May-June 2020
ISSN: 0193-4120 Page No. 6750 - 6755

Estimation of Ground Water Levels by Using Linear Regression and Logistic Regression Algorithms

Samineni Sai Asish, B.Tech at GITAM University, Bengaluru, India
Mr. Bhavin Kumar S, Assistant Professor in CSE Department, GITAM School of Technology, Bengaluru
Shaik Jeelan Basha, B.Tech at GITAM University, Bengaluru, India
Shashank Kaundinya S, B.Tech at GITAM University, Bengaluru, India
Pujari Venkata Kishore, B.Tech at GITAM University, Bengaluru, India

Article Info
Volume 83
Page Number: 6750 - 6755
Publication Issue:
May-June 2020

Abstract:
An indispensable resource that is given naturally to every living being on this planet is water. The Blue Planet contains 71 percentage of water. Water is available in many forms which are not suitable for consumption purpose. The 98% of groundwater that is available on earth is fresh and relevant for drinking. To estimate the amount of water present inside the earth at any particular zones which will help us to manage usage of water for the future generations. In this paper, we are implementing a machine learning model which has the potential to estimate the water levels inside the ground. This machine learning model is trained by using Logistic Regression Algorithm, and it is tested by using a linear regression algorithm which gives more accurate results than any other machine learning algorithms. The model is trained by taking various conditions like the recharge from rainfall during monsoon season and non-monsoon season, other sources to recharge during monsoon season and non-monsoon season, annual replenishable resource, net annual ground water availability, draft due to irrigation needs, domestic and industrial water supply needs, total annual draft, projected demand for domestic and industrial uses up to 2025, ground water availability for future irrigation use, stage of ground water development in real time to predicts the water availability in particular area..

Article History
Article Received: *19 November 2019*
Revised: *27 January 2020*
Accepted: *24 February 2020*
Publication: *18 May 2020*

Keywords: *Machine Learning, Decision Tree, Linear Regression, Ground Water levels*

I. INTRODUCTION

Surface water is always one of the biggest sources for supply of water for national, metropolitan, farming, and industrial uses in many zones, particularly in arid and semi-arid regions. In addition, several problems arise as a result of over-exploitation of underground water and inefficient use and procurement of underground water, such as enormous decreases in water levels, well drying, water quality deterioration, increased pumping charges, land sedimentation, loss of pumping in residential water supply wells, and underground water evaporation.

More such challenges in developing nations, are really a serious issue nationally. Sustainable management of aquifers resources along with surface water has now become the immediate emergency need for another hour in order to preserve water for the upcoming years. Reliable and efficient underground water prediction is a key element for solving this issue, specifically in lakes and rivers in arid and semi- regions which are far

May-June 2020
ISSN: 0193-4120 Page No. 6750 - 6755

more vulnerable to excessive hydrological conditions in the form of drought conditions.

A statistic model which predicts and do the binary classification of type yes or no can be done by using a logistic regression. This regression technique can also be used for multiclass classification. Here in this paper the logistic regression is used to train the machine learning model. Another statistical model which also predicts binary classes in which the target output will be binary format is linear regression method which is used for testing the model that is trained by logistic regression method.

The supply and efficiency of urban drinking water services, the vast number of urban drinking water services have strained proposed water supplies and remote communities are left out. However, exponential development in India's metropolitan areas has exhausted policy strategies, undermined by over-privatization. Because of drinking water improvements, most other water bodies are also polluted with bio and chemical toxins, and water-related illnesses make up more than 21 percent of the population. In addition, access to basic sanitation is only 33 per cent for the world.

Twenty-one Indian cities will run out of groundwater by next year, including the capital New Delhi and Bengaluru's Information Technology Centre, according to a report by the central government think tank NITI Aayog. Two hundred thousand Indians die each year because they do not have a healthy source of water, the study said. A staggering 600 million people are suffering water tension "moderate to severe".

