Social Determinants of Health & the COVID-19 Pandemic:

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

Kristin Cooper | July, 2021

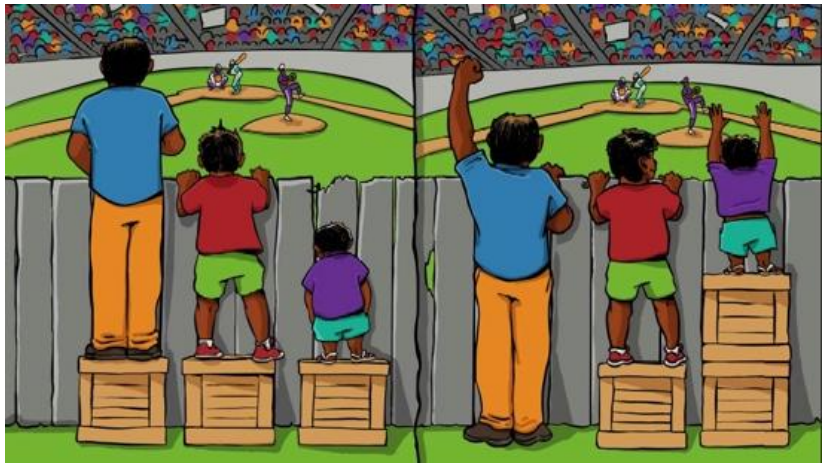
Social Determinants of Health

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



Reference: US Dept HHS, Healthy People 2030

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

...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

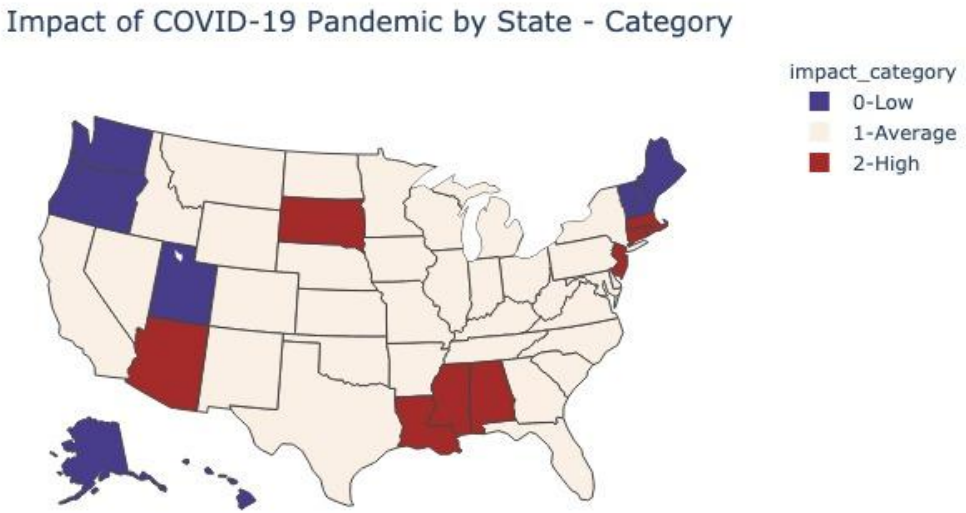
Three impact categories - “High,” “Average,” or “Low” - were calculated to represent the extent of community spread and mortality that occurred in a given state.

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

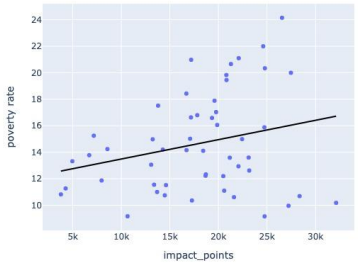


Correlated Features

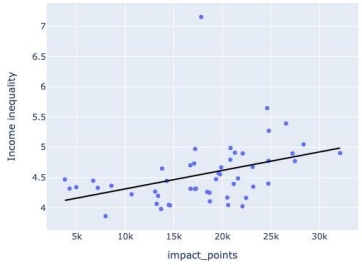
As Impact Score (on the x-axis) increases...

