

Predicting COVID-19 Vaccination Intention: Using A Machine Learning Approach

Xin Zhang, Ph.D. (University of North Carolina at Pembroke)

Zhixin Kang, Ph.D. (University of North Carolina at Pembroke)

Guanlin Gao, Ph.D. (Chaminade University of Honolulu)

Xinyan Shi, Ph.D. (University of North Carolina at Pembroke)

Literature Review

- Demographic and socioeconomic vaccination determinants: e.g. Al-Amer et al., 2022; Carrieri et al., 2021; Detoc et al., 2020, Fernandes et al., 2021; Sherman et al., 2021; Huynh et. al., 2021; Wang et. al., 2022.
- Sources of vaccination information: Latkin et. al., 2021, Piltch-Loeb et. al., 2021, Purvis et. al., 2021, Qiao, 2020.
- Knowledge, altitude and belief: Bailey, 2020; Gao & Li, 2022; Lee, Kang & You, 2021; Rahman et al., 2022
- Machine Learning approach in studying COVID-19: Carrieri, Lagravinese, and Resce,2021; Fernandes et al., 2021; Iwendi, 2020; Mellado et al., 2021; Muhammad et al. 2020; Ong et al., 2020; Ritonga et al., 2021

Objectives

- To predict the participants' and their children's vaccination intention by examining the following factors
 - Socio-demographic factors
 - COVID-19 testing related factors
 - COVID-19 vaccine knowledge
 - COVID information source
 - COVID vaccine behavior
 - COVID-19 vaccine attitude and belief

Hypothesis

Hypothesis 1: People's socioeconomic status, knowledge about COVID-19, attitude and belief about COVID-19 contribute to vaccination behaviors.

Hypothesis 2: People who are more likely to get tested also tend to be vaccinated.

Hypothesis 3: People who are overconfident about their COVID-19 knowledge are less likely get vaccinated.

Survey Design

Socio-demographic questions: gender, age, race, education, family composition, employment, income, area of residence, insurance.

COVID-19 testing questions: the number of testing received, the reasons to get tested, testing locations, barriers to get tested, the difficulty of getting tested, whether they were tested positive.

COVID information questions: the sources of information to learn about COVID-19 vaccine, and how easy or difficult to obtain information about COVID vaccines, to sign up for a COVID vaccine, and to find a time and location.

Survey Design

COVID-19 vaccine knowledge questions:

e.g. The only people who can get severely ill from COVID-19 are the elderly and people with certain medical conditions; People who have COVID-19 and no symptoms can still spread it to others; If someone tests negative for COVID-19, this person can still get infected in the future.

Knowledge confidence questions about self and others: 5-point scale ranging from “Not confident at all” to “Extremely confident.”

Survey Design

COVID vaccine behavior questions:

- their vaccination status and plan for themselves
- Factors that influence their vaccination decision
- plan to get their children vaccinated and
- whether they received flu shots in the last 12 month.

COVID-19 information questions: sources of information to learn about COVID-19 vaccine, and how easy or difficult to obtain information about COVID-19 vaccines, to sign up for a COVID-19 vaccine, and to find a time and location.

Survey Design

COVID-19 vaccine attitude and belief: e.g. *“I am or will be more relaxed once I can get the vaccine”*; *“I may not wear a mask after fully vaccinated”*; and *“It is my responsibility to get vaccinated so that our whole community is protected”*.

- The questionnaire was developed online using Qualtrics.
- Distributed through the Lumbee Tribe of North Carolina’s official website between the end of July and the beginning of August 2021.

Data Analysis

- Two linear regression models: one to predict the participants' intention to vaccinate themselves, and the other to predict the participants' intention to vaccinate their children.
- We performed statistical analysis using Weka 3.8.6 for all the machine learning models: Logistic regression (LR), Decision Trees (DT), Artificial Neural Networks (ANN), and Random Forests (RF).
- We randomly divide the dataset as 90% (890 for adults and 619 for children) for training and 10% (99 for adults 69 for children) for out-of-sample testing set.

Summary Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
Fully vaccinated	929	.757	.4293	0	1
Child vaccination	695	.813	.390	0	1
Testing frequency	997	3.056	2.271	0	10
Asian	1208	.024	.153	0	1
Black	1208	.104	.305	0	1
Hispanic	1208	.035	.183	0	1
Indigenous	1208	.1965	.397	0	1
White	1208	.499	.500	0	1
Age≥30	1208	.589	.492	0	1
Female	1208	.420	.477	0	1
Annual household income < \$25,000	981	.201	.401	0	1
College degree or above	1002	.520	.500	0	1
Employed (full and part-time)	1208	.702	.458	0	1
Have insurance	963	.872	.334	0	1

