2023 Eighth International Conference on Science Technology Engineering and Mathematics (ICONSTEM) | 979-8-3503-4779-1/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICONSTEM56934.2023.10142495

Comparison of Linear Regression and Logistic Regression Algorithms for Ground Water Level Detection with Improved Accuracy

C. Gnaneswar Raju[1], V. Amudha[2] and Sajiv G[3]

[1]*Research Scholar, Department of Electronics and Communication, Saveetha School of Engineering, SIMATS, Chennai, India.*
[2,3]*Department of Electronics and Communication, Saveetha School of Engineering, SIMATS, Chennai, India.*

Email: amuthav.sse@saveetha.com, 3sajivg2005.sse@saveetha.com

Abstract— Comparison of the innovative Linear Regression and Logistic Regression Algorithms for Ground Water Level Detection with Improved Accuracy is the goal of this study, which was designed to investigate that question. A total of 30 Specimens are split up into their respective groups. Every person received 15 different samples. The Novel Linear Regression Algorithm is used for Group 1, whereas the Logistic Regression Algorithm is used for Group 2. The accuracy of the model generated by the Novel Linear Regression Algorithm is (93.27%), which is higher than the accuracy generated by the Logistic Regression Algorithm, which is (86.5%). It is determined using an independent sample T-test, and the Significance Value is 0.439, which indicates that the hypothesis is not significant. This is shown by the fact that p>0.01 is returned. Therefore, the accuracy of the Novel Linear Regression Algorithm, which was found to be 93.23%, is discovered to be greater than the accuracy of the Logistic Regression Algorithm, which was found to be 86.5%.

Keywords: Novel Linear Regression Algorithm, Logistic Regression Algorithm, Groundwater Level, Irrigation.

I. INTRODUCTION

As part of the hydrologic cycle, water travels from higher elevations (surface water) to lower elevations, moving downward. This movement takes place from top to bottom. This takes place from the most highest point to the very lowest (groundwater). In order to replenish groundwater, one of two distinct approaches may be utilised: either the natural or the artificial one. In their own unique ways, natural and artificial means of recharging can each contribute to the process of replenishing resources. Sand, rocky strata with fissures, perennial rivers and streams, and wooded regions are all advantageous for natural recharge due to the fact that precipitation is able to penetrate the soil more quickly via these types of topographical features. This allows for a greater capacity for natural recharge. Sand is beneficial for natural recharging in its own right. [1] This occurrence takes place at a particular point, which is referred to as the "water level," and the word "water level" is used to designate this specific spot in the environment. Agriculture is one of India's most important industries and has a significant impact on the country's overall economy. This is because sixty percent of the country's population relies on the land as their major source of food. As a result, the land has become increasingly degraded. [2] The area that is the subject of this research has a total land area of roughly 91,000 ha, all of which is irrigated by water that is transported from canals that are situated in the surrounding area. It is necessary to have a total of 1,368 unique irrigation tanks in order to provide water to the 47,000 hectares of land that are under their care. Taking readings of the groundwater level and making notes about them is a standard technique that is followed when mapping and surveying an aquifer. This is done in order to create an accurate map of the aquifer. [3].

II. LITERATURE SURVEY

The topic of assessing the level of groundwater has been the subject of more than 1700 different publications that have been authored and published. These publications have been distributed in numerous locations across the globe. The amount of surface water that can be obtained will continue to fall, which will lead to the extraction of groundwater becoming an increasingly important practise [4]. [Citation needed] The planning and implementation of all of the preventative measures that were taken were carried out with the utmost attention paid to ensuring that the groundwater survey that was

conducted was carried out without incident. In this study, we address the issue of predicting water levels by carrying out a comparative analysis based on a novel combination of linear and logistic regression [5]. The results of this analysis show that the novel combination of these two types of regression can produce more accurate predictions. Our investigation makes use of an innovative hybrid approach that combines linear and logistic regression. Based on the findings of our analysis, it appears that combining the methodologies of these two unique forms of regression could result in more accurate findings. Throughout the course of our investigation, we make use of a cutting-edge hybrid methodology that combines linear regression and logistic regression. This allows us to obtain more accurate results. In order to accomplish the goal that is being sought after in terms of making accurate predictions, the ground-breaking method that is known as linear regression uses independent variables as the primary component in the construction of its framework. This allows the method to achieve the desired result. It will first identify how various variables are compared to one another, and then, depending on the results of those comparisons, it will provide projections [6]. This is its primary aim.

