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 techniques and theories from broad areas of mathematics, statistics and computer science for extracting information from large and complex data. 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 reflect our behavior. For example, social scientists used data from users of social network media such as Facebook and Twitter to understand online social relations. However, they are facing challenges handling and processing this data due to the amount of data. In order to manipulate these data, data science is very helpful. Data science is in part data analysis, which 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 enables us explore and examine data from multiple disparate sources. As botanic epidemiologists, we explore the causes of plant disease epidemics. To investigate the causes, we analyze the data by combining weather or climatic data (e.g., temperature, relative humidity) and relevant variables that we think can be used to explain why disease occurs. This process creates more data, and consequently may lead to a new extensive result. Because either the data or the tools to analyze the data didn't exist before, data science potentially gives botanic epidemiologists the opportunities to discover previously hidden insights into why diseases occur. In this seminar, Sith will present the how scientists apply the ideas and the process of data science to their research, and will give some examples of the results from research that data science involved. Finally, Sith will show the possible ways that data science can involve in plant disease epidemiological research.