

Behavioral Insights into H1N1 and Seasonal Flu Vaccine Refusal

by Daniel Pereira De Abreu and Viktoria Stoeva

Individuals of all ages are at risk of developing severe illnesses from the H1N1 virus or seasonal flu. Vaccines prevent specific infectious diseases from spreading uncontrollably by enhancing your immune system. The H1N1 Influenza virus, more commonly known as Swine Flu caused a pandemic in 2009. A vaccine was developed that combatted the spread and limited the overall impact of the disease. Seasonal flu vaccines safeguard individuals from the most common viruses for the upcoming flu season. Yet, there are various ways individuals justify refusing both the H1N1 and seasonal flu vaccines. “The 4 categories are religious reasons, personal beliefs or philosophical reasons, safety concerns, and a desire for more information from healthcare providers.” (Chephra McKee, 2016)

The aim of this report is to provide answers to two questions focusing on individual and group trends in order to increase mass immunity. What behavioral factors can be identified forecasting H1N1 and seasonal flu vaccine refusal among the different demographic groups, and how this insight can be used to increase vaccination rates? What are the underlying social and psychological factors influencing individuals who opt to receive the H1N1 and seasonal flu vaccines, and how can this be utilized to allow for better communication and therefore drive vaccine acceptance? Determining the probability of person receiving the H1N1 and Seasonal Flu vaccine based on specific factors is a multi-classification task. It can be solved using various Machine Learning (ML) models to predict an outcome based on gathered data.

A phone survey was conducted to collect the data on whether candidates had received vaccines for H1N1 or seasonal flu. This information is used as the dependent variables, which need to be predicted based upon the independent variables: demographic questions asked during the survey. The dataset has a total of 38 features with 2 of them being our target predictions. 12 of the remaining features are categorical and the other 26 numerical. There is a data imbalance with every instance having over 50% of the data missing and a skewed class for the H1N1 vaccine. Variables such as age_group, race and hhs_geo_region hold high imbalance which could lead to bias.



The data wrangling and pre-processing steps for the dataset involve cleaning which includes the removal of outliers using quantile ranges and irrelevant instances, alongside type conversion of unique values. Missing data is imputed using mode substitution for categorical data, whereas numerical data is dealt with using mean imputation. The categorical data is encoded using label encoding and scaled with standardisation by removing the mean and scaling to its unit variance. Upon visualising the data, patterns of positive behavioural based features correlating with individuals choosing to take a vaccine can be observed.

References

Chephra McKee (2016) Exploring the Reasons Behind Parental Refusal of Vaccines. Available at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4869767/> (Accessed: 03 March 2023)