The Development of an Artificial Intelligence Model to Predict Weekly COVID-19 Cases Using Important Socioeconomic Variables

Pasha, Kinnoree (School: Clovis North High School)

This project was outlined to identify important socio-economic variables and their influence in COVID-19 spread. This information can be beneficial to minimizing virus spread as it'll allow decision makers to identify and target areas that may be contributing more to virus spread with mitigation methods. An ANN model was also developed that is capable of predicting weekly COVID-19 spread with high accuracy as demonstrated by correlation coefficients for training set 0.99, validation set 0.97, and testing set 0.91. The project was designed to utilize a numeric approach that includes single linear regression (SLR), multilinear regression (MLR), nonlinear powered function (NLR) based on least square regression, principal component analysis (PCA), and Artificial Neural Network (ANN) to analyze and develop predictive models for COVID-19 spread. First, SLR, MLR, and NLR were applied to identify the correlation between forty three (43) socio-economic variables and COVID-19 spread. PCA was then used to analyze system variability and to identify principal components that capture most of the variabilities for the purpose of prediction. The principal components subset that contributes more to system variability was utilized in a machine learning algorithm to develop a predictive model using ANN to predict COVID-19 spread for 31 counties in California. Results of the SLR, MLR, and NLR suggest that variables related to occupation, total population, household income, and transportation are more important than the others. The PCA found that six principal components captured 90% of system variability. Using these six principal components, an ANN model was developed.