By comparing and contrasting the linear and logistic regression approaches, the goal of this study is to explore the efficacy of various models for estimating the amount of ground water. [7] Specifically, the linear and logistic regression approaches will be compared and contrasted. When it comes to the completion of actions that need regression, the application of innovative linear regression is used. On the other hand, logistic regression [8] is utilised as a tool to help in the completion of classification duties. The prediction of forthcoming occurrences or results is one of the most common and widely used applications of supervised machine learning, which is a method that falls under the umbrella of the field of machine learning and is considered to be one of the most common and widely used applications of machine learning. This is one of the most widespread applications of supervised machine learning and one of the most common as well. [9] It is required to make use of datasets that have been labelled in the appropriate manner in order to learn and produce accurate predictions. Using datasets that have been labelled in the suitable manner may be found here. Both the extensive information that our team possesses and the extensive research experience that it has contributed to the production of high-quality publications. [17]–[25].

III. Methodology

This experiment was carried out in the Artificial Intelligence Laboratory of the Department of Artificial Intelligence, which is situated inside the Saveetha School of Engineering, which is located inside the Saveetha Institute of Medical and Technological Sciences. Over the course of this research endeavor, the participants were arbitrarily divided into two separate groups at various points. The Novel Linear Regression method and the Logistic Regression Algorithm are the two groups that are being compared in terms of how accurately they predict groundwater levels using the datasets that have been provided. The Novel Linear Regression method has an accuracy rate of 93.23 percent, while the Logistic Regression Algorithm has an accuracy rate of 86.5 percent. The Novel Linear Regression approach has a success rate of 93.23 percent in terms of accuracy. An accuracy rate of 86.5% may be expected from the Logistic Regression Algorithm. Due of the even distribution of members between the two groups, there are now 15 samples in each of the two groups, bringing the total number of samples that are shared by both groups to 30. In the first group are the first 15 instances of correct answers, and in the second group are the remaining 15 cases of correct responses. Both groups have 30 total instances of correct answers. Both teams have a total of 30 accurate instances in their collections. According to the results of the prior research [10], the procedures for enrolling individuals and collecting samples are carried out in a manner that is consistent with a confidence interval of 95.

It is essential to have a dataset that is composed of datasets that are required in order to instruct the Groundwater Level prediction framework in order to achieve the objectives of this project. A minimum of 8 GB of random-access memory is required for the framework in order to ensure that direct access is granted to the central processing unit (CPU). When it comes to the central processing unit (CPU), it is recommended that you choose with either the Intel i3 or the Intel i5 model that is from the third generation. It is essential to have at least 250 gigabytes of free space on the hard disc in order to store the code, save the necessary photographs of the dataset that are downloaded from www.kaggle.com,

and install the plug-in or other software. The photographs of the dataset can be downloaded from www.kaggle.com. If the hard drive does not have enough free space, then it will not be possible to effectively finish these tasks. It is recommended to use a graphics card that has a minimum of 2 gigabytes of random access memory (RAM) in order to reduce the workload placed on the central processor unit (CPU) of the computer and to enable the processing of pictures to take place at a faster rate. Kaggle is an online development environment that may be used to assist in laying the groundwork for object-oriented programming.

a) Linear Regression Algorithm: Using a new linear regression approach, we were able to forecast future values using statistical data that was compiled based on historical values. b) Linear Regression Database: An innovative use of the linear regression approach used to forecast future values using statistical information and information based on historical values. Regression models find prediction values using independent values. The purpose of forecast analysis, which strives to identify links between the variables that have been stated and the values that are predicted, is aided in some measure by the information provided by this factor.

An Algorithm Steps for Linear Regression:
Step 1: Constructing a Linear Regression Model Using Libraries.
Step 2: Read the input data set, which should contain an example of forecast data with prices from the prior year.
Step 3: Train the model using the values that have been set for Predictions
Step 4: Feed the Model with the Accumulated Test Information
Step 5: Similarities between the Train and Test data sets are computed.
Step 6: Give the value from the Forecast to the function that does the Predict.
Step 7: Write down the results.

Algorithm For Logistic Regression
Logistic regression, often known as LR, is one of the most well-known algorithms utilised in the field of machine learning. Logistic regression (LR) is classified as a type of supervised learning technique. It helps in the prediction of the dependent variable by making use of a few elements that are independent of one another. It is helpful in making predictions about the outcomes of categorical dependent variables. This algorithm can build probability factors to categorise new datasets, which makes it a very significant prediction algorithm that is most commonly utilised for the purpose of solving classification difficulties.

