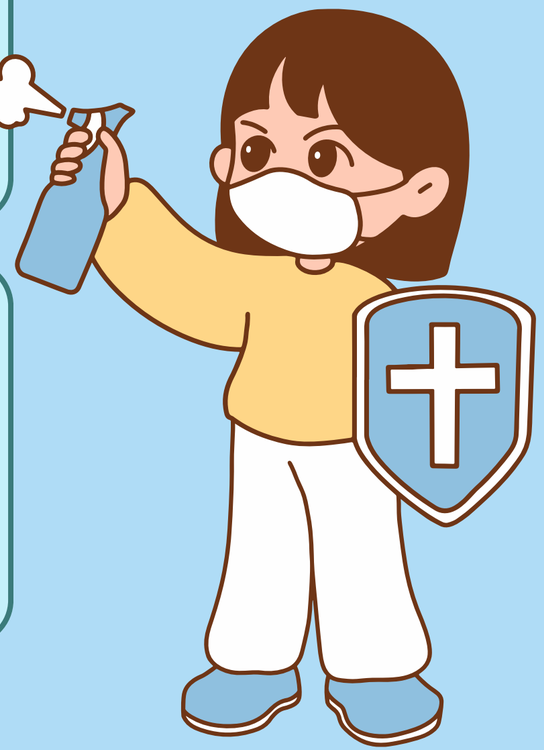
Predictive Modelling for H1N1 Vaccine Acceptance

- By Evangeline Ngunjiri

- 20 October, 2023

Contents



Project Overview

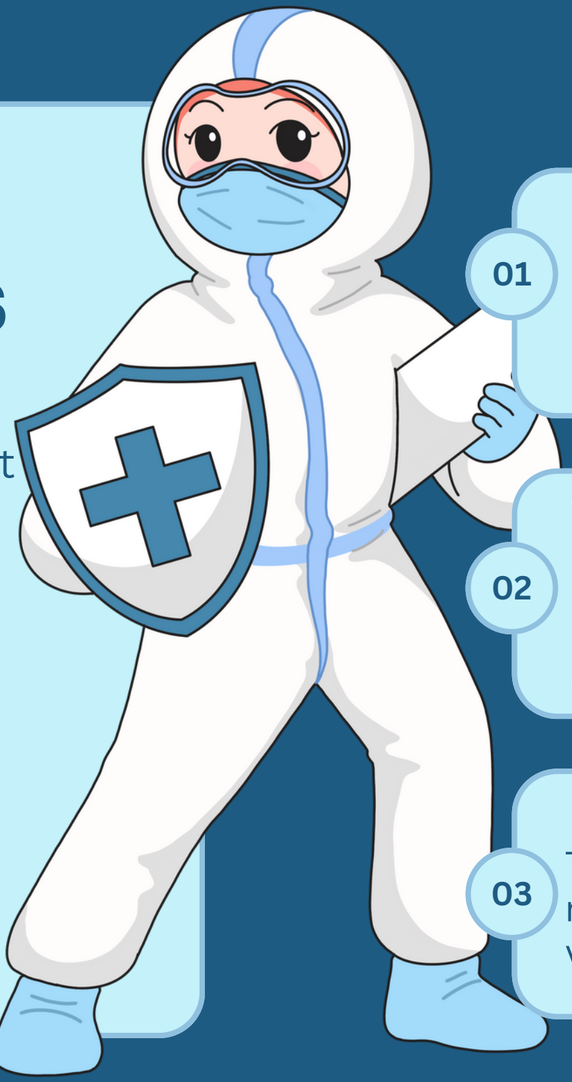
- This project will primarily focus on creating a predictive model to determine whether a survey respondent received the H1N1 vaccine. This will be based on the information collected in the National 2009 H1N1 Flu survey.
- This seeks to provide a pattern/ correlation between people's diverse cultures, backgrounds, religion, opinions and health behaviours with the H1N1 vaccination decision. This project therefore aims to offer valuable insights that can inform and guide public health initiatives on vaccination preferences and also increase preparedness in case of future health pandemics.

Project objectives

1)To investigate how well different machine learning models perform in predicting vaccine uptake

2)To identify which model is the best for predicting whether an individual will be vaccinated

3)To provide recommendations to ensure more people decide to be vaccinated



01 The study aims to provide insights into which machine learning models are most accurate and reliable in this predictive task, which can have significant implications for public health initiatives and vaccination campaigns.

