasible but not practical as the water so sourced may be enough for just a single family's domestic sage living the livestock without any reserves for use.

e) Rain water harvesting from Roofs- this is nonexistent since practically all the housing units have no corrugated iron-sheets. Traditional grass material; or equivalents have been used to build houses.

Received: 2022-09-12
Accepted: 2022-12-10
Published Online: 2022-12-30

International Journal of Multidisciplinary Research and Explorer (IJMRE)
E-ISSN: 2833-7298, P-ISSN: 2833-7301

Table 2: Strengths And Weaknesses Of Different Water Structures

S/No	Source-priority |Number-and Name	Strengths /weakness	Estimate cost	rank
1	earthpan	Material has lots of clays which favors the excavations of small pans. Water quality is fresh.	For a **20,000**cubic meters facility, one may **need Kshs 8,000,000.**	1-cheap and has good quality water.
2	Groundwater deep borehole	Deep aquifers have water from the distal Merti storage. Water is saline	For drilling and equipping a 300m well, one my use up to Kshs 8-10 millions	2-one may use solar to operationalize
3	Spring	Occur on river banks whenever it rains. Water is little and unreliable after rain season	One may use as little as just Kshs 50000	4-very few such points along the River Jilango Profile
4	Sand wells	Water can be obtained at shallow depths shortly after rainfall seasons and even six months afterwards	Costs Nothing at all	3-Not every river cross section has the potential to bear the shallow aquifers
5	Roof harvesting	All one needs to do is buy a storage and attach it to roof materials to harvest rainfall water. Houses have no roofs made of iron sheets	Cost of buying storage tanks only-uptoKshs 30,000	5-Not practical for reasons forestated.

VI. Water Resources Data Analysis

Part Two: Estimating the Water Quality in Terms of Total Dissolved Solids in Shabel-Dulla Area

To undertake this, simple dataframes of GIS info (longitudes, latitudes, elevations, distance to flow course) and the distance to Laghdera stream, as well as the resistivity data, was employed. The study involved prior estimation of salinity using TDS as the dependent variable, after which the next phase involved predicting the depths to water table. The resistivity data was analysed using IPI2WIN softwares and gave approximated depths into aquifers.

i) The dataframe assembled was rendered in excel csv format and was prepared in such a way as to present class of well TDS for the existing borehole whose TDS value existed alongside all the other parameters like longitudes and latitudes, alongside elevations.

ii) Decision Tree was used as the main algorithm for use, based on previous experience with data in the Merti aquifer. This data was then called into the R software from which analysis was

Received: 2022-09-12
Accepted: 2022-12-10
Published Online: 2022-12-30

International Journal of Multidisciplinary Research and Explorer (IJMRE)
E-ISSN: 2833-7298, P-ISSN: 2833-7301

undertaken to generate the predictive model deemed viable for use to estimate the TDS class of a newly proposed well site.

iii) The data was partitioned into training and testing subsets for the purpose of generating prediction analytics and assessing the accuracy performance of the Decision Tree algorithms.

iv) If found to have between 80 to 90 percent levels of accuracy, the model was deemed favorable. If not, a different algorithm was sourced. The present study found the DT classifier to be such a super performer. The model was named 'modelDT'

v) Raw field data comprising longitudes, latitudes, elevations and distance to EwasoNg'iro flow course were now prepared for all the points earmarked for surveys and drilling, if found favorable.

vi) Any study data fed into the model was thus predicted as either favorable or otherwise.

