International Conference on Engineering Science and Technology 2019 (ICEST2019) IOP Publishing
IOP Conf. Series: Materials Science and Engineering **807** (2020) 012027 doi:10.1088/1757-899X/807/1/012027

Modelling Uranium in Vicinity of Groundwater Population by Neural Networks of Multilayers Perceptron

Iing Lukman[1], Noor Akma Ibrahim[2], Natalina[3]

[1]Dept. of Accounting, Faculty of Economics, Universitas Malahayati, Bandar Lampung, Indonesia
[2]Dept. of Mathematics, Faculty of Science, Universiti Putra Malaysia, Serdang Selangor D.E. Malaysia
[3]Dept. of Environmental Engineering, Faculty of Engineering, Universitas Malahayati, Bandarlampung, Indonesia

Corresponding e-mail: iing@malahayati.ac.id; deiingofthelukman@gmail.com

Abstract. The existent of Uranium in vicinity of groundwater population can give a threat to the water supplier for human consumption. The objective of the research was to find the most important variables to the existence of the Uranium. This paper shows some modelling process for above matters by applying Neural Networks of Multilayers Perceptron. Data taken from US Department of Energy. Neural Networks used in this study were learning the representation of the model inside the data, and how best it relation with the output variable that we obtained from prediction. The results showed that the training samples was 87 out of 127, and the testing samples was 40 out of 127. The results were not giving indication that a mathematical model obtained. The conclusion was Conductivity becoming the most important variable to the existence of Uranium, which followed by the second importance that was Arsenic, the third importance was Selenium, the fourth important was Total Alkalinity.

1.0 Introduction

The Radionuclide and the plausible human carcinogens were emit by Uranium [1]. Groundwater sources were relied by rural areas residents as their primary drinking water sources, for example in South Carolina around 40% of its residents regularly use it [2], and in the 10 Southeast Asian and Pacific nations [10]. Colorectal, breath, kidney problems, and total cancer incidences may be increase in region with frequent groundwater use and elevated groundwater Uranium [1]. Uranium is one of the dangerous heavy metals due to its toxic and radioactive, and it is required from its distribution in the environmental region [3]. In fact, many types of mathematical model required the modeller to know things about the system that are unfortunately generally impossible to find [4]. The objective of the research was to find the most important variables to the existence of the Uranium.

IOP Conf. Series: Materials Science and Engineering **807** (2020) 012027 doi:10.1088/1757-899X/807/1/012027

2.0 Methodology

The methodology of the Multi-layer Perceptron applied for data analysis. The multilayer perceptron is the most known and most frequently used type of neural network [11]. One perceptron is a linear classifier and an algorithm that classifies input by separating two categories with a straight line. Input is typically a feature vector x multiplied by weights w and added to a bias b so that y = w * x + b. One perceptron produces a single output based on several real-valued inputs by forming a linear combination using its input weights (and sometimes passing the output through a nonlinear activation function). The equation is as follows:

$$y = \varphi(\sum_{i=1}^{n} \omega_i + b) = \varphi(w^T x + b) \tag{1}$$

Where w denotes the vector of weights, x is the vector of inputs, b is the bias and phi is the non-linear activation function. Rosenblatt [6] built a single-layer perceptron. That is, his hardware-algorithm did not include multiple layers, which allow neural networks to model a feature hierarchy.

3.0 Multilayer Perceptron (MLP)

Subsequent work with multilayer perceptron has shown that they are capable of approximating an XOR operator as well as many other non-linear functions. Just as Rosenblatt [6] based on the perceptron on a McCulloch-Pitts neuron, conceived in 1943, so that, perceptron themselves are building blocks that only prove to be useful in such larger functions as multilayer perceptron.

A multilayer perceptron (MLP) is a deep, artificial neural network [7]. It is composed of more than one perceptron. They are composed of an input layer to receive the signal, an output layer that makes a decision or prediction about the input, and in between those two, an arbitrary number of hidden layers that are the true computational engine of the MLP. MLPs with one hidden layer are capable of approximating any continuous function [7].

Multilayer perceptron often applied to supervised learning problems [8]: they train on a set of input-output pairs and learn to model the correlation (or dependencies) between those inputs and outputs. Training involves adjusting the parameters, or the weights and biases, of the model in order to minimize error. The adjusting parameters is like competing risk in the survival analysis, where the predominantly parameter will have it position in the final output [8]. Applying Backpropagation to make those weigh and bias adjustments relative to the error, and its error be measured in a variety of ways, including by root mean squared error (RMSE).

