Ground Water Level Prediction

Srigurulekha K.[1] & Dhivya S.[2]

[1]PG Student, [2]Assistant Professor, Department of Computer Science and Engineering,
Sri Shakthi Institute of Engineering and Technology, Coimbatore, India.

ABSTRACT

The groundwater is one of the biggest asset. The continuous decline of underground water is one of the crucial factors that affect the development of economy and society. Based on the availability the water is to be exploited. The population growth and emerging climate change lead to water scarcity problem. This would have increased the demand for ground water. But the groundwater is not distributed equally. In our proposed system we state the annual ground water availability for future use. It can be classified by machine learning algorithms such as logistic regression, random forest and decision tree. The input variables used for modelling is based on the recharge of water and usage of water in various season. The regression task needs less training data set and can achieve good performance. The aim of this paper is to compare and analyse the machine learning algorithm for ground water level along with their accuracy. In this paper the machine learning algorithm is used to classify and detect the ground water availability levels in India.

Keywords: Ground water, Logistic regression, Decision tree, Random forest.

1. Introduction

The groundwater is a hydrologic process where water moves downwards from surface water to ground water. The groundwater recharge is of two types natural recharge and artificial recharge. In natural recharge the rainwater enters inside the soil through voids and the conditions favourable for natural recharge are sandy, rocky strata with fractures, perennial rivers, streams, forestland. at a certain intensity the pore space in the soil is full of water and the level at which this takes place is called water level. Groundwater gets replenished after every rainfall, this is called as rainfall recharge. After every recharge, the water table raises denoting that the porous media has saturated with more water. While pumping out the water, level of water goes down, continuous pumping of water, beyond the recharge, will make the wells go dry and force to deepen the well.

Nowadays, due to rapid industrialisation, population increase, intensive agriculture, urbanisation the use of water has been considerably increased. For this purpose the search of ground water got increased, due to the non-availability of sources and due to the declining water tables. So the ground water is essential source for social and financial improvement. But the ground water is not uniformly distributed [1]. Overexploitation of ground water has posed severe problems for groundwater managers in India. Water logging, depletion of water tables, drying of aquifers, salinity, saltwater water encroachment are the major consequences of intensive irrigation and overexploitation. In a recent report the water table in many parts of the country is declining at the rate of 1-2m per year [5]. The availability fresh water for irrigation domestic and industrial uses will get affected. Depth of water level generally ranges from 10-4- m bgl in major parts of north western states-Rajasthan, Haryana and Delhi the water level are more than 40 m bgl is recorded. In India about about 85% of the groundwater sources are drawn for irrigation and domestic, industry. Ground water use in Delhi, Haryana, Punjab and Rajasthan is more than 100% and in states of Gujarat, Karnataka, Tamilnadu and Uttarpradesh it is more than 70%. India is the highest

groundwater user in the world [9]. In future, India is going to face a serious water crisis. to overcome the water scarcity situation the ground water is to be managed and developed. Government has provided various legislative measures for ground water management problems .In this project logistic regression, decision tree classifier and random forest algorithm are used to anticipate the ground water level. To predict the groundwater level, the recharge of water and usage of water is to be represented through the input characteristic . Based on the accuracy obtained for the Algorithms, the efficient algorithm is identified and ground water level is plotted as graph.

2. Literature Survey

Literature review is essential in understanding the machine learning algorithm in ground water level. Groundwater plays a vital part in many countries economy and in this proposed it is anticipated by the machine learning algorithms such as logistic regression, decision tree classifier, and random forest model with little example and data size.

[1] This model utilizes wavelet bundle change and it is improved from the fundamental SVR model. There is difference between the WP-SVR model and the SMO-SVR model. Basically it depends on the information that is given as input.

[2] It gave more precision than backpropogation by utilizing the Gaussian procedure and neural system model.

[3] In this paper, ANFIS model with different structure is implemented and it has the most precision and less error by using the mixture enhancement technique and trapezoidal enrolment capabilities. The outcome of this paper is the groundwater level has the base decrease when the spring water deficiency is equivalent to the siphoning water wells.

[4] Most study uses the CART model to evaluation the relationship between low repetitive data, efficiency of forecast, variables and improved accuracy. The PSO-SVR model uses the CART model for the best fit and to detect the groundwater level.

[5] Artificial Neural System (ANN) is used to get the more accurate results than the various nonlinear relapses and direct regression. Due to low forecast capacity in MLR model it provides level accuracy than the ANN.

