



WHITETREE
ANALYTICS | P3M | TRAINING



Management
Services



Business
Winning



Decision
Support



Training and
Human Capability

Using Machine Learning in Industry, Healthcare and Defence

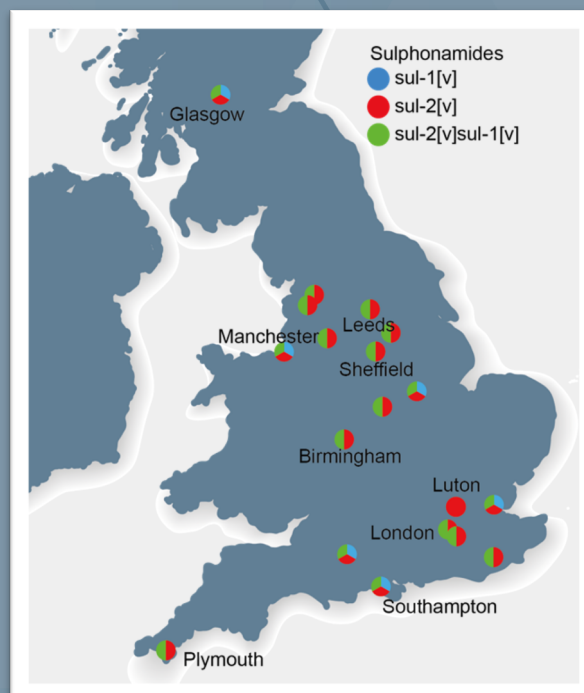
“Machine Learning - the use and development of computer systems that can learn and adapt without following explicit instructions, by using algorithms and statistical models to analyse and draw inferences from patterns in data.”

This article is for anyone interested in reading more into the crossover of science and data, and rejoice, it has nothing to do with Coronavirus!

Instead, we will be focusing on another bacteria, *Shigella sonnei* - which might sound like a spell from Harry Potter - but is a bacterium that causes the intestinal infection, Shigellosis, a leading cause of death in the developing world.

We recently completed a research project looking at antimicrobial resistance genes in *Shigella sonnei* (think specific blocks of Lego in a big Lego model of a squiggly bacteria, that give it resistance to antibiotics like Penicillin) and how we can use data analysis techniques, such as machine learning, data transformation and data visualisations in PowerBI to better understand how to treat infections.

One advantage of using PowerBI is its ability to spatially map the distribution of *Shigella* infections across the UK. From the map below, we can see those major cities such as London, Manchester and Leeds with easy access to international travel, have higher rates of *Shigella* infections. This mapping helps understand that the spread of shigellosis in the UK is driven by travellers from endemic countries.



We used machine learning to take inputs of genomic data to predict the antibiotic resistance genes in each individual *Shigella* strain - don't worry, machine learning isn't the start of the rise of terminator style robots, for now they are safely contained in our laptops!

In this project, a classification supervised model was used and the antimicrobial resistance genes to individual classes were assigned. For example, no antimicrobial genes were class 0, whereas the Sull gene for Sulphonamide resistance was designated class 1.

Contact:
+44 117 4281095
info@whitetree.co.uk

Supervised machine learning models use a dataset of known outputs mapped to a range of input variables to learn and analyse the pattern between inputs and outputs. We can use this learned pattern to predict future versions of the dataset. This is done by splitting the dataset into a training and testing set, where the training set is used to teach the model the pattern and the model used to predict the testing set scores. Since we already know the correct outputs for the testing set, we can test model accuracy by comparing the predicted vs actual outputs. Using this theory, we created a machine learning model on the 2019 UK Shigella sonnei dataset, which saw accuracy scores from 80-99% when predicting the antibiotic resistance genes in specific antibiotics such as Tetracyclines.

Using data wrangling methods (the practice of converting and then plotting data from one 'raw' form into another to make it ready for downstream analytics) across Python and Excel, imbalances in training set classes were adjusted using random oversampling and a feature importance analysis was carried out (allowing us to assign scores to inputs based on how useful they are at predicting the output and therefore gain insight into the driving factors). This project allowed us to combine data techniques with traditional biology to create a model that can accurately predict all the antibiotic resistance genes in a Shigella strain. Seeing the uses of data with biology is fascinating and certainly has an exciting future.

Artificial Intelligence (AI) and machine learning are starting to reshape the defence industry. In 2017, the US Department of Defense spent \$7.4bn on AI and Big Data projects. One notable project was the development of a machine learning model used to predict skills and patient care techniques needed in battlefield healthcare scenarios. Ultimately, the system will be able to identify changes in medical threats, review clinical practice guidelines, assess the roles of different health care providers and provide feedback on needed changes to training. Experts have said the system could become a point-of-care decision-making tool based on identifying patterns that show what care has the best statistical outcomes.

The Whitetree Decision Support team are currently exploring how to embed the benefits of machine learning within our existing projects and are looking forward to seeing how it can be integrated with some of the more 'standard' modelling challenges often encountered.

Get in touch with us at Whitetree to see how our experience with machine learning can support your business.

+44 (0)117 4281095 info@whitetree.co.uk [Nathan D'Souza](#)

About the author



Hi I'm Nathan D'Souza, a Decision Support consultant at Whitetree. I have broad experience in working with data through the variety of projects encountered here at Whitetree. I have leveraged my background in Bioinformatics and extensive experience in Python, Excel and PowerBI, to conduct research and analysis across a variety of different fields.