Random Forest To Predict Dengue Cases and Outbreaks

Manage, Sanuja (School: College Park High School)

With over 400 million annual infections ranging over the Americas, Asia, and Africa, dengue fever is classified as the disease most frequently transmitted by mosquitoes. Proper medication for the treatment of dengue fever is unavailable, thus the prevention of dengue fever is the only viable option to protect from infection. This study implemented tree-based machine learning techniques to predict dengue outbreaks in Sri Lanka using weather factors such as rainfall, temperature, humidity, and wind speed. Specifically, a classification tree and the ensemble learning method of random forest were used. The statistical software R was used to construct these classification models built on data from the District of Colombo, Sri Lanka. As expected, the classification tree largely suffered from the issue of overfitting where the model was fit well for the training set but did not generalize well for the testing set. To overcome this issue, a random forest was constructed based on the same weather factors. The random forest's predictions accuracies of 100% and 74% for the training and testing sets, respectively, are strongly indicative that the random forest can be effectively used as a classification technique to predict dengue outbreaks. The findings of this study will help health officials prepare for potential outbreaks in vulnerable regions, minimizing the severity and fatalities of future dengue epidemics.

Awards Won:
Mu Alpha Theta, National High School and Two-Year College Mathematics Honor Society: Second Award of $1,000
Air Force Research Laboratory on behalf of the United States Air Force: Glass trophy and USAF medal for each recipient
Air Force Research Laboratory on behalf of the United States Air Force: First Award of $750 in each Regeneron ISEF Category