Multiple Regression results

Table 2: Multiple Linear Regressions

Variables	OR Self-vaccination	OR Children vaccination
test frequency	1.140* (0.075)	1.235*** 0.073
vaccine attitude: social responsibility	7.830*** (3.514)	1.623*** 0.225
paid leave	7.122*** (3.408)	2.182** 0.685
flu shot	0.236*** (0.070)	0.333*** 0.083
black	0.159*** (0.064)	
Indegious	0.119*** (0.037)	
knowledge_correct	0.641*** (0.087)	
required	1.946* (0.670)	
vaccine attitude: more relaxed with vaccine	0.565** (0.140)	
vaccine attitude: returns to normal with vaccine)	0.613** (0.153)	
vacc_sign_up	0.523*** (0.116)	
vacc_time_place	1.549** (0.335)	
vacc_info		0.671*** (0.096)
insurance		4.341*** (1.500)
medical provider		0.610* (0.156)
faith leader		0.214*** (0.109)
constant	23.916*** (21.032)	2.283 (1.715)
Observations	879	623
Pseudo R2	0.377	0.231

Standard errors in parentheses
 *** p<0.01, ** p<0.05, * p<0.1

Multiple Regression Results

- **Significant predictors for both self-vaccination and children vaccination intention:** testing frequency, vaccine altitude (social responsibility), vaccine reason: paid leave, flu shot
- **Additional significant predictors for self vaccination intention:** black, indegious, knowledge_correct, vaccination required by the employer, more relaxed once being vaccinated, once vaccine is given to most of the population, our life will return to normal, the difficulty to sign up for a vaccine, the difficulty to find a convenient time and place to be vaccinated
- **Additional significant predictors for children vaccination intention:** insurance, information resource: medical_provider, information resource: faith leader, the difficulty to get information about vaccine

Machine Learning Results

- Model's performance to Predict self-vaccination



	LR	DT	RF	ANN
Accuracy	91.2%	94.0%	96.0%	95.3%
Precision	95.8%	96.3%	96.0%	97.2%
Recall	94.7%	97.3%	99.8%	97.7%
F-Measure	95.2%	96.8%	63.4%	97.5%
MCC	37.0%	51.2%	63.4%	62.2%
True positive rate	94.7%	97.3%	99.8%	97.7%
False positive rate	55.1%	49.3%	55.1%	37.7%
Area under the ROC curve	76.7%	77.7%	95.9%	92.4%
Area under the PRC curve	96.7%	95.7%	99.6%	99.3%

Machine Learning Results

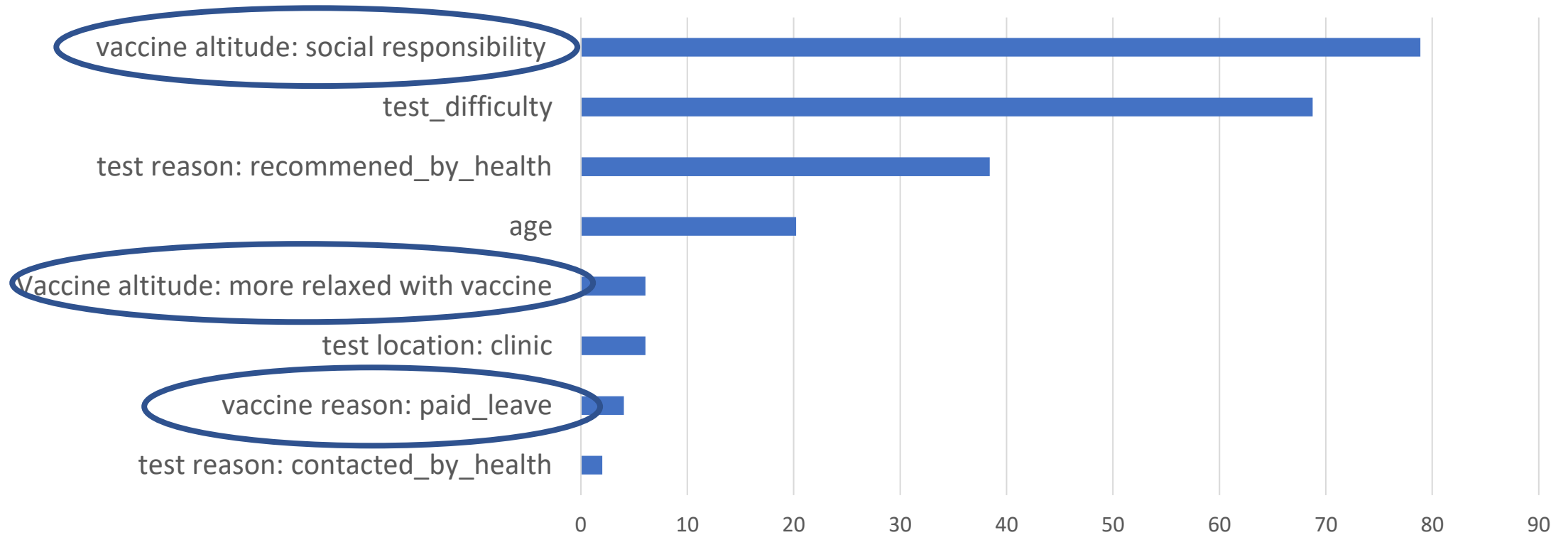
- Model's performance to Predict Children Vaccination tendency



