Artificial Intelligence and machine learning model for spatial and temporal prediction of drought in the Colombia Caribbean region.

Daissy Milenys Herrera Posada* - dmherrerap@unal.edu.co

Edier Aristizábal* - evaristizabalg@unal.edu.co

*National university of Colombia.

Abstract

Drought is one of the most critical hydrometeorological phenomena in terms of impacts to society because it affects soil water content, and consequently, crop production and human diets, in some cases under critical conditions, drought produces starving and people migration. Although Colombia is a tropical country, there are areas of the territory that have periods of drought that cause important economic damages such as fires, death loss in cattle, reduction of the capacity to supply water to persons, impacts to agriculture and fish farming.

Due to recent advances in terms of spatial and temporal resolutions of remote sensing and Artificial Intelligence techniques, it is possible to develop Automatic Learning Models supported on historic information. In this research was built a classifier Random Forest (RF) and Bagged Decision Tree Classifier (DTC) model to predict, spatial and temporal drought occurrence in Colombia, using remote sensing data as land surface temperature, precipitation, soil water content, and evapotranspiration, and macro climatic variables information as ONI, MEI and SOI. It was used the Standardized Precipitation Index (SPI) with 3-month time scale, that allows identifying agricultural drought events. The results showed that Random Forest provides the best outcomes. In terms of recall and precision, RF produced 0.84 and 0.59 and DTC brought a 0.8 and 0.33, respectively, to predict drought. The above, evidence that models could overestimate the number of times where drought occurs, in contrast with normal or humid conditions. On the other hand, False Positive and False Negative rates are important facts for measuring the development of models. In this case, the FP and FN rates are 7.5% and 2% for RF and 21% and 2.5% for DTC respectively, that means that both models made fewer mistakes predicted real drought events, but had more errors forecasting real normal or humid condition, especially, DTC model. RF can provide a better performance predicting drought and normal/humid conditions in contrast with DTC. The implementation of the developed model can allow governmental entities assessment and monitor agricultural drought over time. Taking, in consequence, actions to mitigate the impacts of droughts in their territories.

The results showed that Random Forest provides the best outcomes. Both models made fewer mistakes predicted real drought events, but DTC model had more errors forecasting real normal or humid condition.