"GROUND WATER LEVEL PREDICTION USING MACHINE LEARNING APPROACH"

Biwek Basnet
Computer Science Dept.
Cambridge Institute of Technology
Bengaluru, India

Madan Bohara
Computer Science Dept.
Cambridge Institute of Technology
Bengaluru, India

Roshan Sharma
Computer Science Dept.
Cambridge Institute of Technology
Bengaluru, India

Saikat Chowdhury
Computer Science Dept.
Cambridge Institute of Technology
Bengaluru, India

Yashaswini S
Assistant Professor
Computer Science Dept.
Cambridge Institute of Technology
Bengaluru, India

Abstract - Arid and semi-arid regions face major challenges within the management of scarce fresh resources underneath economic development and temperature change.

Groundwater is often the foremost vital water resource in these areas. correct prediction of formation is a necessary element of appropriate water resources management.

Physically primarily based model square measure typically used to perform groundwater simulation and predications. However, they're not applicable in several arid and semi-arid regions because of information limitations. Data-driven strategies have tried their pertinency in modeling advanced and nonlinear hydrological processes. the main focus of this study is that the application and comparison of data-driven models for statement short groundwater levels. the aim is to develop a brand new experimental methodology for extremely correct formation statement that may be accustomed facilitate water managers, engineers, and stake-holders manage groundwater in a very simpler and property manner. a group of standard datadriven models square measure evaluated and compared, as well as Artificial nerve cell Networks (ANNs), Support Vector Machines (SVMs).

Key words: ANNs,SVMs.

1. INTRODUCTION

In several areas, groundwater is commonly one in all the most important sources of installation for domestic, urban, agricultural and industrial functions, particularly in arid and semi-arid areas. However, several issues happens thanks to overuse of and unsustainable groundwater use and management, like major water-level declines, dehydration of wells, water-quality degradation, redoubled pumping prices, land surface subsidence, loss of pumpage in residential installation wells, and formation compaction. These issues have become a significant issue globally, particularly in developing countries. To secure water for the longer term, the property management of groundwater resources in conjunction with surface water has desperately become the necessity of the hour. correct and reliable prediction of groundwater levels may be a crucial part for achieving this goal, particularly in watersheds in arid and semi-arid regions that square measure additional prone to hydrological extreme events within the kind of droughts.

The data-driven models arrange to determine an instantaneous mapping between the inputs ANd outputs of the system while not reaching an understanding of the interior structure of the physical method. once enjoying abundant success in various hydrologic and water setting applications, like rainfall-runoff modeling and water quality statement, datadriven models square measure currently being applied additional and additional to resolve issues within the space of groundwater. samples of the foremost common strategies employed in data-driven modeling of groundwater levels include: artificial neural networks (ANNs), support vector machines(SVMs). Despite the growing applications and successes of data-driven approaches within the surface water issues, there are solely a couple of studies associated with groundwater in arid and semi-arid regions. This provides AN impetus for this work. the main target of this study is that the application and comparison of 3 datadriven models (i.e., ANN and SVM model tree) for statement short groundwater levels. the aim is to develop a replacement empiric methodology of extremely correct water level statement which will be wont to facilitate water managers, engineers, and stake-holders manage groundwater in an exceedingly simpler and property manner. during this analysis, the flexibility and accuracy of the 2 knowledge driven models square measure investigated by applying them to forecast water level.

1.1.1 EXISTING SYSTEM

The correct prediction of groundwater levels is crucial for property utilization and management of significant groundwater resources. The consumption of water will increase each day with the expansion in population. the bottom water level goes down day by day. In India, groundwater serves concerning eightieth of rural population, five hundredth of urban population and concerning hour of agricultural space. For management of water level a model is needed which might predict the water level in future with the present offered info with manual development.

1.1.2 PROPOSED SYSTEM

The most in style ANN design used for regression or prediction is that the Multi-Layer Perceptron (MLP) network. The MLP network features a bedded design that's comprised of associate input layer, followed by a minimum of one hidden layer associated an output layer. A layer usually consists of variety of neurons. Directed synapses connect every somatic cell in one layer to each neurons within the next layer. every colligation is appointed with a weight. The **“**knowledge**”** concerning the info behavior of a coaching set is keep in terms of the synapses’ weights. conjointly the construct of SVMs is introduced. during this work, support vector regression (SVR) was accustomed describe re
gression with SVM.

