johN-Kioko /
Phase-3-Project-John-Kioko

<> Code   Issues   Pull requests   Actions   Projects   Security   Insights

Vaccine Prediction Model

☆ **0** stars   ⑂ **0** forks   ⊙ **1** watching   〽 Activity

⊕ Public repository

⑂ main ▾   ⋯

⑂ Branches   ⬙ Tags

⬛ **johN-Kioko** Merge branch 'main' of https://github.com/johN-Kioko/Ph...   ...   ✓ 20 hours ago   ⟳ **12**

View code

README.md   ✎

Phase 3 Project John Kioko

Vaccine Prediction Model

1. INTRODUCTION

An NGO in the health sector, wants to learn about the trends in the vaccination space. In the year 2009, a pandemic caused by the H1N1 influenza virus, colloquially named "swine flu," swept across the world. Researchers estimate that in the first year, it was responsible for between 151,000 to 575,000 deaths globally. The Gates foundation wants to find out if they can forecast vaccination of an individual based on specific parameters.

3. BUSINESS PROBLEM

Research question. Can one predict whether a person got seasonal flu or H1N1 vaccine using information they shared about their backgrounds, opinions, and health behaviors?

Main Objective This project aims at getting to know whether a person has received the seasonal flu vaccine or H1N1 using machine learning techniques.

3. DATA UNDERSTANDING

The data in use is from Data driven made up of 26707 rows and 36 columns(12 categorical columns and 24 are numerical.) Namely:

• 'respondent_id'- Unique id • 'h1n1_concern'- the concern one has about the virus. • 'h1n1_knowledge'- knowledge they have about the H1N1 virus. • 'behavioral_antiviral_meds'- If they believe in anti-vaccination. • 'behavioral_avoidance'-do they avoid roaming in public. • 'behavioral_face_mask'- do they wear a face mask. • 'behavioral_wash_hands'- do they regularly wash their hands. • 'behavioral_large_gatherings'- do they tend to be in gatherings. • 'behavioral_outside_home'- are they usually outdoors. • 'behavioral_touch_face'- do they touch their faces often. • 'doctor_recc_h1n1'- • 'doctor_recc_seasonal', • 'chronic_med_condition', • 'child_under_6_months', • 'health_worker', • 'health_insurance', • 'opinion_h1n1_vacc_effective', • 'opinion_h1n1_risk', • 'opinion_h1n1_sick_from_vacc', • 'opinion_seas_vacc_effective', • 'opinion_seas_risk', • 'opinion_seas_sick_from_vacc', • 'age_group'- their age group. • 'education'- level of education • 'race'- their race • 'sex' - their gender • 'income_poverty'- • 'marital_status'- whether they are married or not. • 'rent_or_own'- if they rent or own a house. • 'employment_status'- whether they are employed • 'hhs_geo_region', • 'census_msa'- geographical region • 'household_adults'- number of adults in the house. • 'household_children'-number of children in the house. • 'employment_industry'-industr of employment. • 'employment_occupation'- what they do for a living.

Data Preparation.

• Loading the data. The data set was loaded into the notebook. A data frame was then created and displayed to show content of the data as well as how the variables relate to each other. • Cleaning data. The data was analyzed, checked for duplicates and missing values. Missing values were dropped and irrelevant columns were dropped as well.

Exploratory analysis.

Visualizations were created to show distribution of the variables in our data set. Multivariate analysis was also done to show relationships between some variables. Correlations were also studied between the variables.
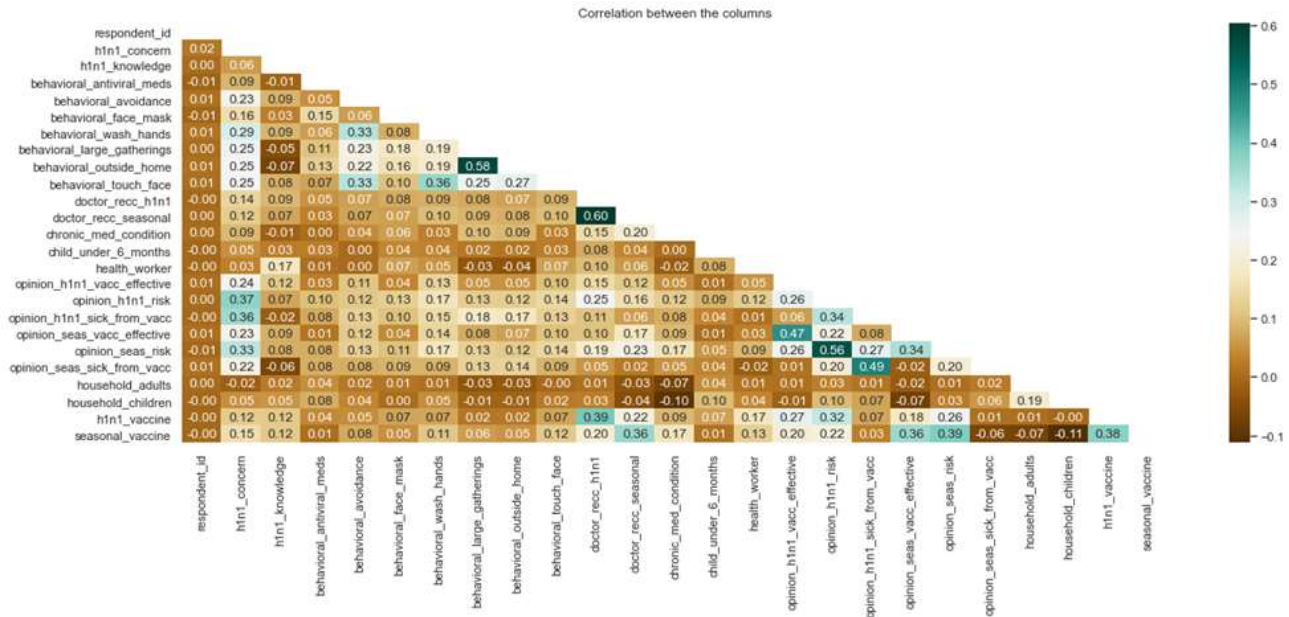
The lighter values indicate a high value of correlation while the darker values indicate low value of correlation.

4. MODELLING

The models that were used are: • Random Forest Classifier. • Logistic Regression Model. • Decision Tree Classifier model. • KNN Model

6. CONCLUSION

• The best working model was the logistic regression with an accuracy of 83.5% . • The accuracy of the model aims in learning the demographics of people to work on to get vaccinated. • People with higher education have a more likelihood of getting the vaccine. • People in the health care industry have a higher chance of getting the vaccine. • Most people who have a more likelihood to get the vaccine are over 65 years.

8. RECOMMENDATIONS

• The foundation should create more awareness by using multiple channels such as social media and websites in order to reach people below 65 years as well. • The foundation should partner with community organizations to provide education about vaccination. • They should perform further investigation on what parameters that need to be added to improve model accuracy

Releases

No releases published
[Create a new release](#)

Packages

No packages published
Publish your first package

Languages

- **Jupyter Notebook** 100.0%