

# **Social Determinants of Health & the COVID-19 Pandemic:**

**Classification Model Predicting the  
Spread & Mortality of a Pandemic in a  
Community**

Kristin Cooper | July, 2021

# Social Determinants of Health

Health outcomes are often determined long before a person seeks care based on where they live & work, their education & income, and the resources they have access to - referred to as **social determinants of health**.

The COVID-19 pandemic was no exception.

## Social Determinants of Health





# Predicting a Community's Vulnerability

This analysis seeks to learn from the COVID-19 pandemic in order to predict and therefore address vulnerabilities U.S. states\* may have in the case of similar future health emergencies.

Investing in healthy communities means investing in the underlying factors that help a person be and stay healthy.

## Do these factors...

COVID-19 Stats	Economic Measures	Health Measures	Social Measures	Demographics
<ul style="list-style-type: none"><li>-Vaccine hesitancy</li><li>-Vaccine rollout concern</li></ul>	<ul style="list-style-type: none"><li>-Per capita income</li><li>-Median household income</li><li>-Income inequality</li><li>-Poverty rate</li><li>-Unemployment rate</li></ul>	<ul style="list-style-type: none"><li>-Life expectancy</li><li>-Premature deaths</li><li>-Smoking &amp; excessive drinking</li><li>-Obesity</li><li>-Poor health days</li><li>-Physical inactivity</li><li>-Preventable hospital stays</li><li>-Ratio of population to primary care physicians</li><li>-Flu vaccinations</li><li>-Uninsured population</li></ul>	<ul style="list-style-type: none"><li>-Housing</li><li>-Internet access</li><li>-Vehicle access</li><li>-Food environment and food insecurity</li><li>-Education</li><li>-Air and drinking water pollution</li></ul>	<ul style="list-style-type: none"><li>-Population and sq mileage</li><li>-Rural vs urban area</li><li>-Racial breakdown</li><li>-Elderly and child populations</li></ul>

## ...influence these outcomes?

**33,471,979**

Total cases in US

**592,400**

Total deaths in US

**1.66%**

Cases resulting in death

\*Future enhancements include detailing at the county level



# Cases, Deaths, and Mortality of High- and Low-Impact States

$$\text{Impact Score} = [(1 * \text{cases per 100k}) + (3 * \text{deaths per 100k})] * \text{mortality rate}$$

## High Impact States

State	Cases per 100k	Deaths per 100k	Mortality Rate
NEW JERSEY	11,534	298	2.58
MASSACHUSETTS	10,401	263	2.53
ARIZONA	12,925	259	2.00
RHODE ISLAND	14,451	258	1.79
MISSISSIPPI	10,801	248	2.30
ALABAMA	11,332	233	2.06
CONNECTICUT	9,762	231	2.37
SOUTH DAKOTA	14,414	235	1.64
LOUISIANA	10,390	230	2.22

## Low Impact States

State	Cases per 100k	Deaths per 100k	Mortality Rate
NEW HAMPSHIRE	7,414	102	1.38
WASHINGTON	6,246	82	1.32
UTAH	13,703	78	0.57
OREGON	5,133	68	1.34
MAINE	5,185	64	1.24
ALASKA	9,583	48	0.51
VERMONT	3,909	41	1.06
HAWAII	2,579	36	1.40

# Findings

Three impact categories - “High,” “Average,” or “Low” - were calculated to represent the extent of community spread and mortality that occurred in a given state.

