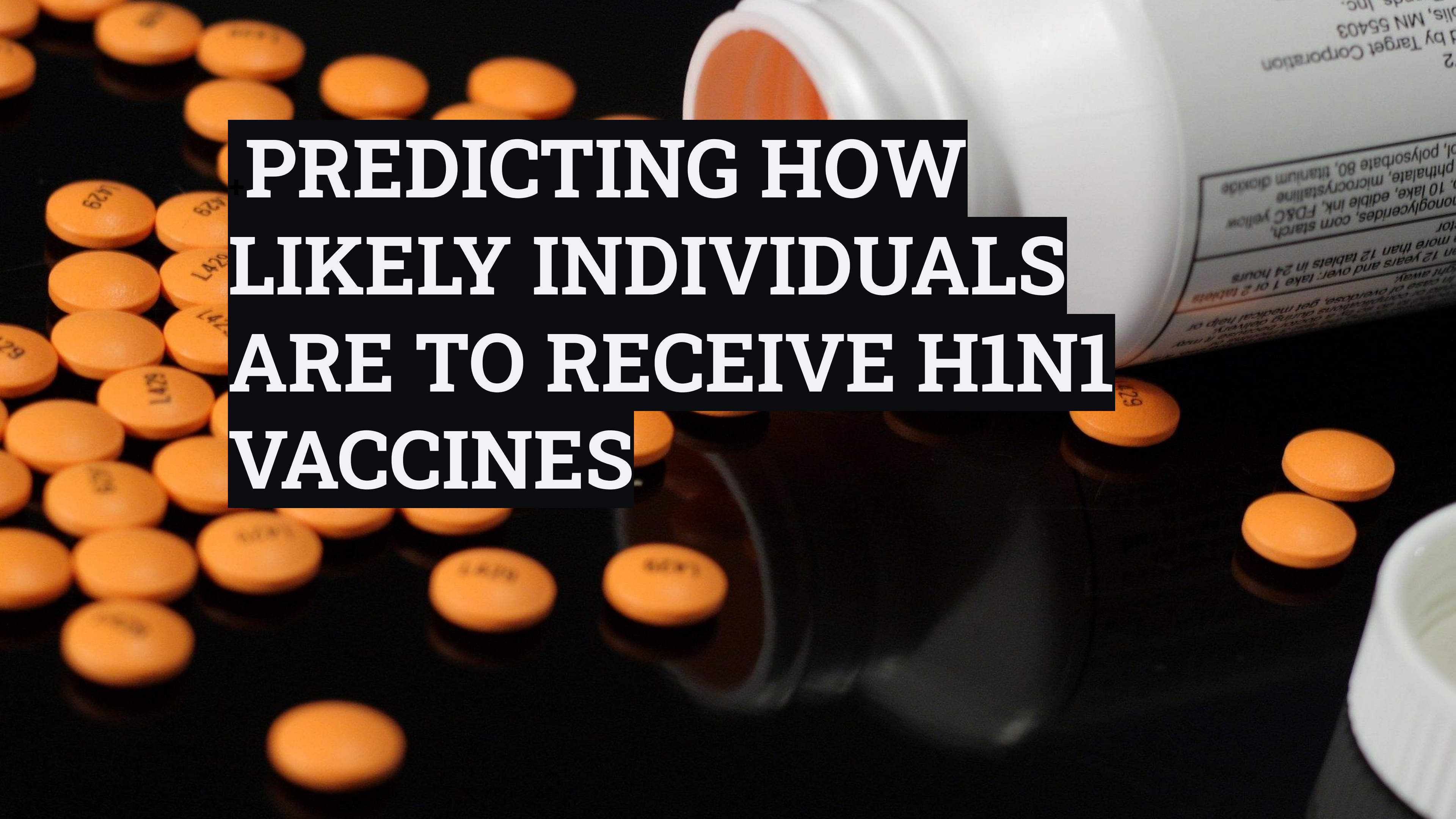
PREDICTING HOW LIKELY INDIVIDUALS ARE TO RECEIVE H1N1 VACCINES



Introduction

The prevention of the spread of infectious illnesses depends heavily on vaccination, which is a critical component of public health. However, not every person reacts to immunizations in the same way. Others may experience side effects including fever, exhaustion, or soreness, while some may not exhibit any negative effects at all.

Our objective is to create a reliable classification model that can correctly evaluate a person’s response to the h1n1 vaccine based on specific characteristics. The outcomes of this initiative will help healthcare practitioners make decisions about the delivery of vaccines and will offer useful information.

