Groundwater Level Prediction Using Modified Linear Regression

Madhuri Kommineni[1], K Veniha Reddy[2], K Jagathi[3], B Dushyanth Reddy[4], Roshini A[5], V Bhavani[6]

[1, 2, 3, 4, 5, 6] *Department of Computer Science and Engineering,*

[1,2,3,4,5,6] *Koneru Lakshmaiah Education Foundation, Vaddeswaram, AP, India,*

[1]*madhuri.cbit@gmail.com*

***Abstract*-- The non-stop decline of the groundwater stage is one of the crucial factors that affect the development of the countrywide economy and society. Based on the linear regression with PLS regression, the predicting model of the groundwater level is presented. The precision of the version is checked using the tracking facts within the Texas vicinity. Using this approach, a clean equation is generated and on evaluating one of a kind algorithms we selected the first-rate configuration and trained the version of the use of the Texas dataset. The case study shows that the precision of the version is instead excessive and its popularization importance is higher than the alternative models and has some practical fee when being used in the dynamic groundwater level analysis. Groundwater is one of the most significant characteristic asset so as to full fill the water necessities from irrigation, domestic, industrial and research needs. Gauge of the ground water level isn't only the pre-basic for long haul forecast of slant strength in repository bank, yet what's more the key for ensuring the sheltered activity of the supply. In this paper, linear regression with PLS regression is a compelling method for forecast of groundwater level in Texas.**

Keywords— Groundwater level, Linear regression, Water, Prediction, PLS Regression.

I. INTRODUCTION

Groundwater is observed below the floor and is to be had through hand pumps and deep wells and so on. The water is observed inside the floor due to rain. In reality, while rainfall happens a few water percolates down through the pores of the soil. At a certain intensity, the pore space in the soil is full of water and the level at which this takes place is called the water level. Groundwater is the essential water source for financial and social improvement, however, the exhaustive exploitation of groundwater contributed to numerous environ geological problems, which include floor subsidence, floor crack and so forth, which affected the improvement of human society significantly. Therefore, the studies on groundwater stage tracking is vital to sensible application. Groundwater is one of the foundations that everyone lives rely on. Groundwater trend is associated with environment precipitation. There is plain seasonality in version of atmosphere precipitation, so the groundwater stage is changed by way of the seasonality and periodicity. In this case, whilst the groundwater level is anticipated, the unique data of groundwater need to be handled in trendy. Several researches had been executed on the prediction of groundwater degree. A decrease in-floor water degree is reduced to some factors including rainfall, evaporation, exploitation. Groundwater dynamic version typically implies the changes of the groundwater measure, standard, and real homes with schedule. It's an antiquated way which shows the groundwater development course, and the groundwater level is the most extreme significant coefficient of groundwater level change process. Groundwater is a fantastically treasured useful resource. Measurement and evaluation of groundwater level is wanted for keeping groundwater availability. The exact forecast of groundwater measure is indispensable for reasonable usage and control of essential groundwater assets. The utilization of water builds each day with the development in the populace. The groundwater measure is going as the day progressed. In India, groundwater serves approximately 80% of the rural populace, 50% of the urban population and about 60% of the agricultural area. For control of groundwater level, a version is needed which could expect the groundwater level in the future with the contemporary to be had records. The level of groundwater is the amount of water found underneath the outside of Earth in soil pore spaces and shake developments break. Essential factors affecting the growth of the national economy and society is the ongoing decrease in groundwater rates.

The prediction of variations in groundwater measure is very significant in the management of water resources. Groundwater level has unpredictable characteristics due to natural and anthropogenic factors influences. We use the groundwater level monitoring data, to calculate in light of the reaction connection between compelling elements, for example, precipitation and populace and the difference in groundwater level, the persuasive elements of groundwater level were chosen. At that point the linear regression was developed and used to anticipate the groundwater level. The modified linear regression form isn't generally the best higher turning out to be and speculation ability, yet additionally hearty advantages in the assessment of groundwater dynamic qualities and the screening of vital factors. It is a powerful technique for the forecast of groundwater level. we selected the best configuration and trained our model using 3 Datasets named as Texas Dataset wells [226 areas], Population_18[226 areas] and test our model with Population_19[226 areas]. The Texas Water Development Board (TWDB) Groundwater Database (GWDB) comprises of data on chose water wells, springs, oil/fuel watches (that had been right off the bat expected to be or were changed over to water wells), water degrees, and water top of the line to profit advisor data around springs in Texas to help water arranging from a close by to a more noteworthy territorial viewpoint.