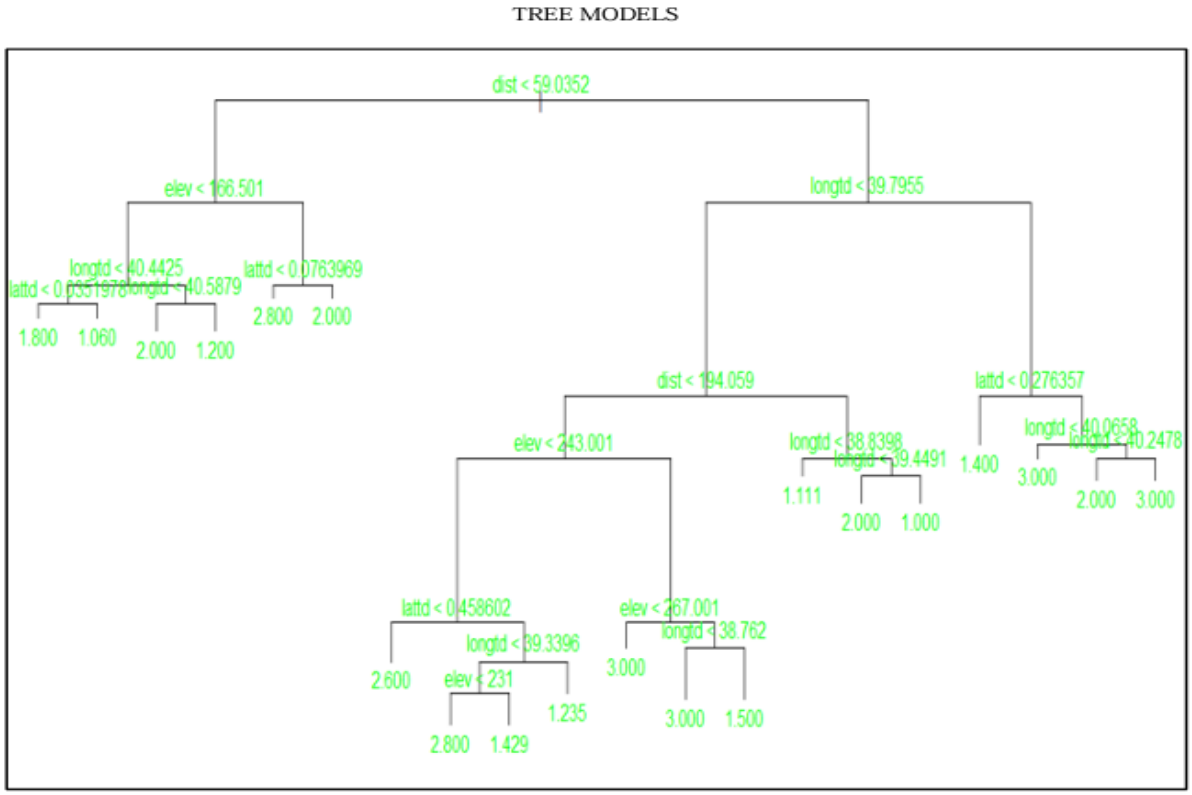


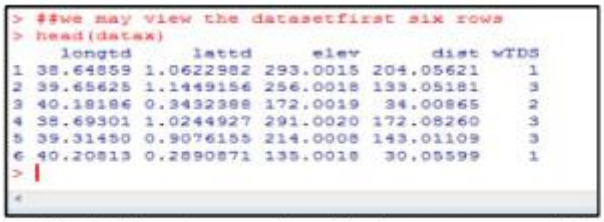
Fig 6.1: Data interpretation of the tree model for Shabel Dulla Studies

the data to be analysed lies in excel csv sheet ## the csv sheet is named innovData datax=read.csv("innovData.csv",header=T,na.strings="NA ") The dataset comprises the GPS coordinates and the distance of the well locations relative to the EwasoNgiro river course. This is because the river course has underground water flow in the subsurface and this is the major

Received: 2022-09-12
Accepted: 2022-12-10
Published Online: 2022-12-30

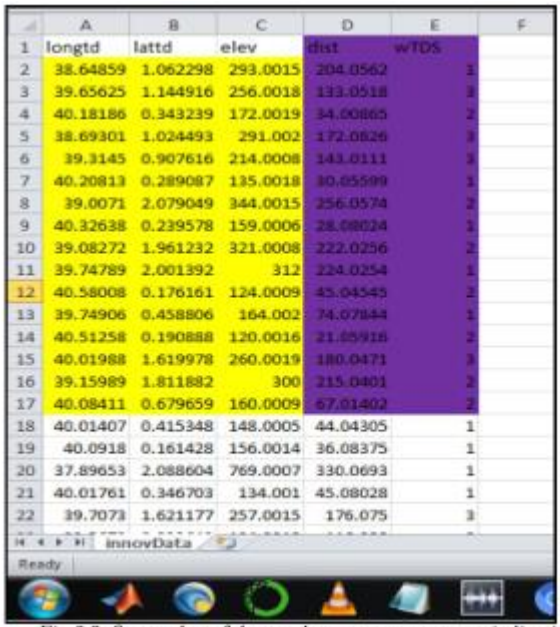
International Journal of Multidisciplinary Research and Explorer (IJMRE)
E-ISSN: 2833-7298, P-ISSN: 2833-7301

