☰ ⬤ Rgmoogachiri / **DSC-phase-3-project**

🔍 Type `/` to search   >_   +▾   ⊙   ⇄   ✉   �⬤

<> **Code**   ⊙ Issues   ⇄ Pull requests   ▷ Actions   ⊞ Projects   📖 Wiki   ⚠ Security   📈 Insights   ⚙ Settings

🔲 **DSC-phase-3-project**   Public

⎇ main ▾   ⇄ **4 Branches**   ⬚ **0 Tags**   ⇄   🏷   🔍 Go to file   t | Go to file | +   About file ▾ | Code | ⋯   ⚙

🔲 **Rgmoogachiri** Update RE... 9b2bd12 · 1 minute ago		🕐 **18 Commits** ⋯
📁 .ipynb_checkpoints	final	7 minutes ago
📁 Data	streamlined model noteb...	2 days ago
📁 Visualizations	visualizations	7 hours ago
📄 GIt repository pdf....	visualizations	7 hours ago
📄 H1N1_and_season...	final	7 minutes ago
📄 H1N1_and_season...	final	7 minutes ago
📄 Predicting - HlN1-...	Final notebook	2 hours ago
📄 README.md	Update README.md	1 minute ago
📄 Visualization1.png	visualizations	7 hours ago

About

No description, website, or topics provided.

📖 Readme

⎓ Activity

☆ 0 stars

👁 1 watching

⑂ 1 fork

Releases

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

Packages

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

Contributors 2

📖 **README**   ✏️ ☰

Rgmoogachiri Richard Gachiri Muriithi

dave-mainag David Githaiga

Predictive Modeling for H1N1 Vaccine Uptake

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 dieases.

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

Languages

● Jupyter Notebook 100.0%

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 dieases like influenza.

Research Questions

1. How may false information regarding vaccines be effectively refuted to encourage factual knowledge and vaccination acceptance of the flu?

2. Which major groups are most likely to be deceived about the flu shot, especially those with lower levels of education? How might focused messaging be developed to allay their fears?

3. How can resource allocation for flu vaccination programs be optimized by the use of predictive modeling, guaranteeing effective distribution to high-risk populations and regions?

4. What programs may be put in place to promote the general adoption of preventative measures like donning face masks, avoiding crowds, and washing your hands properly?

5. How can cooperation with medical professionals be improved to encourage routine physical examinations and support physicians' recommendations for flu shots among high-risk individuals?

Business Objectives

2/13/24, 8:51 PM

Rgmoogachiri/DSC-phase-3-project

1. Analyze how well various communication tactics work to debunk vaccine misconceptions and spread factual information about getting the flu shot.

2. Determine which high-risk groups to target with messaging and interventions, especially those who are less educated and show little concern for the H1N1.

3. Create a predictive model that makes the best use of behavioral, epidemiological, and demographic data to allocate resources for flu vaccine campaigns.

4. Create and put into action policies and initiatives that will increase people's adherence to preventive measures like hand hygiene, avoiding crowds, and wearing face masks.

5. Create efficient channels of communication with medical professionals to enable routine physical examinations and improve physician recommendations for influenza shots in high-risk patients.

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)

https://github.com/Rgmoogachiri/DSC-phase-3-project/tree/main

4/10

- 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.

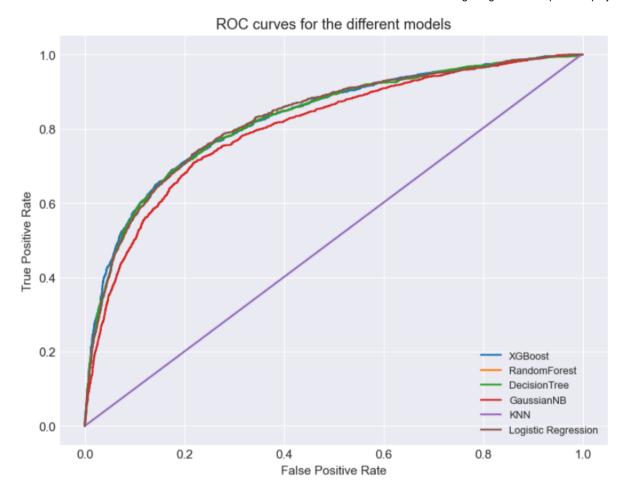# MODELING

Data Preprocessing before model training

- The data preprocessing technqiues 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

ROC curves for the different models

Conclusions

Features that were most important in predicting H1N1 vaccine uptake::

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