[6] This paper states the month to month groundwater ground water level. The outcome show that it has preferred gauging over the Support Vector Machine (SVM).

3. Datasets

The dataset is taken from the data.gov. The data set has 30 data, those data's are states of India and it has 13 attributes. Attributes are Recharge from rainfall Monsoon season, Recharge from other sources, Recharge from rainfall Non-monsoon season, Recharge from other sources (1), Total Rainfall, Natural discharge during non-monsoon season, Net annual groundwater availability, Irrigation, Domestic and industrial uses, Total Usage, Projected demand for domestic and industrial uses up to 2025, Groundwater availability for future irrigation use and situation. The situation is classified as Excess, Moderated and Semi critical. A sample data set is shown in Table 1. The dataset is classified into test data and train data. The dataset were tested through label encoder because

the dataset didn't acquires the disserted accuracy. To overcome the over fitting problem the dummy variable is used in the situation attribute.

Table 1. Groundwater dataset

	0	**1**	**2**	**3**
States	Andhra Pradesh	Arunachal Pradesh	Assam	Bihar
Recharge From Rainfall In Monsoon Season	15.1	3.4	19.0	18.9
Recharge From Other Sources	6.5000	0.0003	2.2000	3.9000
Recharge From Rainfall In Non Monsoon Season	5.5	1.0	8.6	3.4
Recharge From Other Sources (1)	6.7000	0.0004	0.6000	2.4000
Total Rainfall	33.8	4.5	30.4	28.6
Natural Discharge	3.1	0.5	2.5	2.4

During Non Monsoon Season				
Net annual groundwater availability	30.8	4.0	27.8	26.2
Irrigation	12.600	0.002	5.300	9.800
Domestic And Industrial Uses	1.500	0.001	0.700	1.600
Total usage	14.200	0.003	6.000	11.400
Projected Demand For Domestic And Industrial Uses Up to 2025	2.70	0.01	1.00	2.60
Ground Water Availability For Future Irrigation Use	15.9	4.0	21.5	13.9
Situation	Excess	Semi critical	Excess	Moderated
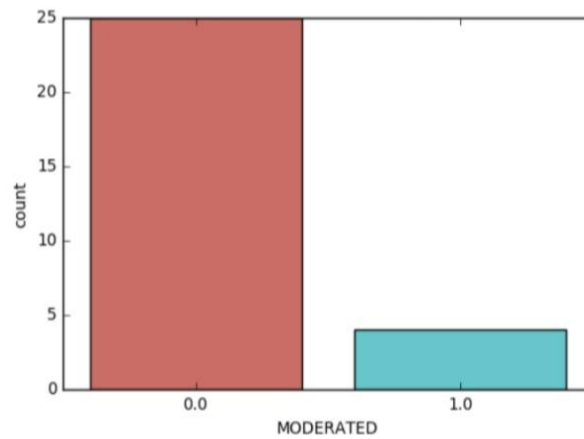
Fig.1. Bar chart for moderated situation
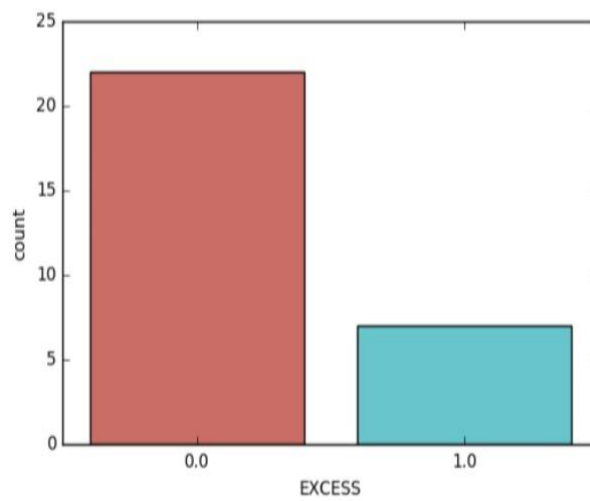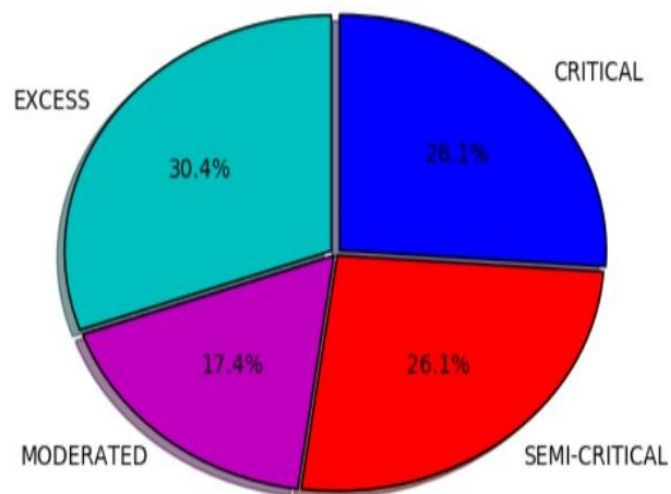

