AP Using Machine Learning Algorithms To Predict Whether Phytochemicals Are Cancerous or Anti-Cancerous

Chetty, Arnav (School: Acton Boxborough Regional High School)

Chemicals found in plants can be beneficial in the treatment and prevention of cancer, however, sometimes they can also be carcinogenic. My project aims to understand if the values of selected physicochemical properties of such phytochemicals can be used to predict their cancer-related biological effects in humans. Leveraging a credible Canadian repository of naturally occurring substances, their properties, and biological effects, I assembled a relevant database for data mining. Using basic statistics and an unsupervised machine learning algorithm, in the form of k-means clustering graphs, I identified causality involving a range of properties such as molecular mass, pH/Pka, electron-donor/acceptor, polarizability, and refractivity. I further developed a deep neural network using a Keras algorithm, which I fitted, trained, and tested to arrive at a product with a predictive accuracy of 88%. My work could have future applications in both diet optimization for cancer treatment & prevention and as an in-silico screening tool for scientists in drug discovery.