II. LITERATURE SURVEY

The prediction of fluctuations in the underground water levels is necessary to manage the water resources for future generations. One such system which predicts the underground water levels by using temperature and precipitation averages is Ground Water Level Prediction Using Artificial Neural Network and M5 Tree Models system. The model is investigated for studying several details by ANN and M5T approach in a particular region in turkey by considering the data of 15 years (i.e., 2000-2015) taken from an observation station which is in Asi-basin a sub-basin region. [1]

To figure out the underground water levels from the geoelectric parameters a nonlinear model is explored by using an Artificial Neural Network (ANN). From the study area all accessible wells water levels data is taken which act as a major output parameter in the ANN model. The parametric Vertical Electrical Soundings (VES) stations of every well will be occupied by using a Schlumberger array configuration which has electrode spacing (AB\2) and has ranging from 1 to 100m. The generated data from VES will generate the parameters which is geoelectric that controls storage and flow of underground water levels in that region by continuous interpretation and this can be used for training the ANN model. Then this nonlinearity approach is applied to predict the underground water levels from the data which had given, seasonal variations will affect the fluctuations of water in wells as a result of the trained model. The Regression Coefficient and Mean Square Error (MSE) is estimated to evaluate the model efficiency. [2]

The biggest challenge for managers and engineers who manages the water resource is to forecast the underground water levels and stream flow of the water. For this, 19 years (i.e., 1996-2015) of rainstorm and underground water levels data is collected and trained. Availability of water resources is predicted by using Artificial Neural Network (ANN) which does regression analysis gives correlation within the range of 0.12-0.97 in Abhanpur block. [3]

The physical parameters which are correlated strongly by the machine learning model will be used as input for the model which can predict underground water level estimation successfully by using a data-driven approach. Fluctuations in the underground water levels are highlighted and confirmed by using various metrics present in this model. [4]

A Neural Network based back propagation Artificial Bee Colony (ABC) optimization algorithm used for predicting underground water levels accurately in overexploited areas which are arid regions. The input for training the model are Rainfall, Recharge, Exploitation, and evaporation. The output of the model which used supervised machine learning technique is ground water level. The stabilization, fitting accuracy, and convergence rate of the results shows the model gives better results than any other BP algorithms. [5]

A Machine Learning based regression model is developed for estimating the ground water recharge in humid and sub-humid regions. The ground water is recharged from precipitation, drainage basins, growing degree days and average basin specific yield. There exists a correlation between recharge, precipitation specific yield which is very much stronger. [6] A model which predicts underground water levels at unconfined shallow aquifer by considering several factors as input for the model to train. The factors that are considered are lake stage, local weather, moving averages of the weather, stream flow data, stage and stream flow data over a period of time. A neural network-based feed-forward time-delay will be used for predicting the underground water levels. [7]

A trend function based residual kriging method which estimates water levels inside the earth surface in the unconfined alluvial aquifer. The trend model auxiliary variable will be considered from the elevation of surface and ground water levels. The vulnerable potential locations can be detected by using kriging method and the spatial correlation is calculated by using classical variogram function. [8] the transient in the ground water levels and groundwater basin is compared and predicted by using Artificial Neural Network (ANN) and Multilayer Linear regression (MLR). The inputs for modelling are: river stage, ambient temperature, rainfall, influential lags of rainfall, river stage, ground water levels, dummy variables. The graphical indicators and residual analysis show that ANN will be the best method for calculating spatio-temporal distribution of water in the underground. [9]

