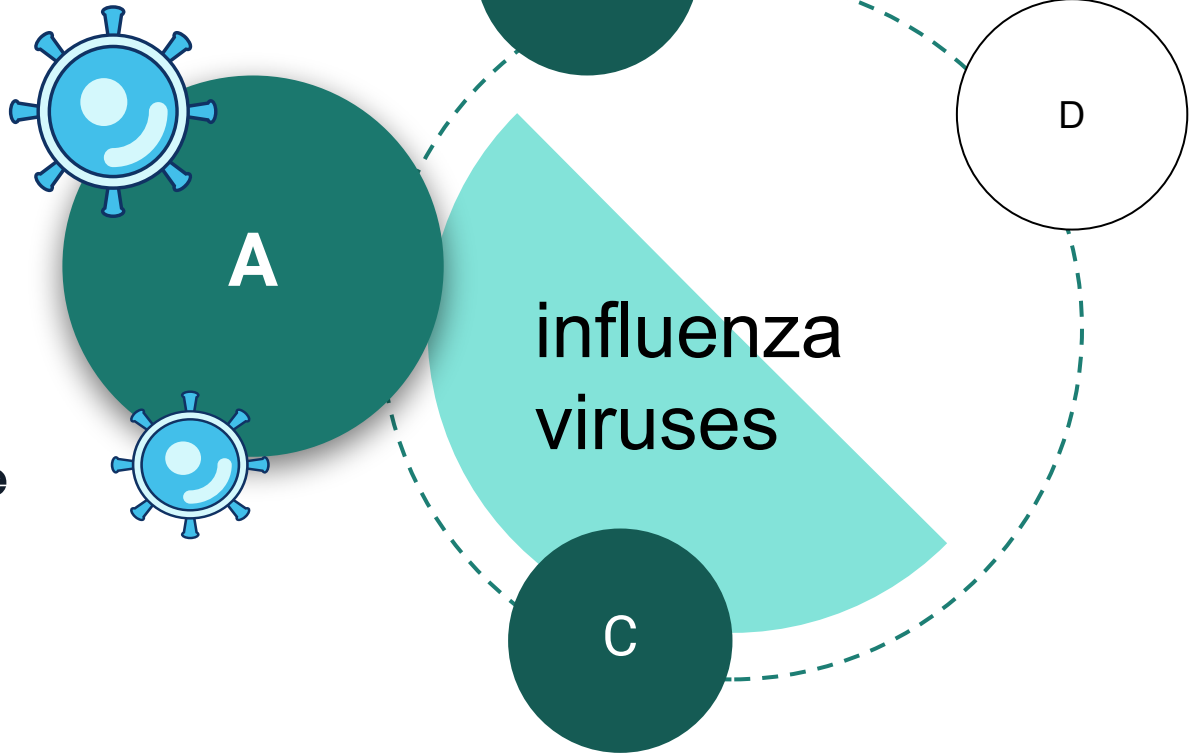
Intro to Data Science 2023
Final Project

H1N1 and Seasonal Vaccines

Team #11

The virus :

A
(H1N1)



Seasonal Influenza (the Flu)

Only influenza type A viruses are known to have caused pandemics

(Source : World Health Organisation)

There are around **a billion cases** of seasonal influenza annually, including **3–5 million cases of severe illness**.

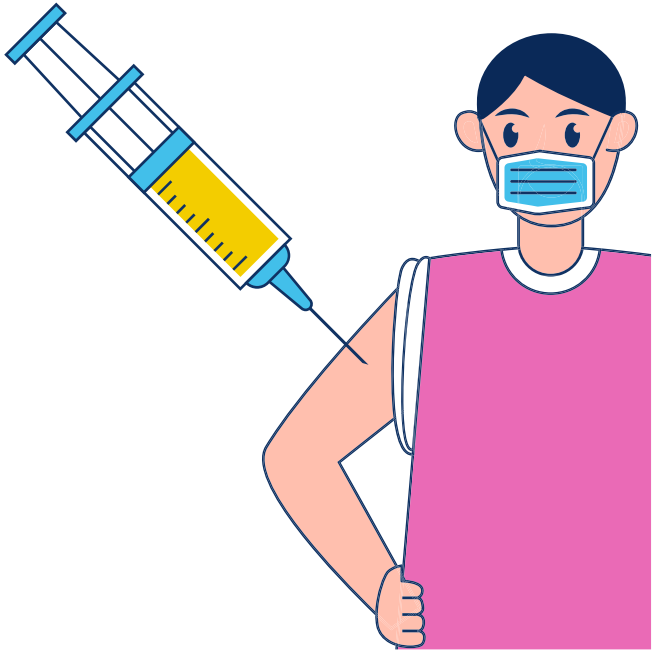
99% of deaths in children under 5 years of age with influenza-related lower respiratory tract infections are in developing countries.