2. LITERATURE SURVEY

Haritha kagita, Malathi. V, Thiyagarajan. A , **“Prediction of Ground Water Level based on machine Learning”**, **International Research Journal of Engineering and Technology (IRJET), Volume: 07 Issue: 02 | Feb 2020**

In this paper work we perform analysis of groundwater level data from various states. We have analyzed this data for the states and developed seasonal models to represent the groundwater behavior. Three different type of models were developed-periodic, polynomial and rainfall models. While periodic and polynomial models capture trends on water levels in observation wells, the rainfall model explores the link between the rainfall levels and water levels. The periodic and polynomial models are developed only using the water level data of observation wells while the rainfall model also uses the rainfall data. All the data and the models developed with a summary of analysis. The larger aim is to build these models to predict temporal changes in water level to aid local water management decisions and also give region specific input to Government planning authorities e.g. Groundwater Survey and Development Agency to flag water status with more information.

Klemen Kenda, Matej Čerin, Mark Bogataj, Matej Senožetnik, Kristina Klemen, Petra Pergar, Chrysi Laspidou and Dunja Mladenić, “Groundwater Modeling with Machine Learning Techniques” MDPI Proceedings2018

In this study a thorough analysis is conducted concerning the prediction of groundwater levels of Ljubljana polje aquifer. Machine learning methodologies are implemented using strongly correlated physical parameters as input variables. The results show that data-driven modeling approaches can perform sufficiently well in predicting groundwater level changes. Different evaluation metrics confirm and highlight the capability of these models to catch the trend of groundwater level fluctuations. Despite the overall adequate performance, further investigation is needed towards improving their accuracy in order to be comprised in decision making processes.

Sasmita Sahoo & Madan K. Jha, “Groundwater-level prediction using multiple linear regression and artificial neural network techniques: a comparative assessment”, Hydrogeology Journal (2013)

The potential of multiple linear regression (MLR) and artificial neural network (ANN) techniques in predicting transient water levels over a groundwater basin were compared. MLR and ANN modeling was carried out at 17 sites in Japan, considering all significant inputs: rainfall, ambient tempera ture, river stage, 11 seasonal dummy variables, and influential lags of rainfall, ambient temperature, river stage and groundwater level. Seventeen sitespecific ANN models were developed, using multi-layer feed-forward neural networks trained with LevenbergMarquardt backpropagation algorithms. The performance of the models was evaluated using statistical and graphical indicators. Comparison of the goodness-of-fit statistics of the MLR models with those of the ANN models indicated that there is better agreement between the ANN-predicted groundwater levels and the observed groundwater levels at all the sites, compared to the MLR. This finding was supported by the graphical indicators and the residual analysis. Thus, it is concluded that the ANN technique is superior to the MLR technique in predicting spatio-temporal distribution of groundwater levels in a basin. However, considering the practical advantages of the MLR technique, it is recommended as an alternative and cost-effective groundwater modeling tool.

**MAO Xiaomin, SHANG Songhao , LIU Xiang, “Groundwater Level Predictions Using Artificial Neural Networks”, ISINGHUA SCIENCE AND TECHNOLOGY
Volume 7, Number 6? December 2002**

The prediction of groundwater level is important for the use and management of groundwater resources. In this paper, the artificial neural networks (ANN) were used to predict groundwater level in the Dawu Aquifer of Zibo in Eastern China. The first step was an auto-correlation analysis of the groundwater level which showed that the monthly groundwater level was time dependent. An auto-regression type ANN (ARANN) model and a regression-auto-regression type ANN (RARANN) model using back-propagation algorithm were then used to predict the groundwater level. Monthly data from June 1988 to May 1998 was used for the network training and testing. The results show that the RARANN model is more reliable than the ARANN model, especially in the testing period, which indicates that the RARANN model can describe the relationship between the groundwater fluctuation and main factors that currently influence the groundwater level. The results suggest that the model is suitable for predicting groundwater level fluctuations in this area for similar conditions in the future

3. SYSTEM ARCHITECTURE

System architecture is that the conceptual design that defines the structure and behavior of a system. An architecture description may be a formal description of a system, organized during a way that supports reasoning about the structural properties of the system. It defines the system components or building blocks and provides an idea from which products are often procured, and

system developed, which will work together to implement the general system.