hydrologic parameter which influences aquifer recharge. Where the river is located less than 25 kilometers away from the Merti aquifer zone earmarked for drilling, the water is found to be fresh, with water TDS ranging between 100 to 900 mg/L, coded as 1 inthe dataframe. Where the site is located between 25 to 40 kilometer away from the Laghdera Flow Course, the water is found to be a fuzzy range of freshness, oscillating between fresh and hard, coded as 2 in thetable of excel sheet. In the event of distance of well location way above 40 kilometers, the water is found to be saline or brackish. This is coded as 3 in thetable of data sheets. ##we may view the datasetfirst six rows head(datax) By running the above code, the table of dataframe comes out. It is this table that may be trained and tested for model precision prior to using it. In the actual work, the model generated predictions for the entire dataset. The GPS locations and the distance away from the giant EwasoNgiro river stand out.



```
> ##we may view the datasetfirst six rows
> head(datax)
    longtd     lattd      elev      dist wTDS
1 38.64859 1.0622982 293.0015 204.05621    1
2 39.65625 1.1449156 256.0018 133.05181    3
3 40.18186 0.3432388 172.0019  34.00865    2
4 38.69301 1.0244927 291.0020 172.08260    3
5 39.31450 0.9076155 214.0008 143.01109    3
6 40.20813 0.2890871 135.0018  30.05599    1
>
```

*Fig 6.2: Shows the screenshot of the data structure as used in the modeling of projected water quality in the **Shabel Dulla** Area of **Laghdera Subcounty***



	A	B	C	D	E	F
1	longtd	lattd	elev	dist	wTDS	
2	38.64859	1.062298	293.0015	204.0562	1	
3	39.65625	1.144916	256.0018	133.0518	3	
4	40.18186	0.343239	172.0019	34.00865	2	
5	38.69301	1.024493	291.002	172.0826	3	
6	39.3145	0.907616	214.0008	143.0111	3	
7	40.20813	0.289087	135.0018	30.05599	1	
8	39.0071	2.079049	344.0015	256.0574	2	
9	40.32638	0.239578	159.0006	28.08024	1	
10	39.08272	1.961232	321.0008	222.0256	2	
11	39.74789	2.001392	312	224.0254	1	
12	40.58008	0.176161	124.0009	45.04545	2	
13	39.74906	0.458806	164.002	74.07844	1	
14	40.51258	0.190888	120.0016	21.05916	2	
15	40.01988	1.619978	260.0019	180.0471	3	
16	39.15989	1.811882	300	215.0401	2	
17	40.08411	0.679659	160.0009	67.01402	2	
18	40.01407	0.415348	148.0005	44.04305	1	
19	40.0918	0.161428	156.0014	36.08375	1	
20	37.89653	2.088604	769.0007	330.0693	1	
21	40.01761	0.346703	134.001	45.08028	1	
22	39.7073	1.621177	257.0015	176.075	3	

Fig 6.3: Screenshot of the csv data structure extract indicating the columns and rows used in this study.

Received: 2022-09-12
Accepted: 2022-12-10
Published Online: 2022-12-30

International Journal of Multidisciplinary Research and Explorer (IJMRE)
E-ISSN: 2833-7298, P-ISSN: 2833-7301

install the ISLR library for Statistical learning in R install.packages("ISLR") library(ISLR) The above library is of help in generating the graphics which will display the branch splits in the model as per the entropy rules employed in decision tree modeling.

