Analysis of Groundwater Level Responses to Precipitation Variability

Anitha R[1] ,Usha J[2]

Department of Computer Applications, RV College of Engineering®, Bengaluru, karnataka, India

Abstract -- Currently Data Mining process is implemented via using diverse equipment and techniques. Those extracting tools are easy to use and provide decisions. These techniques are used and implemented for the effective representation and visualization of data. WEKA is such an easy-to-use tool for data streams of any kind, such as web, medical, spatial, text, environment, etc. Groundwater is the most important source of water which helps in various activities like agriculture, industrial activities, etc. Since the rainfall is uncertain, groundwater level may vary accordingly. Hence this paper aims to analyze the rainfall and groundwater data from the collected input and also to predict the groundwater level based on the rainfall data provided by the user using the tool WEKA. The data is collected from CGWB with the input parameters Rainfall and Groundwater level and the considered region is Kolar. The outcome of the paper is comparison of output along with their displayed results of accuracy of the classifier. The results have been interpreted using graphs. Next section discusses in detail the methods available that helps for analysis.

Key Words:Data Mining, Naive Bayes, KNN, prediction, preprocess, correlation.

I Introduction

Data mining — a mechanism for the detection of various patterns in large data sets which includes at the juncture of machine learning techniques, statistics, etc. It is been applied in various fields such as diagnosis, forecasting, analyzing risks, probability of outcome, etc. The proposed paper focuses towards correlation analysis, prediction and linear regression. Correlation explains the relationship among attributes. While correlating, it is important to discover variant in a single attribute that impact to trade in other characteristic. Regression analysis explains the relationship among two or greater attributes to discover a variable that is depended on different variable.Prediction discovers the quality of the target attribute for the ones unknown items. It is generally named as forecasting. Forecasting as aninformation mining approach givesout the possibility to leverage several belongings of timely collected information, to resource employer preference-makersin order to planning powerfully and derived rate from previous information. Water assets are reducing via way of the growth of population as time goes by.If the rainfall is uncertain and when there is an increased rate of demand for the groundwater, which is a primary supply for groundwater, whichcauses exhaustion of groundwater wells, consequently leads in growth of salinity rate.Hence, firstly analyzing the groundwater level accordingly with the corresponding rainfall data which will help in finding out the relation between rainfall and groundwater level. Next section deals with the various papers which have been referred in order to carry out the work.

II Related Work

Wang.H.Gao JE et al describes that the excessive depth precipitation activities incorporates the majority of rechargeof groundwater.So he aimed at figuring out adjustments in ratio of precipitation and groundwater recharge and acquired the result using MRC approach [1]. Parneet Kaur et al strived at forecasting failure in the subjects which are core with the dataset of student, by utilizing many algorithms. The predicted accurate result of Decision Tree was 93% [2]. Jeerawan saelao et al in their paper, they proposed the analysis of average annual rainfall and used average data of monthly rainfall by considering the year 1993-2012 for analysis at Sansai Districtstation. The trend analysis of rainfall is estrimated based on analysis of descriptive statistics and time series method. The historical data of 20 years that is 1993-2012 was used and equation for trend data were plotted in linear equation. The result showed the extreme average rainfall data of the next 5 years[3]. Adhatus Solichah Ahmadiyah in his paper discussed about the importance of groundwater and also about the analysis and design of the system that could be for accessing groundwater quality status[4]. Erken Turganbaev et al in their paper wherein the intention was in powerful management of groundwater system through the use of current statistics era and also in optimization of statistics management methods on groundwater [5]. Christopher Beckham et al in his paper brings about various regression, classification, and various filter schemes that can be utilized for WEKA that is been implemented in python. He brought various dependencies that exist when applying particular filter and also the pyscript advantages [6]. Rafet Duriqi et al brings out the importance of classification algorithms that can be used for classification or prediction. In this paper they applied classification algorithms on datasets of 3 different types and analyzed the algorithm that is more convenient for a dataset of particular type (classifiers are Naïve Bayes, K* classification and Random Forest) [7]. Anand Kishor Pandey et al in their paper attempted towards comparatively analyzing the several classification algorithms. Such analysis has done using