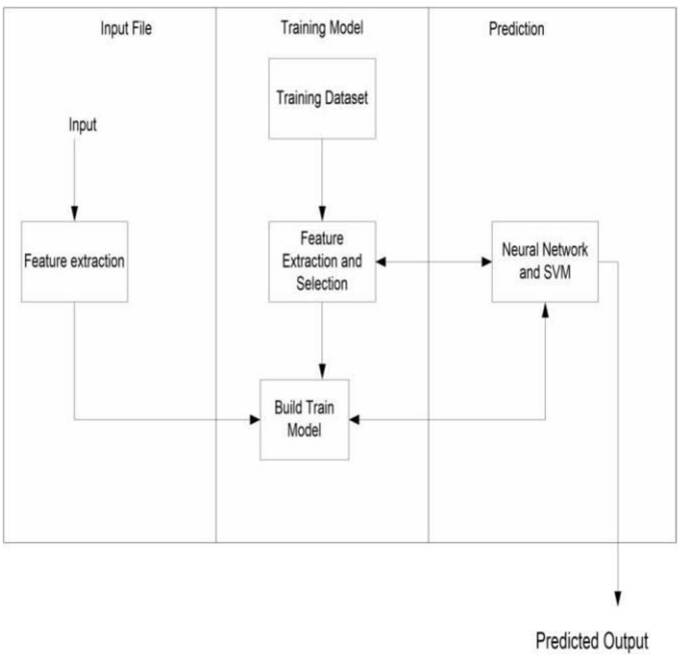


Fig 3.1 Proposed System Architecture

The project has 3 major modules

Data Preprocessing

In this load the dataset, and read the features in the dataset and apply preprocessing, where we do data cleaning and data transformation. This we called as feature extraction and feature selection.

Training

Selected features will be loaded into the ML model to train the data. Trained model gets generated based on the respected ML algorithm

Prediction

We use ML approach, SVM and Neural network to predict the ground water level

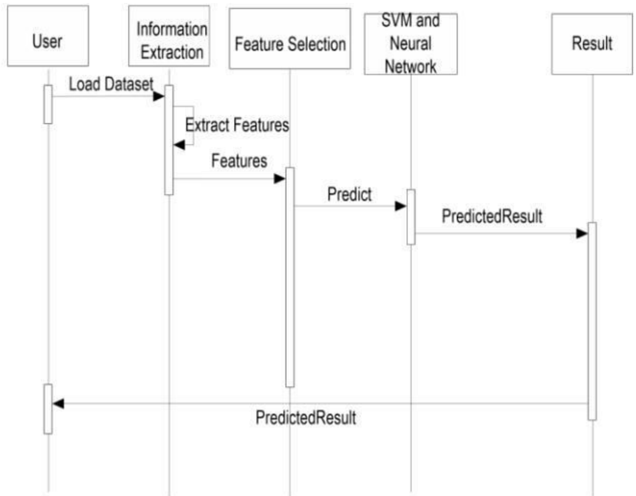


FIG 3.2: SEQUENCE UML DIAGRAM

A sequence diagram in Unified Modeling Language (UML) is a kind of interaction diagram that shows how processes operate with one another and in what order. It is a construct of a Message Sequence Chart

4.METHODLOGY

4.1 SVM algorithm Working

SVM libraries are packed with some popular kernels such as ***Polynomial, Radial Basis Function or rbf***, and ***Sigmoid***. The classification function used in SVM in Machine Learning is SVC. The SVC function looks like this:

sklearn.svm.SVC (C=1.0, kernel= 'rbf', degree=3)

Important parameters

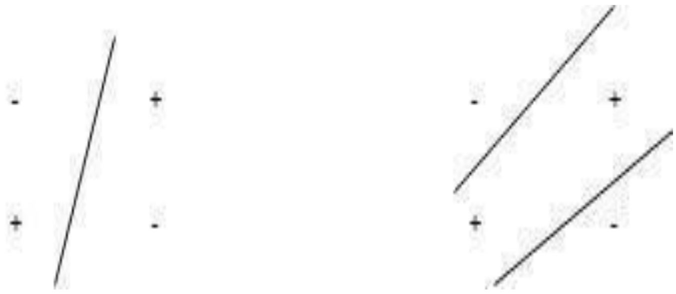
- **C:** Keeping large values of C will indicate the SVM model to choose a smaller margin hyperplane. Small value of C will indicate the SVM model to choose a larger margin hyperplane.
- **kernel:** It is the kernel type to be used in SVM model building. It can be 'linear', 'rbf', 'poly', or 'sigmoid'. The default value of kernel is 'rbf'.
- **degree:** It's only considered in the case of polynomial kernel. It is the degree of the polynomial kernel function. The default value of degree is 3.