Fig.2. Bar chart for excess situation

Fig.3. Situation pie chart representation with percentage value

4. Proposed methodology

This section briefly explains the three classifiers for the ground water level prediction. Logistic regression is a part of the supervised machine learning. Random Forest is the ensemble learning approaches. By analyzing individual trees which builds many decision tree at the time of learning and predicting the output. The decision tree contains the control statements. Mostly it is used in case of possibilities and alternatives. which is used to select the appropriate decision with their consequences.
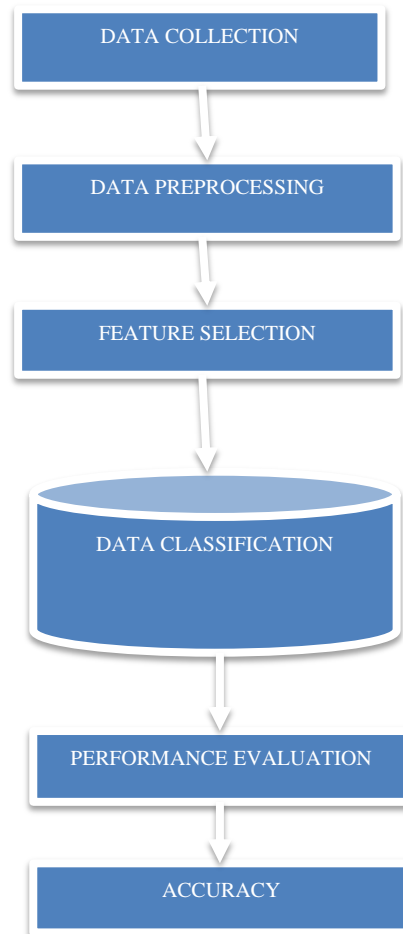

Fig.4. System Architecture

4.1 System architecture

Dataset collection: Dataset is collected explored in the python environment in jupyter notebook.

Data Pre-processing: The set of techniques used before the application of data mining method is named as data pre- processing. It refers to the estimation of missing values in some variables and it is needed as most of the interpretations cannot be done with missing data. The missing values are then replaced with the mean value in case of a continuous variable or with the mode value in case of a categorical variable.

Feature Selection: It is the process of identifying and removing the redundant and unwanted data. It reduces the accidental correlation in algorithms. The goal is to obtain the subset from the original data and describe it appropriately.

Testing: After feature selection 3 classifications algorithms including Logistic

Regression, Decision Trees, Random Forest, were used with the selected feature and a comparison between their prediction accuracy was done using Train/Test split method.

4.2. Algorithms

4.2.1. Logistic regression

It is a technique that can be used for traditional statistics as well as machine learning. Logistic regression is similar to linear regression except the logistic regression predicts whether something is true or false instead of predicting something continuous like size. Logit link function is used, where the data values are fitted for prediction. If a is the probability then the logit function for a is defined as:

$$Logit(a)= \ln(a/1\text{-}a) \qquad (1)$$

4.2.2. Decision tree classifier

It is a non-parametric supervised learning method used for both regression and classification. The ultimate goal of the decision tree is to predict the value of a target variable by creating a model and by learning simple decision rules inferred from the data features. More complexity occurs in decision rules and fitting the model when the tree gets deeper. The data sample is divided into a subsets based on notable attribute value test. It is a typical inductive approach for classification.