WEKA. They brings out the comparison of Decision Tree , J48, Random Tree, Naïve Bayes, Random Forest, etc.algorithms with the assist of techniques that encompass effectively categorized, incorrectly labeled, Accuracy and plenty of other parameters[8]. Satyavati Shukla et al in their paper made a try to comprehend the outcomes of terrestrial adjustments of Upper Bhima watershed on GWL which showed the increase and decrease rates[9]. Mohamed Y. A. Toure et al in their paper examined the relationship between groundwater data in situ extracted from existing wells, and those derived from Gravity Recovery And Climate Experiment (GRACE) [10]. Chandrasegar Thirumalai et al in their paper discussed on heuristic prediction of rainfall the using machine learning strategies. Also measures the specific categories of statistics via linear regression approach in metrics and considered historic datasets that has diverse seasons for analysis. Linear regression technique enables in prediction and additionally suggested the lower correlation among the numerous crop seasons wherein the consequences are primarily based on previous year's records [11]. P.Kalaiyarasi and Mrs.A.Kalaiselvi for the concept of weather forecasting they have used classification clustering, neural network and decision tree. Sudden climatic changes bring out the problem for farmer. So they aimed at prediction using Numerical Weather Prediction method [12]. Razeef Mohd et al in his paper bring out the importance of rainfall prediction. The aim is to compare the performance of the algorithms that provides more accurate results. They have taken into consideration handiest five attributes which were applicable to rainfall prediction and concluded that J48 set of rules has top level of accuracy compared to Linear Regression, IBK, Random Forest and J48 [13]. Aaditya Gupta et al of their paper they concentratedtowards tracking the groundwater quality with an aid of the wirelessly related stress sensors related to thesystem which helped them to get right of entry to groundwater stage facts with much less cost and attempt. They made use of movementsto detectwireless communication infrastructure, technologies and integrated circuit[14]. Swati H. Mhaskar et al in their paper used Multiple Linear Regression (MLR) and Artificial Neural Network (ANN) for the assessment of recharge of groundwater in single catchment Shivade.Therecharge of groundwater was assessed using hydrological data and used data from 2002-2010 for the analysis. The models were tested and the performance showed that the Artificial Neural Network (ANN) was first-rate which showed better values in terms of correlation coefficient and measures of error in lower values [15]. Dodi Devianto et al in their paper discussed about the time series rainfall model and observed the intensity of rainfall were modeled using the Markov Switching Autoregressive (MSAR) as consequence from integrating the Markov chain and the classic Auto regressive model in the information mining analysis[16]. Alvin Lal in their paper, they applied gadget getting to know algorithms that is, Gaussian Process Regression (GPR) and Genetic Programming (GP) modelthat became used to approximate the density set up saltwater intrusion procedures and were anticipating salinity concentrations in an illustrative coastal aquifer machine.The GPR and GP fashions have been educated and validated the use of resulting and pumping salinity attention datasets received by solving numerical 3-d temporary density dependent of finite element based totally coastal aquifer float and transport version [17]. Huang Zhi-fang in his paper, uponseepage idea ofsoil saturated - unsaturated, by using evaluation on infiltration of rainfall version usingsoftware of finite detail analysis, the version of expansive slope of soil saturated - unsaturated rainfall infiltration turned into set up to simulate groundwater price[18]. Kreshna Gopal et al in their paper mentioned about the rainfall and groundwater level in which each exterior and groundwater level handiness have been resoluted by rainfall level, that is thenmotivated by means of massive-scale move styles [19]. Abhishek Sudhakar Jagtap et al of their paper, in which the goal changed into to broaden a machine for tracking the groundwater usage and to set up some policies and criterions based totally at groundwater degree in specific region. A programmed system wasformedthat monitored water stage and treatmentof water of the region. Machine mastering models have been used to compute the brink value in aprecise manner. Based on several factors exclusive edge values were fixed and the equal has been taken [20]. Meriyam Mhammdi Alaoui et al of their paper evolved a smart that turned into capable of reproducing heuristics mechanisms of hydrogeologists' specialists to assist any engineer / researcher of all levels prior to any hydrogeological modeling of the groundwater tool. A understanding-based totally framework changed into constructed adopting the object-orientated method, and the use of an Expert System shell. Also affords the number one steps of the framework development, its top-degree item hierarchy, and corresponding heuristic guidelines. [21]. Sorja Koesuma et al of their paper, wherein the purpose have become to decide the intensity of groundwater, decide potential sources of water based at the aquifers thickness, and plot the intensity of groundwater circulation the usage of GISapplication software [22]. Denisa Pospisilova et al in their paper analyzed the precipitation traits of Brno and its surroundings which have been executed with datasets from Czech Hydrometeorological Institute also additionally deals with spatiotemporal precipitation variability totals in Brno and its surroundings [23]. Madhuri Kommineni et al focused at prediction of groundwater level using PLS regression and modified linear regression. They suggested a higher technique to estimate groundwater stage for the assumed information through the use of Modified Linear Regression method [24]. Yogesh Kumar Joshi et al in their paper, they made use ofdata visualization techniques for the prediction which helped in analyzing large datasets. The graphs and plots helped them in higher analyzing which labeled the rainfall into excessive, lowest and common rainfall stage and topmost ten utmost and lowermost rainfall in States or Union Territories of India, maximum and minimum rainfall within the province of West Rajasthan and Coastal Karnataka on the aspect of the once a year rainfall has additionally been showed in their paper. Result confirmed the inconsequentialdeclining fashion in yearly common rainfall inside the location of Coastal Karnataka and West Rajasthan [25]. In subsequent section we're going to talk about the materials and method that describes the hassle.

