# VACCINATION PREDICTION

# BUSINESS UNDERSTANDING

## *Business problem*

The flu and H1N1 vaccination rate in the United States is relatively low at 50.2%. This presents a public
health concern, particularly given the high availability and proven effectiveness of flu vaccines.As the world struggles to vaccinate the global population against COVID-19, an understanding of how people's backgrounds, opinions, and health behaviors are related to their personal vaccination patterns can provide guidance for future public health efforts.

Understanding these factors and modelling them as a data scientist will be pertinent in infroming the the public health stakeholders in the the World Health Organization.

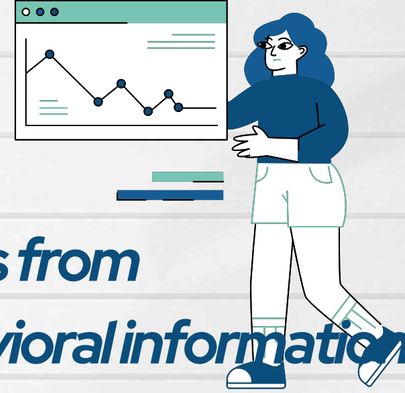
# *Proposed Solution*

The primary objective of this algorithm is classification. It should analyze attributes from the National 2009 H1N1 Flu and Seasonal Flu Survey dataset to determine if a respondent received the seasonal flu vaccine or the H1N1 flu vaccine. The model's effectiveness will be measured by how accurately it predicts vaccination status based on the respondent's behaviors and motivational factors.

# OBJECTIVES

- .To determine whether opinion is associated with higher vaccination rates for both H1N1 and seasonal flu.
- 
- . To develop a learning algorithim that predicts vaccination status based on the respondent behaviours.
- 
- .To recommend strategies to improve vaccination rates based on the identified key factors and behavioral patterns.
- 