III. PROPOSED METHOD ARCHITECTURE

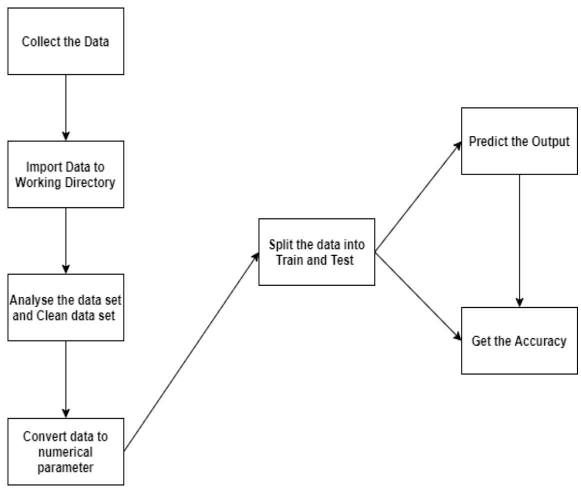


Fig 1: Architecture Diagram of Proposed System

IV. WORKING

The data related to rainfall during monsoon season and non-monsoon season, other sources to recharge during monsoon season and non-monsoon season, annual replenishable resources, annual natural discharge, net annual ground water availability, draft due to irrigation needs, draft due to domestic and industrial water supply needs, total annual draft, projected demand for industrial and domestic usage up to 2025, Ground water availability for future irrigation use, stage of ground water development is collected and import into working directory.

These data are given to machine learning model to train the model. This is trained by using logistic regression algorithm. How the logistic regression algorithm works will be explained in detail. Classification of data can be done by using a hypothesis called linear regression hypothesis.

May-June 2020
ISSN: 0193-4120 Page No. 6750 - 6755

$$h_\theta = \theta^T x$$

The generalization of these linear regression hypothesis gives us logistic regression hypothesis which uses logistic function.

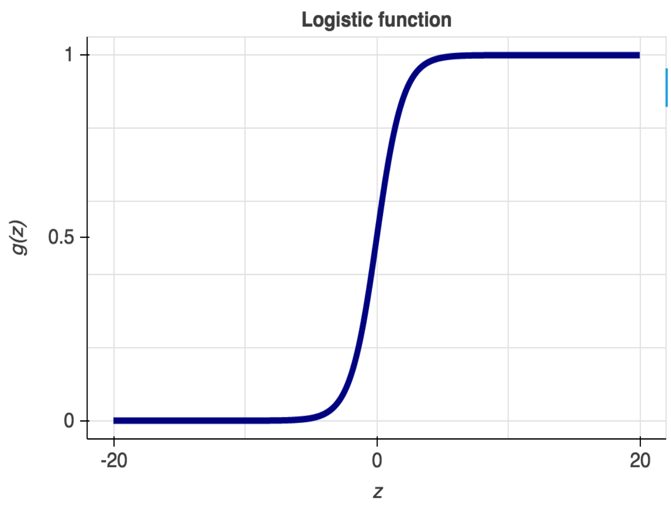
$$h_\theta(x) = g(\theta^T x)$$

$$g(x) = \frac{1}{1+e^{-2}}$$

The analysis of the logistic regression hypothesis is

$$h_\theta(x) = \frac{1}{1+e^{-\theta^T x}}$$

The logistic function which is also called as sigmoid function is given by **g(z)** which has the asymptotes at 0 and 1 that crosses the axis at 0.5.



To train the model we will consider only few parameters or features for which the logistic regression hypothesis is given by

$$h_\theta(x) = g(\theta_0 + \theta_1 z_1 + \theta_2 z_2)$$

The prediction using logistic regression can be done only if:

$$\theta_0 + \theta_1 x_1 + \theta_2 x_2 \geq 0$$

Because, threshold value for the logistic regression is set at **g(z)**=0.5. for the dataset we are using the values for θ will be

$$\theta = \begin{pmatrix} 0.69254 \\ -0.49269 \\ 0.19834 \end{pmatrix}$$

$$\theta^T x$$

$$y = \beta 0 + \beta 1X1 + \beta 2X2 + \ldots\ldots + \beta nXn$$

$$Error = \sum_{i=1}^{n} (Predicted\_output - Average\_of\_actual\_output)\wedge 2$$

$$\theta = \begin{pmatrix} 0.69254 \\ -0.49269 \\ 0.19834 \end{pmatrix}$$

A manual prediction can be done by using these coefficients for computing the vector product.