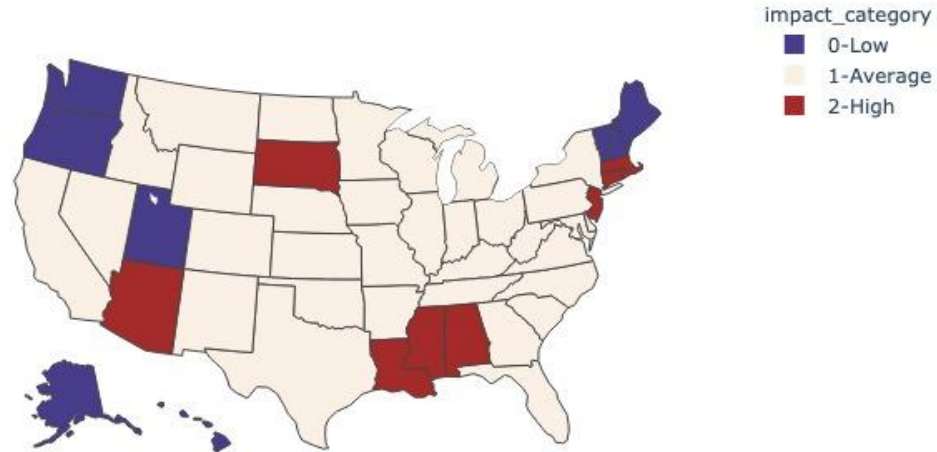
## The following 10 states experienced “high” impact of the pandemic:

- Arizona
- South Dakota
- Louisiana
- Alabama
- Mississippi
- Pennsylvania
- New Jersey
- Connecticut
- Rhode Island
- Massachusetts

## The following 8 states experienced a “low” impact:

- Alaska
- Hawaii
- Washington
- Oregon
- Utah
- Vermont
- New Hampshire
- Maine

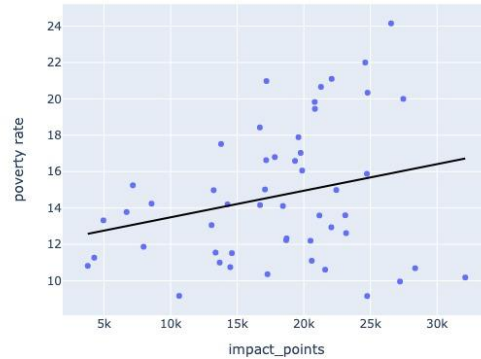
Impact of COVID-19 Pandemic by State - Category



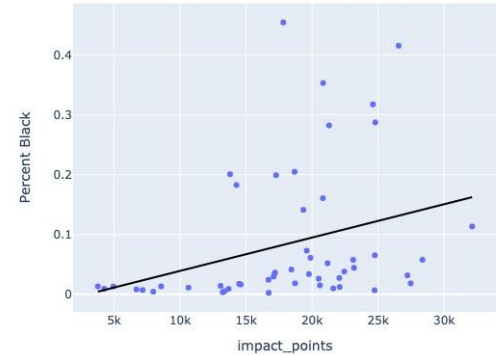
# Correlated Features (1/2)

As Impact Score (on the x-axis) increases...

