Machine Learning:

**Logistic Regression:**

Accuracy: 77.09%

Accuracy CV 10-Fold: 70.43%

The Accuracy score of Logistic Regression is 77.09% which predicts that COVID-19 impacts states and counties at a socio-economic level

**Naïve Bayes:**

Accuracy: 98.4%

Accuracy CV 10-Fold: 96.79%

The model behind Naive Bayes Classifier has something to do with probability distributions. The aim is to maximize the probability of the target class given the x features. As such, based on our data and analysis, the accuracy score predicts that there is a high chance that COVID-19 impacts the states and counties at a socio-economic level.

**K Nearest Neighbor:**

Accuracy: 79.47%

Accuracy CV 10-Fold: 60.56%

* The KNN algorithm assumes that similar things exist in proximity. In other words, similar things are near to each other. Our KNN implementation above relies on structured data. It needs to be in a table format. Additionally, the implementation assumes that all columns contain numerical data and that the last column of our data has labels that we can perform some function on. So, the KNN accuracy score of 79.47% here predicts that our data and variables are related. To better explain this, socio-economic factors in our communities are impacted by COVID-19.

**Decision Tree:**

Accuracy: 100%

Accuracy CV 10-Fold: 90.37%

Decision Tree algorithms predicts to make decision making easier. A higher accuracy score means that the model is considering all variables and predicts a strong relationship between variables.

**Random Forest:**

Accuracy: 99.15%

Accuracy CV 10-Fold: 93.38%

Random forests is considered as a highly accurate and robust method because of the number of decision trees participating in the process. Because the accuracy score is high, our data analysis model implies that variables are related and can be highly impacted by each other.

The combination of the COVID-19 pandemic, a drop-in unemployment, increasing number of cases and deaths have the potential to seriously impact stability state and county wide, and will strain public health and social security systems in the short to medium terms. This analysis uses machine learning models' accuracy scores to predict relationship between covid-19 cases/deaths and several socio-economic factors like poverty, household income, unemployment, education and race. The high accuracy scores of each model indicate that there is a strong relationship and all dependent variables can be highly impacted by changes in independent variables, such as covid-19 cases/deaths.