When a Plant Disease Epidemiologist works with data science

Sith Jaisong

Plant Disease Management Group, CESD, IRRI

Los Baños, Philippines

s.jaisong@irri.org

Data science is the science dealing data (large dataset) with statistics and computer science. Data generally are raw and need to be cleaned in order to extract meaning from them or solve the problems. There are many methods in data processing that botanic epidemiology can borrow from data science. Epidemiologists often deal with more than one factor possibly causing or contributing to plant disease. Building the models to capture the those relationships and prediction is a useful approach. Temperature, moisture, soil pH, soil type, plant varieties, crop density are some examples of variables considering the causes of plant disease development. When this number of variables are added for a study, the dataset becomes massive. However, handling and analysis the large data set combining with various variables indeed is challenge. Realistically, handling and analyzing the large data set requires the computer's capability to process the computations. Statistics is on these processes also. It served to determine the causal effect from data, fit the model, for example. The data science process thus server to open new opportunities for future approaches to studies of botanical epidemiology.