Abstract

Water is one of the main sources of life on Earth. As a result of the progress made in the field of industry and technology, water has become one of the most important wealth that must be preserved. Studies indicate that the world is heading toward a crisis in the percentage of available water by the year 2025 as a result of the scarcity of water sources, the increase in pollution rates, and the increased use of water. On the other hand, water refining is a very expensive method. Therefore, it was necessary to go to computer methods characterized by high accuracy to know the percentage of water quality scale and the possibility of using it in different places other than drinking before resorting to the refining process. This paper presents a model for predicting a water quality scale based on twelve concentrations called (IM^{12}CP-WQI) that is based on the use of the concept of intelligent data mining that combines the construction of two algorithms, namely (DWM-Bat and DMARS). The DWM-Bat worked to find the number of DMARS models in addition to the weights of each of the concentrations used in this study. The DMARS algorithm has found a mathematical model that combines these concentrations to predict the percentage of water quality. The MARS algorithm was developed by replacing its kernel with four functions: [linear, RBF, sigmoid, and polynomial]. The proposed model consists of four basic stages that included: the first stage is data collection and preliminary treatment to put it within the same ranges, which are [0, 1], as well as finding the correlation between concentrations to find out the direct or inverse correlation between those concentrations and their relationship with the water quality coefficient WQI. The second stage included building an optimization algorithm called DWM-Bat to find the optimal weights for each of the twelve concentrations, as well as the optimal number of M models for DMARS. The third stage included building a mathematical model that combines these concentrations, based on DMARS and benefiting from the results of the previous stage, DWM-Bat. The last stage included evaluating the results that were reached using three types of measurements (R^2, NSE, D) on the basis of which the WQI value was determined based on four cases. The first case if the WQI value is less than 25, it can be used for the purpose of drinking, the second case if it was between (26–50) and it is used in fish lakes, the third case if it was between (51–75) and it could be used in agriculture, the fourth case if the WQI value is higher than 75 and then the water needs a refining process. Also, the results of the proposed model called (IM^{12}CP-WQI) were compared with the results of MARS after it was developed by using different kernel functions. By applying the proposed model, it was found using DWM-Bat that the optimal number of M related to the winter and summer data sets is 9. And the best weight for each concentration was as follows: PH = 0.247, NTU = 0.420, TDS = 0.004, Ca = 0.028, Mg = 0.042, Cl = 0.008, Na = 0.011,

K = 0.175, SO_4 = 0.008, NO_3 = 0.042, $CaCO_3$ (TA) = 0.011, and $CaCO_3$ (TH) = 0.004. On the other hand, the study demonstrated a high correlation between WQI, and the following concentrations are k = 0.985, TH = 0.86, NO_3 = 0.761, TDS = 0.55, Na = 0.415, PH = 0.371, TA = 0.37, Cl = 0.362, and Ca = 0.317. The results showed that the predictor IM^{12}CP-WQI is a good indicator compared with other techniques represented by MARS-linear, MARS-Sig, MARS-RBF, and MARS-Poly. Thus, the proposed model IM^{12}CP-WQI is considered one of the most promising techniques in the field of water quality measurement despite the different concentrations that cause water pollution .