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 adopting the techniques and theories from broad areas of mathematics, statistics and computer science for extracting information from data, which are large and complex. Scientists in many disciplines are paying attention to massive data sets because the data that are generated simultaneously by devices in realtime from our living can reflect our behavior. For example, social scientists used data from user of social network media such as Facebook, Twitter, etc. to understand online social relations, but they are facing challenges with handling and processing big data. In order to manipulate these data, data science is very helpful. Additionally, in part of data analysis, it involves a range of tools to perform predictive analytics including statistical modeling, classification and regression. Whereas a traditional data analyst may look only at data from a single source (a set of experiments), data science will enable us explore and examine data from multiple disparate sources. As a botanic epidemiologist, we struggle to search for the causes of plant disease epidemics. To investigate the causation, we analyze the data by combining the climatic data (e.g., temperature, relative humidity) and relevant variables from several areas where plant diseases spread with the plant disease epidemiological data. This process will create more data, and consequently they may lead to new extensive result. Because either the data or the tools to analyze the data didn't exist before, data science potentially give botanic epidemiologists the opportunities to discover a previous hidden insight from many angles.