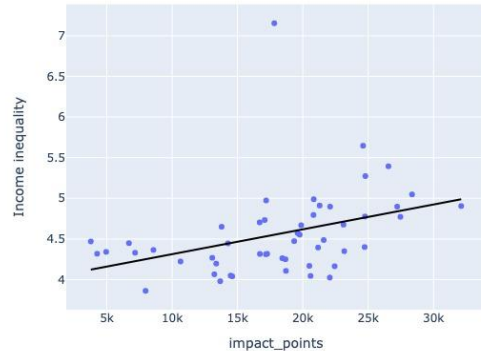
...poverty rate increases



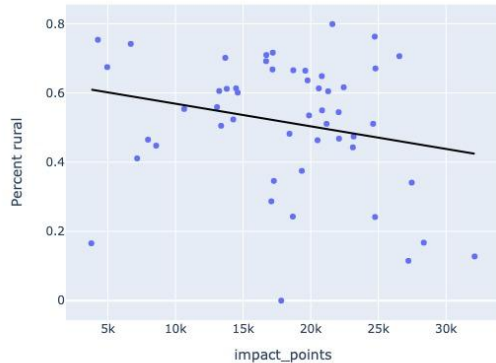
...percent Black increases



...income inequality increases



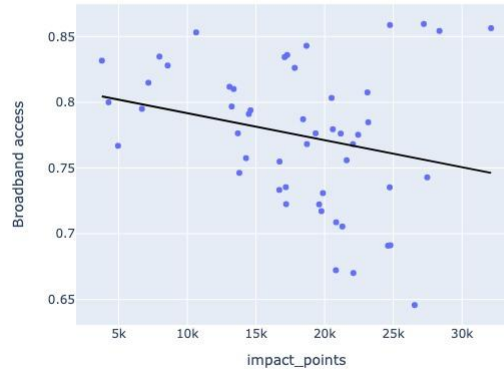
...percent rural decreases



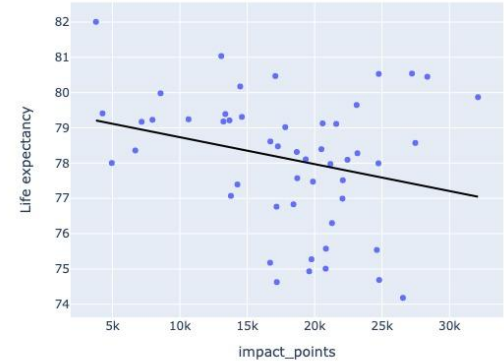
# Correlated Features (2/2)

As Impact Score (on the x-axis) increases...

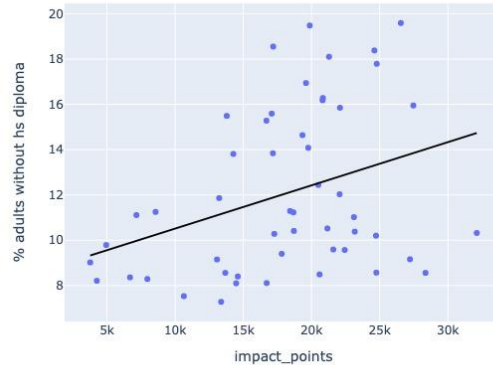
...access to internet decreases



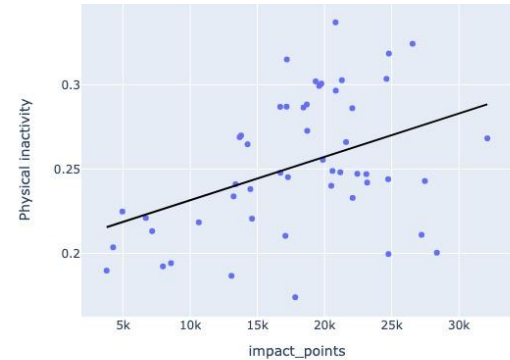
...life expectancy decreases



...percent undereducated adults increases



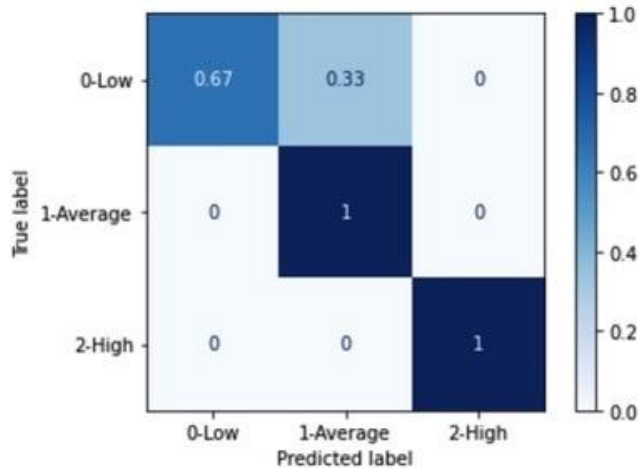
...physically inactive lifestyle increases