The algorithm for logistic regression is as follows:
Step 1:Begin entering the values.
Step 2: Import the dataset that will be used for training.
Step 3: Importing Python Libraries.
Step 4: Investigate and organise the data.
Step 5: Converting the Categorical Variables into Dummy Variables.
Step 6: Separate the Training Datasets from the Test Datasets.
Step 7: Scaling, which involves transforming the numerical variables.

A. Step 8: Logistic regression model fitting.
B. Step 9: Conduct an Evaluation of the Model.
C. Step 10: Interpret the Results.
D. ***The Analyses of Statistics***
E. The IBM SPSS Version 26 software was utilised to carry out the statistical analysis. The Independent Sample T-test is carried out by comparing the mean accuracies of the samples being tested. Viewing angle and image size are considered to be dependent variables, while accuracy and the number of photographs taken are considered to be independent factors (Kevin Zhou 2015).

IV. RESULTS

The Accuracy of the Novel Linear Regression algorithm is 93% and the accuracy of the Logistic Regression algorithm is 85%.

Table 1: shows the sample data of the accuracies of the Novel Linear Regression algorithm and Logistic Regression algorithm

S. No	LIR	LOR
1	93.50	86.53
2	91.70	85.70
3	90.60	85.20
4	92.80	87.80
5	89.70	84.90
6	84.40	83.80
7	92.80	84.86
`8	91.80	89.60
9	93.50	82.80
10	95.80	84.98

Table 1 shows the sample data of the accuracies of the Novel Linear Regression algorithm and Logistic Regression algorithm.

Table 2: Group statistics comparison for accuracy of Sample outputs. shows the group statistics for size N= 30, which contains mean accuracies and standard deviations for Novel Linear Regression and Logistic Regression methods.

Group Statistics

	GROUP	N	Mean	St. Deviation	St. Error Mean
ACCURACY	LIR	10	91.6600	3.0576	.96726
	LOR	10	85.6170	1.95431	.61801

Table 2 shows the group statistics for size N= 30, which contains mean accuracies and standard deviations for Novel Linear Regression and Logistic Regression methods

Table 3: Shows Independent Sample Test between LIR and LOR algorithm. Shows the Independent sample T-test that compares the proposed as well as the existing algorithm with a confidence interval of 95%.

GROUP		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	T	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
ACCUR-ACY	Equal variances assumed	.627	.439	5.265	18	.000	6.04300	1.14784	3.63148	8.45452
	Equal variances not assumed			5.265	15.298	.000	6.04300	1.14784	3.60059	8.48541

Table 3 Shows the Independent sample T-test that compares the proposed as well as the existing algorithm with a confidence interval of 95%. Table 4 shows the Bi-variate correlation of the accuracies of both the Linear Regression algorithm and the Logistic Regression Algorithm.

Figure 1 is a graph that shows the comparison between the mean accuracy of the Novel Linear Regression Algorithm and the mean accuracy of the Logistic Regression algorithm along with the error bars

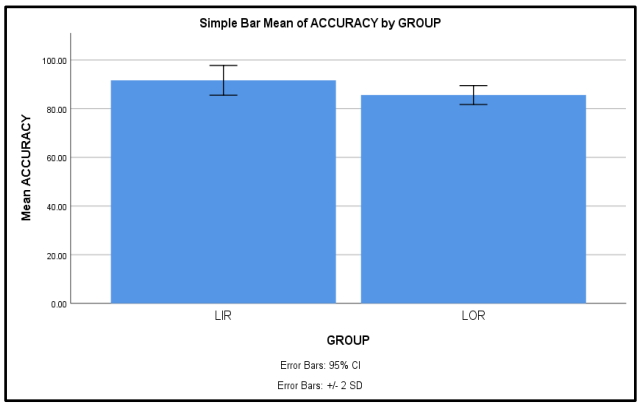

Fig 1: Shows mean accuracy comparison graph that shows the comparison between the mean accuracy of the Novel Linear Regression Algorithm and the mean accuracy of the Logistic Regression algorithm along with the error bars

V. Discussion

According to the findings, the method used for Novel Linear Regression performed appreciably better than that used for Logistic Regression [11]. This can be deduced from the fact that the aforementioned conclusion is supported by the evidence. This becomes abundantly evident when one considers the fact that the method of Innovative Linear Regression was described in the previous section. It has been revealed that the accuracy percentage of logistic regression is 86.5%, but the accuracy percentage of new linear regression is found to be 93.7% [12].