It causes **290 000 to 650 000** respiratory **deaths annually**.

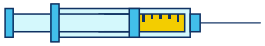
Source : World Health Organisation (Oct, 2023)

Influenza spreads easily between people when they cough or sneeze.

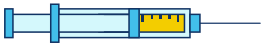
Vaccination is the best way to prevent the disease.



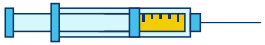
Annual vaccination is recommended for:



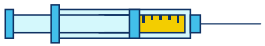
pregnant women



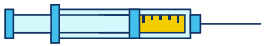
children aged 6 months to 5 years



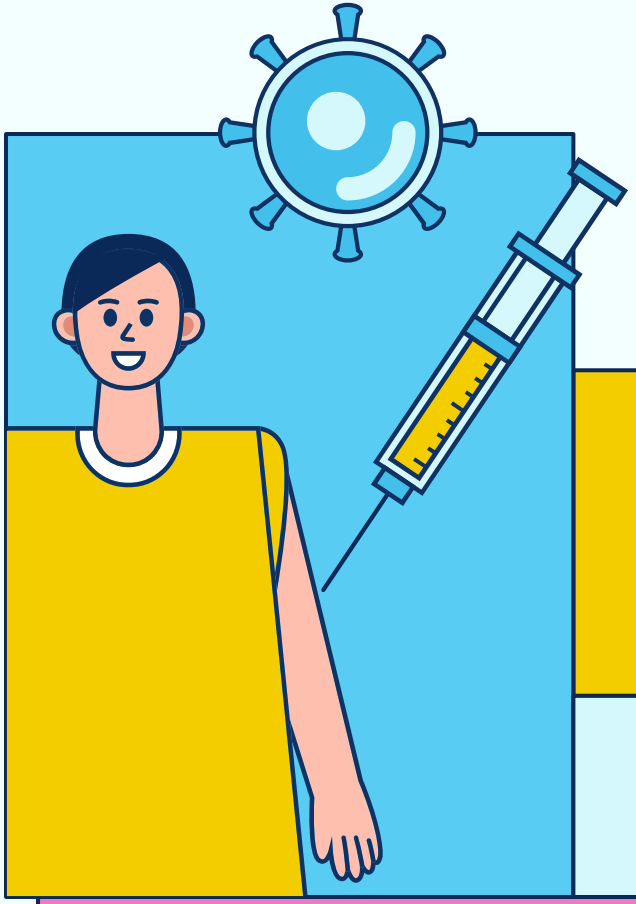
people over age 65



people with chronic medical conditions



health workers.