III. Materials and Process

Various Data Mining equipment that is existing are weka, Rapid miner, Tanagra and Orange. This document gives the information of information mining device WEKA. For Mining duties, Weka is a tool which is the group of gadget getting to knowonce the installation has been executed, the GUI carries 4 applications including Explorer, Experimenter, information go with the flow and easy CLI.The explorer application has classify panel, preprocessing panel, cluster panel, Association panel, visualization panel and selection attributes panel.Preprocess panel facilitate for importing the facts within the shape of ARFF (attribute relational flat record) and preprocessing the information through normalization. The Classify panel affords the user with the features to use regression system, Gaussian manner, selection regulations and selection tress to the ensuing statistics set and in estimating the accuracy of the predictive versionand in visualizing the statistics in margin curves, threshold curves and value advantage analysis and so on. The Cluster panel offers get right of entry to to the clustering algorithms along with Cobweb, easy okay-manner, hierarchical cluster; Density based totally cluster, and many others. The Associate panel provides Apriori set of rules for frequent item set mining. The Selection panel affords algorithms for insertion of the maximum predictive features in a datasets. Visualize panel indicates scatter plot matrix in which man or woman scatter plots can be decided, enhanced and considered in addition using different preference operators. Next segment deals with the method for the chosen hassle.

IV. Methodology

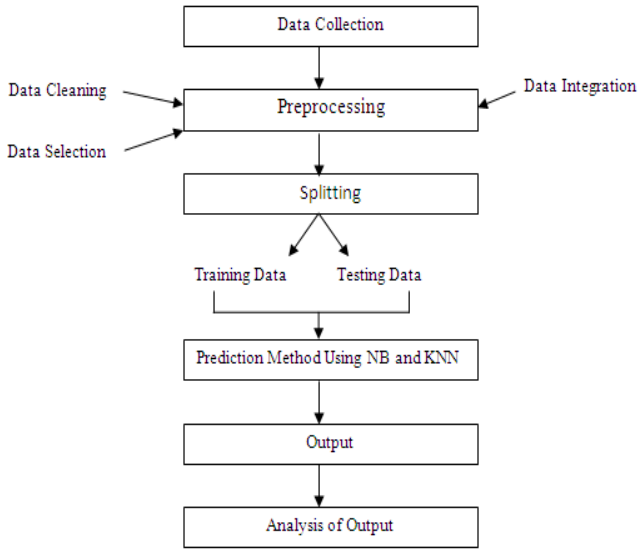

Fig. 1 Flow Diagram that shows the steps involved in the system

The above figure explains the steps involved in the process to predict the groundwater which takes the input data and preprocesses then splits the processed data into training set and test set which then predicts using Naïve Bayes and KNN and finally the output is displayed along with the comparison.

