A Look at Damian Mingle at WPC Healthcare Data Scientists Become the New Epidemiologists

Nov 13, 2015 | Posted by Sarianne Gruber | Healthcare Delivery, HITA Feed

In a recent conversation with WPC Healthcare's Chief Data Scientist, Damian Mingle shared with me how he developed an algorithm for detecting West Nile disease, which will be presented in Sri Lanka at the Global Public Health Conference in December 2015. I have to say after hearing about his work, it made me think about our modern technologically advanced world, and how data science is being applied to the tenets of classical epidemiological research. The founding of epidemiology starts with the story of John Snow, a British physician, who ended the cholera epidemic in London, England in 1854. At the time, the theory for contracting disease was thought to be spread via "miasma" in the air. Snow's hypothesis was that the cholera reproduced in the human body and was contracted through contaminated water.   In the Soho district, not far from his home, he mapped 13 public wells and noted the clustering of many cases around one pump. Snow collected his data himself, going from door to door to find out whether cholera stricken households had collected their drinking water from this particular pump. Viewing the well water samples under the microscope, he confirmed the presence of an unknown bacterium. As the story goes, the handle was removed from the Broad Street pump and the outbreak subsided.

Following in the footsteps of the 'father of public health', Mingle's team took on project on to fight West Nile disease locally for an international crisis. The WPC data science team built an algorithm to predict the presence of West Nile disease. This disease is transmitted by infected mosquitoes that had previously fed on infected birds and animals. Mingle shared how excited they were to work on the international health problem for West Nile virus, with a goal of trying to predict outbreaks.

DAMIAN MINGLE
Chief Data Scientist
WPC Healthcare

The Chicago Department of Public Health provided the WPC team with data, and to that data set they were able to aggregate quite a bit of interesting information.   They had results from the mosquito traps that the city set up in suspect areas. Mosquitoes would be caught in these traps and then sent to the lab to be tested for the presence of West Nile. During the same study time period, they would also track weather data. They examined how the city chose to spray to kill for mosquitoes as well as the patterns of spraying. With the availability of geospatial data, trap locations were mapped as to how close in proximity they may be to high risk areas like water or low trafficked areas.

Mingle described the data collection as a very interesting process. He noted that they really handled

the West Nile problem much like they do at WPC Healthcare.   Mingle explained, "We just went through our four step process. We try to describe the data to ourselves so we can start to understand elements. We explore the data to try to understand a little bit about the outliers and see what's present in the data. And in many cases, just like with health care data we end up having to do some inference modeling. We might have to try to fill in the gaps where (for whatever reason) the data is not present but we think it should be present."

I was curious about his imputation method for missing data. He informed me that gaps in the data or 'missingness' are based on an inference that data should be present. He described that an imputation algorithm could be derived from averages, median or mean, or some other Bayesian technique that they might explore. Mingle credits the success of their West Nile algorithm on having created multiple data sets from the original dataset with different features sets.   Mingle explained that they wanted to essentially understand the variation like you would typically do in a data science problem. They tried to group those instances together with similar mosquito trap information, with mosquito species and then, the particular date and time. They created several new features that they believed would express the information better for the learning algorithm. After that, they blended all the results from various models, and they were able to increase the accuracy over to 80% – which was pretty significant.

Mingle commented on the current prevention methods for West Nile. He expressed, "In terms of what's going on right now, is kind of randomized, most cities like Chicago and around the world, would deal with this problem in a retro or reactive way. If somebody sees a bunch of mosquitoes around a place of residence, they may call in and schedule some sort of spray. This would make that person happy but it not define whether or not the mosquito had any kind of West Nile. Not every mosquito has West Nile virus. It is a very shot gun approach that is being exhibited in most cities around the world today compared to this [our] more laser focused approach."

I asked Mingle if the city of Chicago was able to take a more proactive approach with these algorithms. He replied, "I don't really know. All I can do is come up with a solution and try to make it as useable as possible. We certainly hope that we did this work in a way that might benefit the world. We are speaking about this model in Sri Lanka coming up in December for the World Health Symposium. We are putting it out there in the public domain for those who do want to use it." Mingle believes his role is to experiment large data sets, constantly come up with techniques that attempt to communicate findings in understandable, actionable ways. Perhaps that is what Dr. John Snow felt about collecting water samples to fight a disease, a few blocks from his home, that eventually had a worldwide, timeless impact.

To discover more about Damian Mingle, Chief Data Scientist visit [WPC Healthcare](#).