Predictive Modelling for Water Borehole Drilling Using Machine Learning

Mr.P.Mariappan[1],Ms.K.Priyadharshini[2],Mr.R.Navaneetha Krishnan.[3],S.Santhosh[4]
[1]*Assistant Professor, Department of Computer Science and Engineering, Velalar College of Engineering and Technology, Erode, Tamil Nadu, India.*
[2,3,4]*Student, Department of Computer Science and Engineering, Velalar College of Engineering and Technology, Erode, Tamil Nadu, India.*

ABSTRACT– The proposed work integrates supervised machine learning models to identify the best model that predicts the parameters like water table depth, drilling duration, soil layer type, drilling cost, water quality and soil erosion rate with better accuracy. The proposed system involves the steps such as data collection and preprocessing, model training and output prediction, model evaluation, comparison of models and model deployment. The machine learning algorithms such as Random Forest, Gradient Boosting, Extreme Gradient Boosting were incorporated in order to predict the output parameters with better accuracy. To evaluate the proposed model we had performed comparative analysis using scatter plots. The primary objective is to predict the parameters that help the end user during water borehole drilling using a precise machine learning algorithm. Prediction of these parameters help the industries and end user in the management of water resources efficiently and ensuring a sustainable supply and reducing the risk of water-related issues. The proposed model can provide assistance to an environmental planning organization in selecting suitable locations for urban planning and infrastructure development. This project offers an approach to predict the borehole drilling parameters, paving the way for enhanced decision making. **KEYWORDS: object detection, YOLO, digital image processing**

I. INTRODUCTION

Clean water plays an essential role in achieving industrial and economic development, whether used for food production, drinking, or domestic use. Groundwater is the primary water source necessary for agriculture, irrigation, industrial activities, and drinking around the globe. Drilling machines are used to dig a bore well by drilling a borehole in the ground in search of water. The availability of groundwater is influenced by the water table depth, which varies significantly across various regions. Borehole placement on a drilling site with hard soil composition renders expensive machinery, skilled workforce, and time budget compared to a soft underground soil layer. To meet the global water demand a huge number of bore wells are being drilled, resulting in over-exploitation of scarce groundwater resources. In the past few decades, advanced technologies have been employed to speed up the drilling process. Factors like soil hardness, water table depth, and number of days spent on the drilling process in certain regions are essential to be considered before starting the drilling process. The predictive analytics help the drilling companies and hydro-geological resource managers in effective planning to estimate drilling cost and drilling resources in advance. To perform the analysis and extracting the information from the rich hydro geological data-sets, machine learning method are highly preferred due to their exceptional performance. Therefore, the prediction of the water table depth, soil layer type, drilling duration, drilling cost, water quality and soil erosion rate using advanced machine learning techniques paves a better way to carry out the bore hole drilling process.

II.RELATED WORK

1.WATER LEVEL PREDICTION USING LINEAR REGRESSION

Water level estimation is essential for efficient groundwater management and achieving sustainable development goals. The authors estimated the water level using temperature and monthly mean precipitation using Multiple Linear Regression (MLR). Experiments are conducted on seasonal variables including temperature, rainfall, evaporation and transpiration.

2. GROUND POTENTIAL CLASSIFICATION USING ENSEMBLE LEARNING MODEL

Predicting the groundwater potential is highly imperative for effective groundwater resource management. It proposed an ensemble learning model based on Logistic Regression. The proposed work considered sixteen groundwater such as slope, topographic wetness index, elevation, distance from river network and elevation etc as independent variables for modeling.

3.MACHINE LEARNING BASED WATER TABLE PREDICTION

Groundwater table prediction plays an essential role in the planning and management of ground resources. Here we compared the performance of RF and XGB in forecasting the water table depth for cranberry field farms in Canada. Experimental findings state that XGB achieved better results than RF for water table depth forecasting. The authors developed ANFIS based model to forecast the groundwater table.

4.MACHINE LEARNING FOR DRILLING RATE OF PENETRATION PREDICTION.

The drilling rate of penetration prediction is adapted to optimize drilling performance. In a study the authors used ANN, SVM, and Hybrid Multi-Layer Perceptron for drilling rate of penetration prediction. A hybrid ANN with a Simulated Annealing (SA), Invasive Weed Optimization Algorithm, Firefly Algorithm (FA), Shuffled Frog Leaping Algorithm, and Standard Back-propagation to learn the weights for drilling rate index estimation is proposed. The experimental results demonstrated that ANN with SA achieved noteworthy performance.