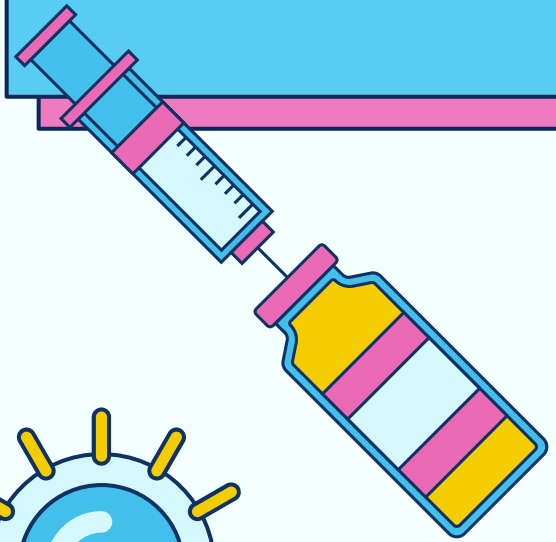


01

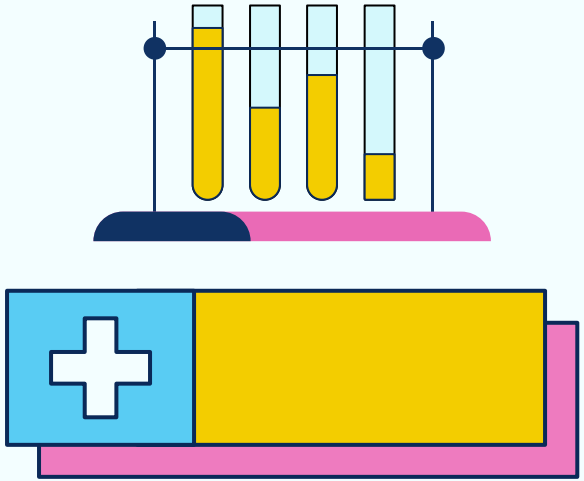
Problem Description

Why is it important?
How can our ML model help?

Problem Statement



The problem we aim to address using data science techniques is the prediction of individuals' likelihood to receive their H1N1 (h1n1_vaccine) and seasonal flu vaccines (seasonal_vaccine).



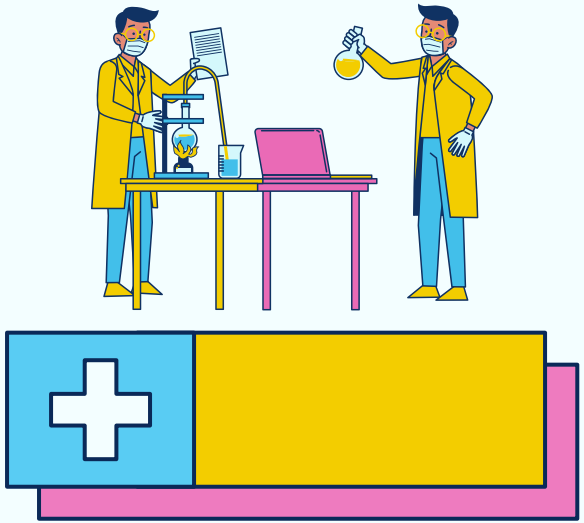
Importance of the problem

Disease Reduction: Higher vaccine coverage reduces the spread of infectious diseases like the flu, ultimately saving lives and reducing the burden on the system.

Disease Prevention: Healthcare providers can reach out to individuals at risk of not getting vaccinated, addressing concerns.

Cost Savings: A predictive model can lead to cost savings by optimizing the distribution of vaccines .

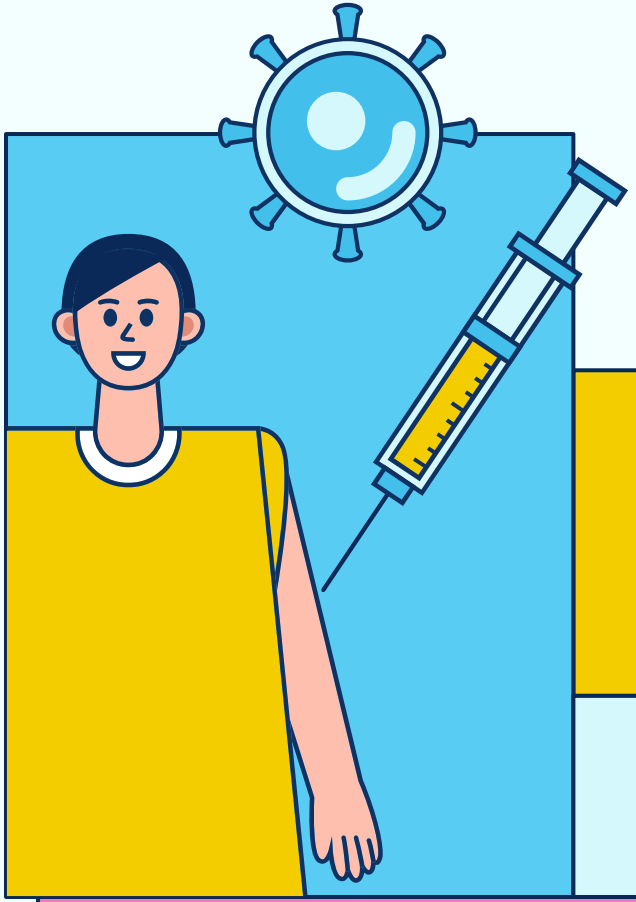
ML Model Benefits



Allocating vaccination resources can be done more efficiently, ensuring that vaccines are distributed where they are needed the most.