Now install decision tree library in R library(tree) The above library 'tree' will display the model expected from the tasks at hand and will be the basis of the predictions of the class, denoted as wTDS or "Water TDS level class". ## generate the model now model = tree(wTDS~., data=datax) summary(model) This will be displayed as thus shown:

```
-1.6000 -0.1111 -0.0597  0.0000  0.0000  1.6000
> ##we may view the datasetfirst six rows
> head(datax)
    longtd     lattd      elev      dist wTDS
1 38.64859 1.0622982 293.0015 204.05621    1
2 39.65625 1.1449156 256.0018 133.05181    3
3 40.18186 0.3432388 172.0019  34.00865    2
4 38.69301 1.0244927 291.0020 172.08260    3
5 39.31450 0.9076155 214.0008 143.01109    3
6 40.20813 0.2890871 135.0018  30.05599    1
>
> # Now install decision tree library in R
> library(tree)
> model = tree(wTDS~., data=datax)
> summary(model)

Regression tree:
tree(formula = wTDS ~ ., data = datax)
Number of terminal nodes:  20
Residual mean deviance:  0.1074 = 32.21 / 300
Distribution of residuals:
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-1.6000 -0.1111 -0.0597  0.0000  0.0000  1.6000
>
```

Fig 6.4: The screenshot of the decision tree model used above using R codes.

The model calculator is generated by running the two lines above and shall be indicative of the predictions for the entire dataset. ## generate the Model split graphics plot(model,col="green") text(model, pretty = 0,col="orange") set.seed(101) train=sample(1:nrow(datax), 320) The preceding codes shall generate the splits possible from the entropy rules and is displayed as thus shown:

Received: 2022-09-12
Accepted: 2022-12-10
Published Online: 2022-12-30

International Journal of Multidisciplinary Research and Explorer (IJMRE)
E-ISSN: 2833-7298, P-ISSN: 2833-7301

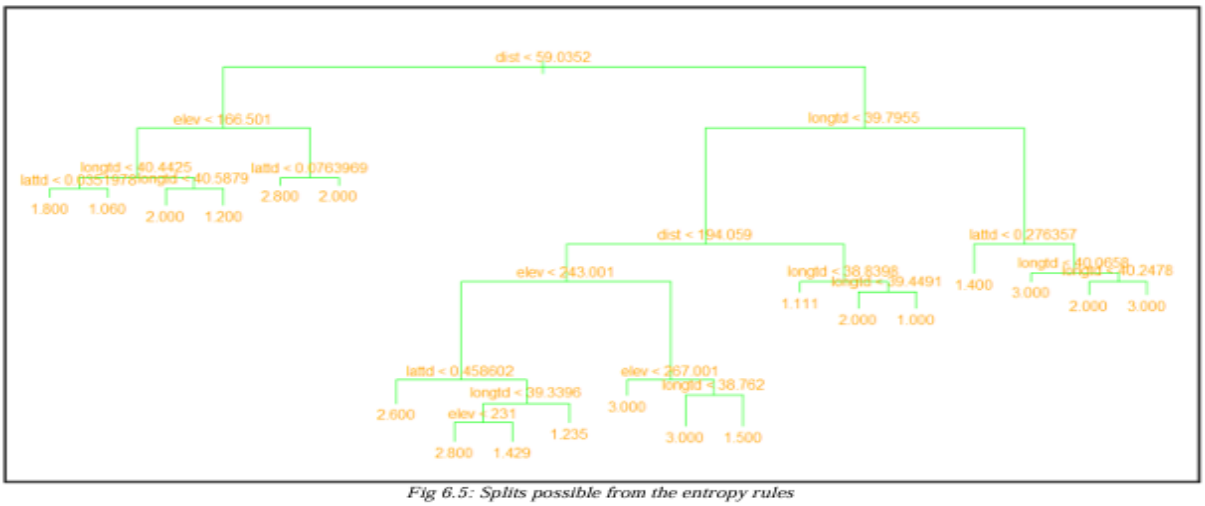


Fig 6.5: Splits possible from the entropy rules

#testing dataset for new sites for Total dissolved Solids #dataT=read.csv("amimoj2.csv",header=T,na.strings="NA ") #dataT In the above datasets, a selection of the rows in the study distal Merti were picked randomly and these have known water quality. Only one is brackish or saline as per the history of the wells. This was done to deliberately use the model to generate predictions in R. The data is displayed in the screenshot shown/displayed thus:

	A	B	C	D	E
1	longtd	lattd	elev	dist	wTDS
2	39.74906	0.458806	164.002	74.07844	1
3	40.51258	0.190888	120.0016	21.05916	2
4	40.01988	1.619978	260.0019	180.0471	3
5	39.15989	1.811882	300	215.0401	2
6	40.08411	0.679659	160.0009	67.01402	2
7	40.01407	0.415348	148.0005	44.04305	1
8	40.0918	0.161428	156.0014	36.08375	1
9	37.89653	2.088604	769.0007	330.0693	1
10					