# DATA UNDERSTANDING

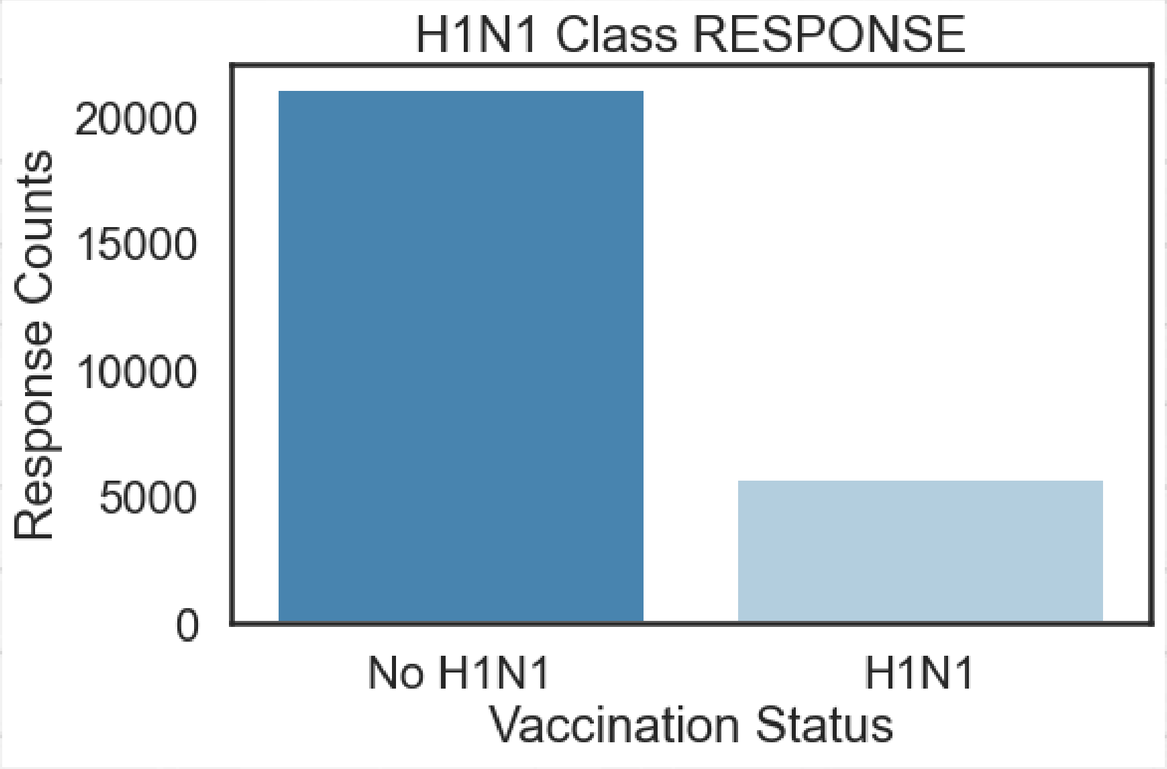


- *The data comes from the National 2009 H1N1 Flu Survey (NHFS) and contains answers from respondents regarding demographic information, opinions on vaccinations, and behavioral information*
- 
- *The data was collected via a list-assisted random-digit-dialing telephone survey of households. The target population was all persons 6 months or older living in the United States. The survey produced estimates of vaccine coverage rates for both the monovalent pH1N1 and trivalent seasonal influenza vaccines.*
- 
- *DrivenData. (n.d.). Flu Shot Learning: Predict H1N1 and Seasonal Flu Vaccines. DrivenData from https://www.drivendata.org/competitions/66/flu-shot-learning/data/*
- 
- 

# EXPLORATORY DATA ANALYSIS

## VACCINATION OF H1:

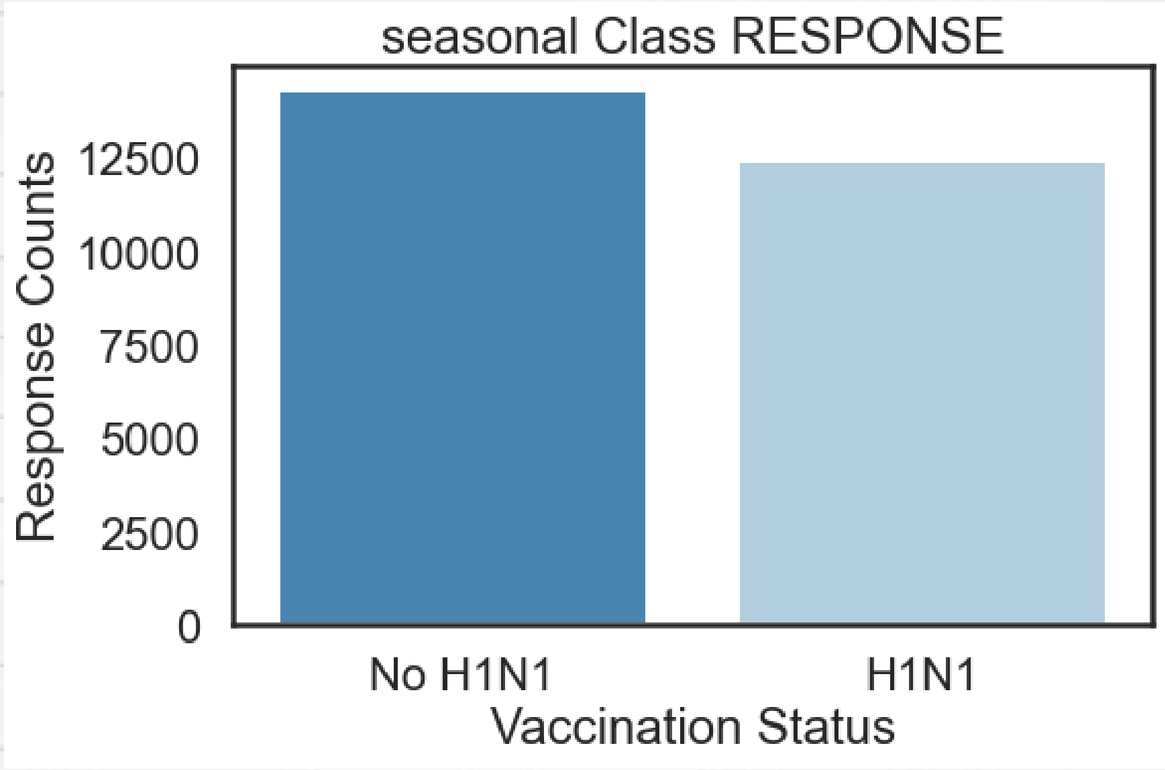
The visualization shows the vaccinated people for h1n1 flu and seasonal vaccine