The data source for this research was 127 groundwater samples (cases or row) and 12 measurements (variables) taken from [9] which was collected by Department of Energy of United States of America. Samples data consist of horizon producing codes (P_GC) by 5 types of P_GC namely TPO (Ogallala Formation), TRD (Dockum Formation), POQ (Quartermaster Group), PGWC (Whitehorse and Cloud Chief Group), PGEB (El Reno Group and Blaine Formation). Variables are consist of Uranium, (U), Arsenic (AS), Boron (B), Barium (BA), Molybdenum (MO), Selenium (SE), Vanadium (V), Sulphate (SO4), Total Alkalinity (T_AK), Bicarbonate (BC), Conductivity (CT), and pH. Table 1 show the layout of groundwater data.

Table 1. Layout of Groundwater data.

Sample Number	**Latitude**	**Longitude**	**P_GC**	**U**	**AS**	**B**	**BA**	**MO**	**SE**	**V**	**SO4**	**T_AK**	**BC**	**CT**	**pH**
2201	33.21	101.445	TPO	7.99	17.6	300	150	-4	0.4	100	35	278	157	640	7.60
………	………	………	………	………	………	………	………	………	……...	……...	……...	………	………	………	………	……….
2210	33.164	100.608	PGWC	3.1	4.0	625	750	-4	-0.2	25	-30	175	101	540	7.56

IOP Conf. Series: Materials Science and Engineering **807** (2020) 012027 doi:10.1088/1757-899X/807/1/012027

4.0 Results and Discussion

Data analysed by SPSS Neural Network or Multiple Perceptron give as follows: it seems that the real value and its predicted value had big different or high deviation. The different between predicted to the real value was bigger, for example, as seen on the Figure 1, that if the real value was closed to 80, then the predicted value was closed to 30 or the deviation was around 50. The outlier value of the real data between 120 and 140 did not give predicted value at all, thus this Multiple Perceptron was not that fully suited for this data. Training a multilayer perceptron is often quite slow, requiring thousands or tens of thousands of epochs for complex problems [11].

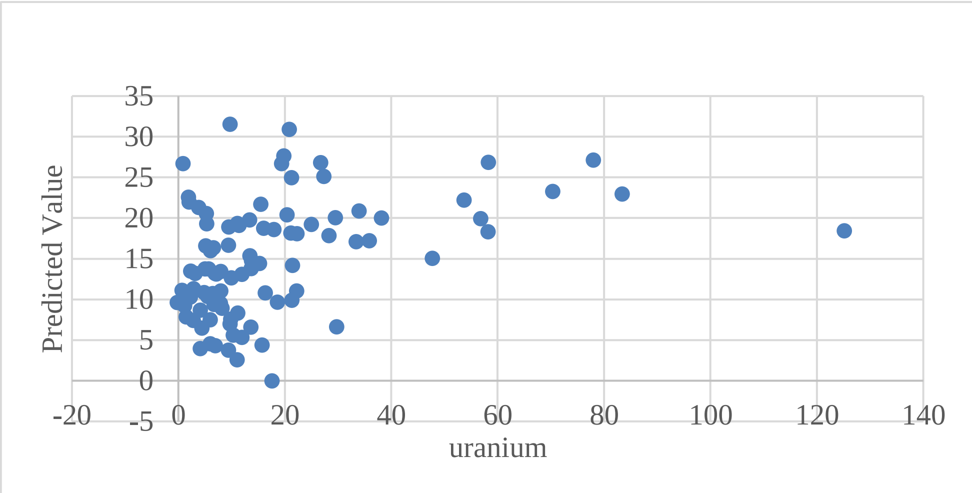

Figure 1. Uranium real and its predicted value

Figure 2 described that the residual of predicted value of Uranium from normalised value of 0 to 15 were mostly around 0 which was good for the prediction. However, some of the residual value of normalised value of 17 to 30 were generally far from zero. Therefore, generally the residual of the predicted value not closed to zero, which then giving problem to the precision of the model.