***Fig 6.6**- Data screenshot of the randomly picked points for testing using the tree model, showing the expected class*

The same data was subsequently stripped of thewTDS codes so that the tree model was now used to help predict the same. See below:

Received: 2022-09-12
Accepted: 2022-12-10
Published Online: 2022-12-30

International Journal of Multidisciplinary Research and Explorer (IJMRE)
E-ISSN: 2833-7298, P-ISSN: 2833-7301

	A	B	C	D
1	longtd	lattd	elev	dist
2	39.74906	0.458806	164.002	74.07844
3	40.51258	0.190888	120.0016	21.05916
4	40.01988	1.619978	260.0019	180.0471
5	39.15989	1.811882	300	215.0401
6	40.08411	0.679659	160.0009	67.01402
7	40.01407	0.415348	148.0005	44.04305
8	40.0918	0.161428	156.0014	36.08375
9	37.89653	2.088604	769.0007	330.0693
10				

Fig 6.7: The selected Spots which are to be predicted

prediction of these new spots for TDS newPred=predict(model,dataT) newPred newPred=as.data.frame(newPred) newPred The foregoing coding was run in R and the output screenshot is as shown thus:

```
-1.6000 -0.1111 -0.0597  0.0000  0.0000  1.6000
>
> ## generate the Model split graphics
> plot(model,col="green")
> text(model, pretty = 0,col="orange")
> set.seed(101)
> train=sample(1:nrow(datax), 320)
> ## prediction of these new spots for TDS
> newPred=predict(model,dataT)
> newPred
       1        2        3        4        5        6        7        8
1.235294 2.000000 3.000000 2.000000 2.000000 1.059701 1.059701 1.111111
>
> newPred=as.data.frame(newPred)
> newPred
   newPred
1 1.235294
2 2.000000
3 3.000000
4 2.000000
5 2.000000
6 1.059701
7 1.059701
8 1.111111
>
```

Fig 6.8: The Predicted Values

The classes predicted are hundred percent true for the sample testing dataset taken from distal Merti aquifer. With this kind of performance, the Shabel dulla site data was now taken into the model for predictions. This would tell us more about water quality expected from the site before the actual drilling.

Received: 2022-09-12
Accepted: 2022-12-10
Published Online: 2022-12-30

International Journal of Multidisciplinary Research and Explorer (IJMRE)
E-ISSN: 2833-7298, P-ISSN: 2833-7301

Fig 6.9: Google Map Image of Shabel Dulla

	A	B	C	D
1	longtd	lattd	elev	dist
2	0.554242	39.285	258	75
3				
4				

Fig 6.10: *The csv Data Image of Shabel Dulla Area Located Approximately 75 Kilometers Away From Laghdera Flow Course nearHabaswein.*

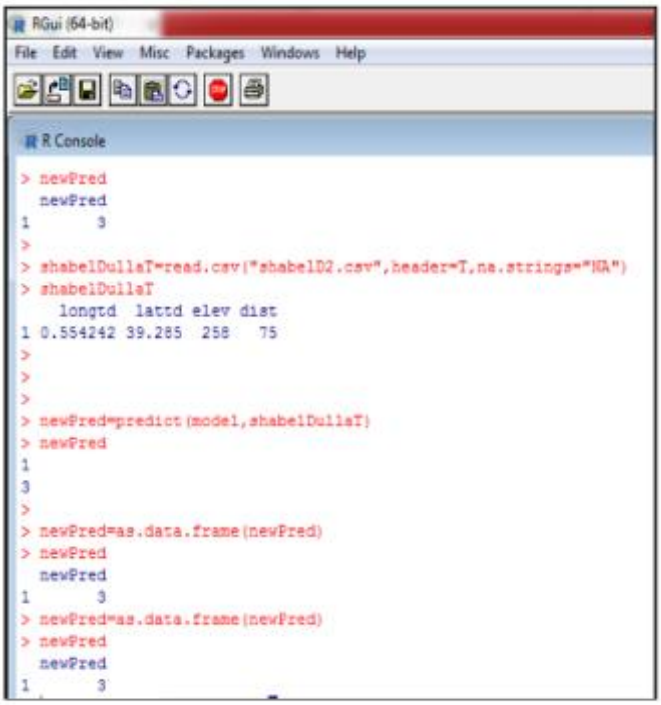


Fig 6.11: The predicted value

Received: 2022-09-12
Accepted: 2022-12-10
Published Online: 2022-12-30

International Journal of Multidisciplinary Research and Explorer (IJMRE)
E-ISSN: 2833-7298, P-ISSN: 2833-7301

The Class Predicted for Shabel dulla Data is 3, Meaning Saline Water is Expected here.