Content

Business Understanding

Exploratory Data Analysis

Modeling

Conclusion

Recommendations and Future Steps

Business Context

Problem Statement - Vaccination has become a key public health measure that is used to fight and in most cases curb infectious diseases. The goal of this project is to build a model that can predict the response of individuals to a vaccine based on certain features, such as age, sex, health status, and their knowledge on H1N1 vaccine. This information can help healthcare professionals make informed decisions about who should receive the vaccine and how to best manage its administration.

Main Objective

To build a classification model that can predict the response of individuals to a vaccine based on certain features.

Specific Objectives

Identify which factors affect individuals’ response to vaccines.

Accurately predict the general response of individuals’ response to a new vaccine.

Build a classification model that accurately predicts the response of individuals to new vaccines and provides actionable insights on how to reduce the spread of contagious infections.

Metric for Success

The final model will be considered a success if it has an accuracy and f1 score of not less 75%. The goal is to make as accurate as possible predictions, that is why the choice of success metrics is the accuracy score and f1 score.

Exploratory Data Analysis

Univariate Analysis

The aim of this analysis is to establish the proportion of the target variable h1n1_vaccine. That is, what percentage of the population that was surveyed got the vaccine. A bar graph representing this information is shown below.

Univariate Analysis (cont..)

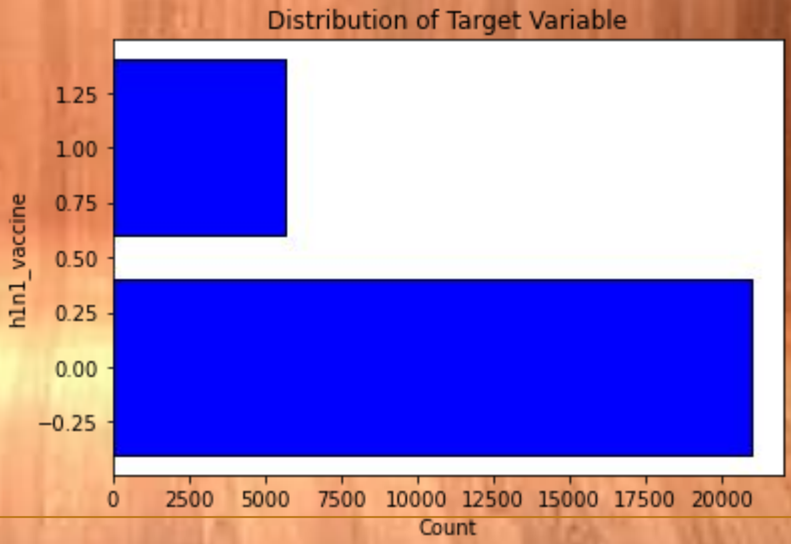
The graph shows that 5000

people

out 20000 got the vaccine.,

that is

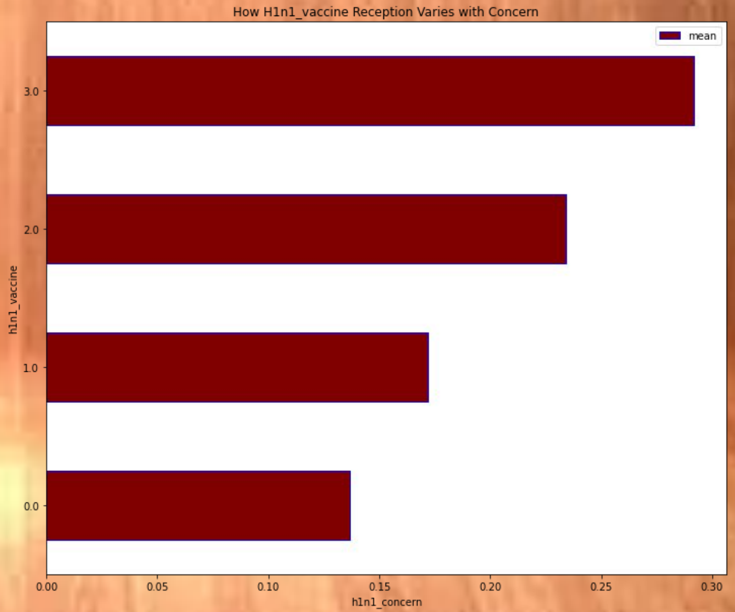
20% of the population.