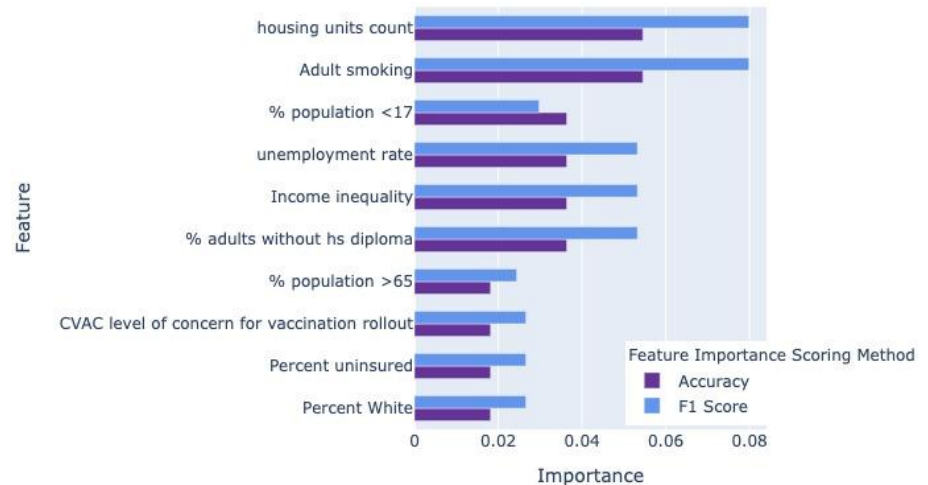
# Model Performance

An iterative modeling process revealed the best model performed with **91% accuracy**, only misclassifying some Low-Impact states as Average-Impact.

The following features were most important to making accurate predictions:



Top 10 Most Important Features to KNN Model (Full Feature Set)







# Recommendations (1/2)

If we don't learn from history, we are doomed to repeat it.

Based on what COVID-19 data tells us, **state and local** organizations should:

→ **Measure what matters.**

States should have accurate, frequent measurement plans in place for each of the features shown here to increase their community's vulnerability to extensive spread of illness and mortality. What you measure, you can manage.

→ **Extend the reach of health budgets to invest in socioeconomic and access barriers.**

When planning public health budgets, consider what social determinants - unemployment, education, income inequality, etc. - may play a role in health outcomes. Seek funding for social, economic, and environmental projects from health budgets.

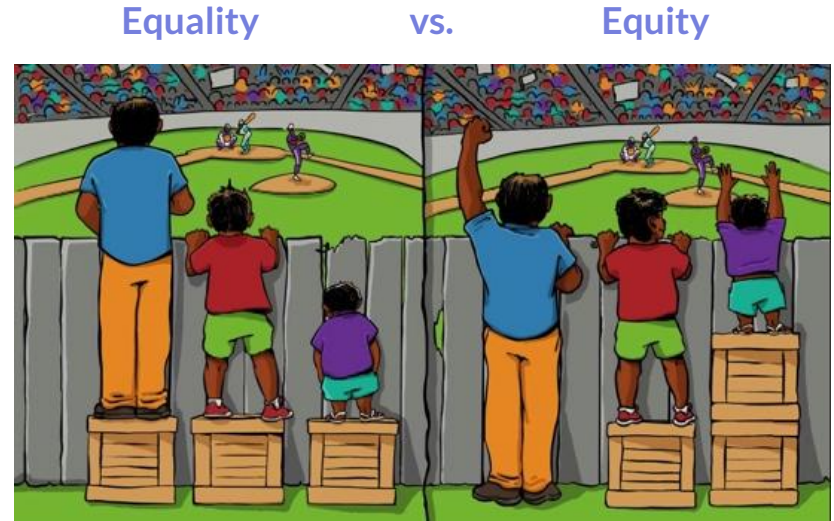
# Recommendations (2/2)

If we don't learn from history, we are doomed to repeat it.

Based on what COVID-19 data tells us, the U.S. **federal** government should:

- **Ensure epidemic procedures at all levels of government enable resources to be allocated based on vulnerability.**

In the case of a pandemic, organizations should be able to quickly decide and allocate emergency funds equitably - according to vulnerability - rather than equally - according to population - in order to achieve the most benefit from limited resources.





# Future Enhancements

Despite strong efforts by The New York Times, there are still gaps in consistent, timely reporting of COVID-19 data by state and local governments that hinder US-wide, county-specific analysis.

