

Flu Shot Learning: Predict H1N1 and Seasonal Flu Vaccines

Machine Learning Project Proposal

APC 2022/2023

Team: Jorge Pais, Pedro Duarte and José Baptista

Class: 1MEECT01 and 1MEECT02

1 Data Set

For this project, we'll use a data set from a DrivenData competition on H1N1 and Seasonal Flu Vaccine administration prediction. The data was taken from the United States National 2009 H1N1 Flu Survey and includes 35 features from each individual regarding behaviour towards transmission mitigation, opinions concerning the viruses/vaccines and social-economic background.

2 Project Idea

With the recent COVID-19 epidemic, vaccination played a major role in the immunization of populations against this virus. A similar event happened in 2009 during the H1N1 Flu epidemic. With this project we'll try to predict how likely a person is to get vaccinated based on the features of the competition's dataset.

The objective is to label each of the survey entries as having taken either the H1N1 vaccine or the Seasonal Flu vaccine, both or neither. Thus, this is a multi-label classification problem. Taking the feature and label training data from the competition we'll develop machine learning models for predicting these probabilities and compare each of the models using the competition's performance metric.

3 Software

For this particular problem we'll be using logistic regressions in order to identify the different classes. As such we'll be using Python and machine learning libraries like *scikit-learn* to implement these models. For the training sets, the data preprocessing, as we're dealing with a categorical dataset, should be appropriate for this. From the papers we've analysed, a one-hot encoder is typically used. For increasing the performance of the models, a gradient boosting library can also be utilized, such as *CatBoost* which can handle categorical features and from what we gather, features very high performance on categorical datasets.

4 Papers to read

- [1] S. Inampudi, G. Johnson, J. Jhaveri, S. Niranjana, K. Chaurasia, "Machine Learning Based Prediction of H1N1 and Seasonal Flu Vaccination", *Advanced Computing*, 2021. 139–150.
- [2] Xue, Hongxin & Bai, Yanping & Hu, Hongping. (2017). Influenza Activity Surveillance Based on Multiple Regression Model and Artificial Neural Network. *IEEE Access*
- [3] S. S. Ayachit, T. Kumar, S. Deshpande, N. Sharma, K. Chaurasia and M. Dixit, "Predicting H1N1 and Seasonal Flu : Vaccine Cases using Ensemble Learning approach," 2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN), 2020, pp. 172-176