III.PROPOSED SYSTEM

The proposed work uses supervised machine learning algorithm such as Random Forest, Gradient Boosting, Extreme Gradient Boosting (XGB) to predict the borehole related parameters. A key innovation in this project is the development of a user-friendly web application. By comparing these supervised machine learning algorithms, the project aims to develop a robust and accurate model for predicting the water table depth, soil layer type, drilling duration, drilling cost, water quality and soil erosion rate.

1.DATA PREPROCESSING

1.1. DATA COLLECTION:

Data collection in machine learning is the process of gathering relevant information or samples that will be used to train, validate, or test a machine learning model. This step may involve downloading datasets, extracting information from documents and storing it in excel sheet.

Latitude	Longitude	Elevation	Precipitation	Temperature	SoilMoisture	SoilPermeability	pH	Nitrate	Chloride	duration	soil_layer	water_table_depth	drilling_cost	water_quality	erosion_rate
37.1	-120.9	227.1	31.03	10.1	21	0.3	6.4	3.5	16.3	60	1	300	300	1	0.148
38.2	-119.34	229.4	31.96	10.15	21.76	0.35	6.9	3.53	16.322	61.8	1	303	303	1	0.153
39.3	-117.78	231.7	32.89	10.2	22.52	0.4	6.902	3.54	16.542	63.6	1	306	306	1	0.158
40.4	-116.22	234	33.82	10.25	23.28	0.45	6.904	3.55	16.762	65.4	1	309	309	1	0.163
41.5	-114.66	236.3	34.75	10.3	24.04	0.5	6.906	3.56	16.982	67.2	1	312	312	1	0.168
42.6	-113.1	238.6	35.68	10.35	24.8	0.55	6.908	3.57	17.202	69	1	315	315	1	0.173
43.7	-111.54	240.9	36.61	10.4	25.56	0.6	6.91	3.58	17.422	70.8	1	318	318	1	0.178
44.8	-109.98	243.2	37.54	10.45	26.32	0.65	6.912	3.59	17.642	72.6	1	321	321	1	0.183
45.9	-108.42	245.5	38.47	10.5	27.08	0.7	6.914	3.6	17.862	74.4	1	324	324	1	0.188
47	-106.86	247.8	39.4	10.55	27.84	0.75	6.916	3.61	18.082	76.2	1	327	327	1	0.193
48.1	-105.3	250.1	40.33	10.6	28.6	0.8	6.918	3.62	18.302	78	1	330	330	1	0.198
49.2	-103.74	252.4	41.26	10.65	29.36	0.85	6.92	3.63	18.522	79.8	1	333	333	1	0.203
50.3	-102.18	254.7	42.19	10.7	30.12	0.9	6.922	3.64	18.742	81.6	1	336	336	1	0.208
51.4	-100.62	257	43.12	10.75	30.88	0.95	6.924	3.65	18.962	83.4	1	339	339	1	0.213
52.5	-99.06	259.3	44.05	10.8	31.64	1	6.926	3.66	19.182	85.2	1	342	342	1	0.218
53.6	-97.5	261.6	44.98	10.85	32.4	1.05	6.928	3.67	19.402	87	1	345	345	1	0.223
54.7	-95.94	263.9	45.91	10.9	33.16	1.1	6.93	3.68	19.622	88.8	1	348	348	1	0.228
55.8	-94.38	266.2	46.84	10.95	33.92	1.15	6.932	3.69	19.842	90.6	1	351	351	1	0.233
56.9	-92.82	268.5	47.77	11	34.68	1.2	6.934	3.7	20.062	92.4	1	354	354	1	0.238
58	-91.26	270.8	48.7	11.05	35.44	1.25	6.936	3.71	20.282	94.2	1	357	357	1	0.243
59.1	-89.7	273.1	49.63	11.1	36.2	1.3	6.938	3.72	20.502	96	1	360	360	1	0.248
60.2	-88.14	275.4	50.56	11.15	36.96	1.35	6.94	3.73	20.722	97.8	1	363	363	1	0.253
61.3	-86.58	277.7	51.49	11.2	37.72	1.4	6.942	3.74	20.942	99.6	1	366	366	1	0.258
62.4	-85.02	280	52.42	11.25	38.48	1.45	6.944	3.75	21.162	101.4	1	369	369	1	0.263
63.5	-83.46	282.3	53.35	11.3	39.24	1.5	6.946	3.76	21.382	103.2	1	372	372	1	0.268

dataset

Figure 1:Borehole Dataset

1.2 DATA PREPROCESSING

The collected data is typically divided into two subsets: training dataset and test dataset. The training set is used to train the model and the test set is used to evaluate the model's performance on unseen data. Along with that a detailed description of the input and output parameters are provided for better understanding of the end user. Users are able to join data files together and use preprocessing to filter any unnecessary noise from the data which can allow for higher accuracy. Users use Python programming scripts accompanied by the pandas library which gives them the ability to import data from a comma-separated values as a data-frame. Pandas (software) which is a powerful tool that allows for data analysis and manipulation; which makes data visualizations, statistical operations and much more, a lot easier.