Future enhancements to the model and report include:

- Analyze and model at the county level of detail to capture more specific community-level disparities
- Conduct more rigorous feature selection to narrow down the features that most influence pandemic vulnerability
- Incorporate ICU/hospital capacity and economic measures (job loss, change in economic activity, bankruptcy, etc) into Pandemic Impact calculation

# Thank You!

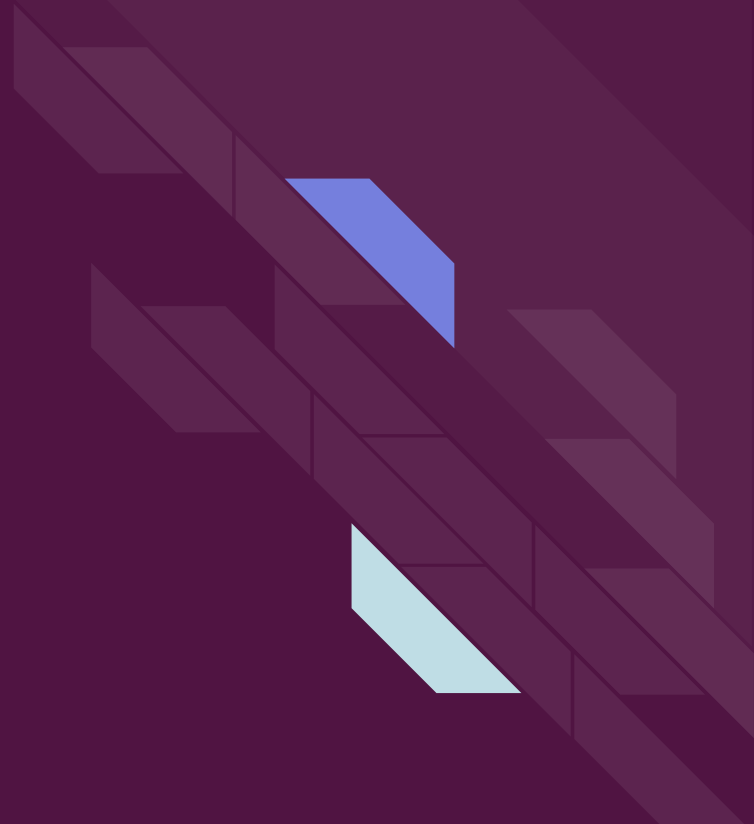
**Full Report:**

<https://github.com/kcoop610/capstone-pandemic-us-health-inequities>

**Email:** [kcoop610@gmail.com](mailto:kcoop610@gmail.com)

**LinkedIn:** <https://www.linkedin.com/in/kristincooper16>

# Appendix





# Data

Data for this analysis came from the CDC, the New York Times, and the University of Wisconsin Population Health Institute, accessed on July 7th, 2021.

## CDC's Social Vulnerability Index 2018

Individual and aggregated measures in the following themes:

- Socioeconomic status
- Household composition & disability
- Minority status & language
- Housing type & transportation

## CDC's Vaccine Hesitancy

Predictions of hesitancy rates based on responses from the U.S. Census Bureau's Household Pulse Survey (HPS).

- Racial breakdown
- % vaccine hesitancy

## U of Wisconsin Population Health Institute's County Health Rankings

30+ factors that influence how long and how well we live:

- Environment
- Health behaviors
- Access to resources
- Economic measures

## The New York Times COVID-19 Data

COVID-19 data by county compiled by the New York Times from state and local governments and health departments.