VII. Recommendations And Conclusions

From the foregoing synthesis of the Project Areas hydrology, geophysics, hydrogeology and stratigraphy, the sites investigated bear moderate groundwater potential. All the sites investigated possess reliable borehole potential. Other recommendations are as thus:

(i) A borehole may be drilled upto a depth of 285m bgl and be used for watering sanitation , livestock and domestic purposes, apart from cooking and drinking.

(ii) The few buildings that have corrugated iron roofs may be used to harvest the little rain water that comes every season, to have drinking water storage

(iii) Spring intake development is of no impact here as the few spring points are not perennial as desired. (iv) A series of sand dams may be done, within intervals of hundred and fifty or so meters , along the ephemeral Jilango river flow course

(v) The deep aquifer water is little and is saline, albeit ideal for livestock use and also for sanitation purposes.

(vi) Pans may be done in the area for impounding fresh rainfall flow run-offs for the purpose of harvesting the flow for drinking water. The depth of such a pan may be upto 6m bgl

(vii) Machine Learning may be used in prospecting for water in study area as it is a useful tool complementing the existing equipment and methods and tools in water resources mapping.

Acknowledgements

The publication of this paper immensely benefitted from facilitation, insight and technical input of the Ag CEO of the Northern Water Works Development Agency. Mr. Andrew Rage facilitated the study by proofreading the material and sharing insights that are of socioeconomic benefits to the project, in terms of resources availability, relative to the priority ranking schemes used. He is a Certified CPA and holds an MBA in Finance from Kenyatta University, Kenya.

Received: 2022-09-12
Accepted: 2022-12-10
Published Online: 2022-12-30

International Journal of Multidisciplinary Research and Explorer (IJMRE)
E-ISSN: 2833-7298, P-ISSN: 2833-7301

References

[1] Carretero, S. C., Capítulo, L. R., & Kruse, E. E. (2020). Decision tree as a tool for the management of coastal aquifers of limited saturated thickness. Quarterly Journal of Engineering Geology and Hydrogeology, 53(2), 189-200.

[2] Dauji, S., &Keesari, T. (2021). Decision tree for estimating groundwater contaminant through proxies considering seasonality and soil saturation. Environmental Monitoring and Assessment, 193(12), 1-23

. **[3]** Khader, A. I., Rosenberg, D. E., & McKee, M. (2013). A decision tree model to estimate the value of information provided by a groundwater quality monitoring network. Hydrology and Earth System Sciences, 17(5), 1797-1807.

[4] Krhoda, G. O., & Amimo, M. O. (2019). Groundwater quality prediction using logistic regression model for Garissa County, African Journal of Physical Sciences.

[5] Mirhashemi, S. H., Mirzaei, F., & Panahi, M. (2020). The study of environmental and human factors affecting aquifer depth changes using tree algorithm. International Journal of Environmental Science and Technology,B 17(3), 1825-1834.

[6] Stumpp, C., Żurek, A. J., Wachniew, P., Gargini, A., Gemitzi, A., Filippini, M., &Witczak, S. (2016). A decision tree tool supporting the assessment of groundwater vulnerability. Environmental Earth Sciences, 75(13), 1-7.

[7] Singha, S. S., Singha, S., Pasupuleti, S., & Venkatesh, A. S. (2022). Knowledge-driven and machine learning decision tree-based approach for assessment of geospatial variation of groundwater quality around coal mining regions, Korba district, Central India. Environmental Earth Sciences, 81(2), 1-13.

[8] Yoo, K., Shukla, S. K., Ahn, J. J., Oh, K., & Park, J. (2016). Decision tree-based data mining and rule induction for identifying hydrogeological parameters that influence groundwater pollution sensitivity. Journal of Cleaner Production, 122, 277-286.