Economic

...poverty rate increases

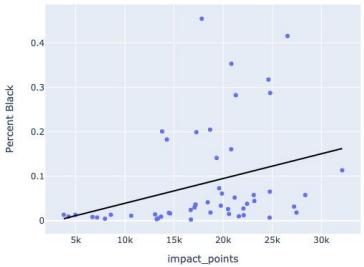


...income inequality increases

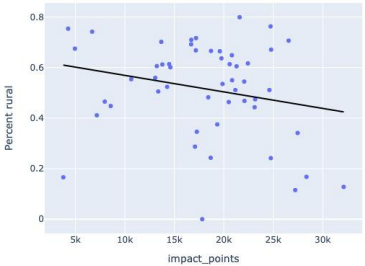


Demographic

...percent Black increases

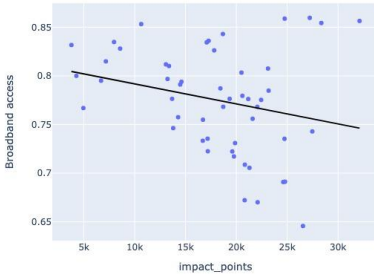


...percent rural decreases

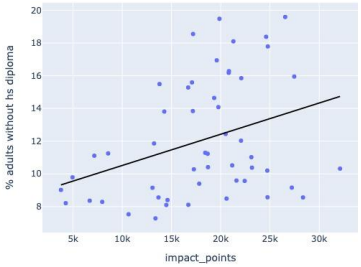


Access & Community

...access to internet decreases

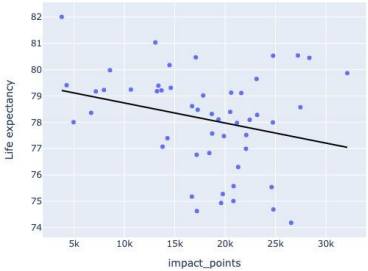


...percent undereducated adults increases

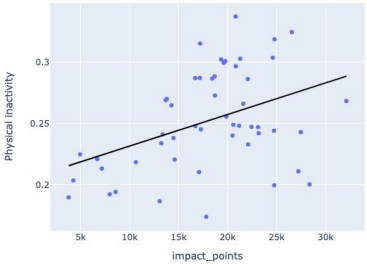


Health

...life expectancy decreases



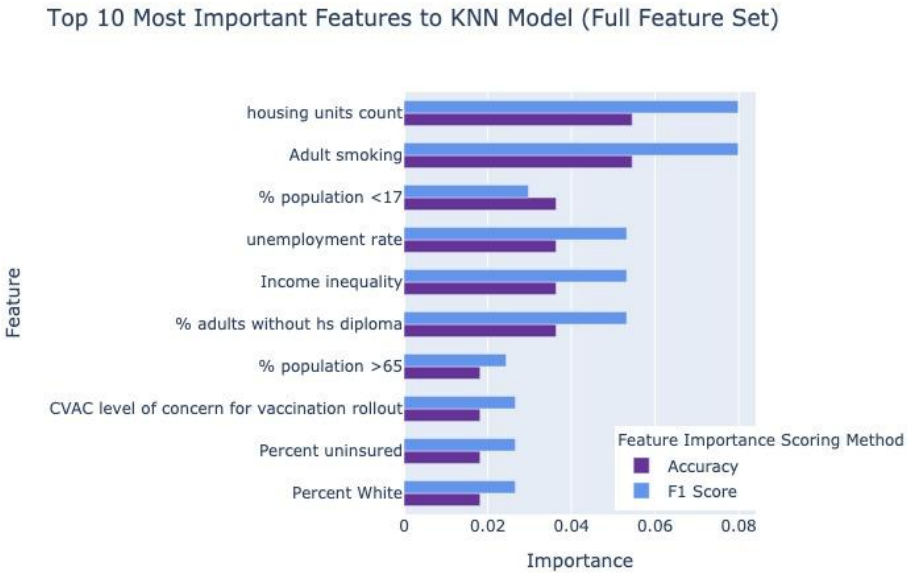
...physically inactive lifestyle increases