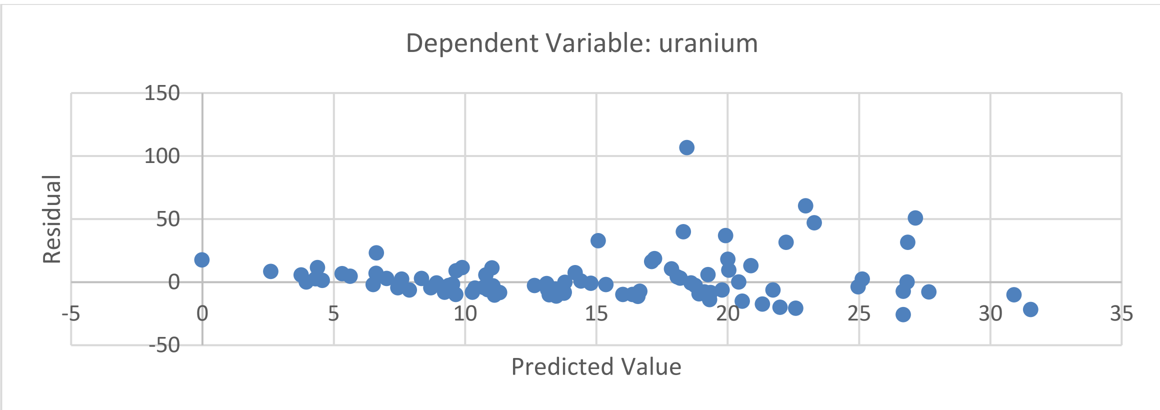

Figure 2. Uranium: it predicted and residual value

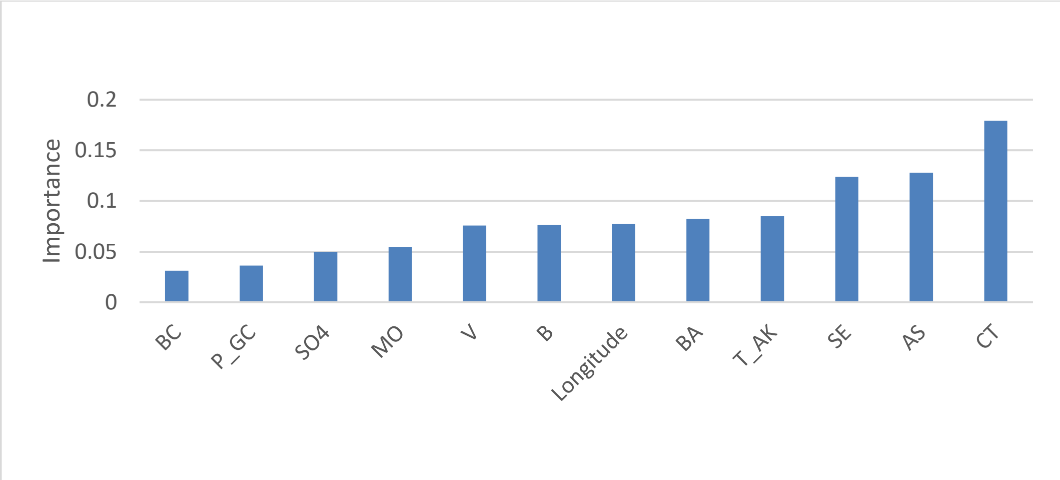

Figure 3. Uranium was a dependent variable against the rest of the variables in the observation

Figure 3 showed that for the dependent variable Uranium, the most independent variable was CT (0.179) followed by AS (0.128) , SE (0.124), T_AK (0.085), BA (0.082), Longitude (0.078), Boron (0.76), MO (0.055), SO4 (.050), horizon (.036). Table 2 showed the importance of the variables and its normalized scale.

Table 2. Independent Variable Importance

	Importance	Normalized Importance
horizon	.036	20.3%
Longitude	.078	43.3%
arsenic	.128	71.5%
boron	.076	42.6%
BA	.082	46.0%
MO	.055	30.4%
SE	.124	69.3%
V	.076	42.3%
SO4	.050	27.9%
T_AK	.085	47.5%
BC	.031	17.3%
CT	.179	100.0%

As seen on Table 2 that CT (conductivity) was the most important variable to the existence of the Uranium followed by Arsenic, Selenium, Total Alkalinity etc. This was in accordance with Hin [12].Then the least important is Bicarbonate. The computation of percentage of importance is as follows: the scale of CT is the biggest one then label it as a hundred percent importance. Afterwards, the normalized scale lower than that then be compared to normalized scale of CT which then give the results as seen on Table 2.