Early predictions can provide a warning system for potential vaccination gaps, allowing for timely responses and proactive interventions.

The insights from the model can inform policy decisions, helping authorities design more effective public health campaigns.



02

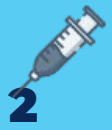
Exploratory Data Analysis

Exploring, analyzing and visualizing meaningful patterns in the data

Key insights about the data

26707

Total observations in our dataset



2 Targets/Labels

Our data consists of multiple labels

35 Features

We have 35 independent variables

Doctor's recommendation

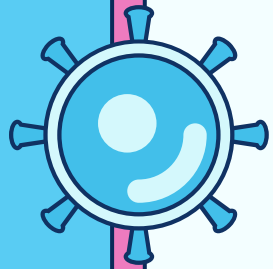
Corresponding feature has the highest correlation with getting an H1N1 vaccine

5 features

Only 5 of them didn't have any missing values initially

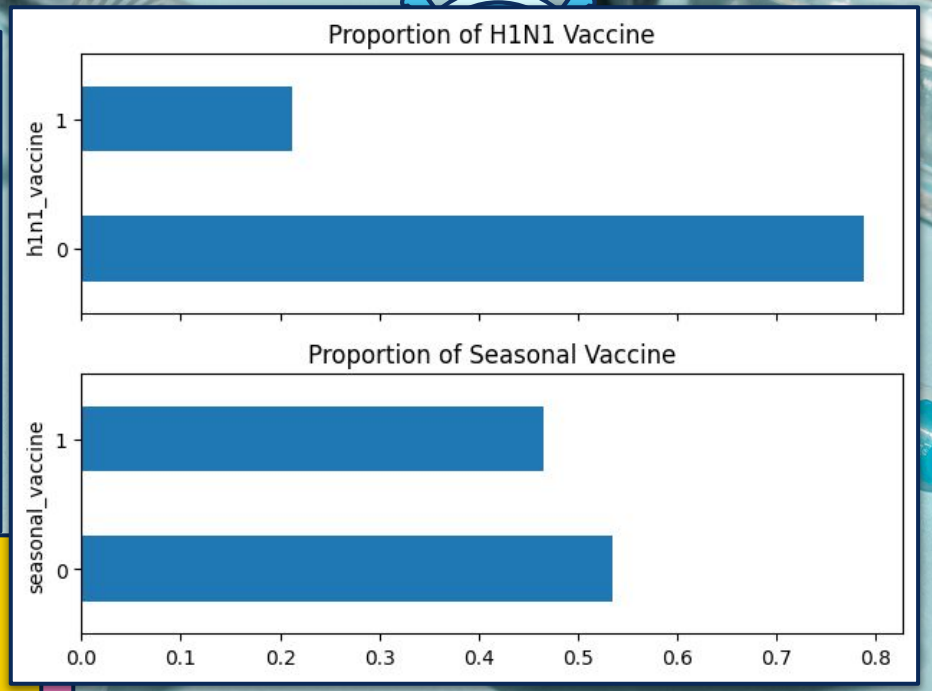
Feeling safer

Mostly, people who are vaccinated for H1N1 are also vaccinated for seasonal flu

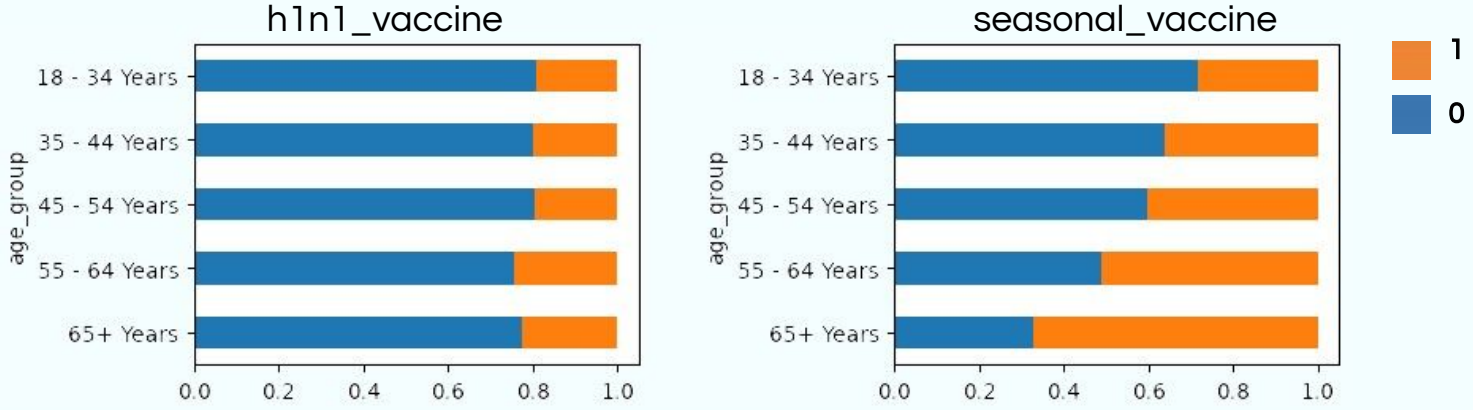


- The **H1N1 vaccination** target has a considerable class **imbalance**, with around 20% population receiving the H1N1 flu vaccine.
- The **seasonal vaccine** target has roughly equal representation in both classes, making it **balanced**.

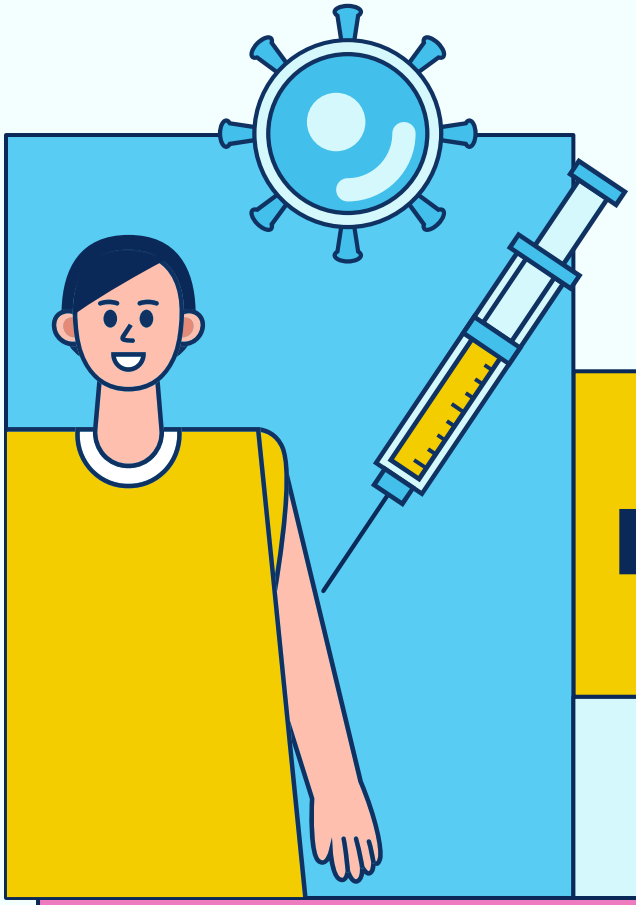
A graphic is worth a thousand words



A closer look to age dependence



- According to the data, older individuals are more likely to be prepared against seasonal flu due to their awareness of being more fragile against it.
- Also, it turns out that H1N1 has an interesting relationship with age: even though older people have higher risk of complications, they were less likely to get infected! All in all, it seems like the risk factors ended up being reflected in the vaccination rates.



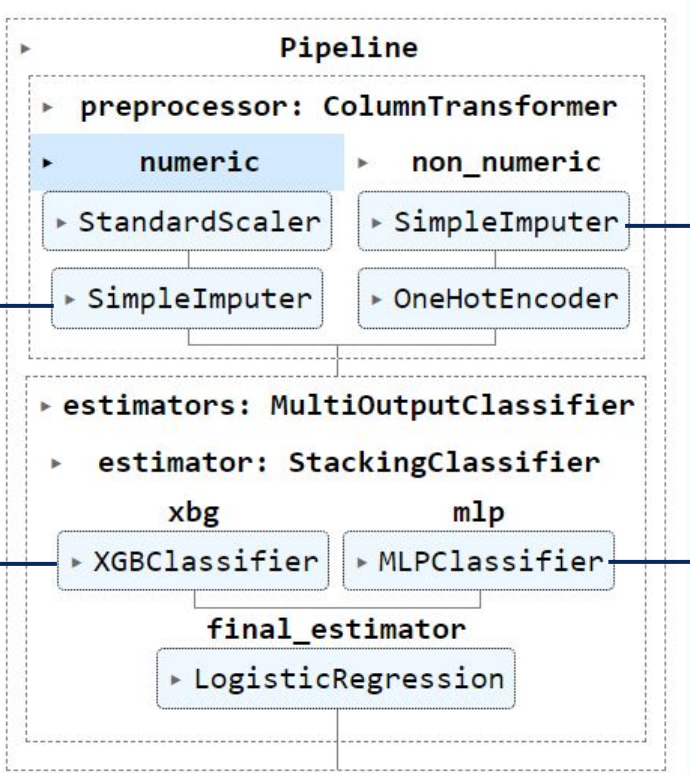
03

Processing

Data preprocessing
Training & Evaluating model
Fine-tuning model

Pipeline

fill missing values with the ‘mean’ value of the feature



fill missing values with the constant value 'missing'

eXtreme Gradient Boost

is a boosting algorithm based on gradient boosted decision trees algorithm

Multi-Layer Perceptron

relies on an underlying neural network to perform the task of classification

Optimizing models

Best trial for XGBoost:

Params:

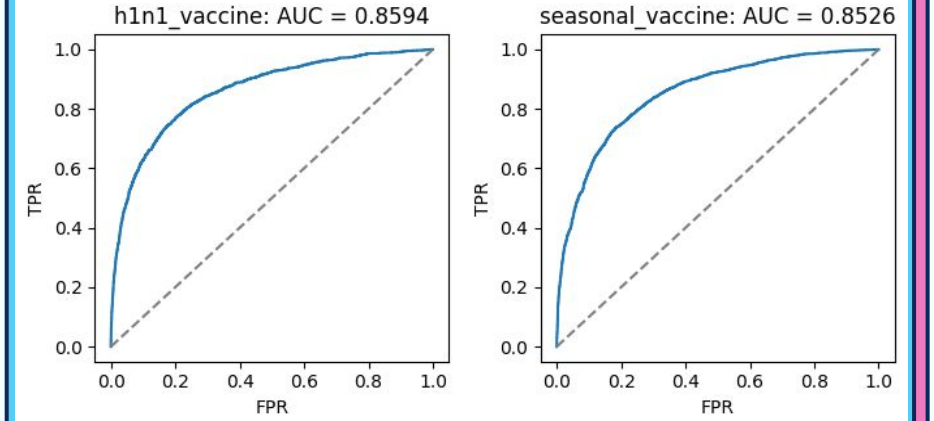
- learning_rate = 0.031,
- n_estimators = 448,
- max_depth = 6,
- min_child_weight = 5,
- gamma = 0.934,
- subsample = 0.601,
- colsample_bytree = 0.472,
- reg_alpha = 0.955,
- reg_lambda = 18.649

