

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



Equality

vs.

Equity





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
Vaccine hesitancy Vaccine rollout concern	Per capita income Median household income Income inequality Poverty rate Unemployment rate	Life expectancy Premature deaths Smoking & excessive drinking Obesity Poor health days Physical inactivity Preventable hospital stays Ratio of population to primary care physicians Flu vaccinations Uninsured population	Housing Internet access Vehicle access Food environment and food insecurity Education Air and drinking water pollution	Population and sq mileage Rural vs urban area Racial breakdown Elderly and child populations

...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

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.

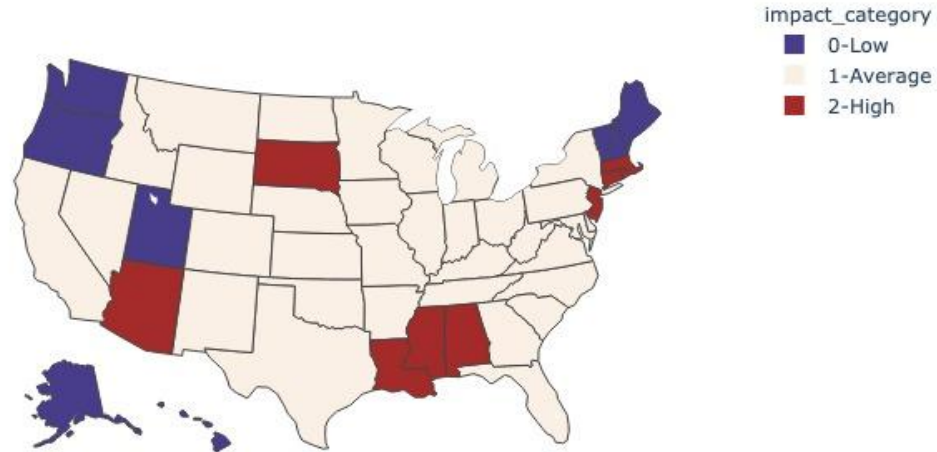
The following states experienced "high" impact of the pandemic:

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

The following states experienced a “low” impact:

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

Impact of COVID-19 Pandemic by State - Category

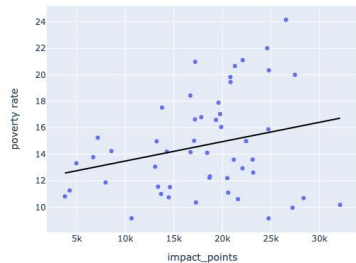


Correlated Features

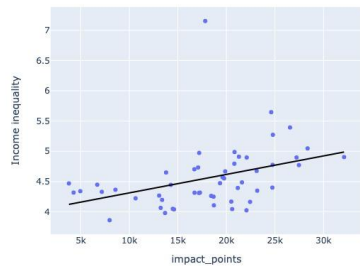
As Impact Score (on the x-axis) increases...