A. Data Collection

The data is collected from CGWB (Central Ground Water Board) which contains numerous attributes and with various districts and timestamps. Many input parameters like Rainfall, Groundwater, Windspeed, Humidity, Temperature, etc. Also the input is obtained with multiple formats like csv, excel, etc. The raw data which helps in analyzing and in predicting that includes various algorithms or classifiers for predictions and comparisons along with the decision making.

B. Preprocessing

The raw data that has been collected contains various attributes and along with various districts also with many file formats. So here, the raw data has been categorized according to timestamp and district wise and finally considering only the required attributes i.e. here only Kolar district has been taken into consideration with the two input parameters Rainfall and Groundwater.

For the Null Analysis, the data that is collected is provided as an input forWEKA in Attribute-Relation File Format (ARFF). File contains training data and the attributes considered are Rainfall and Groundwater with the ranges being defined i.e. high-medium-low. Filter options are to be modified to remove those values having Null. Here, ReplaceMissingWithUserConstant option has been selected which fills the user provided constant in place of missing values. For those records containing duplicate values, can be achieved through distinct records.

C. Splitting

Once the data is processed, the data is been distributed into Training set and Test set i.e. 20% of input records is taken as test set.

Training set is used in fitting the parameters to the model and test set to evaluate the overall performance of the model that is generalization of final chosen model.

D. Prediction Method

After selecting .arff file under classification choose Naive Bayes and select training set option and for the output predictions choose plain text then start. The output shows the instances that have been correctly classified and if any error, even those are displayed. Along with the RMSE, RAE, RRSE, Precision, Recall, etc are displayed then save the model.

Then supply the test set by selecting supplied test set and choose the input (.arff) file and in more options uncheck all the checked boxes and load the saved model which has been previously saved and Re-evaluate the model on test set. The output generated shows the output that is predicted groundwater results for the actual input. Along with that error prediction also been displayed.

E. Output

The output of Naïve Bayes shows the result with the RMSE of 0.0027 along with the precision, recall, F-Measure values and Confusion Matrix.

F. Analysis of Output

In order to compare the efficiency, another classifier has been considered i.e. IBK(K nearest). The RMSE value of KNN is 0.0189 which is slightly greater than Naïve Bayes.

TABLE. I. Experimental Analysis of the Classifiers

Classifier Used	Parameter			
	Root Mean Squared Error	Recall	Precision	F-Measure
Naïve Bayes	0.0027	1.000	1.000	1.000
KNN	0.0189	1.000	1.000	1.000

The above table shows the outcomes of classifiers and is evaluated on the source of RMSE, Recall, Precision, F-Measure. Accuracy of the predicted outcomeand overall performance procedures of implemented prediction models which are based on the input are displayed in above table with the classifiers.
In next section we are going to analyze with the results and discussions.

V. Result and Discussions

For information mining tool such as WEKA, predictions can be achieved for the unknown values for the given input based on the training data. The result obtained from the testing data shows the accuracy of the model i.e. in this paper we have used Naïve Bayes and KNN classifier to predict the groundwater level for the given input. Naïve Bayes generates the output with the RMSE of 0.0027 whereas KNN with the RMSE of 0.0189. RMSE is the degree of dissimilarities amongst values anticipated with the aid of a model and the values certainly found.It is been said that RMSE of 0.7 is small and smaller the RMSE, the better result level can be obtained. Here the results are acceptable and the data is said to be acceptable since the RMSE values are lower than the mentioned and the correlation is expressed through the plot which is been generated according to the given input and the obtained result.
The next section provides the conclusion about the work and the future implementations which can be enhanced with the present work.

