Factors Impacting US COVID-19 Vaccine Adoption: A Statistical Exploration

Amogh Ananda Rao, Ashita Jawali, Hamidreza Akhbariyoon, Anh-Thu Pham, Wrootchit Mishra
Advisor: Dr. Peter Freeman, Statistics & Data Science Department, Carnegie Mellon University

INTRODUCTION

Vaccine hesitancy is a complex phenomenon characterized by a range of attitudes and behaviors towards vaccination. It is defined as a delay in acceptance or refusal of vaccination despite the availability of vaccination services (1). Politics, culture, healthcare professionals, employment, vaccine attitudes and beliefs, social networks, and media play significant roles in shaping vaccine decision-making among hesitant adopters (2).

OBJECTIVE

To build a model to predict primary drivers of vaccine adoption and offer insights to aid policymakers, healthcare professionals, and public health campaigns.

The DATA

Data has been collected from Carnegie Mellon University's COVID cast project by the Delphi research group and from the Kaiser Family Foundation. The dataset contains 50 observations and 17 features as summarized in Table 1.

Feature Name	Description
uninsured	Percentage of Uninsured
total_private_health_insurance_spending	Total Private Health Insurance Spending (2014)
number_of_births	Number of Births (2019)
infant_mortality_rate	Infant Mortality Rate (2018)
firearms_death_rate	Firearm Death Rate per 100,000 Residents (2018)
median_annual_household_income	Median Annual Household Income (2019)
governor_political_affiliation	Governor Political Affiliation
state_senate_majority_political_affiliation	State Senate Majority Political Affiliation
state_house_majority_political_affiliation	State House Majority Political Affiliation
total_gross_state_product	Total Gross State Product (millions of current dollars)
unemployment_claims	Unemployment Claims, Week of 8/28/2021
average_monthly_snap_participants	Average Monthly SNAP Participants 2019
smoking	Percent of Adults Who Smoke (2017)
drug_overdoses	Drug Overdose Death Rate (per 100,000, 2019)
hospital_inpatient_day_expenses	Hospital Adjusted Expenses per Inpatient Day (2019)
population	Total US Population (2019)
vaccinated_or_accept	Vaccine acceptance among COVIDcastsurvey respondents (August 2021)

Table 1: The Features and Descriptions from the COVIDcast Dataset

ANALYSIS

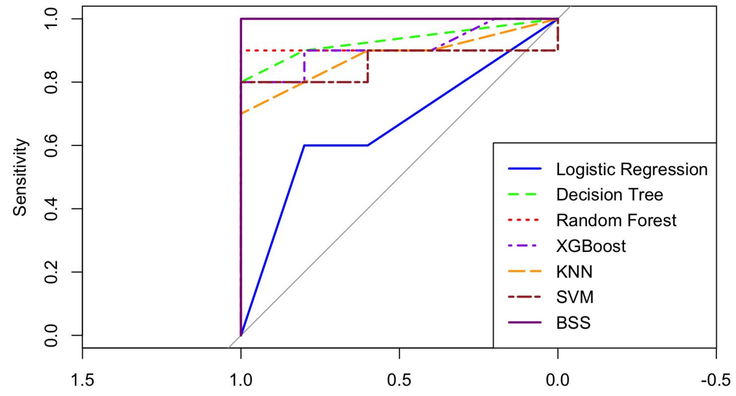
After looking for missingness and outliers in the data, we decided not to remove any observation from the analysis dataset. We used square-root and logarithmic transformations to quantitative variables after understanding the distributions of all the features. To prevent overfitting the models, we performed a 70:30 division of the dataset to create two distinct subsets: one for training and the other for testing machine learning models.

The model performances have been summarized in Table 2. The decision tree was the best performing independent model with a ROC-AUC = 0.930 (J=0.857), followed by random forest, XGBoost, k-nearest neighbors, and support vector machine. Logistic regression, with a J=0.99, had the lowest ROC-AUC = 0.66. However, owing to the nature of the model, we have introduced explainability to this analysis in Figure 2. The best subset model had the best performance with a ROC-AUC = 1, and zero misclassification.

Model	Threshold	Misclassification Rate	ROC-AUC
Logistic Regression	1.00	0.267	0.660
Decision Tree	0.857	0.867	0.930
Random Forest	0.490	0.067	0.900
XGBoost	0.796	0.867	0.910
K-nearest Neighbors	0.786	0.800	0.880
Support Vector Machine	0.466	0.867	0.860
Best Subset Model	0.500	1.00	1.00

Table 2: Summary of Model Performances.

Figure 1: ROC Curve for all the Models



		Actual	
		ACCEPT	REJECT
Predicted	1	10	0
	0	0	5

Confusion Matrix for the Best Subset Model

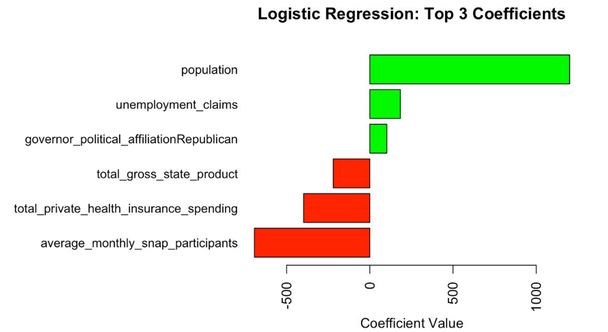


Figure 2: Top 3 Coefficients of Logistic Regression

CONCLUSION

The Random Forest model appears to be the best as it has both a low misclassification rate and a high ROC-AUC value (0.900). The Decision Tree has a high ROC-AUC but a very high misclassification rate, which is not ideal for our dataset. Along with the feature explanations from logistic regression, this study provides a framework for a better understanding of vaccine adoption

References

1. Nuwarda RF, Ramzan I, Weekes L, Kayser V. Vaccine Hesitancy: Contemporary Issues and Historical Background. Vaccines (Basel). 2022 Sep 22;10(10):1595. doi: 10.3390/vaccines10101595. PMID: 36298459; PMCID: PMC9612044.
2. Purvis RS, Moore R, Willis DE, Hallgren E, McElfish PA. Factors influencing COVID-19 vaccine decision-making among hesitant adopters in the United States. Hum Vaccin Immunother. 2022 Nov 30;18(6):2114701. doi: 10.1080/21645515.2022.2114701. Epub 2022 Sep 7. PMID: 36070518; PMCID: PMC9746519.