In logistic regression, the assumption of linearity between the dependent variables and the independent variables is established. This results in a lower design cost, which is one of the many benefits of using logistic regression [13]. The method of Innovative Linear Regression makes the process of estimate a great deal simpler, and perhaps even more importantly, it makes the applicability of these linear equations at the modular level abundantly clear.

Because it requires the production of a collection of model predictions that are difficult or impossible to undo in the event that a particular model fails, the process of training new linear regression algorithms is time-consuming and expensive [14]. This is because the process requires producing predictions that cannot be undone.

In the beginning, those people who would later be referred to as the Generation Even though there is no groundwater monitoring well data available, you can still make use of our findings to fill in the data gaps for shallow groundwater conditions, investigate the potential implications on GDEs, and improve sustainable groundwater management [16]. You will be able to accomplish this goal with the cooperation of water authorities at both the state and local levels[15]. [For additional details, please click here.]

VI. Conclusion

Based on the findings, we were able to determine without a doubt that the Novel Linear Regression method has the highest accuracy rate of 93% out of all the calculations that have been carried out. The accuracy of the Logistic Regression approach is lower than that of the Linear Regression technique, which has an 86% success rate. When it comes to making reliable predictions about groundwater levels, the Novel Linear Regression method, as opposed to the Logistic Regression approach, is the method of choice.

References

[1] M. Demirci, F. Üneş, and S. Körlü, "MODELING OF GROUNDWATER LEVEL USING ARTIFICIAL INTELLIGENCE TECHNIQUES: A CASE STUDY OF REYHANLI REGION IN TURKEY," Applied Ecology and Environmental Research, vol. 17, no. 2. pp. 2651–2663, 2019. doi: 10.15666/aeer/1702_26512663.

[2] N. S. Raghavendra, N. Sujay Raghavendra, and P. C. Deka, "Forecasting monthly groundwater level fluctuations in coastal aquifers using hybrid Wavelet packet–Support vector regression," Cogent Engineering, vol. 2, no. 1. p. 999414, 2015. doi: 10.1080/23311916.2014.999414.

[3] S.-F. Yan, S.-E. Yu, Y.-B. Wu, D.-F. Pan, and J.-G. Dong, "Understanding groundwater table using a statistical model," Water Science and Engineering, vol. 11, no. 1. pp. 1–7, 2018. doi: 10.1016/j.wse.2018.03.003.

[4] N. Vetrivel and K. Elangovan, "Prediction of Groundwater Level Fluctuation by ANN and ANFIS Techniques," Asian Journal of Research in Social Sciences and Humanities, vol. 6, no. 5. p. 616, 2016. doi: 10.5958/2249-7315.2016.00140.4.

[5] S. Emamgholizadeh, K. Moslemi, and G. Karami, "Prediction the Groundwater Level of Bastam Plain (Iran) by Artificial Neural Network (ANN) and Adaptive Neuro-Fuzzy Inference System (ANFIS)," Water Resources Management, vol. 28, no. 15. pp. 5433–5446, 2014. doi: 10.1007/s11269-014-0810-0.

[6] S. Sahoo, T. A. Russo, J. Elliott, and I. Foster, "Machine learning algorithms for modeling groundwater level changes in agricultural regions of the U.S," Water Resources Research, vol. 53, no. 5. pp. 3878–3895, 2017. doi: 10.1002/2016wr019933.

[7] M. U. Farooq, A. M. Zafar, W. Raheem, M. I. Jalees, and A. A. Hassan, "Assessment of Algorithm Performance on Predicting Total Dissolved Solids Using Artificial Neural Network and Multiple Linear Regression for the Groundwater Data," Water, vol. 14, no. 13. p. 2002, 2022. doi: 10.3390/w14132002.

[8] A. Madani, M. Hagage, and S. F. Elbeih, "Random Forest and Logistic Regression algorithms for prediction of groundwater contamination using ammonia concentration," Arabian Journal of Geosciences, vol. 15, no. 20. 2022. doi: 10.1007/s12517-022-10872-2.

[9] P. Theodoridou, E. Varouchakis, and G. Karatzas, "Regression Analysis and Risk Assessment of Groundwater Levels," EWaS3 2018. 2018. doi: 10.3390/proceedings2110641.

[10] C. Reinders, H. Ackermann, M. Y. Yang, and B. Rosenhahn, "Object Recognition from very few Training Examples for Enhancing Bicycle Maps," 2018 IEEE Intelligent Vehicles Symposium (IV). 2018. doi: 10.1109/ivs.2018.8500469.

