

ASSESSMENT OF THE WATER QUALITY IN NORTHERN AREA OF SRI LANKA BASED ON THE STATISTICAL TECHNIQUES

D.M.T.K. Dasanayaka^{1*}, H.T.K. Abeyesundara¹ and R. Chandrajith²

¹*Department of Statistics and Computer Science, University of Peradeniya, Peradeniya, Sri Lanka.*

²*Department of Geology, University of Peradeniya, Peradeniya, Sri Lanka.*

*tharindukavinda1999@gmail.com

Groundwater contamination is a serious environmental concern, particularly in urban and coastal areas. This study provides a comprehensive assessment of water quality issues in the northern part of Sri Lanka, with a focus on the critical dependence on groundwater resources. A total of 258 water samples were analysed for 18 water quality parameters, including chemical constituents (cations and anions) and physical properties such as pH. The aim was to identify regions more suitable for habitation based on groundwater quality. Multivariate statistical techniques and machine learning algorithms were applied to analyse spatial variations in water quality. The Random Forest algorithm was employed to identify the most significant parameters contributing to water quality variations across seven regions. Model performance was improved using 800 trees, with evaluation based on sensitivity and specificity metrics. Variable importance was measured using the Mean Decrease Gini Index. Cluster analysis was then performed on the significant parameters to identify groups with similar patterns. *K*-means clustering revealed four groups ($K = 4$), with the Elbow Method used for selection, and a silhouette score of 0.3 obtained. The Water Quality Index (WQI) was calculated for each water sample and compared with WHO guidelines to assess suitability for drinking. Clusters were categorised as normal, good, medium-good, and poor-quality water. Results indicated that Vavuniya, Kalpitiya, and Jaffna had a higher proportion of poor-quality water samples, whereas Mannar, Mullaitivu, and Jaffna-Mullaitivu showed more balanced distributions. These findings highlight the need for continuous groundwater monitoring in Northern Sri Lanka to ensure safe water use.

Keywords: Cluster analysis, Ground water quality, Machine learning algorithm, Multivariate statistical techniques, Random forest