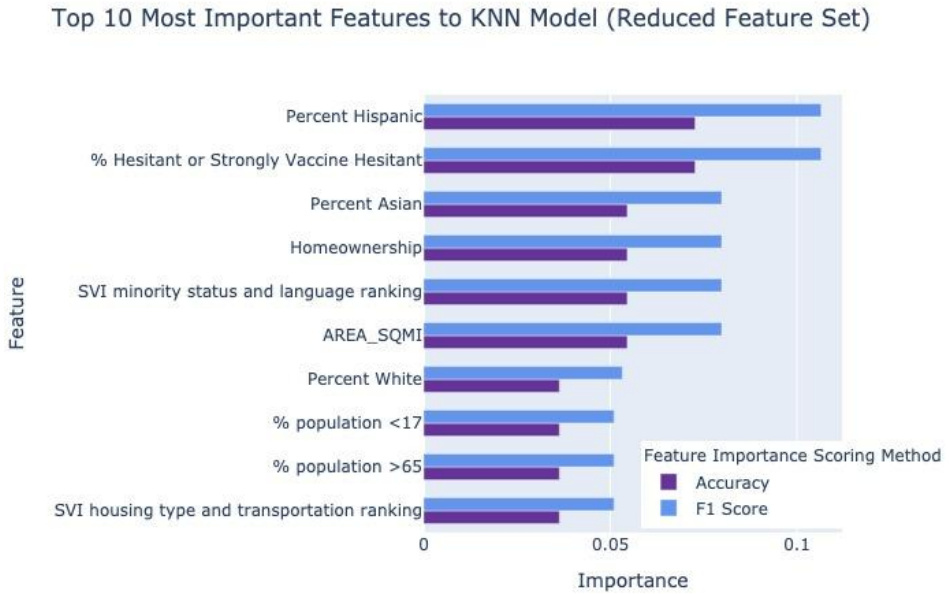


Model Performance

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:





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:

- Analysis and modeling at the county level of detail to capture more specific community-level disparities
- 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

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

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

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)