	LR	DT	RF	ANN
Accuracy	82.7%	81.5%	86.5%	83.0%
Precision	88.8%	87.4%	86.2%	87.9%
Recall	90.2%	90.4%	99.3%	91.8%
F-Measure	89.5%	88.9%	92.3%	89.8%
MCC	40.6%	34.6%	46.8%	38.8%
True positive rate	90.2%	90.4%	99.3%	91.9%
False positive rate	50.8%	57.9%	70.6%	56.3%
Area under the ROC curve	79.1%	65.0%	87.5%	82.6%
Area under the PRC curve	91.6%	84.7%	97.0%	94.5%

Feature Ranking Results

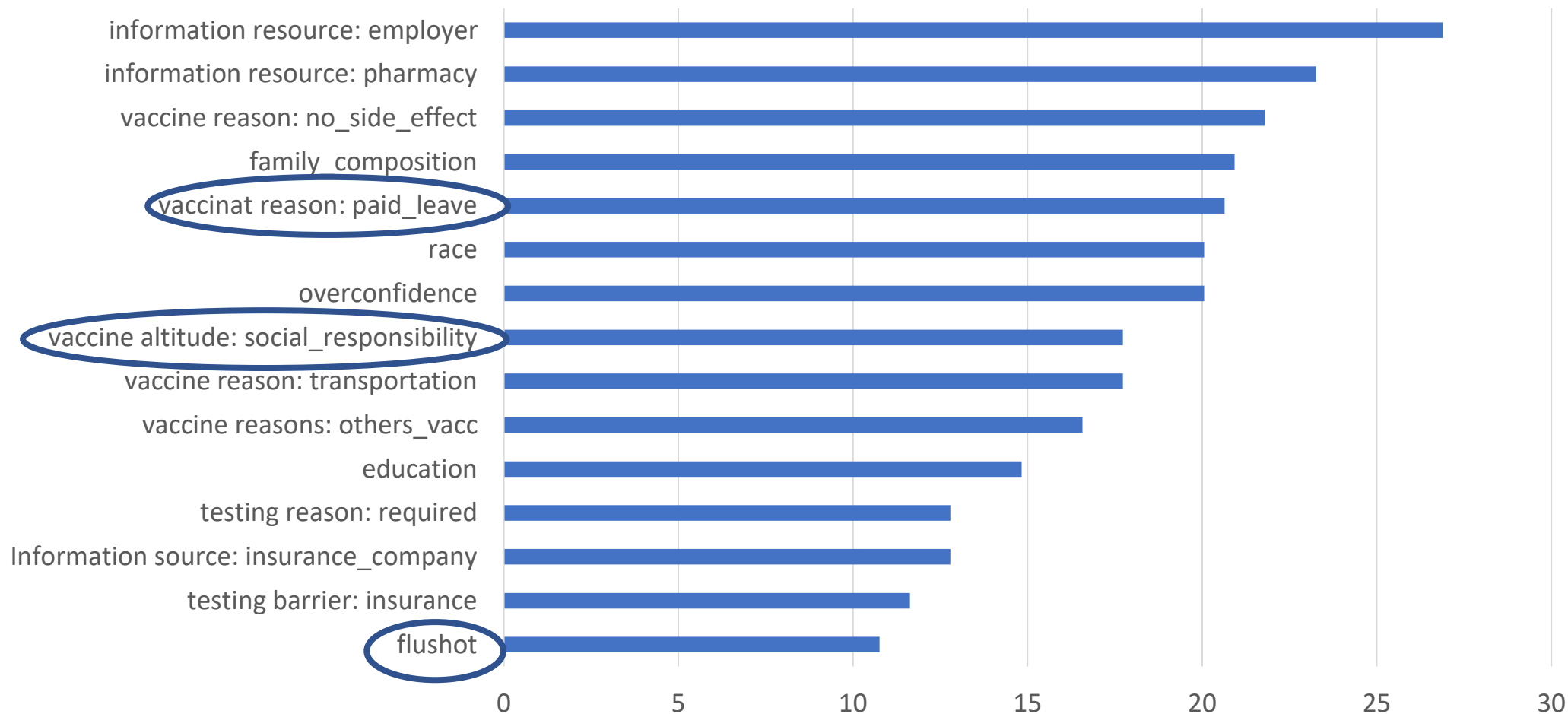
Figure 1: Feature importance to predict self vaccination intention by RF model



Note: The circled features are the significant factors by linear regression model.

Feature Ranking Results

Feature importance to predict children vaccination intention by RF model



Note: The circled features are the significant factors by linear regression model.

Main Results

- Social responsibility and paid leave are strong predictors of people's vaccination decisions for themselves and their children.
- Although test frequency and flu shot are significant in linear regression models, they were not recommended by Machine Learning Models.
- The sources of information where people learn about COVID-19 vaccines have a strong influence in their decision making for vaccinating their children.

Policy Implications

- Allocate more resources to communities with a lower testing rate
- Leverage existing testing sites to educate and promote the social responsibility of vaccination
- Incentivize employers to offer paid leave to their workers
- Invest in public health information structure to make the learning more accessible.