02 This objective is centered on evaluating different machine learning models to determine which one is most effective in predicting whether an individual will choose to be vaccinated or not.

03 The objective stated aims to formulate recommendations to encourage and increase vaccine uptake among individuals.

Data Pre-processing

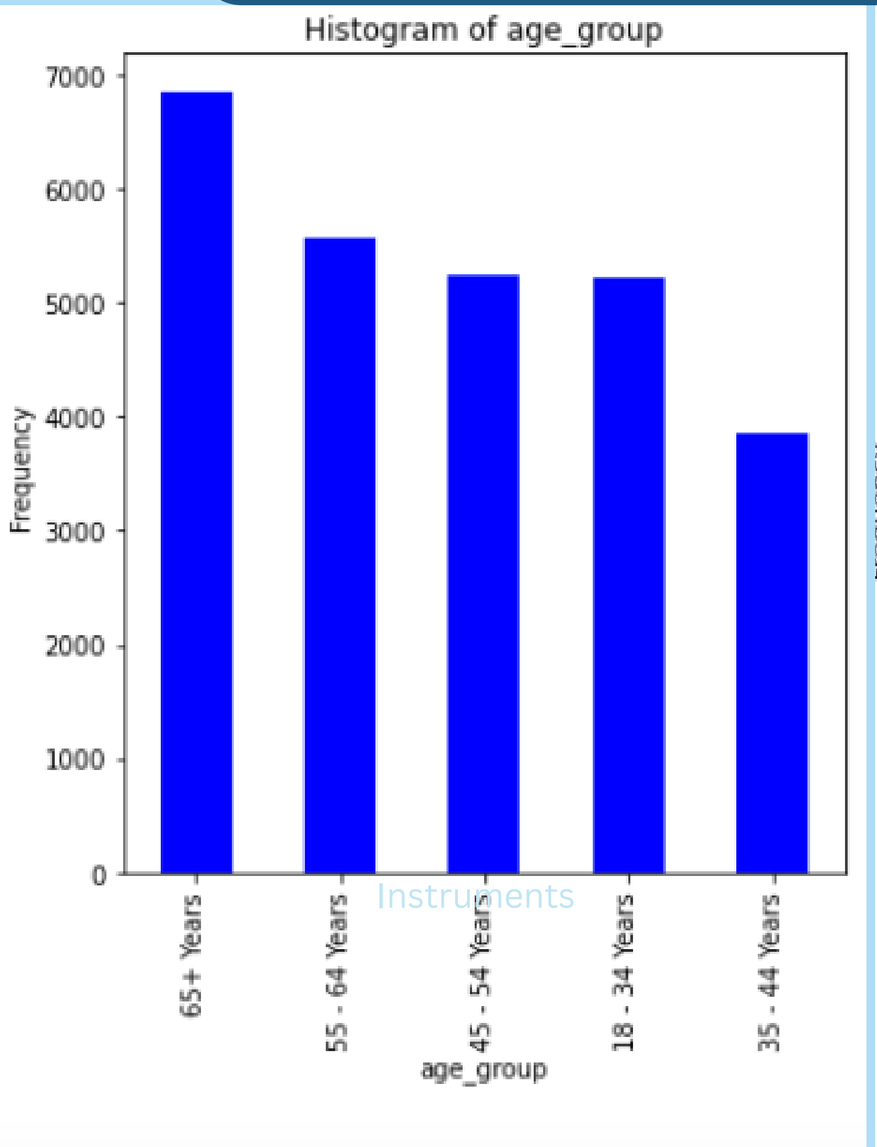
Data cleaning

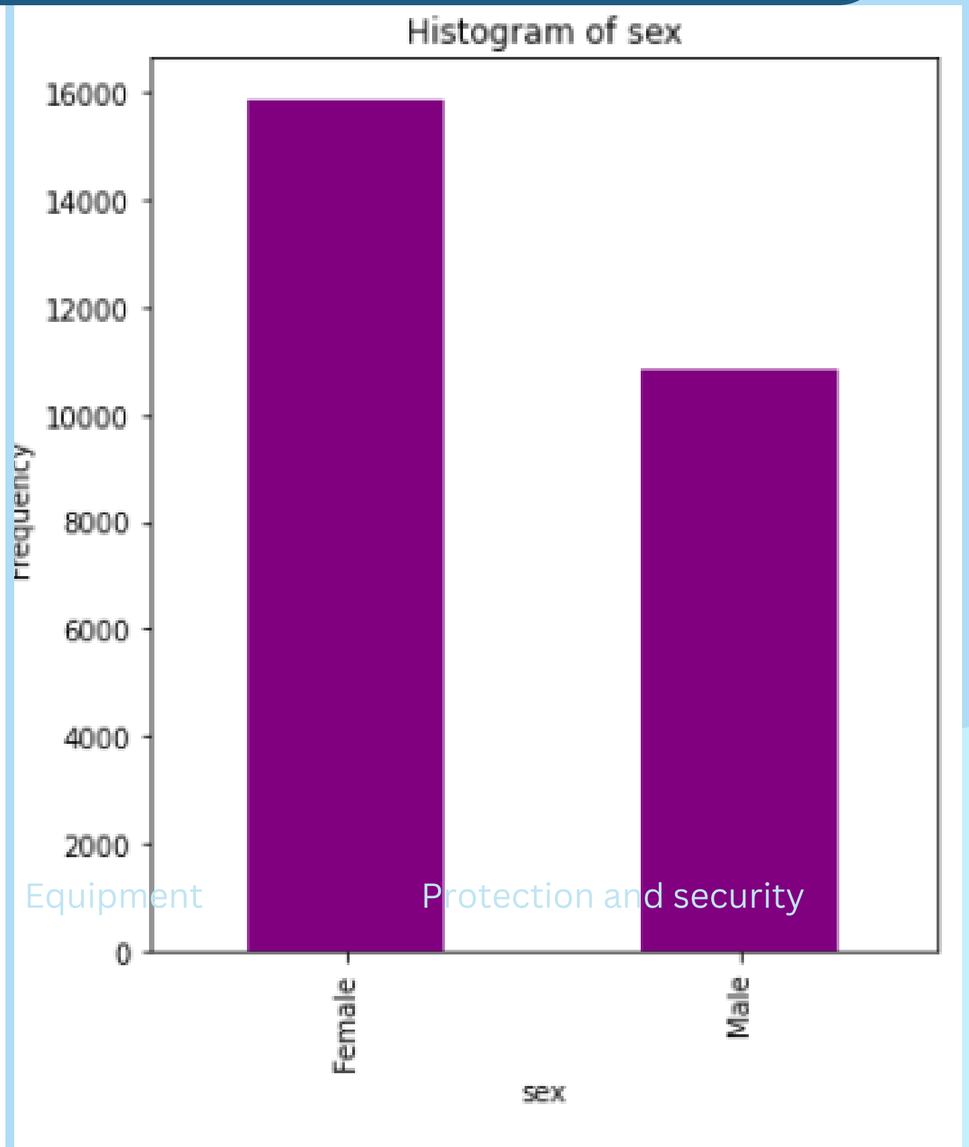
This is to prepare the data in a format that is god to feed to the models. It involves the following steps:

* Dropping irrelevant rows
* Checking for missing values
* Checking for duplicates

Exploratory Data Analysis

After data cleaning, let us take a look at how our data is distributed.





From the histogram, we can conclude that:

1. people above 65 Years were the highest consumers of the vaccine.
2. Most females decided to be vaccinated in comparison to males

Modeling

Model 1

Logistic Regression

Logistic regression is commonly used for binary classification tasks where the target variable is binary (e.g., yes/no or 0/1).

Model 2

Decision Trees

Decision tree focuses on building a tree like structure with our data
In a decision tree, data is split into subsets based on the most significant attributes or features, leading to a tree-like structure where nodes represent decisions, branches represent possible outcomes, and leaves represent final predictions or classifications.

Model 3

Random Forest

In a Random Forest model, multiple decision trees are built and combined to make predictions. It is a versatile and powerful machine learning ensemble technique that is widely used for both classification and regression tasks. It's an extension of decision tree modeling and offers several advantages..