Table 3. Case Processing Summary

		N	Percentage (%)
Sample	Training	87	68
	Testing	40	31
Valid		127	100
Excluded		0	
Total		127	

Table 3 showed that from the total samples of 127, 87 taken for training, and 40 taken for testing. All the 127 samples are valid. Table 4 showed that the horizon producing codes P_GC as a factor, and the rest of the variables as covariates except longitude and latitude.

Table 4. Network Information

Input Layer	Factors	1	P_GC
	Covariates	1	AS
		2	B
		3	BA
		4	MO
		5	SE
		6	V
		7	SO4
		8	T_AK
		9	BC
		10	CT
		11	PH
	Number of Units[a]		16
	Rescaling Method for Covariates		Standardized
Hidden Layer(s)	Number of Hidden Layers		1
	Number of Units in Hidden Layer 1[a]		4
	Activation Function		Hyperbolic tangent
Output Layer	Dependent Variables	1	U
	Number of Units		1
	Rescaling Method for Scale Dependents		Standardized
	Activation Function		Identity
	Error Function		Sum of Squares

a. Excluding the bias unit

Table 5. Model Summary

Training	Sum of Squares Error	33.435
	Relative Error	.778
	Stopping Rule Used	1 consecutive step(s) with no decrease in error[a]
	Training Time	0:00:00,07
Testing	Sum of Squares Error	27.543
	Relative Error	.513

Dependent Variable: U

a. Error computations were based on the testing sample.

Table 5 showed the sum squares error of the training and the testing samples, which suggested that the testing sample has less sum square error than that of training one, and it relative error was smaller than that of training one.

5.0 Conclusion

CT or conductivity was the most important variable to the existence of the Uranium followed by Arsenic, Selenium, Total Alkalinity. The least important was Bicarbonate.

Acknowledgments

Authors wishing to acknowledge assistance or encouragement from colleagues to name a few Rina Febrina, Devi Oktarina, Dewi Fadila Sari, Emy Khikmawati,Weka Indra Darmawan, Marcellya Wijaya, Heri Wibowo Kristianto, and Rani from Engineering Faculty of Universitas Malahayati, Bandarlampung.

References

[1] ATSDR 1999 *Toxicological profile for uranium. US Department of Health and Human Srvices, Public Health Service; Atlanta: Agency for Toxic Substances, Disease Registry*
[2] SCGWA 2009 South Carolina Groundwater Association http://scgwa.org/pages/industrynews/sconomy.html. Accessed on 20 July 2018
[3] Nichols, C.E., Kane V.E., Browning, M.T. and Cagle, G. W. (1976). Northwest Texas Pilot Geochemical Survey, Union Carbide, Nuclear Division Technical Report.
[4] Farlow, S. J. 1984 GMDH Algorithm *Self-Organizing Method in Modeling: GMDH Type Algorithm Stanley J. Farlow, Editor pp 1-24*
[6] Rosenblatt, F 1958 *The Perceptron: A Probabilistic Model for information Storage and Organization in the Brain*
[7] Minsky, M and Papert, S 1988 *Perceptrons: An Introduction to Computational Geometry The MIT Press Cambridge*
[8] Lukman, I, Noor Akma Ibrahim, Isa Bin Daud, Fauziah Maarof, and Mohd Nasir Hassan 2002 Advantage of the modified Lunn-Mcneil technique over Kalbfleisch- Prentics technique in competing risks *Proceedings of SPIE Vol.4730 (2002) Data Mining and Knowledge Discovery: Theory, Tools, and Technology IV, Belur V. Dasarathy, Editor. Pp 203-211*
[9] Andrews, D.F., and Herzberg, A.M. 1985 *Data: A Collection of Problems from Many Fields for the Student and Research Worker. Springer-Verlag.*
[10] Carrard, N., Foster, T., Willetts, J. 2019 Groundwater as a Source of Drinking Water in Southeast Asian and the Pacific: A Multi-Country Review of Current Reliance and Resource Concerns *Water 2019, 11(8), 1605; https:doi.org/10.3390/W11081605, Accessed 15th January 2020*
[11] Popescu, M., Balas, V. E., Popescu, L. P., Mastorakis, N. 2009 Multilayer Perceptron and Neural Networks *WSEAS Transactions on Circuits and Systems*
[12] Hin, Celine 2018 Thermal Conductivity of Metallic Uranium. DOI: 10.2172/1433931