Building a Machine Learning Algorithm to Predict Cardiovascular Disease and Coding an Application to Provide Personalized Feedback

Kolan, Shreyas (School: T.C. Jasper High School)

Cardiovascular Diseases are conditions which block blood transportation to bodily organs. Using an online dataset and machine learning methods, it was tested how heart diseases can be predicted early and easily with the highest accuracy. The data was imported to Google Colab, an editor where the coding was done. Categorical variables were numerized to feed to the algorithm. The data was then split into training and testing portions. Various types of machine learning models were trained and tested on this data, to compare run times and accuracy of their predictions. The best functioning model was then used to generate graphs and tables which describe variable correlation and model performance. Graphs were also made to compare the different models. Then, an application was coded to allow a user to input their height, weight, age, etc. to find if they were probable for a positive diagnosis, and take any validation to improve and add to the training data. In the end, it was found that the Logistic Regression model functioned best and with an acceptable run time. It was able to predict the disease at around 92% of the time. In conclusion, the Logistic Regression model fits this use case best, and this information can be expanded to develop large and even more effective algorithms to make heart disease diagnosis easier and more accurate than ever.

Awards Won: