Predictive Modeling of H1N1 Vaccination Uptake

Apoorva Anand Kulkarni
Department of Computer Science and Engineering
PES University
Bangalore, India
apoorvakulkarni2001@gmail.com

Alheena Mujeeb
Department of Computer Science and Engineering
PES University
Bangalore, India
alheenamujeeb2002@gmail.com

***Abstract*— Vaccination is a key public health measure to fight infectious diseases. They provide immunization and help to prevent further spread of the disease. Vaccines for H1N1 were first publicly available in the United States in October 2009, when the United States government began a nation-wide vaccination campaign. Understanding the impact of socioeconomic background, behavior and health-related opinions on vaccination patterns can prove to be useful for future public health efforts.**

Keywords—Pandemic, H1N1, Vaccination, Immunization, Machine Learning

I. Introduction

The H1N1 virus was first detected in 2009 and quickly became a significant public health issue, resulting in approximately 700 million cases worldwide. The H1N1 virus is highly contagious in nature and can also potentially be life-threatening if the patient has other comorbidities. The most useful technique to counter this virus is mass immunization and vaccination. Vaccines for H1N1 were first available in the United States in October 2009, when the United States government began a vaccination campaign. Vaccines provide immunization for individuals, and enough immunization in a community can further reduce the spread of diseases through herd immunity. Immunization efforts have had a tremendous impact on the health and well-being of entire populations. Multiple studies have shown that social, economic and demographic background have an impact on vaccination patterns. Apart from this, opinions on risks of illness and vaccine effectiveness, and behaviors towards mitigating transmission also play a significant role. A better understanding of how these characteristics are associated with vaccination can provide guidance for future public health efforts. In this paper, we aim to understand the impact of these factors on vaccination rates and build a predictive model for H1N1 vaccination rates.

II. Review of Literature

There has been in-depth research carried out to understand the various factors that impact vaccination. Studies regarding both H1N1 vaccination and COVID-19 vaccination were analyzed. In one study, the authors attempted to determine the association between sociodemographic factors and vaccine uptake across counties in the United States. [1] COVID-19 related data was sourced from several online databases such as the US Centers for Disease Control and Prevention and the US Census Bureau COVID-19 Site. XGBoost model was applied on the dataset and an accuracy of 62% was achieved. Location, education, ethnicity, income, and household access to the internet were identified as the most critical sociodemographic features. Another paper attempts to identify predictors of willingness to accept a COVID-19 vaccine, and compares them to predictors of previous acceptance of an influenza vaccine among adults in the southeastern US. [2] Attributes related to health and healthcare utilization, COVID-19 testing, results, perceived risk of infection, COVID-19 vaccine willingness, and influenza vaccine acceptance were included. A series of univariate logistic regressions were employed to investigate and compare predictors of respondents' (1) willingness to accept a COVID-19 vaccine, and (2) previous acceptance of the influenza vaccine for the current flu season. This paper establishes a relationship between behaviors indicative of reducing the spread of COVID-19, education, income, recent healthcare utilization and vaccine uptake The goal of this paper studied was to examine state-level features and policies that are most important in achieving a threshold level vaccination rate to curve the effects of the COVID-19 pandemic [3]. Chi-squared Automatic Interaction Detection (CHAID), a Decision Tree algorithm was used to do the same. K-fold cross validation technique was utilized to determine algorithms' generalization error. 88% accuracy and a sensitivity score of 92.5% was achieved. Workplace travels, political affiliation of the governor, and the vaccine mandate in schools were the top three features to predict Covid 19 vaccine uptake. In the fourth paper, the authors investigate factors associated with seasonal and 2009 H1N1 influenza vaccine uptake among middle- and high-school teachers in a rural, low-income setting [4]. Chi-square test was used to correlate seasonal and 2009 vaccine uptake. Six psychosocial variables were assessed to determine participants' attitudes and beliefs regarding seasonal and H1N1 influenza vaccinations. The paper concluded that teachers have a heightened risk of both contracting and spreading influenza to others, which can place their students and families at risk.

The main aim of [5] is to identify factors influencing non-hospital HCP H1N1 influenza vaccine compliance. Surveys were used to assess H1N1 influenza vaccine compliance and examine factors that predicted H1N1 influenza vaccine uptake. Logistic regression was used. Attributes included extent to which H1N1 vaccination was mandated or encouraged, perceived importance of vaccination, access to no-cost vaccine provided on-site, no fear of vaccine side effects, and trust in public health officials when they say that the influenza vaccine is safe. The conclusion was that non-hospital-based HCP versus hospital-based HCP reasons for H1N1 vaccine uptake differed. Targeted interventions are needed to increase compliance with pandemic-related vaccines. In the final paper [6], the authors attempted to identify the various mathematical models used in prediction of H1N1 outbreaks, as well as to compare the usefulness of these models in providing magnitude of H1N1 outbreaks. To do so, data was collected from past influenza pandemics and the 2009 influenza A (H1N1). Statistical methods such as significant t-test were used along with stochastic (probabilistic) and deterministic (compartmental models, non probabilistic). Among all the mathematical models, it was found that the SIER model produced the best results.

III. Proposed Method

The aim of the paper is to predict whether an individual will receive H1N1 vaccination, based on many parameters.The factors taken into account include social, economic and demographic background. Apart from this, opinions on risks of illness and vaccine effectiveness, and behaviors towards mitigating transmission have also been considered.

A. Dataset

Data has been obtained from the National 2009 H1N1 Flu Survey (NHFS). The dataset contains around 26000 observations and 33 attributes. These attributes can be split into 3 categories- Socioeconomic background, Opinions on illness and vaccination and Behaviour.

The attributes taken into consideration are: Level of concern about the H1N1 flu, Level of knowledge about H1N1 flu, consumption of antiviral medications, Has avoided close contact with others with flu-like symptoms, Has bought a face mask, Has frequently washed hands or used hand sanitizer, Has reduced time at large gatherings, Has reduced contact with people outside of the household, Has avoided touching eyes, nose, or mouth, H1N1 flu vaccine was recommended by the doctor, Has chronic medical condition, Has regular close contact with a child under the age of six months, Is a healthcare worker, Has health insurance, Respondent's opinion about H1N1 vaccine effectiveness, Respondent's opinion about risk of getting sick with H1N1 flu without vaccine, Respondent's worry of getting sick from taking H1N1 vaccine, Age group of respondent, Self-reported education level, Race of respondent, Sex of respondent., Household annual income of respondents, Marital status of respondent, Housing situation of respondent, Employment status of respondent, Respondent's residence, Number of adults in household, Number of children in household, Occupation of respondent.

B. Exploratory Data Analysis

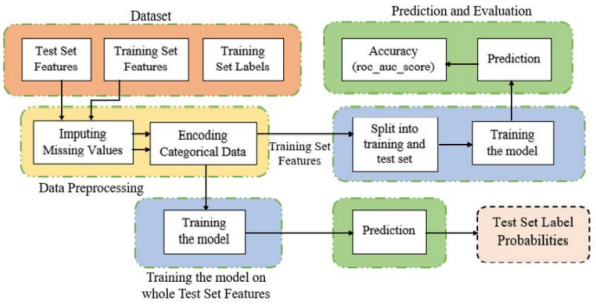
Exploratory data analysis has been carried out on the dataset. Columns with more than 10000 missing values were dropped. In other columns, missing values were imputed using the mode of that attribute. Mode was used in place of mean as most attributes in the dataset are categorical in nature. Apart from this, bar charts were plotted to understand the impact of each explanatory variable on the response variable.

C. Proposed Models and Modeling Technique

In our proposed method, the attributes taken into account are far more extensive as compared to previous work. We also propose to carry out predictive modeling in two stages as compared to only one step in other research.

Stage one is to identify the most important subset of attributes from the set of 33 attributes. To do so, different attribute selection methods can be utilized. We propose to use a decision tree, similar to [3], in order to select the most important predictors.

After doing so, in stage two, predictive modeling can be carried out. The dataset is split into training (80%) and testing (20%) data. Various models can then be used.



System Architecture

The initial proposed approach is to use Logistic Regression as the response variable is binary in nature. Different explanatory attributes can be utilized to predict whether an individual will take the H1N1 vaccine or not. Logistic Regression has been proven to be effective in papers [2] and [5]

Other possible approaches include using machine learning models such as XGBoost, which has been used in [1].

D. Evaluation of result:

There are various evaluation methods available to measure the performance and the quality of the prediction made by the model. In our implementation, we propose to use ROC-AUC score as this metric is useful for binary classification tasks.

References

[1] Cheong Q, Au-yeung M, Quon S, Concepcion K, Kong J Predictive Modeling of Vaccination Uptake in US Counties: A Machine Learning–Based Approach J Med Internet, Res2021;23(11):e33231
URL: https://www.jmir.org/2021/11/e33231
DOI: 10.2196/33231

[2] Burch, A.E., Lee, E., Shackelford, P. et al. Willingness to Vaccinate Against COVID-19: Predictors of Vaccine Uptake Among Adults in the US. J Primary Prevent 43, 83–93 (2022). https://doi.org/10.1007/s10935-02

[3] Syed Muhammad Ishraque Osman, Ahmed Sabit,Predictors of COVID-19 vaccination rate in USA: A machine learning approach, Machine Learning with Applications, 2022, 100408, ISSN 2666-8270, https://doi.org/10.1016/j.mlwa.2022.100408.

[4] Seasonal and 2009 H1N1 influenza vaccine uptake, predictors of vaccination and self-reported barriers to vaccination among secondary school teachers and staff by Lisa M. Gargano, Julia E. Painter, Jessica M. Sales, Christopher Morfaw, LaDawna M. Jones, Dennis Murray, Gina M. Wingood, Ralph J. DiClemente & James M. Hughes

[5] Peiris, JSM, Poon, LLM, Guan, Y. Emergence of a novel swine-origin influenza A virus (S-OIV) H1N1 virus in humans. J Clin Virol 2009;45(3):169–173.

[6] International Journal of Community Medicine and Public Health,,Mutalik AV. Models to predict H1N1 outbreaks: a literature review. Int J Community Med Public Health 2017;