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. Scientists in many disciplines are paying attention to massive data sets which are dramatically generated, accumulated and stored via the technology such as high-throughput genomic sequence, smart phone etc. In order to handling these data, computer science is very helpful. Additionally, statistics will be applied for making the decision. Whereas a traditional data analyst may look only at data from a single source (a set of experiment), data science will enable us explore and examine data from multiple disparate sources. As a plant epidemiologist, we struggle to search for the causes of disease epidemics. To investigate the causation, we analyze the data by combining the record data of temperature and relative humidity from difference locations with the disease epidemic. Practically, data analysis process involves data manipulation, sometimes called wrangling or munging, which is the part of computer science and then may come up with the modeling. Because either the data or the tools to analyze didn't exist before, the technique of data science potentially enable plant epidemiologists discovering a previous hidden insight from many angles.