$$\theta^T x$$

The final linear regression equation used for computing is

$$y = \beta 0 + \beta 1X1 + \beta 2X2 + \ldots\ldots + \beta nXn$$

The output data is visualized as

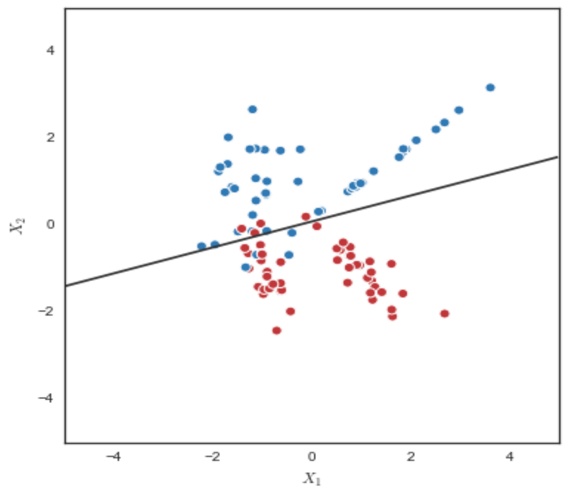


Fig 2: Decision Boundary for Visualization of output

The trained model is tested by using linear regression algorithm which predicts output with minimum error.

Y (pred)=b0+b1*x

To minimize the error rate b0 and b1 should be chosen carefully. Sum of squared error is considered

May-June 2020
ISSN: 0193-4120 Page No. 6750 - 6755

for evaluating the model which reduces the error in better way.

$$Error = \sum_{i=1}^{n}(actual\_output - predicted\_output)**2$$

The intercept and coefficient for prediction is given by using the formulae:

$$b_0 = \bar{y} - b_1\bar{x}$$

$$b_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

The error in calculating the regression is calculated by using the formulae:

$$\text{Error} = \sum_{i=1}^{n}(\text{Predicted\_output} - \text{average\_of\_actual\_output})\text{^}2$$

The total error formula is given by:

$$Error = \sum_{i=1}(Predicted\_output - Average\_of\_actual\_output)\text{^}2$$

R^2 = 1 – (SSE/SSTO)

The correlation coefficient is calculated by using the formula:

r = (+/-) sqrt(r²)

The final regression line for the data that is tested is given in visual manner:

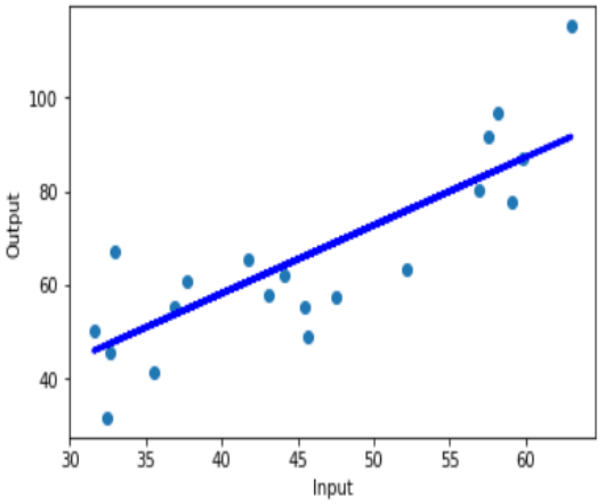


Fig 3: Final regression for the tested data

These data sets are analysed, cleaned and converted into numerical parameters. With this dataset 70 percentage of data is used for training the model, 30 percentage of data is used for testing purpose and predict the output and get accuracy.