978-1-7281-5197-7/20/$31.00 ©2020 IEEE

Modified linear regression with PLS regression to find the groundwater level. PLS regression, in each stage a variable is considered for expansion or subtraction from set of illustrative factors dependent on some recommended criteria. We utilize the groundwater level observing information, to figure dependent on the reaction connection between persuasive factors, for example, precipitation and repository level and the difference in groundwater level.

II. LITERATURE SURVEY

Directing literature review preceding start on position paper is essential in understanding the Machine Learning algorithms**.** Missing data is a given within the scientific domain, so gadget studying models ought to have great performance even if lacking facts takes place. Groundwater level is the vital point for many countries economy and is anticipated by linear regression model with little examples and data size. In this paper we present knowledge engineering apparatuses with linear regression model based on the satellite data of water levels and population of the respective areas in last two years. The consequences of this trial show that our proposed system is superior to anything other knowledge engineering ways to deal with foresee the two-year by and large water level. Several predictive fashions based on conventional statistical techniques and device studying strategies were said, be that as it may, no controlling to variety in execution has not been outfitted to date Therefore, in this investigation, we perform linear regression method in the knowledge engineering to ad lib the effectiveness of the model. The exhibition of this regression method is assessed related to a few element choice methodology and the effect of the element determination on execution is future assessed.

[6] The model is improved from the fundamental SVR model by utilizing Wavelet bundle change. There was a contrast between the proposed WP – SVR model and the essential SMO-SVR model for each well area. The skill of any relapse exam basically depends on the nature of the information that is fed into it.

[1] It was seen that straight relapse with neural system calculations has been executed for better exactness where the Linear recurrence works admirably at anticipating the low groundwater levels in late spring and neural networks to predict the high groundwater levels in winter.

[16] A certain article proposes Artificial Neural Systems ANN are more accurate than various nonlinear relapses and different direct regression. After analysing the effects of MLR and ANN, it showed that the created ANN models give more accurate groundwater level expectations than the MLR models due to MLR's low forecast capacity.

[8] It utilizes Gaussian procedure order which gave more precision than backpropagation and counterfeit neural systems model.

[17]. This examination article proposes an adjusted test for prescient exactness of stochastic direct relapse model by utilizing Modified Lagrange Multiplier. This model is utilized to work with cutting edge numerical apparatuses. It incorporates indicating both the desire work and the qualities of the blunder.

[12] Proposes IK and RK techniques are utilized for anticipating the hazard evaluation of groundwater levels and sees that the utilization of the two previously mentioned strategies, went by the fitting spatial reliance capacity can give valuable and solid outcomes in request spatial changeability and potential consumption danger of the spring level to be recognized.

[3] This Suggests aligned enduring state groundwater table model is utilized for discovering groundwater level variances

[15] Lstm is contrasted and MLP and straight relapse models, where lstm is beaten. In return, mlp model beat the straight relapse model, which doesn't show better precision.

[2] Comparisons showed that in groundwater rate prediction, the SVM – RBF and SVM-PK had the similar accuracy. Overall, the MLR design has the lowest rate of achievement. Models ANFIS and RBNN yielded preferred results over the MLR method.The after-effects of the contextual analysis are agreeable and show that SVM-RBF and SVM-PK models in the groundwater hydrology region can be a valuable predictive device.

[11] This Proposes preparing the information with Levenberg Marquardt preparing calculation and utilize Artificial neural system for better exactness.

[4] The present examination completed to locate the ideal execution of ANN model through various pre-preparing procedures. On examination with Con-ANN, CWTFT-ANN expectation during preparing and testing process is seen to be acceptable. Additionally the created CWTFT-ANN predicts the definite variances example of the examination zone which is additionally used to gauge the change to one time feeling of anxiety.

[13] Results of several studies show that the CART model can evacuate the relationship between variables, lower repetitive data, and improve accuracy and efficiency of forecasts. In contrast to the PSO-SVR model, it is possible to use CART model with better fit and prescient speculation capacity to anticipate groundwater level avalanche.