Input Parameters	Description
Latitude	Geographical coordinate specifying the north-south position in degrees. Units: Degrees
Longitude	Geographical coordinate specifying the east-west position in degrees. Units: Degrees
Elevation	Height above sea level. Units: Meters
Precipitation	Amount of rainfall. Units: Millimeters
Temperature	Ambient air temperature. Units: Celsius
SoilMoisture	Moisture content in the soil. Units: Percentage
SoilPermeability	Ability of soil to transmit water. Units: Cubic meters per second
pH	Measure of acidity or alkalinity. Units: Dimensionless
Nitrate	Concentration of nitrate in water. Units: Milligrams per liter
Chloride	Concentration of chloride in water. Units: Milligrams per liter
Duration	Time taken for drilling. Units: Hours
SoilLayer	Type of soil layer. Units: Categorical (1-Silt, 2-Sand, 3-Clay)
WaterTableDepth	Depth of the water table below the ground surface. Units: Meters
DrillingCost	Cost associated with drilling. Units: Currency (e.g., dollars)
water_quality	water_quality. Units:liter(1-Good, 2-Avg, 3-Poor)
ErosionRate	Rate of soil erosion. Units: Cubic meters per year

Figure 2. Input Data Description

Output Parameters	Description
Duration	Time taken for drilling. Units: Hours
SoilLayer	Type of soil layer. Units: Categorical (1-Silt, 2-Sand, 3-Clay)
WaterTableDepth	Depth of the water table below the ground surface. Units: Meters
DrillingCost	Cost associated with drilling. Units: Currency (Per Feet)
water_quality	water_quality. Units:liter(1-Good, 2-Avg, 3-Poor)
ErosionRate	Rate of soil erosion. Units: Cubic meters per year

Figure 3. Output Data Description

2.MODEL TRAINING AND OUTPUT PREDICTION

Model training is the process of teaching a machine learning model to recognize patterns and make predictions from the data it's provided. Here three machine learning models are trained with the training dataset which was in csv format. The supervised machine learning algorithms such as Random Forest algorithm, Gradient Boosting algorithm, Extreme Gradient Boosting algorithm are used to train the model. During training, input data is passed through the model, producing predictions. These predictions are compared to the actual target values using a scatter plot graph, which quantifies the model's performance. Along with that the MSE, SSI, SNR values are plotted through bar graph for model comparison.

2.1 RANDOM FOREST REGRESSOR

Random Forest is a popular ensemble learning

algorithm used in both classification and regression tasks. It operates by constructing a multitude of decision trees during training and outputs the prediction of the individual trees. It is known for its robustness, scalability, and ability to handle high-dimensional data with a large number of features. It's less prone to overfitting compared to individual decision trees.

2.2 GRADIENT BOOSTING

Gradient Boosting is a powerful boosting algorithm that combines several weak learners into strong learners, in which each new model is trained to minimize the loss function such as mean squared error or cross-entropy of the previous model using gradient descent. In each iteration, the algorithm computes the gradient of the loss function with respect to the predictions of the current ensemble and then trains a new weak model to minimize this gradient. The predictions of the new model are then added to the ensemble, and the process is repeated until a stopping criterion is met.

2.3 EXTREME GRADIENT BOOSTING

XGBoost, or Extreme Gradient Boosting, is a state-of-the-art machine learning algorithm renowned for its exceptional predictive performance. XGBoost builds a strong predictive model by aggregating the predictions of several weak learners, usually decision trees. By adding a regularization term and utilizing a more advanced optimization algorithm, XGBoost goes one step further and improves accuracy and efficiency.

3.MODEL EVALUATION

Model evaluation is a critical step in the machine learning workflow to assess how well a trained model generalizes to unseen data. It involves measuring the performance of the model using various metrics and techniques. In these regression tasks, metrics like mean squared error (MSE), Signal-to-Interference Ratio(SSI) and Signal-to-Noise Ratio(SNR) are commonly used for evaluation of the models. It helps to make informed decisions about model selection, optimization, and deployment, leading to more effective and reliable machine learning solutions.

