Robinzulu /
H1N1-Vaccine-And-Seasonal-Vaccine-Predict

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

⭐ 0 stars   ⑂ 0 forks   ⊙ 1 watching   〜 Activity

🌐 Public repository

⑂ master ▾

⑂ Branches   🏷 Tags

ROBINZULU First commit   ...   8 minutes ago   🕘 1

View code

☰   **README.md**   ✎

Phase 3 Group Project

Table of contents

Business Problem Understanding

Introduction

The world has recently experienced the impact of major flu outbreaks like the COVID-19, Swine Flu (H1N1) and the Avian Flu(H5N1). The effect of any flu outbreak depends on the type of flu and its respective variants, the population demographics like age, and other underlying health conditions of the individual. In the US, seasonal flu places a substantial burden on the health of people in the United States each year. CDC estimates that flu has resulted in 9 million – 41 million illnesses, 140,000 – 710,000 hospitalizations and 12,000 – 52,000 deaths annually between 2010 and 2020. The results obtained from CDC website also shows that the vaccination rates have remained low with an overall average of 57.8% as of 2022.

Despite the availability and effectiveness of flu vaccines, there are still significant portions of the population who choose not to get vaccinated. To address this problem, it is crucial to investigate the reasons behind these decisions and identify the key factors driving individuals' opinions, perceptions, and behaviors related to flu vaccination.

Target Variables

`h1n1_vaccine` - Whether respondent received H1N1 flu vaccine. (0 = No; 1 = Yes)

`seasonal_vaccine` - Whether respondent received seasonal flu vaccine.(0 = No; 1 = Yes)

Business goals

The main aims is to analyze and uncover the factors that influence individuals' decision-making processes regarding getting vaccinated against flu. By gaining insights into these factors,governments and healthcare organizations can develop targeted strategies and interventions to increase vaccination rates and improve public health outcomes.

Project goals

The objective of this project is to understand how the following factors affect H1N1 & Seasonal flu vaccination rates.

- Opinions & Perceptions
- Demographics - Age, Education, Employment status & Income levels
- Behavioral - Handwashing, masking and avoiding large crowds.
- Doctors Recommendation

Potential Project Outputs

possible deliverables from this project may include;

- **Identification of key factors influencing vaccination decisions** - This information can help in guiding the development of targeted campaigns and interventions.
- **Segmentation of the target population:** Segmentation can help tailor messages and interventions to specific groups, considering their unique characteristics and needs.
- **Prediction models:** These model will assist in forecasting future vaccination rates and evaluating the effectiveness of intervention strategies.

The outputs of this project will provide actionable insights for businesses and healthcare organizations to increase vaccination rates.

Data Understanding

The dataset used in our project had the following columns/features. Our initial data had 26707 rows and 37 columns

`seasonal_vaccine` - Whether respondent received seasonal flu vaccine.(0 = No; 1 = Yes)

`h1n1_vaccine` - Whether respondent received H1N1 flu vaccine. (0 = No; 1 = Yes)

`respondent_id` - Unique and random identifier.

`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.

Data processing

This stage focussed on data cleaning in preparation for analysis. This included;

- Checking for duplicates.
- Checking for missing values. Columns with very high number of missing values were dropped
- Replacement of categorical variables with numeric values
- Populating the NULLs/NaNs with suitable substitutes