[9] This study was performed to analyze the capability of two methodologies guided by data, MLR and ANN, to re-establish / anticipate transient groundwater levels over a groundwater bowl using relevant genuine information. It used the traditional ANN and MLR conventions displaying just like all the relevant and powerful information factors to achieve this goal where the created ANN models provide more precise groundwater level forecasting than the MLR models.

[5] This paper shows that when using trapezoidal enrollment capacities and mixture enhancement technique, implementing ANFIS model with different structures has the most precision and less error. Therefore, the ANFIS system was used to forecast groundwater utilizing three conditions in the next 2 years. The outcomes demonstrated that the groundwater level

has the base decrease when the siphoning pace of water wells is equivalent to the water deficiency of the spring.

[10] Proposes joining the neural network of Elman with wavelet analysis, this paper develops the model organized by Elman wavelet. This study also demonstrates the Elman wavelet arrange model's preparation procedure and applied this model to Naolihe bowl's groundwater expectation.

[7] Observes the Hybrid prefix achieves greater accuracy of standards than ones with different (untransformed) inputs. The HANN model's general implementation is higher than cross breed direct relapse (HMLR) and half breed nonlinear relapse (HMNLR) models.

[14] This examination assessed the capacity of another outrageous learning machine technique to anticipate month to month groundwater levels. The outcomes show that the Extreme learning machine has preferred gauging over the Support Vector Machine.

III. DATA SET

The dataset used for this is taken from Texas water level data where you can get the well water data and the population of 2018 and we can predict the water level for 2019 based on 2019 population data. The water levels are being affected by many variables such as temperature, rainfall, population density and water usage. The state of Texas has been currently recovering from the drought. The rainfall levels in Texas in 2018 is as follows

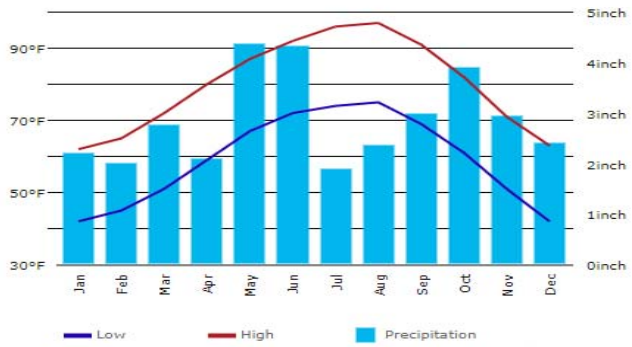

Figure 1. Rainfall and temperature levels in each month in 2018

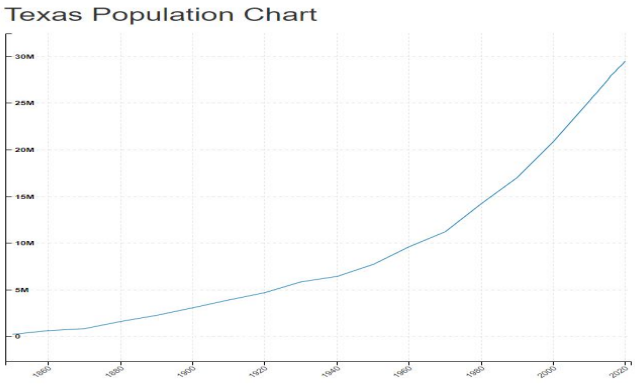

Figure 2. Raise of population in Texas State

The population of Texas is 28,701,845 in 2018 and 29,087,070 in 2019 with a growth of 1.34%. This also fairly effects the water usage because studies prove that nearly 60% of the water which is being used daily is ground water used in the state.

The above figure displays clear population growth in the recent years. As we can see, there has been a significant raise in the population over the past two decades where the Texas state has seen an accelerated growth by 73.5%.

IV. METHODOLOGY

The Linear Regression Algorithm has been a widespread algorithm and we are attempting to

implement it using the regression method of Partial Least Squares (PLS), for imposing restrictions on discriminatory analysis, main component regression and canonical correlation we are extending multiple linear regression. In addition, the correlation coefficient calculation shows the dependence between variables.

Step 1: Calculate the correlation between water level with each other variable.

All variables b, $a_1,...,a_j,.,a_p$ are assumed to be centered. The PLS regression model with m components is written as

$$b = \sum_{h=1}^{m} c_h \left(\sum_{j=1}^{p} w_{hj}^{*}.x_j\right) + residual$$

If we consider a sample set of data with response variables as B (in matrix form) and a large number of predictor variables as A (in matrix form), some of which are highly correlated. We calculate the value of factor score matrix T = XW, where W is the appropriate weight matrix. Then using the T value the linear regression model B = TQ+E is constructed, Where Q is the regression coefficient matrix of T and the error (noise) term is E.

