

MACHINE LEARNING APPROACH FOR FLOOD PREDICTION IN RATNAPURA, SRI LANKA

Chathurangi K.A.A.^{1*}, Chathuranga L.L.G.¹ and Rathnayake R.M.K.T.²

¹Department of Computing and Information Systems, Faculty of Applied Sciences, Sabaragamuwa University of Sri Lanka, Sri Lanka

²Department of Physical Sciences and Technology, Faculty of Applied Sciences, Sabaragamuwa University of Sri Lanka, Sri Lanka

*kaayesha@appsc.sab.ac.lk

Most countries in the world are facing natural disasters such as floods, hurricanes, landslides and droughts. Floods significantly result in destruction of human life, the economy and property. Damages cause by floods can be minimize from a proper pre-warning prior to the flood. In this study, propose a new flood forecasting model for Ratnapura town in Sri Lanka. The main reason for the flood in Ratnapura town is Kalu River. According to previous flood records and geographical maps, Department of meteorology in Sri Lanka monitored that rainfall of five major areas are affecting to the water level of the Kalu River, which affected to floods in Ratnapura town. Such as Alupola station, Hapugasthenna station, Guruluwana station, Lelopitiya station and Ratnapura station. Dataset for this study was collected from the Department of Irrigation, Sri Lanka. The methodology of this study is divided into two main phases. In the first phase, K-mean clustering is used to cluster the water level of the Kalu River according to the rainfall of five meteorological stations. Distortion curve method was used to identify optimal initial number of clusters. Core relation analysis with heat map was used to feature selection. Results of both clustering and feature selection confirmed that rainfall of these five-stations is important to predict water level of Kalu river. In the second phase, the Artificial Neural Network (ANN) model is successfully implemented for forecasting flood in Ratnapura town according to the rainfall of above mentioned five stations. Two models accuracy standards were employed. Such as Mean Absolute Error (MAE) and Mean Squared Error (MSE). Model accuracy for training dataset and testing dataset is 97% and 96% respectively while the consecutive percentage of MAE and MSE of this model is 2.93% and 2.07%. When comparing other machine learning algorithms namely Logistic Regression, Naive Bayes, Decision Tree, Support Vector Machine, Random Forest and Liner Regression; the ANN model gives highest accuracy in the case of flood prediction. So, this proposed model will help to reduce social and economic losses caused by the flood.

Keywords: *Flood prediction, Artificial neural network, K-mean clustering*