A comparison of machine learning techniques for predicting downstream acid mine drainage

Terence L van Zyl

EOSIT, Meraka Institute, CSIR, Pretoria, South Africa

Abstract

We consider the challenge of providing downstream predictions of water quality using a time-series of upstream insitu measurements and a time-series of remote sensed precipitation data from the Tropical Rainfall Measuring Mission (TRMM). We use a windowing approach over historical values to generate a prediction for the current value. We evaluate a number of Machine Learning techniques as regressors including Support Vector Regression, Random Forests, Stochastic Gradient Decent Regression, Linear Regression, Ridge Regression and Gaussian Processes. We show that overall we are able to attain R2 values above 0:80 (0:16) for most target variables and that Random Forests are the most effective Machine Learning technique in this predictive task.