VI. Conclusion and Future Work

In present work, we performed an investigational approach to analyze and equateprevalent facts mining classifiers to predict groundwater with the usage of several performance metrics over Rainfall and Groundwater information of Kolar district. The exceptional measuring attributes played a crucial position in determining groundwater prediction. The Naïve Bayes and KNN models and helps in comparing the efficiency or accuracy of the model. The prediction of groundwater through Naïve Bayes has given the better result than the KNN. The RMSE value of Naïve Bayes is 0.0027 and the RMSE value of KNN is 0.0189. According to this, the RMSE of KNN is slightly greater than Naïve Bayes. In present scenario, Naïve Bayes method proves to be an ideal technique for prediction of groundwater. The accuracy level and predicted result distinctly be contingent on the facts getting utilized as an input for prediction and classification. Every single algorithm deals with its very own benefits and boundaries and it is miles hard to pick out the excellent algorithm.By considering input and the outcome, the best approach can be considered. In future, I would like to add the climatic changes due to the effect of rainfall by observing the temperature, humidity and other weather conditions. The following section contains list of references.

VII. References

[1] Wang H, Gao JE, Zhang MJ, Li XH, Zhang SL, Jia LZ. “Effects of rainfall intensity on groundwater recharge based on simulated rainfall experiments and a groundwater flow model”, Catena. 2015 Apr 1;127:80-91.

[2] Parneet Kaur, Manpreet Singh and Gurpreet Singh Josan, “Classification and Prediction Based Data Mining Algorithms to Predict Slow Learners in Education Sector”, 3rd International Conference on Recent Trends in Computing, Procedia Computer Science 57, Elsevier, pp. 500-508, 2015.

[3] Jeerawan saelao, Nopporn Patcharaprakiti, “A Rainfall Trend Analysis for Agriculture Irrigation System Management Case Study: Sansai District - Chiangmai Province”, International Conference on Science and Technology, RMUTT, 2015

[4] Adhatus Solichah Ahmadiyah, "Analysis and Design of Groundwater Quality Monitoring Application", International Conf. on Advanced Mechatronics, Intelligent Manufacture, and Industrial Automation 2015 (ICAMIMIA 2015), Surabaya, Indonesia, on October 15-17, 2015.

[5] Erken Turganbaev, Zhanar Beldeubayeva, Saule Rakhmetullina, Vladimir Krivykh, "Information system of efficient data management of groundwater monitoring the Republic of Kazakhstan", 9th International Conference on Application of Information and Communication Technologies (AICT), 2015.

[6] Christopher Beckham, Mark Hall, Eibe Frank, “WekaPyScript: Classification, Regression, and Filter Schemes for WEKA Implemented in Python” Journal of Open Research Software, 4: e33, 2016.

[7] Rafet Duriqi, Vigan Raca, Betim Cico “Comparative Analysis of Classification Algorithms on Three Different

Datasets using WEKA" , 5th Mediterranean Conference on Embedded Computing, MECO 2016.
[8] Anand Kishor Pandey, Dharmveer Singh Rajpoot, "A comparative study of classification techniques by utilizing WEKA", IEEE 2016.

[9] Satyavati Shukla, M. V. Khire and S. S. Gedam, "REMOTE SENSING AND GIS BASED IMPACT ASSESSMENT OF LAND TRANSFORMATIONS ON GROUND WATER LEVEL OF A RAPIDLY URBANIZING WATERSHED OF INDIA", IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 2016.

[10] Mohamed Y. A. Toure, Kalifa Goïta, Ramata Magagi, and Ally M. Toure, "COMPARISON OF IN SITU AND GRACE ESTIMATED GROUNDWATER IN THE CANADIAN PRAIRIES", IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 2016.