The constructed model is similar to A = BX+E, where the variable X can be used as a method of predictive regression by using X = WQ.

Step 2: Apply the linear regression model on the variables which were found suitable after calculating the correlation using the PLS method. The model uses the formula

$$b_i = \beta_0 + \beta_1 a_{i1} + \ldots + \beta_p a_{ip} + \varepsilon_i$$

Where these equations can be written in matrix format after stacking as,

$$B = A\beta + \varepsilon ,$$

Where,

$b = [b_1,$

$b_2,$

.

.

$b_n]$

$$A = \begin{bmatrix} A_1^T \\ A_2^T \\ . \\ . \\ A_n^T \end{bmatrix} = \begin{bmatrix} 1 & a_{11} & . & . & a_{1p} \\ 1 & a_{21} & . & . & a_{2p} \\ . & . & . & . & . \\ . & . & . & . & . \\ 1 & a_{n1} & . & . & a_{np} \end{bmatrix}$$

$$A = \begin{bmatrix} \beta_0 \\ \beta_1 \\ . \\ . \\ \beta_p \end{bmatrix} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ . \\ . \\ \varepsilon_n \end{bmatrix}$$

Step 3:
After getting the relation between the variables we can predict the ground water level for the testing data.

start
function MLR
Estimate initial w
Define W to be a matrix and do the same for p
matrix and q vector
From the above variables define the variables which have high correlation
With those variables train the dataset
Predict the results for testing dataset
Compare the values
Print accuracy
Stop

Algorithm 1. PLS-LR Model

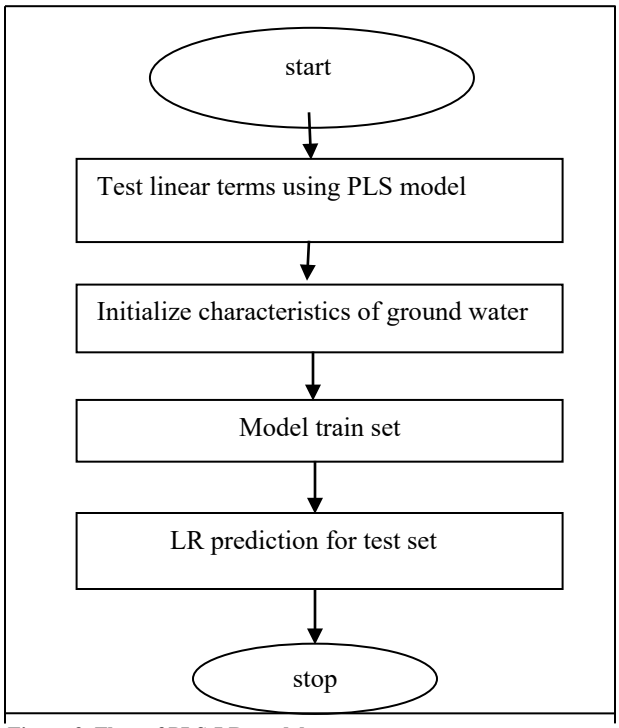

Figure 3. Flow of PLS-LR model

V. PERFORMANCE AND RESULT ANALYSIS

Table 1. comparision Table

Architecture	ME	MAE
PSO-SVR	1.53	21.11
CART	0.88	17.92
Modified LR	0.68	13.26

The above table shows the comparison between the existing models like CART and modified Linear Regression model. The Mean Error has decreased significantly when compared to the remaining models stated.

The plot shows the predicted values and the original values. The difference between both of them is not so big so you can say that most of the values are predicted accurately.