Economic

...poverty rate increases

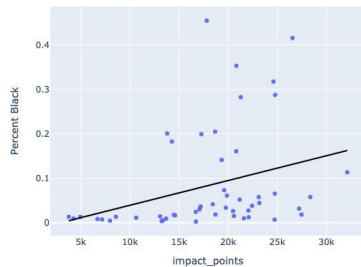


...income inequality increases

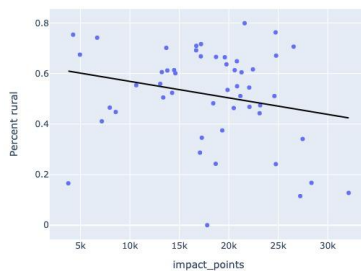


Demographic

...percent Black increases

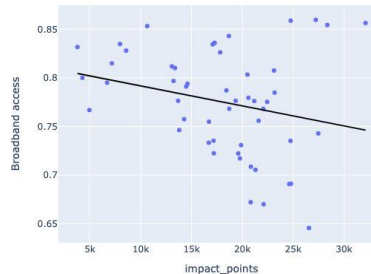


...percent rural decreases

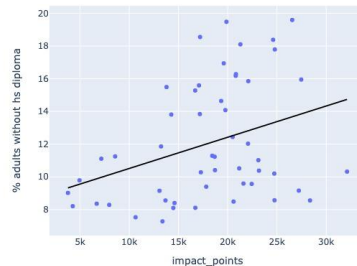


Access & Community

...access to internet decreases

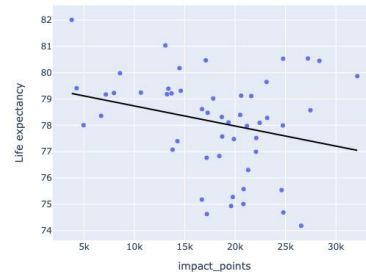


...percent undereducated adults increases

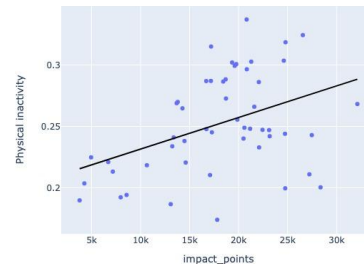


Health

...life expectancy decreases



...physically inactive lifestyle increases

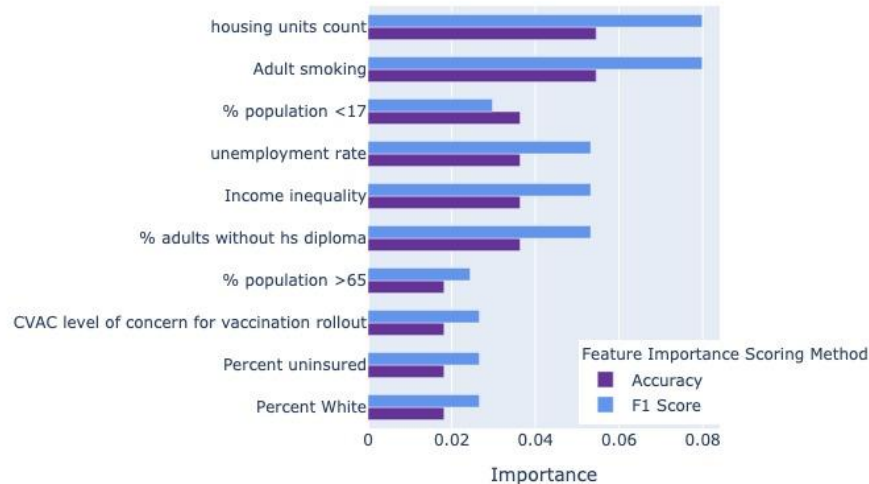


Model Performance

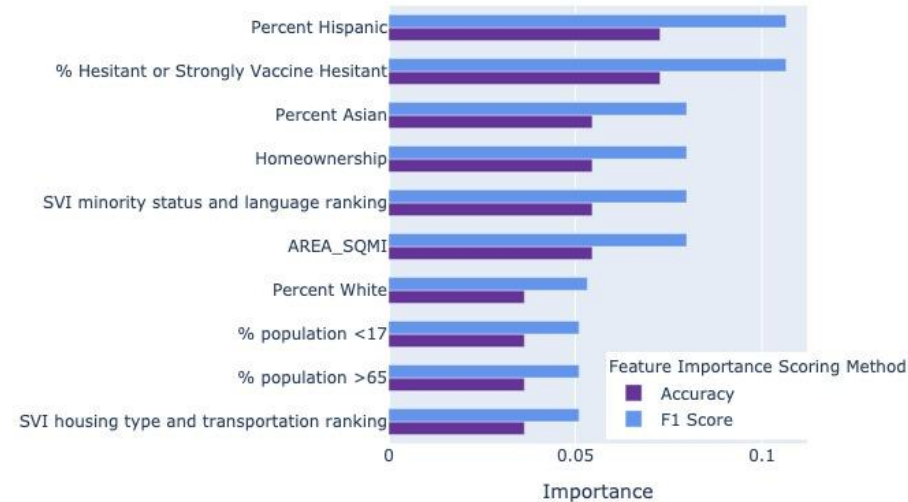
An iterative modeling process revealed the best model performed with **91% accuracy**, only misclassifying some Low-Impact states as Average-Impact.

The following features were most important to making accurate predictions:

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



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





Recommendations

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

Based on what COVID-19 data tells us, the U.S. government 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.

→ **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!

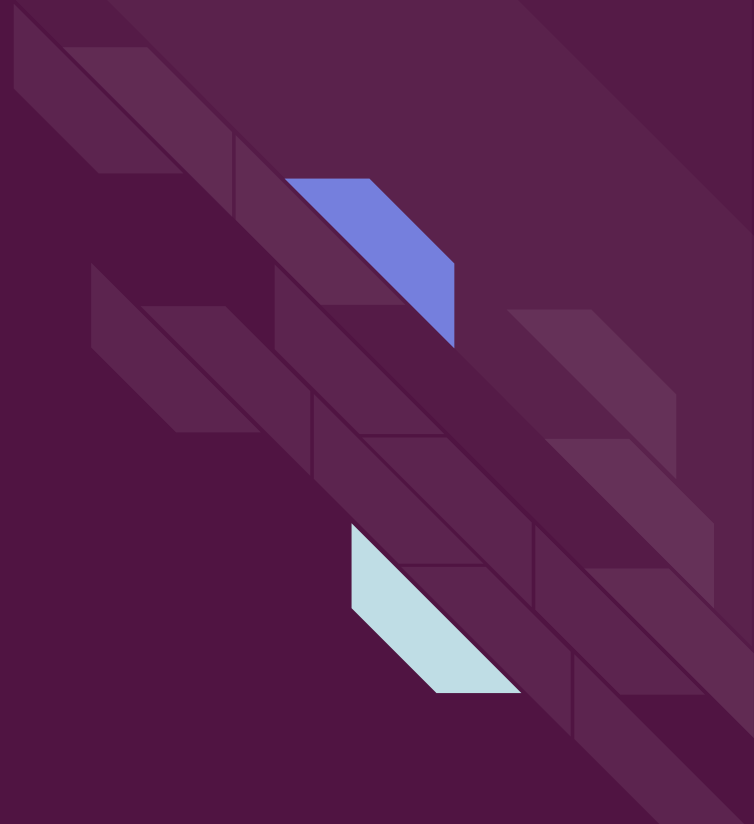
Full Report:

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

Email: kcoop610@gmail.com

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

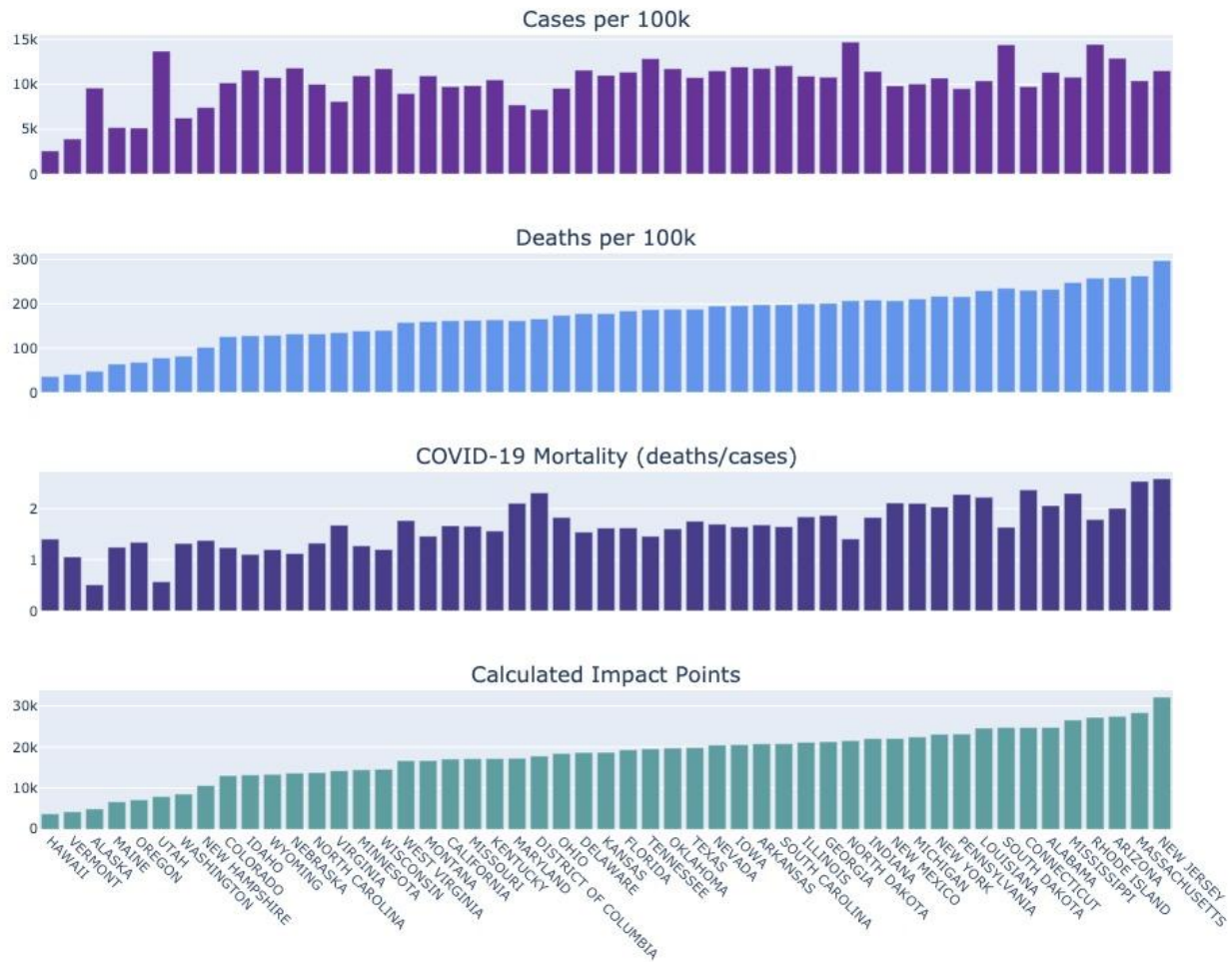
Multiple algorithms with various parameters were iteratively trained on the sample data in order to develop the best possible model.

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