Alright, let us dive right into the hands-on of SVM in Python programming language.

4.2 MLP NN algorithm Working

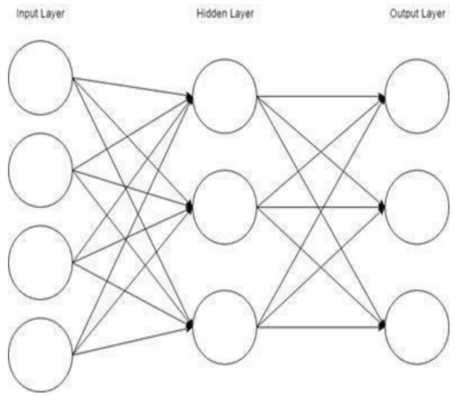
International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 05 Issue: 05 | May - 2021 ISSN: 2582-3930

The perceptron is very useful for classifying data sets that are linearly separable. They encounter serious limitations with data sets that do not conform to this pattern as discovered with the XOR problem. The XOR problem shows that for any classification of four points that there exists a set that are not linearly separable.



The MultiLayer Perceptron (MLPs) breaks this restriction and classifies datasets which are not linearly separable. They do this by using a more robust and complex architecture to learn regression and classification models for difficult datasets.

The Perceptron consists of an input layer and an output layer which are fully connected. MLPs have the same input and output layers but may have multiple hidden layers in between the aforementioned layers, as seen below.



The algorithm for the MLP is as follows:

1. Just as with the perceptron, the inputs are pushed forward through the MLP by taking the dot product of the input with the weights that exist between the input layer and the hidden layer (WH). This dot product yields a value at the hidden layer. We do not push this value forward as we would with a perceptron though.
2. MLPs utilize activation functions at each of their calculated layers. There are many activation functions to discuss: rectified linear units (ReLU), sigmoid function, tanh. Push the calculated output at the current layer through any of these activation functions.
3. Once the calculated output at the hidden layer has been pushed through the activation function, push it to the next layer in the MLP by taking the dot product with the corresponding weights.
4. Repeat steps two and three until the output layer is reached.
5. At the output layer, the calculations will either be used for a backpropagation algorithm that corresponds to the activation function that was selected for the MLP (in the case of training) or a decision will be made based on the output (in the case of testing).

MLPs form the basis for all neural networks and have greatly improved the power of computers when applied to classification and regression problems. Computers are no longer limited by XOR cases and can learn rich and complex models thanks to the multilayer perceptron

5. RESULT

The following snapshots define the results or outputs that we'll get after step-by-step execution of all the modules of the system

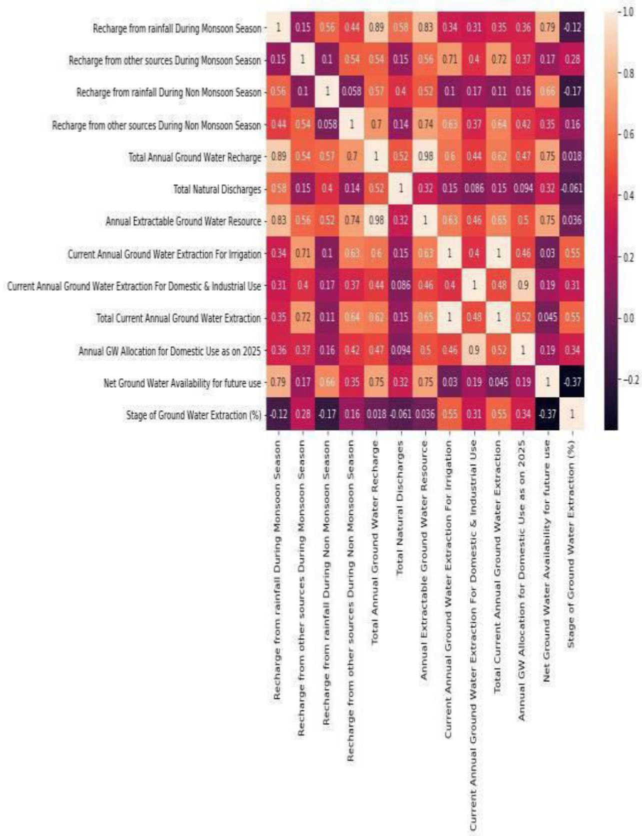


Fig: **Heatmap Graph for Input Data**

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 05 Issue: 05 | May - 2021 ISSN: 2582-3930