[11] Chandrasegar Thirumalai, M Lakshmi Deepak, K Sri Harsha, K Chaitanya Krishna, "Heuristic Prediction of Rainfall Using Machine Learning Techniques", International Conference on Trends in Electronics and Informatics (ICEI), 2017.

[12] P.Kalaiyarasi, Mrs.A.Kalaiselvi "Data Mining Techniques Using To Weather Prediction" International Journal of Computer Science Trends and Technology (IJCST) – Volume 6 Issue 3, May - June 2018.

[13] Razeef Mohd, Muheet Ahmed ButT, MajidZaman Baba, "Comparative Study of Rainfall Prediction Modeling Techniques" Asian Journal of Computer Science and Technology, ISSN: 2249-0701 Vol.7 No.3, 2018, pp. 13-19.

[14] Aaditya Gupta, Cheshta Bansal, Agha Imran Husain, "Ground Water Quality Monitoring Using Wireless Sensors and Machine Learning, International Conference on Automation and Computational Engineering (ICACE - 2018) Amity University Greater Noida Campus, U. P., India, Oct 3-4, 2018.

[15] Swati H. Mhaskar, Nikhil M. Ingale, Shrikant Charhate, "Application of Soft and Hard Computing Tools for the Estimation of Groundwater Recharge", International Conference on Smart City and Emerging Tecchnology(ICSCET), 2018.

[16] Dodi Devianto, Maiyastri, Uqwatul Alma Wisza, Michelin Wara, Putri Permathasari, Rika Okda Marlina Zen, "Time Series of Rainfall Model with Markov Switching Autoregressive", International Conference on Applied Information Technology and Innovation (ICAITI), 2018.

[17] Alvin Lal, Bithin Datta, "Genetic Programming and Gaussian Process Regression Models for Groundwater Salinity Prediction: Machine Learning for Sustainable Water Resources Management", IEEE Conference on Technologies for Sustainability (SusTech), 2018.

[18] Huang Zhi-fang, "Analysis on the Impact of Expansive Soil Slope Stability by Rainfall Infiltration", International Conference on Engineering Simulation and Intelligent Control, 2018.

[19] Kreshna Gopal, C Prakash Khedun, Anoop Sohun, "Modeling the Influence of Large-Scale Circulation Patterns on Precipitation in Mauritius", International Conference on Intelligent and Innovative Computing Applications (ICONIC), 2018.

[20] Abhishek Sudhakar Jagtap, K. V. N. Kavitha, Ali Dakhel Hussein, "Monitoring of Groundwater level and Development of Control Mechanism based on Machine Learning Algorithm", International Conference on Vision Towards Emerging Trends in Communication and Networking (ViTECoN), 2019.

[21] Meriyam Mhammdi Alaoui, Ilias Kacimi, Driss Ouazar, "A knowledge based framework for groundwater modeling", 5th International Conference on Optimization and Applications (ICOA), 2019.

[22] Sorja Koesuma, Ilham Robiansyah Akbar, Mohtar Yunianto, Sunaryo, Doddy Sutarno, Ari Handono Ramelan, "Preliminary Estimation of Groundwater Level With Geoelectric Method in North Part of Surakarta City",IEEE Asia-Pacific Conference on Geoscience, Electronics and Remote Sensing Technology (AGERS), 2019.

[23] Denisa Pospisilova and Petr Kolar, "Precipitation Characteristics of Brno and Surroundings in the 2015–2016 Period: Case Study of Crowdsourcing in Meteorology", International Conference on Military Technologies (ICMT) 2019 May 30 – 31, Brno, Czech Republic, 2019.

[24] Madhuri Kommineni, K Veniha Reddy, K Jagathi, B Dushyanth Reddy, Roshini A, V Bhavani, "Groundwater Level Prediction Using Modified Linear Regression",6th International Conference on Advanced Computing & Communication Systems (ICACCS),2020.

[25] Yogesh Kumar Joshi, Udit Chawla, Shipra Shukla, "Rainfall Prediction Using Data Visualisation Techniques",10thInternational Conference on Cloud Computing, Data Science and Engineering (Confluence) ,2020.