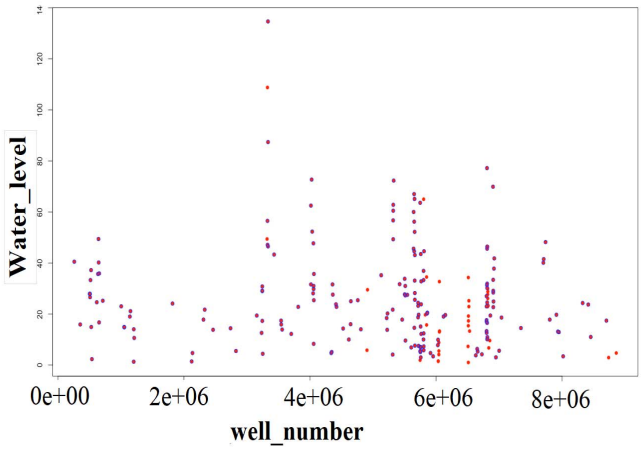

Figure 4. Comparison between predicted and original values

VI. CONCLUSION

This paper focused on the groundwater level prediction by using a modified linear regression method with PLS regression. We proposed a better method for the estimating groundwater level for the given data set using the Modified Linear Regression method. The predicting version can train itself and modify the performance of have a look at robotically inside the process of samples training. The performance does not rely upon what people already understand. The experimental results show that the predicting version has excessive precision. It can characterize the complicated system of the groundwater level. Using this method, a clear equation is generated and on comparing different algorithms we selected the best configuration and trained our model using Texas Dataset wells[226 areas], Population_18[226 areas] and test our model with Population_19[226 areas] and we noticed the highest accuracy. The results show that applying the Modified Linear regression model with Stepwise regression has the most accuracy and less error.

REFERENCES

[1] Predicting groundwater levels using linear regression by Sara Matta (2011)
[2] Modeling of groundwater level using artificial intelligence techniques : A case study of Reyhanli region in Turkey DEMIRCI, M,– ÜNEŞ, F. – KÖRLÜ, S. (2019)
[3] Understanding groundwater table using a statistical model Shao-fengYan, Shuang- enYu, Yu-baiWu, De-fengpan, Jia-gen dodge (2018)
[4] Pprediction and forecasting of groundwater level fluctuation by ANN technique . N.Vetrivel, K.Elangovan (2016)
[5] "Prediction of the groundwater level of bastam palin (Itan) by Artificial Neural Network(ANN) and Adaptive and Adaptive Neuro-Fuzzy Inference Syatems(ANFIS)". Samd Emamgholizadeh, Khadija Moslemi, Ghola Mhosein Karami (2014)
[6] "Forecasting monthly groundwater level fluctuations in coastal aquifers using hybrid Wavelet packet–Support vector regression", Sujay Raghavendra.N, Paresh Chandra Deka (2015)
[7] "Machine learning algorithms for modeling groundwater level changes in agricultural regions of the U.S". S.Sahoo, T.A.Russo, J.elliott, I.Foster (2017)
[8] Saeed Azimi, Mehdi Azhdary Moghaddam, and Seyed Arman Hashemi Monfared "Large-scale association analysis of climate drought and decline in groundwater quantity using Gaussian process classification" (2018)
[9] Groundwater-level prediction using multiple linear regression and artificial neural network techniques:a comparative assessment. Authors: Sasmitha Sahoo, Madan K.Jha (2013)
[10] "Research on groundwater level prediction of naoli river basin based on Elman Wavelet Neural Networks". Peng Sheng-min, Huang Jia- Xin, Fu Qiang (2011)
[11] Groundwater level prediction using artificial neural network model. Author: Pallavi Porte, Rajendra Kumar Isacc,kipoo Kiran Singh Year of Publish: 2018
[12] "Regression Analysis and Risk Assessment of Groundwater Levels", Panagiota Theodoridou, Emmanouil Varouchakis and George Karatzas (2018)
[13] "Groundwater level prediction of landslide based on classification and regression tree" by Yannan Zhao, Yuan Li, Lifen Zhang, Qiuliang Wang (2016)
[14] Basant Yadav, Sudheer ch, Shashi Mathur, Jan Adamowski "Assessing the suitability of extreme learning machines for groundwater level prediction". (2016).
[15] Rebecca Doble, Guobin Fu, Nan Zhang "Groundwater Recharge Prediction Using Linear Regression, Multi-Layer Perception Network, and Deep Learning" (2019).
[16] Prediction of groundwater levels from lake levels and climate data using ANN approach. Ahmet Dogan, Husnu Demirpence, and Murat Cobaner (2008)
[17] "A Study on Misspecification and Predictive Accuracy of Stochastic Linear Regression Models." C. Narayana , B. Mahaboob , B.Venkateswarlu , J. Ravi sankar and P. Balasiddamuni.(2018)