Using Environmental Factors to Predict COVID-19 Vaccine Acceptance

By: Jaeyoung Kang, Alexander Kroll, Tianhan Ling, Pham Thanh Ngoc, Jie Sing Yoo, Alice Zilberberg
Project Advisor: Peter Freeman

Introduction & Background

The development of the COVID-19 vaccine was a vital step in the fight against the global pandemic. The number of cases in a given geographical area are directly affected by the number of individuals choosing not to receive the vaccination. Predicting acceptance of the vaccine could help adequately prepare for the rise or fall of cases by location.

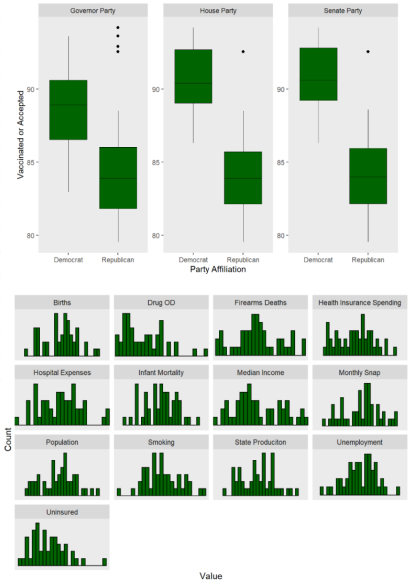
Our goal is to develop a model that can use environmental factors to predict whether an individual will accept or reject the COVID-19 vaccine.

Data Pre-Processing/EDA

The dataset has a sample size of 49 with 16 predictor variables and 1 qualitative response variable that is acceptance of the vaccine.

Variable	Type
Vaccinated/Accepted	Qualitative Response
Uninsured	Quantitative Predictor
Health Insurance Spending	Quantitative Predictor
Births	Quantitative Predictor
Infant Mortality	Quantitative Predictor
Firearms Deaths	Quantitative Predictor
Median Income	Quantitative Predictor
Governor Party	Qualitative Predictor
Senate Party	Qualitative Predictor
House Party	Qualitative Predictor
State Production	Quantitative Predictor
Unemployment	Quantitative Predictor
Monthly Snap	Quantitative Predictor
Smoking	Quantitative Predictor
Drug OD	Quantitative Predictor
Hospital Expenses	Quantitative Predictor
Population	Quantitative Predictor

Distribution of Predictor Variables After Transformation and Removal of Outliers



Log transformed variables:

- monthly snap
- births
- population
- state production
- insurance spending
- unemployment

Methods

- Used various tuning methods to find optimized parameters for multiple linear regression, random forest, and k-nearest neighbors models.
- Compared the tuned models using mean-squared error.

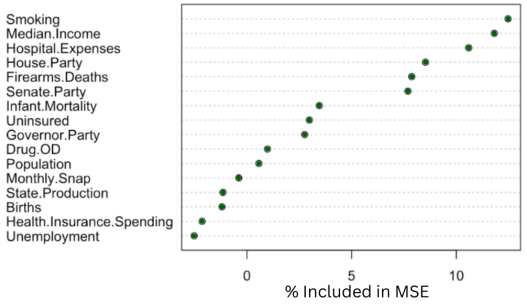
Analysis and Results

Performance of Various Tuned Models

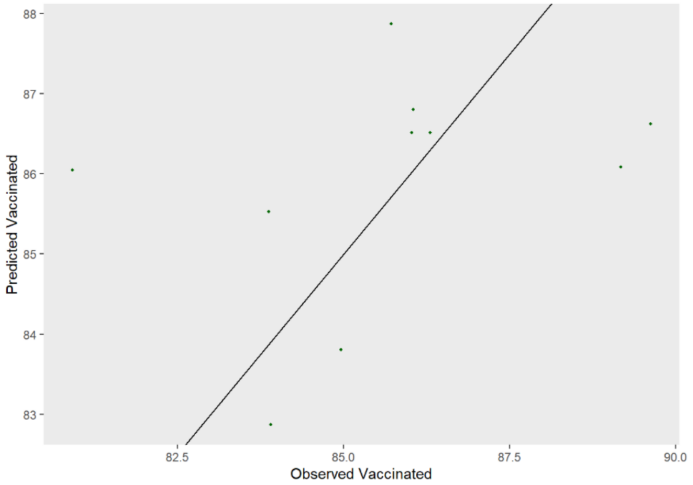
Model	Mean-Squared Error
Multiple Linear Regression	11.45
Random Forest	5.54
K-Nearest Neighbors	10.46

- The random forest model predicted state vaccination acceptance rates with the highest accuracy.

Random Forest Model Percentage Included MSE



Predicted vs Observed Vaccine Acceptance (Random Forest



- State smoking rates, median household income, and hospital expenses were the most significant predictors
- Smokers show higher rates of negative attitudes towards vaccines compared to non-smokers which follows with the research conducted by Jackson et al.

Conclusion

- The random forest algorithm outperforms linear regression and k-nearest neighbor algorithms in predicting vaccination acceptance rates by state
- State smoking rates were the most significant predictor of vaccination acceptance rate
- Public health policies targeting smokers could be important in addressing equitable vaccine distribution in regional populations of the United States

References

- Group, Delphi. “Covidcast Dashboard.” Go to the Main Page, https://delphi.cmu.edu/covidcast/.
- Jackson, Sarah E., et al. “Negative Vaccine Attitudes and Intentions to Vaccinate against Covid-19 in Relation to Smoking Status: A Population Survey of UK Adults.” MedRxiv, Cold Spring Harbor Laboratory Press, 1 Jan. 2020, https://www.medrxiv.org/content/10.1101/2020.12.17.20248396v1.full.