4.2.3. Random forest

Random forest is a ensemble of decision tree. It can be trained with bagging method. The objective of bagging method is to increase the overall end result. It also have the same hyper parameters like decision tree. Random forest algorithm has a collection of decision tree by using the various random vector. It searches the best feature instead of the important feature in a dataset during node splitting. This results in a better model.
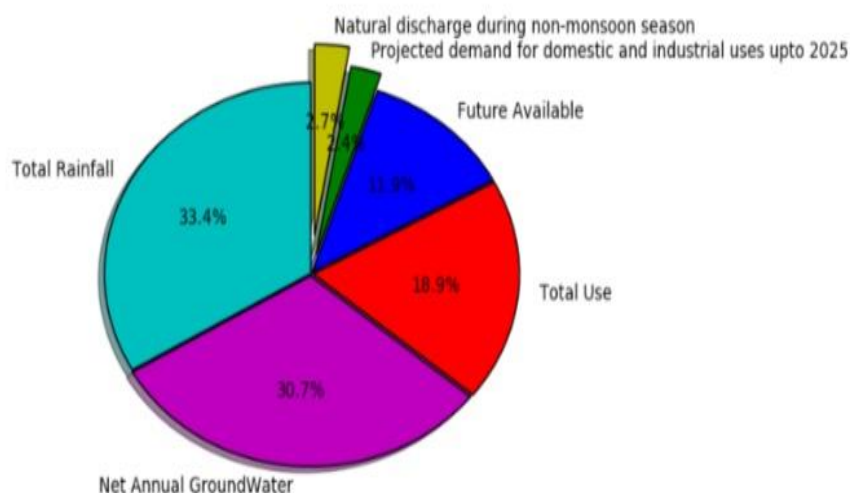
5. Experimental Results

Fig.5. Water usage pie chart representation with percentage value

Table 2. Comparison Table

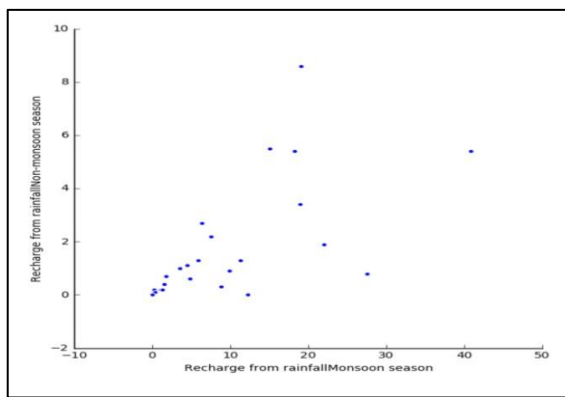|	**TEST DATA**	**TRAIN DATA**
LOGISTIC REGRESSION	0.93	0.92
RANDOM FOREST	0.90	1.00
DECISION TREE	0.93	1.00

Fig.6. Comparison between rainfall in monsoon and non monsoon season
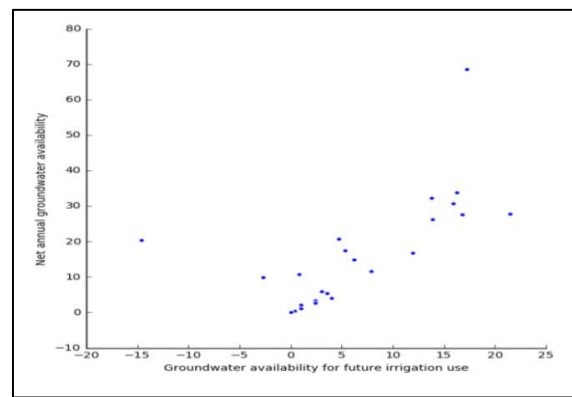