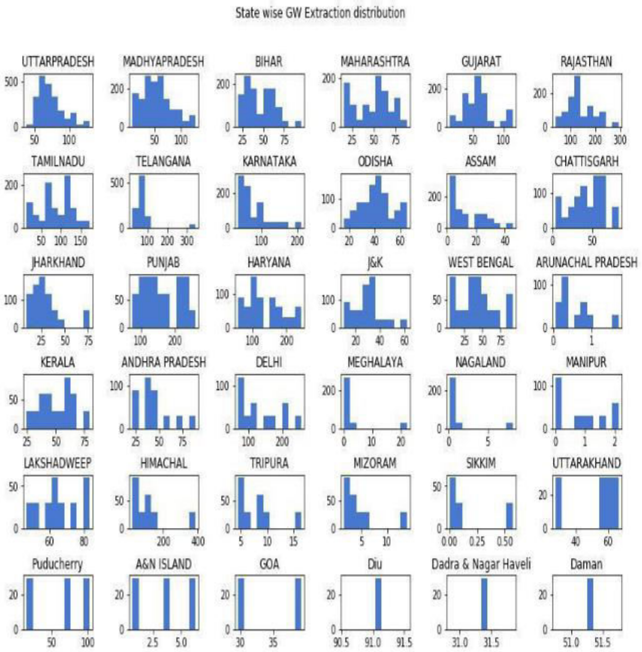
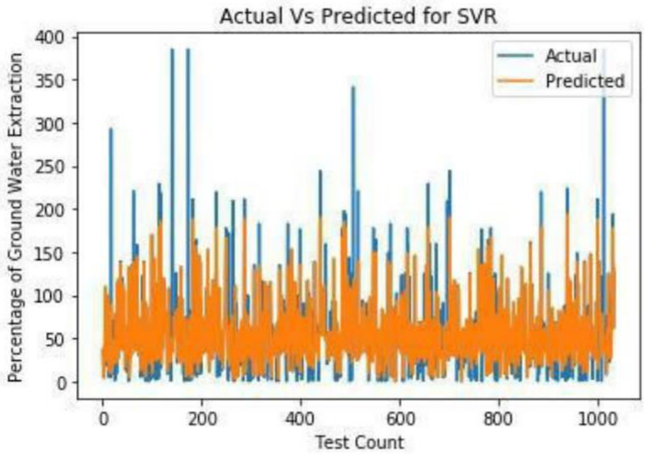


Fig: **Statewise groundwater Extraction**

Following tells the output of SVM Regression Model with Comparision Graph and Performance Meric.

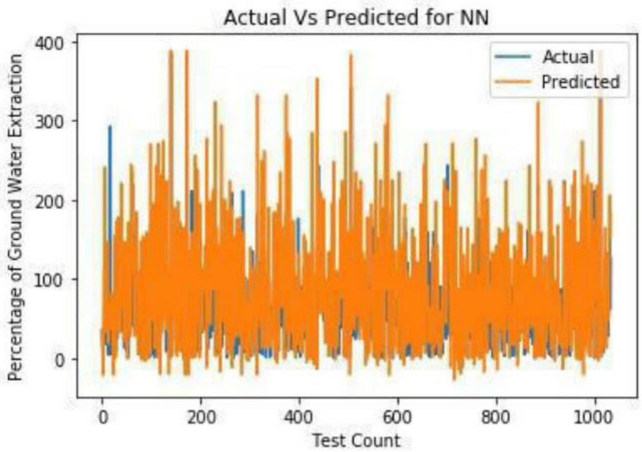
```
actual ouptut
     Stage of Ground Water Extraction (%)
10772              36.51
17795               9.86
18652               4.85
2182                3.91
2072               36.87
                ...
13777             101.99
1841              117.39
490               193.99
19798              61.73
18238             126.63

[1034 rows x 1 columns]
predicted output
[ 35.64584588  17.70681537   4.94968247 ... 178.07040772  61.72687114
 131.22989167]
```



MSE for SVM Regression
738.5329621748415
MAE for SVM Regression
9.123781578471023
RMSE for SVM Regression
27.175962948437384
Accuracy for SVM Regression
0.6823120175087418