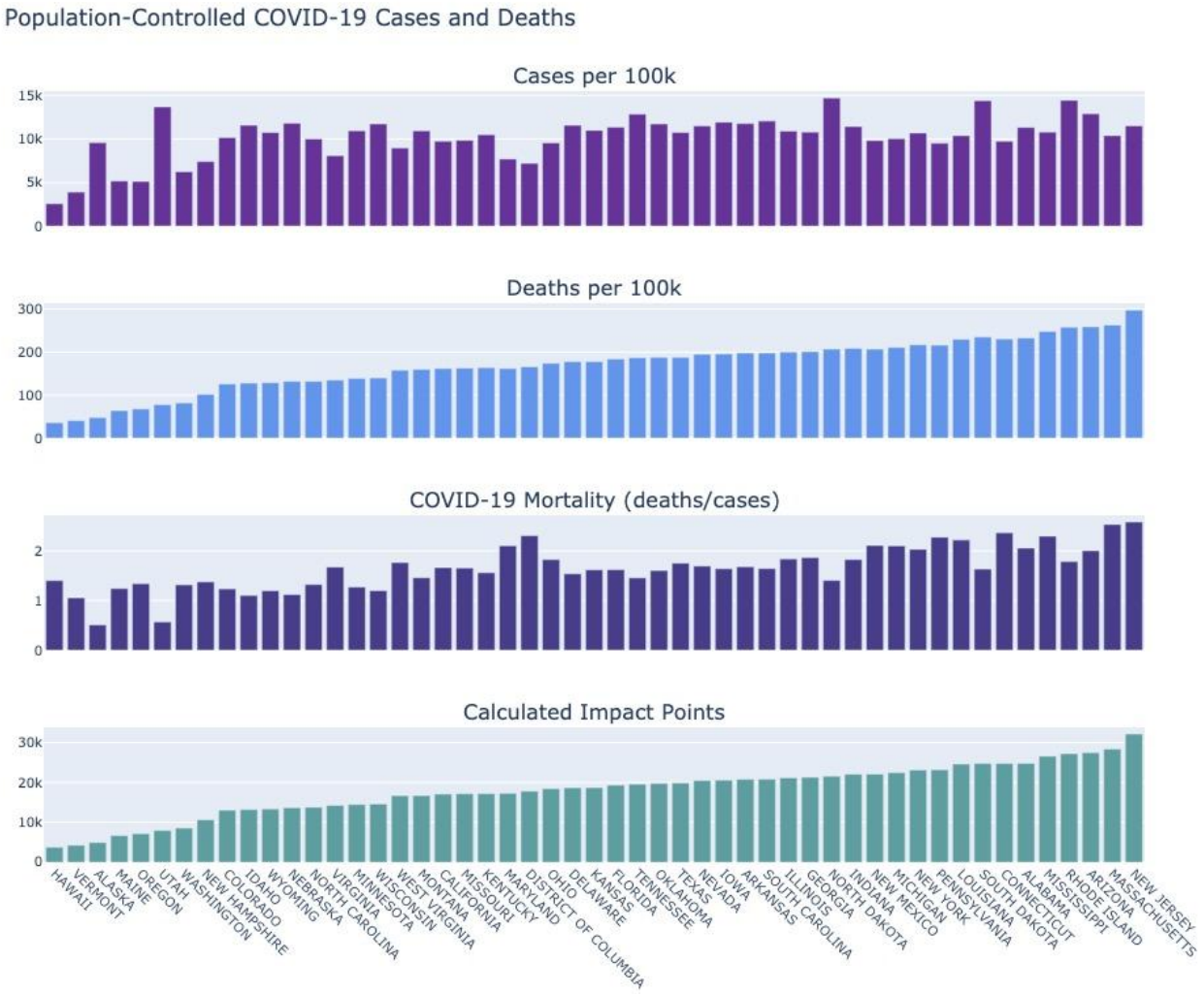
Modeling Process

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
Cases per 100k
15k
10k
5k
0
Deaths per 100k
300
200
100
0
COVID-19 Mortality (deaths/cases)
2
1
0
Calculated Impact Points
30k
20k
10k
0
HAWAII
VERMONT
ALASKA
MAINE
OREGON
UTAH
WASHINGTON
NEW HAMPSHIRE
COLORADO
IDAHO
WYOMING
NEBRASKA
NORTH CAROLINA
VIRGINIA
MINNESOTA
WISCONSIN
WEST VIRGINIA
MONTANA
CALIFORNIA
MISSOURI
KENTUCKY
MARYLAND
DISTRICT OF COLUMBIA
OHIO
DELAWARE
KANSAS
FLORIDA
TENNESSEE
OKLAHOMA
TEXAS
NEVADA
IOWA
ARKANSAS
SOUTH CAROLINA
ILLINOIS
GEORGIA
NORTH DAKOTA
INDIANA
NEW MEXICO
MICHIGAN
NEW YORK
PENNSYLVANIA
LOUISIANA
SOUTH DAKOTA
CONNECTICUT
ALABAMA
MISSISSIPPI
RHODE ISLAND
ARIZONA
MASSACHUSETTS
NEW JERSEY

Best Model Performance

	precision	recall	f1-score	support
0-Low	1.00	0.67	0.80	3
1-Average	0.86	1.00	0.92	6
2-High	1.00	1.00	1.00	2
accuracy			0.91	11
macro avg	0.95	0.89	0.91	11
weighted avg	0.92	0.91	0.90	11