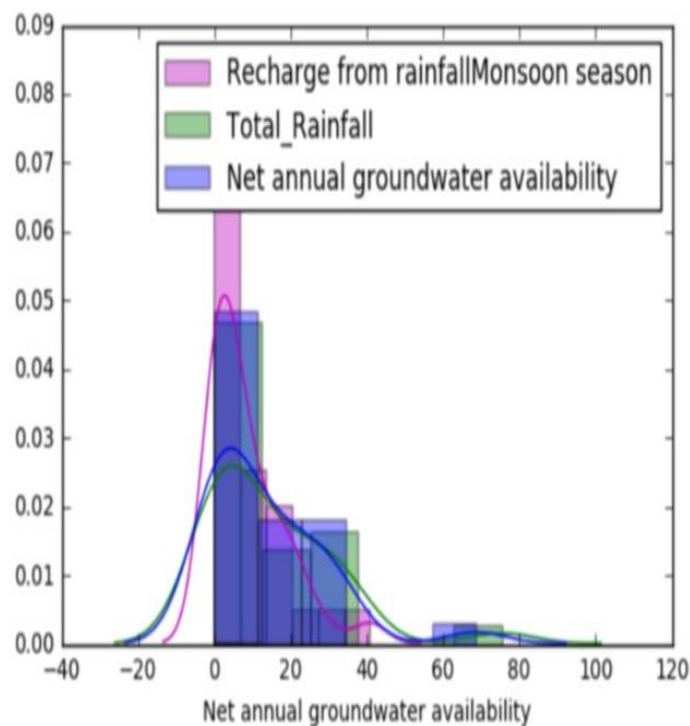**Fig.7.** Comparison amid ground water availability for future use and net annual groundwater availability
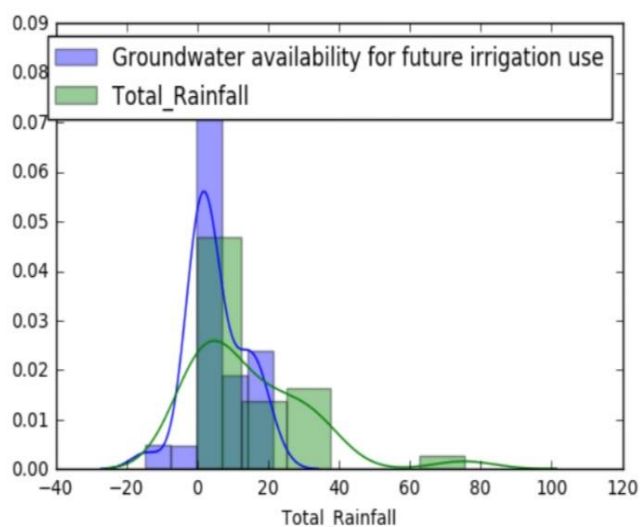

Fig.8. Net annual ground water availability

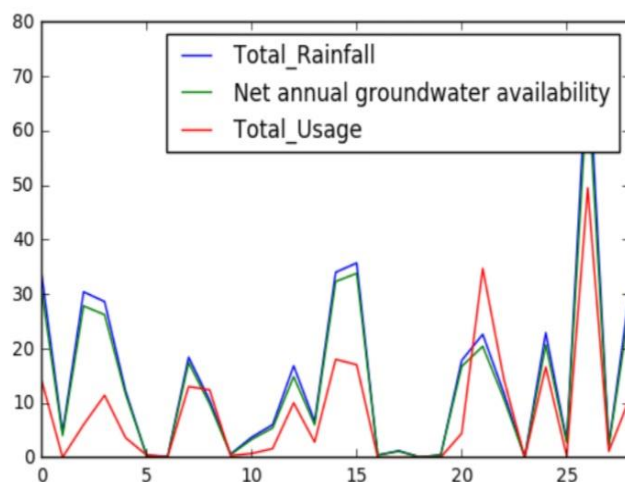**Fig.9.** Compaison between total rainfall and ground water availability

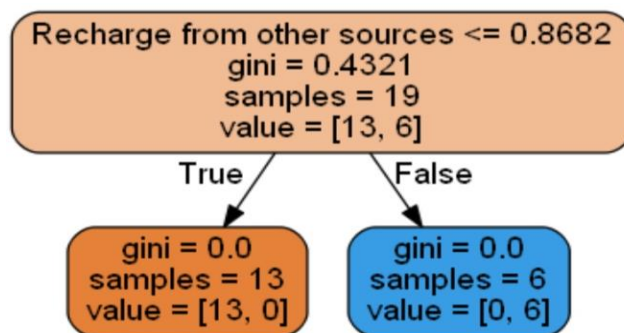**Fig 10.** Total usage of water

Fig.11. Recharge from other sources using decision tree classifer

6. Conclusion

In this project , the ground water level prediction is implemented using machine learning techniques. We proposed a good method for estimating the ground water level using logistic regression, random forest and decision tree . On

comparing these algorithm by using ground water availability in India dataset, we obtain the best accuracy with less error in all the three algorithms. In that, decision tree obtain more accuracy when compared logistic regression and random forest. And the results are plotted as a graph by using the datasets.

Declarations

Source of Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Competing Interests Statement

The authors declare no competing financial, professional and personal interests.

Consent for publication

We declare that we consented for the publication of this research work.

Code availability

The programming code that we have used for this research is available and authors are willing to share when it is required.

References

[1] Predicting groundwater levels using linear regression, Sara Matta (2011).

[2] Modeling of groundwater level using artificial intelligence techniques: A case study of Reyhanli region in Turkey DEMIRCI, M, (ù 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] Prediction 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, 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, Pallavi Porte, Rajendra Kumar Isacc, kipoo Kiran Singh (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, 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).