V. RESULTS

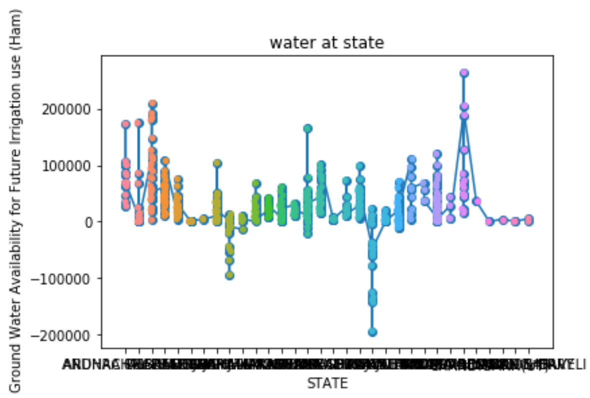


Fig 4: Water Levels Available as Per State Statistics

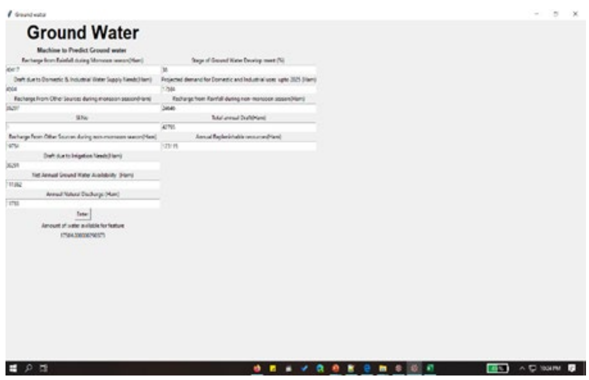


Fig 5: Application Developed for Predicting Water Levels

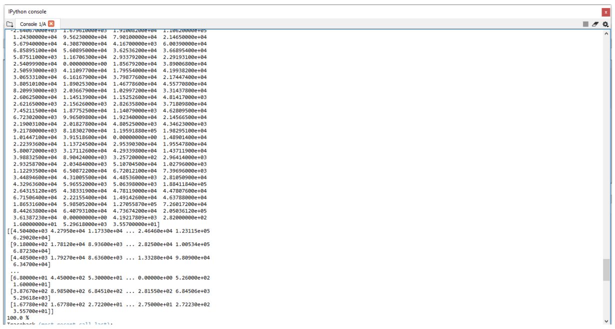


Fig 6: The Accuracy Obtained by Testing the Data

VI. CONCLUSION

This machine learning model predicts the underground water levels which is the major need for every living being on this earth. The algorithms used to train and test the model will give more accurate result than any other machine learning algorithm. In future, this model also used for agriculture and household needs just by training the model by using the dataset.

REFERENCES

1. Y. Z. Kaya, F. Üneş, M. Demirci, B. Taşar, H. Varçin, "Groundwater Level Prediction Using Artificial Neural Network and M5 Tree Models".
2. K. A. N. Adiat, O. F. Ajayi, A. A. Akinlalu, I. B. Tijani, "Prediction of Groundwater Level In Basement Complex Terrain Using Artificial Neural Network: A Case Of Ijebu Jesa, Southwestern Nigeria", Applied Water Science, 30 October 2019.
3. Pallavi Porte, Rajendra Kumar Isaac, Kipoo Kiran Singh Mahilang, Khilendra Sonboier, Pankaj Minj, "Groundwater Level Prediction Using Artificial Neural Network Model", IJCMAS, Vol 07, No 2, 2018.
4. Xuanhui Wang, Tailian Liu, Xilai Zheng, Hui Peng, Jia Xin, Bo Zhang, "Short term prediction of groundwater level using improved random forest regression with a combination of random features", AWS, 24 july 2018.
5. Huanhuan Li, Yudong Lu, Ce Zheng, Mi Yang, Shuangli Li, "Groundwater Level Prediction for the Arid Oasis of Northwest China Based on the Artificial Bee Colony Algorithm and a Back-propagation Neural Network with Double Hidden Layers", MDPI, 24 April 2019.
6. Y David L. Lorenz, Geoffrey N. Delin, "A Regression Model to Estimate Regional Ground Water Recharge", National Ground Water Association, 2006.
7. Sara Maatta, "Predicting groundwater levels using linear regression and neural networks", December 15, 2011.
8. Panagiota Theodoridou, Emmanouil Varouchakis, George Karatzas, "Regression Analysis and Risk Assessment of Groundwater Levels", MDPI, 24 August 2018.
9. Sasmita Sahoo, Madan K. Jha, "Groundwater-level prediction using multiple linear regression and artificial neural network techniques: a comparative assessment", Hydrogeology Journal, October 2013.