- Case counts (confirmed & probable)
- Death counts (confirmed & probable)



# Modeling Process

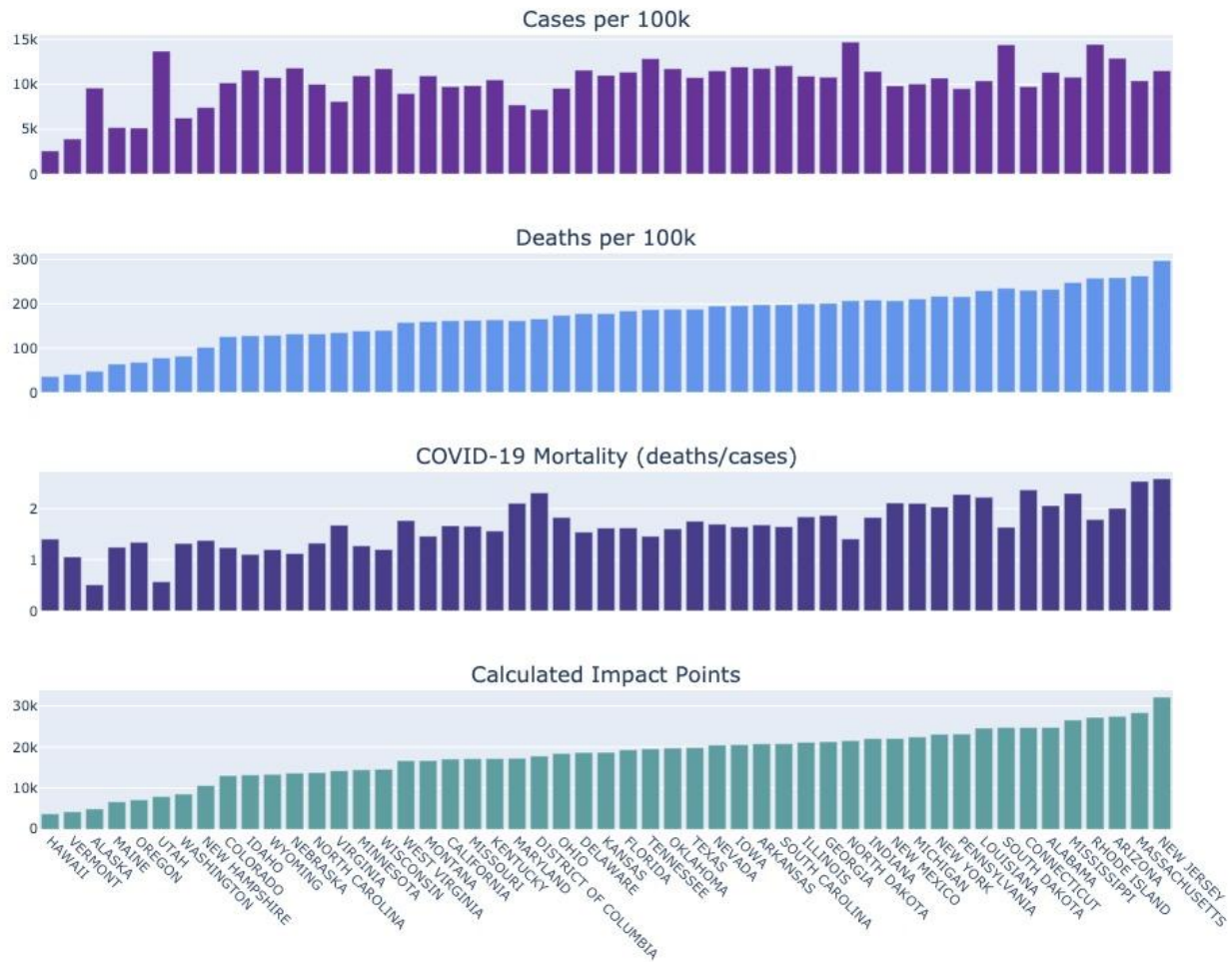
Multiple algorithms with various parameters were iteratively trained on the sample data in order to develop the best possible model.

Methods conducted:

- Logistic Regression with Cross-Validation
- K-Nearest Neighbors
- Decision Trees and Random Forest
- Boosting Ensemble Methods including AdaBoost, GradientBoost, & XGBoost
- Support Vector Machine
- Grid Search

Model complexity was also iterated, with one set of models leveraging all predictors available and another set focusing on non-health related measures.

## Population-Controlled COVID-19 Cases and Deaths





# Best Model Performance

