





 Rgmoogachiri / DSC-phase-3-project




[Code](#) [Issues](#) [Pull requests](#) [Actions](#) [Projects](#) [Wiki](#) [Security](#) [Insights](#) [Settings](#)




☆ 0 stars 🍴 1 fork 👁 1 watching 🌿 4 Branches 🏷 0 Tags 🔄 Activity

🌐 Public repository

🌿 main ▾ 🌿 4 Branches 🏷 0 Tags 🌿 🔍 Go to file t Go to file + Add file ▾ Code ⋮

 Rgmoogachiri	Update README.md	3 minutes ago	⋮	🕒
📁 .ipynb_checkpoints	read me & visualization	6 minutes ago		
📁 Data	streamlined model notebook	2 days ago		
📁 Visualizations	read me & visualization	6 minutes ago		
📄 H1N1_and_seasonal_flu_vaccines_...	read me & visualization	6 minutes ago		
📄 README.md	Update README.md	3 minutes ago		
📄 Visualization1.png	added ROC plot	1 hour ago		

 README ✎ ☰

Predicting H1N1 Vaccination Status

Overview

As the world struggles to vaccinate the global population against COVID-19, an understanding of how people's backgrounds, opinions, and health behaviors are related to their personal vaccination patterns can provide guidance for future public health efforts. Such findings can guide policymakers and public health professionals develop public health efforts to improve vaccine uptake to mitigate spread of preventable communicable diseases.

This project utilises data from a United States' conducted National 2009 H1N1 Flu Survey to predict whether someone reviewed H1N1 flu vaccines. Gaining deeper insights into how these attributes correlate with individual vaccination behaviors can offer valuable direction for upcoming public health initiatives.

Business Understanding

The National 2009 H1N1 Flu Survey data was downloaded [DrivenData](#) and the purpose of this project is to use data to forecast whether or not a person received the H1N1 flu vaccination, using features such as social, economic, and demographic background, opinions on risks of illness and vaccine effectiveness, and behaviors towards mitigating transmission, etcetera. The findings would also be applicable for use by the Kenya's Ministry of Health to more effectively target public health initiatives that boost vaccination rates and localise for other communicable diseases like influenza.

Business Objectives

1. Our objective is to develop a predictive model to identify individuals who are likely to exhibit vaccine hesitancy. Our goal is to provide insights that can inform targeted vaccination campaigns and interventions aimed at addressing vaccine hesitancy and increasing vaccine uptake rates.
2. Identify common factors associated with increased uptake of vaccines. Understanding these factors can inform the development of targeted interventions and public health strategies aimed at promoting vaccination uptake and improving overall immunization rates
3. Examine the influence of socio-economic factors, such as income and education, on vaccination decisions.