Following tells the output of MLP Regression Model with Comparison Graph and Performance Metric.

```
actual ouptut
     Stage of Ground Water Extraction (%)
10772              36.51
17795               9.86
18652               4.85
2182                3.91
2072               36.87
                ...
13777             101.99
1841              117.39
490               193.99
19798              61.73
```



MSE for NN Regression
3477.9581215704666
MAE for NN Regression
41.6552230999589
RMSE for NN Regression
58.974215735102966
Accuracy for NN Regression
0.4960814959119324

6. CONCLUSION

In this study, we tend to square measure victimisation Prediction of water table amendment proven to be a more robust approach compared to the prediction of absolute groundwater levels. All models tested square measure variable and therefore the predictors inserted are: Name of State, Name of District, Recharge from downfall throughout Monsoon Season, Recharge from different sources throughout Monsoon Season, Recharge from downfall throughout Non Monsoon Season, Recharge from different sources throughout Non Monsoon Season, Total Annual spring water Recharge, Total Natural Discharges, Annual extractible spring water Resource, Current Annual spring water Extraction For Irrigation, Current Annual spring water Extraction For Domestic & Industrial Use, Total Current Annual spring water Extraction, Annual GW Allocation for Domestic Use as on 2025,Net spring water accessibility for future use.

7. REFERENCES

1. Solomatine, D.P.; Ostfeld, A. Data-driven modelling: Some past experiences and new approaches. *J. Hydroinform.* **2008**, *10*, 3–22.

2. Shirmohammadi, B.; Vafakhah, M.; Moosavi, V.; Moghaddamnia, A. Application of Several Data-Driven Techniques for Predicting Groundwater Level. *Water Resour. Manag.* **2013**, *27*, 419–432.

3. Rahmati, O.; Pourghasemi, H.R.; Melesse, A.M. Application of GIS-based data driven random forest and maximum entropy models for groundwater potential mapping: A case study at Mehran Region, Iran. *CATENA* **2016**, *137*, 360–372.

4. Sahoo, S.; Russo, T.A.; Elliott, J.; Foster, I. Machine learning algorithms for modeling groundwater level changes in agricultural regions of the U.S. *Water Resour. Res.* **2017**, *53*, 3878–3895.

5. Rejec Brancelj, I.; Smrekar, A.; Kladnik, D. Podtalnica Ljubljanskega polja. *Geografija Slovenije* **2005**, *10*, 1–222.

6. Vizintin, G.; Souvent, P.; **Veselič,** M.; Curk, B.C. Determination of urban groundwater pollution in alluvial aquifer using linked process models considering urban water cycle. *J. Hydrol.* **2009**, *377*, 261–273.

7. **Janža,** M.; **Meglič,** P.; **Šram,** D. *Numerical Hydrological Model*; Tech. Report Income Water Care EU Life Project; European Commission: Brussels, Belgium, 2011.

8. Auersperger, P.; **Čenčur** Curk, B.; Jamnik, B.A.; Kus, J.; Prestor, J.; Urbanc, J. Dinamika podzemne vode. Podtalnica Ljubljanskega polja, *Geografija Slovenije* **2005**, *10*, 39–61.

9. Vrzel, J.; Ogrinc, N.; **Vižintin,** G. Data preparation for groundwater modelling—Ljubljansko polje aquifer system. *RMZ-M&G* **2015**, *62*, 167–173.

10. Kranjc, M. *Data from National Monitoring*; Technical report Income Water Care EU Life Project; European Commission: Brussels, Belgium, 2011.

11. **Janža,** M. Modelling heterogeneity of Ljubljana polje aquifer using Markov chain and geostatistics. *Geologija* **2009**, *52/2*, 233–240.

12. **Pavlič,** M.U. **Geološko**-Geomehanski Model Zgradbe tal na **Območju** Mesta Ljubljane. Ph.D. Dissertation. University of Ljubljana, Ljubljana, Slovenia, 2016.

13. **Senožetnik,** M.; Herga, Z.; **Šubic,** T.; **Bradeško,** L.; Kenda, K.; Klemen, K.; Pergar, P.; **Mladenić,** D. IoT middleware for water management. In Proceedings of the 3rd EWaS International Conference, Lefkada, Greece, 27–30 June 2018; (accepted).

14. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning*, 2nd ed.; Springer: New York, NY, USA, 2017; pp. 43–100.

15. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32.

16. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* 2000. 29, 1189–1232.