1. Logistic regression

train report:

	precision	recall	f1-score	support
0	0.84	0.94	0.89	15773
1	0.62	0.36	0.45	4257
accuracy			0.82	20030
macro avg	0.73	0.65	0.67	20030
weighted avg	0.80	0.82	0.80	20030

test report:

	precision	recall	f1-score	support
0	0.90	0.80	0.85	5260
1	0.48	0.68	0.56	1417
accuracy			0.78	6677
macro avg	0.69	0.74	0.71	6677
weighted avg	0.81	0.78	0.79	6677

From the logistic regression modelling, it is evident that our accuracy is at 78% even after model tuning to get the best performance

2. Decision Tree

```
Train Accuracy: 0.8285044062638687
Test Accuracy: 0.8274674254904898

Train Recall_score: 0.7517276358333862
Test Recall_score: 0.5137614678899083

Train Precision_score: 0.8880982697925249
Test Precision_score: 0.6112510495382032
```

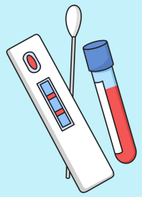
From the decision tree model, we can see that the accuracy is now at 83% after model tuning to increase performance

3. Random forest

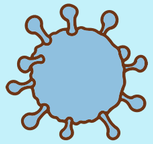
```
Mean Train Score: 83.27%
Mean Test Score: 83.56%
Best Parameter Combination Found During Grid Search:

Train Accuracy: 0.8464303544682975
```

From the Random forest model, we can see that the accuracy is now at 85% after model tuning to increase performance

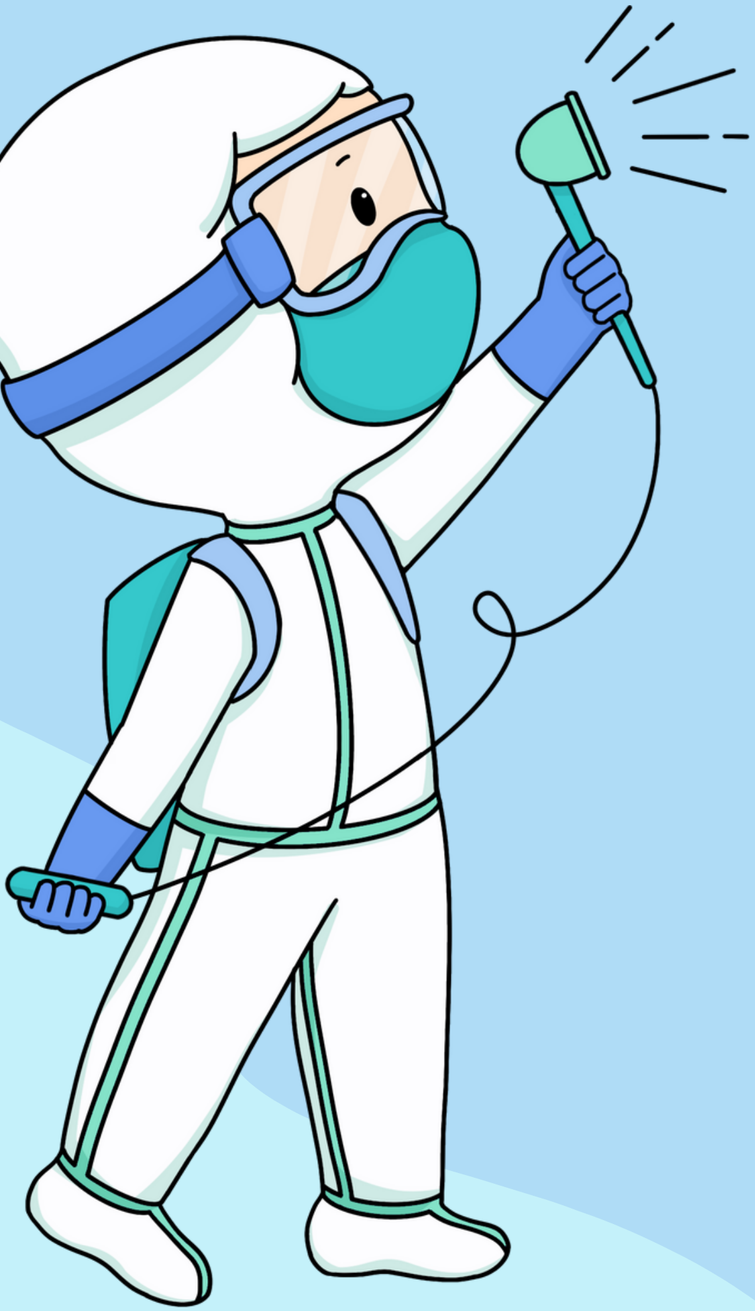
Project Conclusions



The data analysis suggests that a greater number of people chose not to receive the H1N1 vaccine, evident from the prevalence of "0" over "1" values in the dataset.