Best trial for MLPClassifier

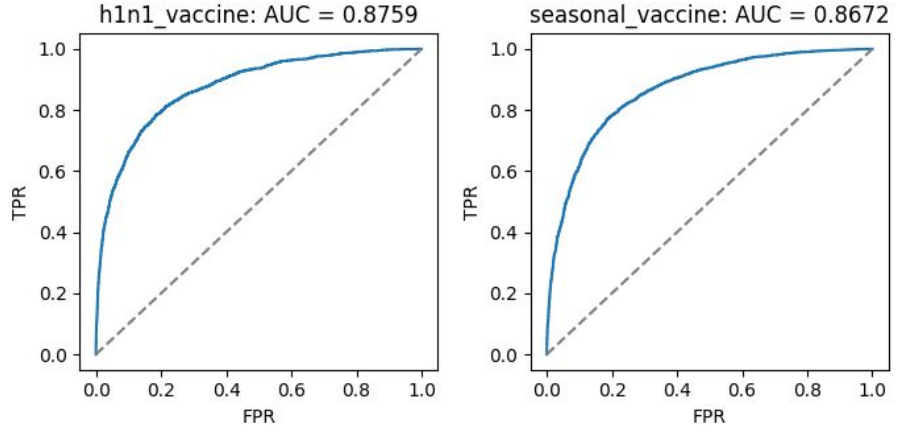
Params:

- activation = relu,
- learning_rate = invscaling,
- learning_rate_init = 0.055,
- alpha = 0.009,
- max_iter = 967

ROC_AUC score



Default models



Fine-tuned models

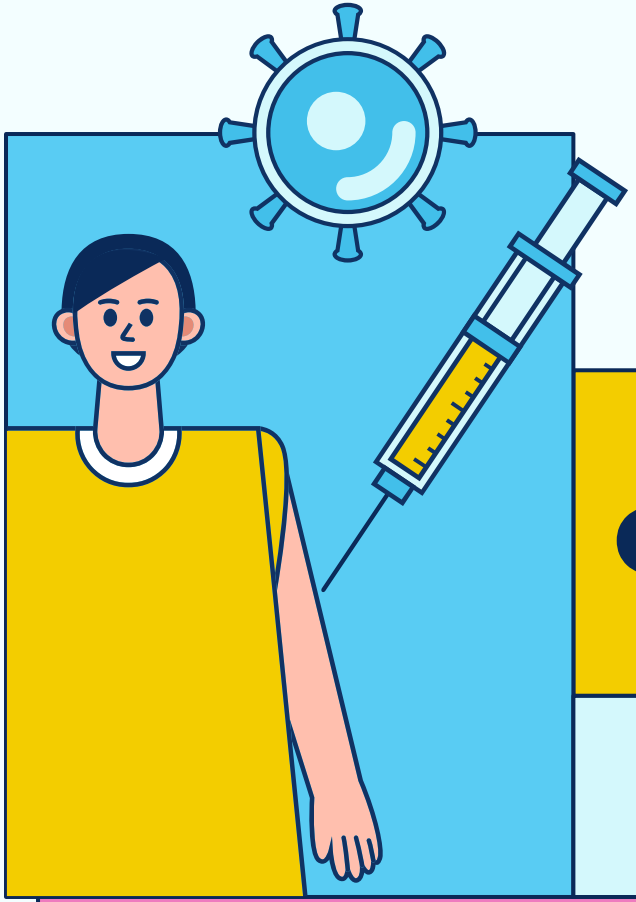
Result

ROC-AUC Score

ID	Vaccine H1N1	Vaccine seasonal	
6368	0.070	0.092	✕
12566	0.906	0.909	+
13996	0.065	0.119	✕
12769	0.828	0.928	+
16915	0.058	0.098	✕

more likely to get vaccinated +

less likely to get vaccinated ✕



04

Conclusion

Conclusion

In conclusion, the machine learning model significantly impacts the problem domain of vaccine uptake prediction.

1) It aids in achieving higher vaccination coverage, optimizing resource allocation, and reducing costs.

2) Moreover, it improves public health outcomes by reducing the spread of influenza.

3) The model's adaptable nature and data-driven approach make it a valuable asset for public health planning and decision-making.

While it may not solve the problem entirely, it offers a robust and dynamic approach to addressing vaccine uptake challenges.

Thanks!

We'll be glad to answer your questions!

Ngoc Bich Uyen Vo
Oguzhan Aksoy
Ahmed Eltwam
Asma Benachour