4.MODEL COMPARISON

After training and evaluating the three machine learning models, MSE of each model is plotted in bar graph using the Matplotlib python library. On Comparing we can able to identify that the random forest regressor predicts the output with less error and better accuracy. In conclusion random forest regression algorithm is the most suitable model for prediction.

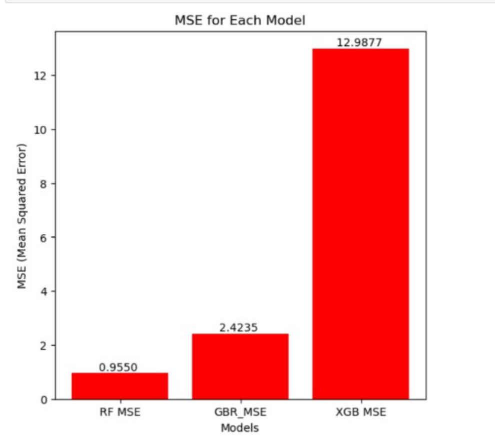

Figure 4. Model comparison.

5.MODEL DEPLOYMENT

In model deployment a local host website was developed as a user interface that uses the random forest model for output predictions.

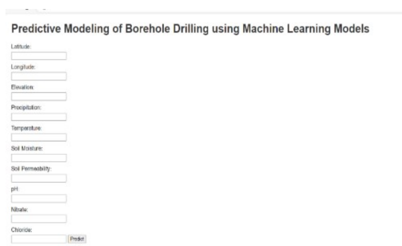

Figure5.Website Application

IV.CONCLUSION AND FUTURE WORK

In conclusion the project Predictive Modeling for Water Borehole Drilling using Machine Learning combined with a website application helps to identify the best suitable surface for water borehole drilling. The website acts as an interface for the end user to provide the input values and get the output predictions on time. In future using advanced machine learning models the output parameters can be predicted with better accuracy and robustness.

REFERENCES

[1] J. D. Mackay, C. R. Jackson, A. Brookshaw, A. A. Scaife, J. Cook and R. S. Ward, "Seasonal forecasting of groundwater levels in principal aquifers of the united kingdom", J. Hydrol., vol. 530, pp. 815-828, Nov. 2015.

[2] C.Nagarajan and M.Madheswaran - ‘Experimental verification and stability state space analysis of CLL-T Series Parallel Resonan Converter’ - Journal of ELECTRICAL ENGINEERING, Vol.63 (6), pp.365-372, Dec.2012.

[3] C.Nagarajan and M.Madheswaran - ‘Performance Analysis of LCL-T Resonant Converter with Fuzzy/PID Using State Space Analysis’- Springer, Electrical Engineering, Vol.93 (3), pp.167-178, September 2011.

[4] C.Nagarajan and M.Madheswaran - ‘Stability Analysis of Series Parallel Resonant Converter with Fuzzy Logic Controller Using State Space Techniques’- Taylor &Francis, Electric Power Components and Systems, Vol.39 (8), pp.780-793, May 2011.

[5] C.Nagarajan and M.Madheswaran - ‘Experimental Study and steady state stability analysis of CLL-T Series Parallel Resonant Converter with Fuzzy controller using State Space Analysis’- Iranian Journal of Electrical &Electronic Engineering, Vol.8 (3), pp.259-267, September 2012.

[6] Nagarajan C., Neelakrishnan G., Akila P., Fathima U., Sneha S. “Performance Analysis and Implementation of 89C51 Controller Based Solar Tracking System with Boost Converter” Journal of VLSI Design Tools & Technology. 2022; 12(2): 34–41p.

[7] C. Nagarajan, G.Neelakrishnan, R. Janani, S.Maithili, G. Ramya “Investigation on Fault Analysis for Power Transformers Using Adaptive Differential Relay” Asian Journal of Electrical Science, Vol.11 No.1, pp: 1-8, 2022.

[8] G.Neelakrishnan, K.Anandhakumar, A.Prathap, S.Prakash “Performance Estimation of cascaded h-bridge MLI for HEV using

[9] SVPWM” Suraj Punj Journal for Multidisciplinary Research, 2021, Volume 11, Issue 4, pp:750-756