Exploratory Data Analysis

below is an overview of some of the factors affecting vaccination rates.
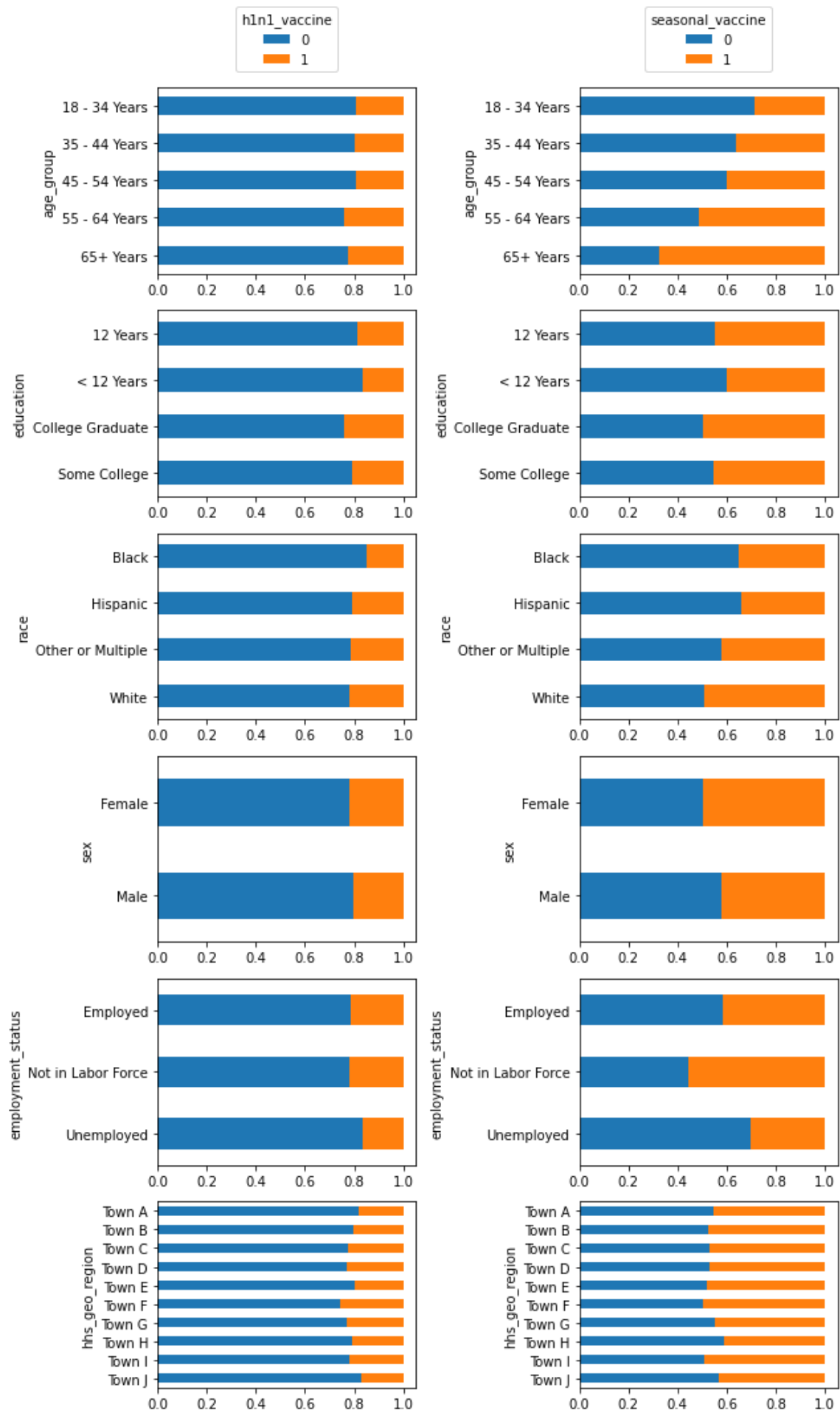
Elderly people were more likely to receive vaccines compared to the rest of the age groups. This can be attributed to decreased immunity in the elderly people resulting in high risk of flu transmissions. On the contrary, younger people have a stronger immunity making them less prone to contracting the flu.

College graduates made the majority of those who received vaccination probably due to increased awareness on the importance of vaccination compared to other educational levels.

Black people are unlikely to receive vaccination compared to other races probably due to cultural differences.

Relatively equal number of men and women received H1N1 vaccine. However, slighty more women received Seasonal Vaccine compared to men. There is not enough evidence to support that women have higher probability of being vaccinated. The slight difference noted can be attributed to errors such as misclassification.

People not in labor force, that is, individuals who are not employed and are not actively seeking employment eg retirees have higher chances of getting vaccinated compared to other workforce. This can be attributed to more awareness of the risks associated with influenza and are more likely to prioritize vaccination to protect their health.

Modelling

We developed several models using different features in the dataset in order to predict the likelihood of an individual going for H1N1 vaccine or Seasonal Flu Vaccines. The following models were built:

1. KNN
2. Decision Trees
3. Logistics Regression
4. Random Forest
5. Naive Bayes

The models had different levels of accuracy. ROC/AUC curves Logistic Regression and the RandomForest had the highest ROC and Accuracy scores. The Logistic Regression had a score of 81% and 84% for H1N1 and Seasonal Vaccine respectively while the Random Forest Classifier had scores of 80% and 83% for H1N1 and Seasonal Vacc respectively. Since the Logistic Regression was the baseline model, we chose to go with Random Forest Classifier which has ROC score of 83%.

To improve the models perfomance we decided to do Hyper Parameter tuning using GridSearch.

Training the model

The model was trained with training features (X_train) and training labels (y_train). The model was given some new data it hasn't seen before (X_test) to evaluate how well it classifies the new data. The data was split into training and testing. 80% of the dataset was used for training and the remaining 20% for testing.

MinMaxScaler was used to transform the dataset by scaling each feature to a range between 0 to 1.

Conclusion

The following top 4 features had the most effect on vaccine uptake

i) Respondent's opinion about risk of getting sick with seasonal flu without vaccine

ii) Age group

iii) Respondent's opinion about seasonal flu vaccine effectiveness

iv) Respondent's opinion about risk of getting sick with H1N1

Recommendations

-**Mass and grassroot sensitization:** Stakeholders should invest in mass and grassroot sensitization about H1N1 and the seasonal flu vaccine led by healthcare professionals. They should be educated on the vaccine efficacy and dangers of not getting it.

-**Healthcare Provider Recommendations:** Encouraging healthcare providers to actively recommend and offer H1N1 and seasonal vaccination to their patients can significantly impact vaccine uptake. The stakeholders can provide training and resources to healthcare professionals to increase their knowledge and confidence in recommending the vaccine.

-**Collaboration and Partnerships:** Collaborating with community organizations, schools, workplaces, and other stakeholders can amplify efforts to promote H1N1 vaccination. Engaging influencers, community leaders, and trusted voices within different communities can help build trust and increase acceptance of the vaccine.

Limitations

The dataset used was obtained eletronically and had some limitations, for example, the geographic location information was encrypted and couldn't be read. We subsituted it with the alphabets A-Z.

Releases

No releases published
Create a new release

Packages

No packages published
Publish your first package

Languages

- **Jupyter Notebook** 100.0%