Based on these insights, I can confidently assert that the Random forest model is the most suitable choice for model deployment.

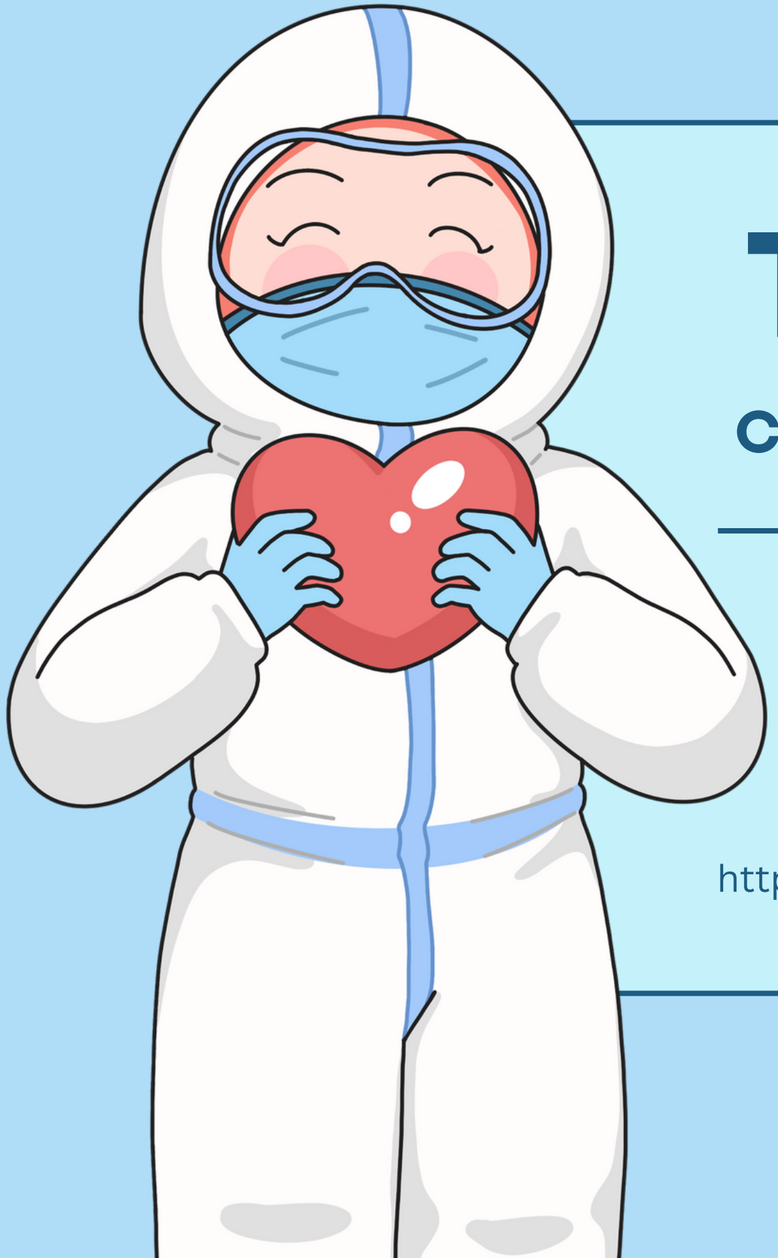


Recommendations

Focused Outreach: Due to the higher prevalence of non-vaccination, it is advisable for public health initiatives to concentrate on precise outreach efforts. This can be instrumental in promoting vaccination, especially within demographics showing a greater reluctance to receive the vaccine.

Promote Education and Awareness: Health authorities should prioritize educational and awareness campaigns to debunk vaccine-related myths and misconceptions. This approach has the potential to boost vaccine acceptance rates.

Initial Model Deployment: I suggest commencing with the implementation of the Random forest model for the prediction of H1N1 vaccine adoption. It is essential to maintain continuous monitoring of its performance and explore the integration of more sophisticated models, like Neural Networks or Bayesian Networks, to enhance predictive accuracy over time.



THANK YOU!

Contact me if you have questions

+254 723236813

evangeline.ngunjiri@student.moringaschool.com

https://github.com/EvangelineNgunjiri/