[10] G.Neelakrishnan, S.N.Pruthika, P.T.Shalini, S.Soniya, “Perfromance Investigation of T-Source Inverter fed with Solar Cell” Suraj Punj Journal for Multidisciplinary Research, 2021, Volume 11, Issue 4, pp:744-749
[11] C.Nagarajan and M.Madheswaran, “Analysis and Simulation of LCL Series Resonant Full Bridge Converter Using PWM Technique
[12] with Load Independent Operation” has been presented in ICTES’08, a IEEE / IET International Conference organized by M.G.R.University, Chennai.Vol.no.1, pp.190-195, Dec.2007
[13] M Suganthi, N Ramesh, “Treatment of water using natural zeolite as membrane filter”, Journal of Environmental Protection and Ecology, Volume 23, Issue 2, pp: 520-530,2022
[14] M Suganthi, N Ramesh, CT Sivakumar, K Vidhya, “Physiochemical Analysis of Ground Water used for Domestic needs in the Areabof Perundurai in Erode District”, International Research Journal of Multidisciplinary Technovation, pp: 630-635, 2019
[15] O. Rahmati, M. Avand, P. Yariyan, J. P. Tiefenbacher, A. Azareh and D. T. Bui, "Assessment of Gini- entropy- and ratio-based classification trees for groundwater potential modelling and prediction", Geocarto Int., vol. 37, no. 12, pp. 3397-3415, Jun. 2022.
[16] A. A. Nadiri, K. Naderi, R. Khatibi and M. Gharekhani, "Modelling groundwater level variations by learning from multiple models using fuzzy logic", Hydrological Sci. J., vol. 64, no. 2, pp. 210-226, Jan. 2019.
[17] Y. Ben, C. James and D. Cao, "Development and application of a real-time drilling state classification algorithm with machine learning", Proc. 7th Unconventional Resour. Technol. Conf., pp. 1-12, 2019.
[18] Y. Z. Kaya, F. Üneş, M. Demirci, B. Taşar and H. Varçin, "Groundwater level prediction using artificial neural network and M5 tree models", Aerul şi Apa Componente ale Mediului, pp. 195-201, 2018.
[19] A. Mosavi, F. S. Hosseini, B. Choubin, M. Goodarzi, A. A. Dineva and E. R. Sardooi, "Ensemble boosting and bagging based machine learning models for groundwater potential prediction", Water Resour. Manage., vol. 35, no. 1, pp. 23-37, Jan. 2021.
[20] S. Pradhan, S. Kumar, Y. Kumar and H. C. Sharma, "Assessment of groundwater utilization status and prediction of water table depth using different heuristic models in an Indian interbasin", Soft Comput., vol. 23, no. 20, pp. 10261-10285, Oct. 2019.
[21] M. A. Mojid, M. F. Parvez, M. Mainuddin and G. Hodgson, "Water table trend—A sustainability status of groundwater development in north-west Bangladesh", Water, vol. 11, no. 6, pp. 1182, 2019.
[22] M. Sabah, M. Talebkeikhah, D. A. Wood, R. Khosravanian, M. Anemangely and A. Younesi, "A machine learning approach to predict drilling rate using petrophysical and mud logging data", Earth Sci. Informat., vol. 12, no. 3, pp. 319-339, Sep. 2019.
[23] A. I. A. Osman, A. N. Ahmed, M. F. Chow, Y. F. Huang and A. El-Shafie, "Extreme gradient boosting (Xgboost) model to predict the groundwater levels in Selangor Malaysia", Ain Shams Eng. J., vol. 12, no. 2, pp. 1545-1556, Jun. 2021.
[24] M. H. Lo, J. S. Famiglietti, J. T. Reager, M. Rodell, S. Swenson and W. Y. Wu, "Grace-based estimates of global groundwater depletion", Terrestrial Water Cycle and Climate Change: Natural and Human-Induced Impacts, pp. 135-146, 2016.
[25] K. Khosravi, M. Sartaj, F. T.-C. Tsai, V. P. Singh, N. Kazakis, A. M. Melesse, et al., "A comparison study of DRASTIC methods with various objective methods for groundwater vulnerability assessment", Sci. Total Environ., vol. 642, pp. 1032-1049, Nov. 2018.
[26] A. M. Alsalama, J. P. Canlas and S. H. Gharbi, "An integrated system for drilling real time data analytics", Proc. SPE Intell. Energy Int. Conf. Exhib., pp. 1-11, Sep. 2016.
[27] J. Zhang, Y. Zhu, X. Zhang, M. Ye and J. Yang, "Developing a long short-term memory (LSTM) based model for predicting water table depth in agricultural areas", J. Hydrol., vol. 561, pp. 918-929, Jun. 2018.
[28] F. Sajedi-Hosseini, A. Malekian, B. Choubin, O. Rahmati, S. Cipullo, F. Coulon, et al., "A novel machine learning-based approach for the risk assessment of nitrate groundwater contamination", Sci. Total Environ., vol. 644, pp. 954-962, Dec. 2018.