[11] P. G. Theodoridou, E. A. Varouchakis, and G. P. Karatzas, "Spatial analysis of groundwater levels using Fuzzy Logic and geostatistical tools," Journal of Hydrology, vol. 555. pp. 242–252, 2017. doi: 10.1016/j.jhydrol.2017.10.027.

[12] A. Dogan, H. Demirpence, and M. Cobaner, "Prediction of groundwater levels from lake levels and climate data using ann approach," Water SA, vol. 34, no. 2. p. 199, 2019. doi: 10.4314/wsa.v34i2.183640.

[13] X. Huang, L. Gao, R. S. Crosbie, N. Zhang, G. Fu, and R. Doble, "Groundwater Recharge Prediction Using Linear Regression, Multi-Layer Perception Network, and Deep Learning," Water, vol. 11, no. 9. p. 1879, 2019. doi: 10.3390/w11091879.

[14] C. Narayana, B. Mahaboob, B. Venkateswarlu, J. Ravi sankar, and P. Balasiddamuni, "A Study on Misspecification and Predictive Accuracy of Stochastic Linear Regression Models," International Journal of Engineering & Technology, vol. 7, no.

4.10. p. 532, 2018. doi: 10.14419/ijet.v7i4.10.21221.
[15] C.-Y. J. Peng, K. L. Lee, and G. M. Ingersoll, “An Introduction to Logistic Regression Analysis and Reporting,” The Journal of Educational Research, vol. 96, no. 1. pp. 3–14, 2002. doi: 10.1080/00220670209598786.
[16] C. Suryanarayana, C. Sudheer, V. Mahammood, and B. K. Panigrahi, “An integrated wavelet-support vector machine for groundwater level prediction in Visakhapatnam, India,” Neurocomputing, vol. 145. pp. 324–335, 2014. doi: 10.1016/j.neucom.2014.05.026.
[17] M. Kamal et al., “Machine Learning and Image Processing Enabled Evolutionary Framework for Brain MRI Analysis for Alzheimer’s Disease Detection,” Comput. Intell. Neurosci., vol. 2022, p. 5261942, Mar. 2022.
[18] V. Mousavi, M. Varshosaz, F. Remondino, S. Pirasteh, and J. Li, “A Two-Step Descriptor-Based Keypoint Filtering Algorithm for Robust Image Matching,” IEEE Transactions on Geoscience and Remote Sensing, vol. 60. pp. 1–21, 2022. doi: 10.1109/tgrs.2022.3188931.
[19] R. T et al., “Hyperspectral Image Classification Model Using Squeeze and Excitation Network with Deep Learning,” Comput. Intell. Neurosci., vol. 2022, p. 9430779, Aug. 2022.
[20] B. T. Geetha, P. Mohan, A. V. R. Mayuri, T. Jackulin, J. L. Aldo Stalin, and V. Anitha, “Pigeon Inspired Optimization with Encryption Based Secure Medical Image Management System,” Comput. Intell. Neurosci., vol. 2022, p. 2243827, Aug. 2022.
[21] Y. Alqahtani, U. Mandawkar, A. Sharma, M. N. S. Hasan, M. H. Kulkarni, and R. Sugumar, “Breast Cancer Pathological Image Classification Based on the Multiscale CNN Squeeze Model,” Comput. Intell. Neurosci., vol. 2022, p. 7075408, Aug. 2022.
[22] M. Saravanan et al., “Intelligent Satin Bowerbird Optimizer Based Compression Technique for Remote Sensing Images,” Computers, Materials & Continua, vol. 72, no. 2. pp. 2683–2696, 2022. doi: 10.32604/cmc.2022.025642.
[23] P. Vasanthkumar, A. R. Revathi, G. Ramya Devi, R. J. Kavitha, A. Muniappan, and C. Karthikeyan, “Improved wild horse optimizer with deep learning enabled battery management system for internet of things based hybrid electric vehicles,” Sustainable Energy Technologies and Assessments, vol. 52. p. 102281, 2022. doi: 10.1016/j.seta.2022.102281.
[24] A. Wongchai, S. K. Shukla, M. A. Ahmed, U. Sakthi, M. Jagdish, and R. Kumar, “Artificial intelligence - enabled soft sensor and internet of things for sustainable agriculture using ensemble deep learning architecture,” Computers and Electrical Engineering, vol. 102. p. 108128, 2022. doi: 10.1016/j.compeleceng.2022.108128.
[25] B. Swaminathan et al., “IOTEML: An Internet of Things (IoT)-Based Enhanced Machine Learning Model for Tumour Investigation,” Comput. Intell. Neurosci., vol. 2022, p. 1391340, Sep. 2022.