Data Description

- h1n1_concern - Level of concern about the H1N1 flu.
- 0 = Not at all concerned; 1 = Not very concerned; 2 = Somewhat concerned; 3 = Very concerned.
- h1n1_knowledge - Level of knowledge about H1N1 flu.
- 0 = No knowledge; 1 = A little knowledge; 2 = A lot of knowledge.
- behavioral_antiviral_meds - Has taken antiviral medications. (binary)
- behavioral_avoidance - Has avoided close contact with others with flu-like symptoms. (binary)
- behavioral_face_mask - Has bought a face mask. (binary)
- behavioral_wash_hands - Has frequently washed hands or used hand sanitizer. (binary)
- behavioral_large_gatherings - Has reduced time at large gatherings. (binary)
- behavioral_outside_home - Has reduced contact with people outside of own household. (binary)
- behavioral_touch_face - Has avoided touching eyes, nose, or mouth. (binary)
- doctor_recc_h1n1 - H1N1 flu vaccine was recommended by doctor. (binary)
- doctor_recc_seasonal - Seasonal flu vaccine was recommended by doctor. (binary)
- chronic_med_condition - Has any of the following chronic medical conditions: asthma or an other lung condition, diabetes, a heart condition, a kidney condition, sickle cell anemia or other anemia, a neurological or neuromuscular condition, a liver condition, or a weakened immune system caused by a chronic illness or by medicines taken for a chronic illness. (binary)
- child_under_6_months - Has regular close contact with a child under the age of six months. (binary)
- health_worker - Is a healthcare worker. (binary)
- health_insurance - Has health insurance. (binary)
- opinion_h1n1_vacc_effective - Respondent's opinion about H1N1 vaccine effectiveness.
- 1 = Not at all effective; 2 = Not very effective; 3 = Don't know; 4 = Somewhat effective; 5 = Very effective.
- opinion_h1n1_risk - Respondent's opinion about risk of getting sick with H1N1 flu without vaccine.
- 1 = Very Low; 2 = Somewhat low; 3 = Don't know; 4 = Somewhat high; 5 = Very high.
- opinion_h1n1_sick_from_vacc - Respondent's worry of getting sick from taking H1N1 vaccine.
- 1 = Not at all worried; 2 = Not very worried; 3 = Don't know; 4 = Somewhat worried; 5 = Very worried.
- opinion_seas_vacc_effective - Respondent's opinion about seasonal flu vaccine effectiveness.
- 1 = Not at all effective; 2 = Not very effective; 3 = Don't know; 4 = Somewhat effective; 5 = Very effective.
- opinion_seas_risk - Respondent's opinion about risk of getting sick with seasonal flu without vaccine.
- 1 = Very Low; 2 = Somewhat low; 3 = Don't know; 4 = Somewhat high; 5 = Very high.
- opinion_seas_sick_from_vacc - Respondent's worry of getting sick from taking seasonal flu vaccine.
- 1 = Not at all worried; 2 = Not very worried; 3 = Don't know; 4 = Somewhat worried; 5 = Very worried.
- age_group - Age group of respondent.
- education - Self-reported education level.
- race - Race of respondent.
- sex - Sex of respondent.
- income_poverty - Household annual income of respondent with respect to 2008 Census poverty thresholds.
- marital_status - Marital status of respondent.
- rent_or_own - Housing situation of respondent.
- employment_status - Employment status of respondent.

- `hhs_geo_region` - Respondent's residence using a 10-region geographic classification defined by the U.S. Dept. of Health and Human Services. Values are represented as short random character strings.
- `census_msa` - Respondent's residence within metropolitan statistical areas (MSA) as defined by the U.S. Census.
- `household_adults` - Number of other adults in household, top-coded to 3.
- `household_children` - Number of children in household, top-coded to 3.
- `employment_industry` - Type of industry respondent is employed in. Values are represented as short random character strings.
- `employment_occupation` - Type of occupation of respondent. Values are represented as short random character strings

Feature Engineering

i) `behavior_score`

- By adding up all the behavioral variables, create a variable that shows how much a person has done behaviorally to avoid the virus, aside from getting vaccinated. These are all binary columns, where 1 denotes YES, indicating that the individual has taken a step to lower their risk of getting the flu. A higher score, calculated by adding the values of these columns, indicates a more circumspect and flu-conscious person.