Bivariate Analysis

How does level of concern affect their choice to get the vaccine?

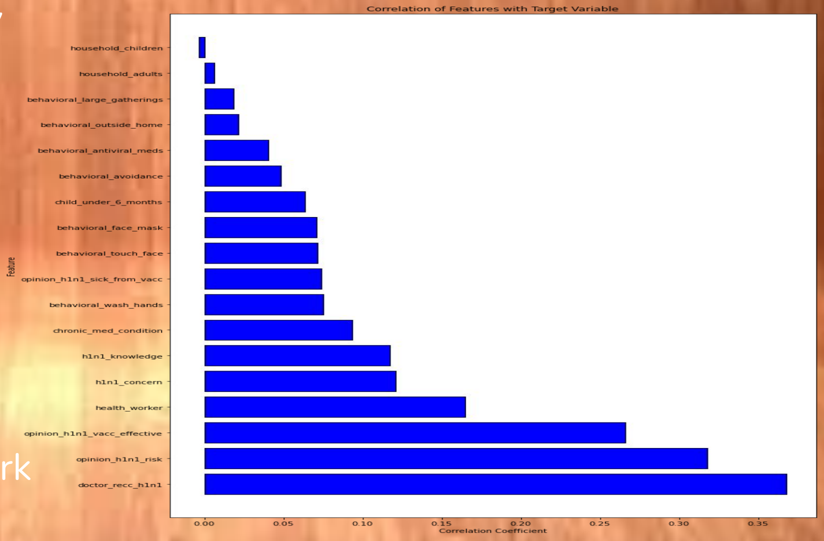
It is obvious that even though the majority
of people do not
receive the vaccine,
those who are more concerned are more
Likely to do so.
Therefore, the more concerned a person is
about their health,
The more likely they are to get the vaccine.



Bivariate Analysis

How are each of the other variables related to h1n1_vaccine?

The correlation coefficient is generally not high for most of the variables. However, there are some that are More correlated to the target variable like doctor_recc_h1n1, Opinion_h1n1_vacc_effective. These seem like great variables to work with to increase prediction accuracy.



Modeling

Model 1 - Baseline Model (Logistic Regression)

Model 1 - Baseline Model (Logistic Regression)

This simple logistic regression model was fitted on the training dataset and this resulted in an accuracy score of 81% and an f-score of 43%. This is not the best model to use for predictions because with an f1-score that low, the model is susceptible to making more wrong predictions than right ones.

Model 2 - K-Nearest Neighbour

Before this model was fitted, feature selection was performed in an attempt to improve the outcome. Features that were not highly correlated with the target variable were dropped and only 10 were left to be used in training and fitting the model. The outcome was an accuracy score of 79% and an f1-score of 36%. This model was performing worse than the previous one.

Model 3 - Decision Tree

The decision tree was fitted to the dataset and the outcome was san accuracy score of 78% and an f1-score of 36%. This model was also not doing well because of the low f1- score.

Hyperparameter Tuning

This was carried out in order to improve the model performance by introducing hyperparameters and tuning them until the best combination was found and applied to the model. Grid search cv was used to carry this out on the decision tree model and the outcome was a model with 80% accuracy score and 78% f1-score. This model surpassed the success metric before and would therefore be a well performing model in terms of prediction accuracy.

Conclusion

After experimenting with different models using different techniques, the final model (the decision tree), which was tuned by hyperparameters was selected. This is because it met the success criteria by recording an accuracy score of 80% and f1 score of 78%. The first three models did not do so well because they were not complex enough to handle the nature of data presented. This is why while the accuracy score was high, the f1 score remained low. By adjusting the hyperparameters, we were able to find the optimal complexity that balanced overfitting and underfitting.

After experimenting with different models using different techniques, the final model (the decision tree), which was tuned by hyperparameters was selected. This is because it met the success criteria by recording an accuracy score of 80% and f1 score of 78%. The first three models did not do so well because they were not complex enough to handle the nature of data presented. This is why while the accuracy score was high, the f1 score remained low. By adjusting the hyperparameters, we were able to find the optimal complexity that balanced overfitting and underfitting.

Recommendations and Future Steps

More research on other vaccines should be carried out to supplement this ons on h1n1.

Future surveys to be carried out in person.

An improvement in data preprocessing techniques is paramount for future future projects.

A research on why people are averse to vaccination should be carried out.

Thankyou!

Questions?