What data science can offer a botanic epidemiologist

Sith Jaisong
Plant Disease Management Group, CESD, IRRI
Los Baños, Philippines
s.jaisong@irri.org

Data science is the science of dealing with data to extract meaning and data product by combining statistics and computer science. Data often are raw and need to be cleaned in order to extract meaning from them or solve 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. Temperature, moisture, soil pH, soil type, plant varieties, crop density are some examples of variables that contribute the causes of and are commonly used to explain why plant disease develops. Because data collection is easier now than ever, the amount of data that we can record and use is larger than ever. This means that the handling and analysis of these large data sets is challenging. Realistically, a large data set requires a powerful computer and a good knowledge of statistics for a meaningful analysis. It served to determine the causal effect from data, fit the model, for example. Data science processes thus serve to open new opportunities for future approaches to studies of botanic epidemiology.