MODELING

Data Preprocessing before model training

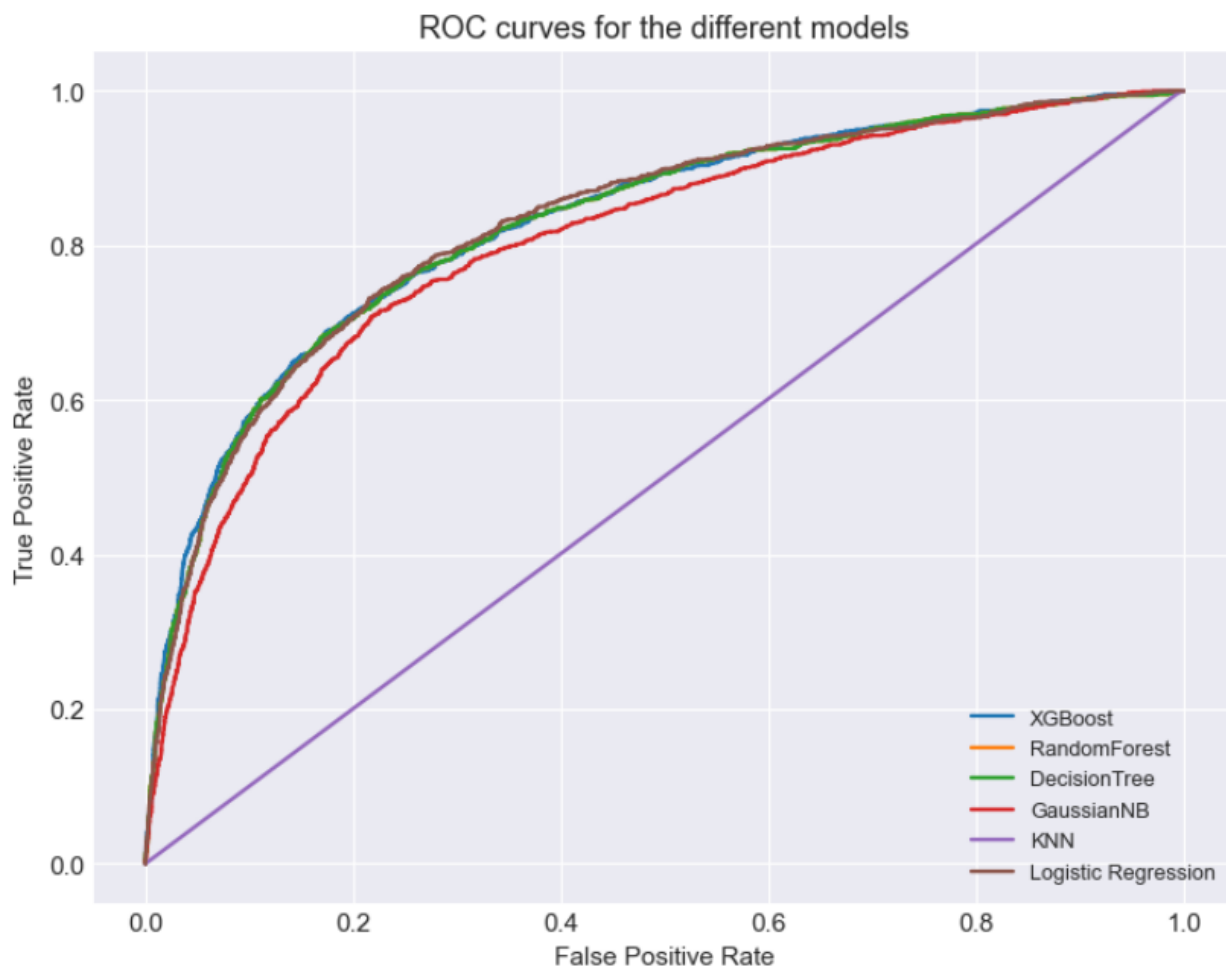
- The data preprocessing techniques such as:
- Imputing Missing Values.
- One Hot Encoding of Categorical variables.
- Ordinal Encoding of ordinal variables.
- We perform the preprocessing and transformations by fitting and transforming our functions on the training data, then simply transforming on the test data in an attempt to prevent **data leakage**.

Summary of all the tuned models in terms of ROC-AUC Score:

1. Logistic Regression - **0.7735**
2. Decision Tree - **0.7814**
3. KNN - **0.7811**
4. Naive Bayes - **0.7528**
5. Random Forest - **0.8288**
6. XGBoost - **0.8405**

- XGBoost is therefore our **best and final model** with the best Accuracy of **0.8405**, as compared to all other models

ROC Curve



Conclusions

Features that were most important in predicting whether someone was vaccinated or not included:

1. Opinion on H1N1 vaccine effectiveness - `opinion_h1n1_vacc_effective`
2. Doctor's recommendation - `doctor_recc_h1n1`
3. H1N1 concern - `h1n1_concern`
4. Gender - `sex`
5. H1N1 Knowledge - `h1n1_knowledge`
6. Perceived side-effects from H1N1 vaccine - `opinion_h1n1_risk`
7. Chronic medical condition - `chronic_med_condition`
8. Number of adults in a household - `household_adults`

Recommendations

1. Dispel vaccine myths and promote preventative measures against the flu.
2. Identify high-risk groups particularly those with lower levels of education and individuals expressing low concern

Releases



No releases published

[Create a new release](#)

Packages

No packages published
[Publish your first package](#)

Contributors 2

-  Rgmoogachiri Richard Gachiri Muriithi
-  dave-mainag David Githaiga

Languages

- Jupyter Notebook 100.0%