# EXPLORATORY DATA ANALYSIS

## seasonal flu :

seaosonal fluvaccines appear to be more appreciated

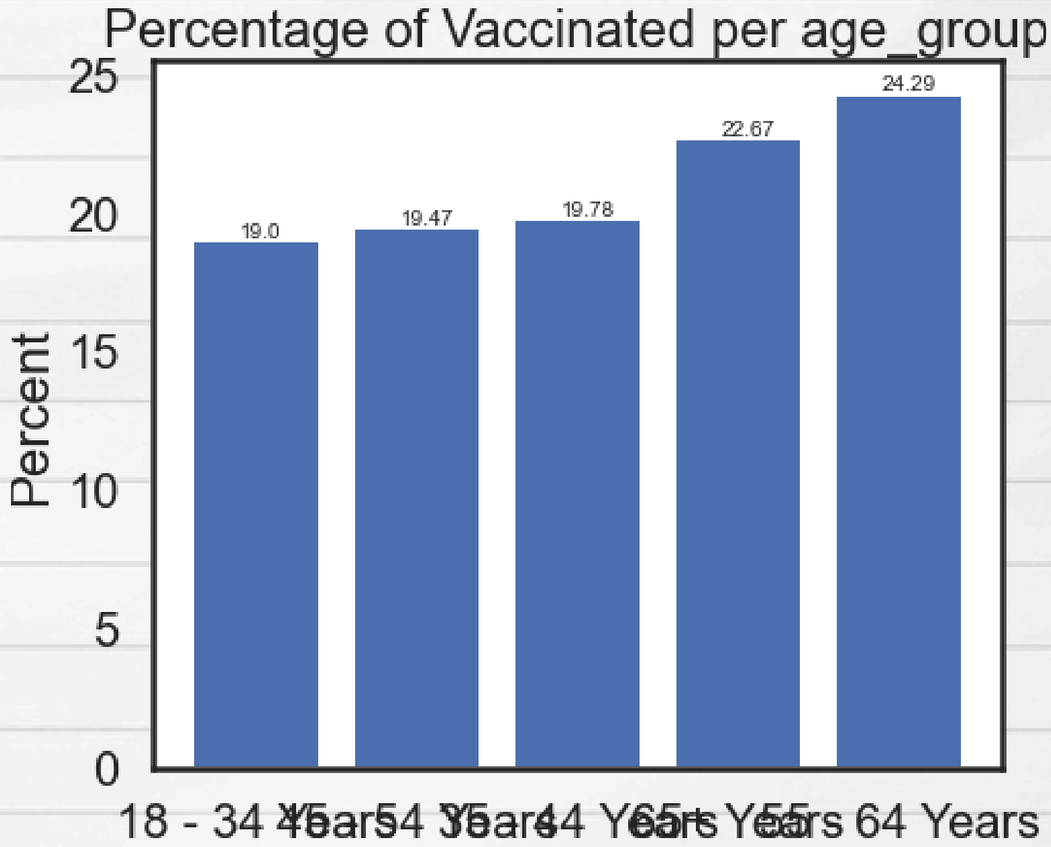


# EXPLORATORY DATA ANALYSIS

## Age:

agegroup effect on response to h1n1

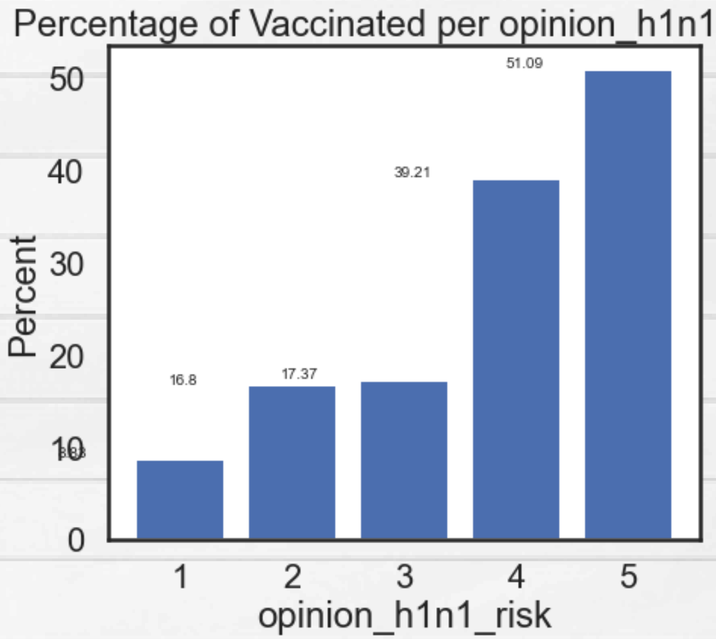
older people  take their vaccines more

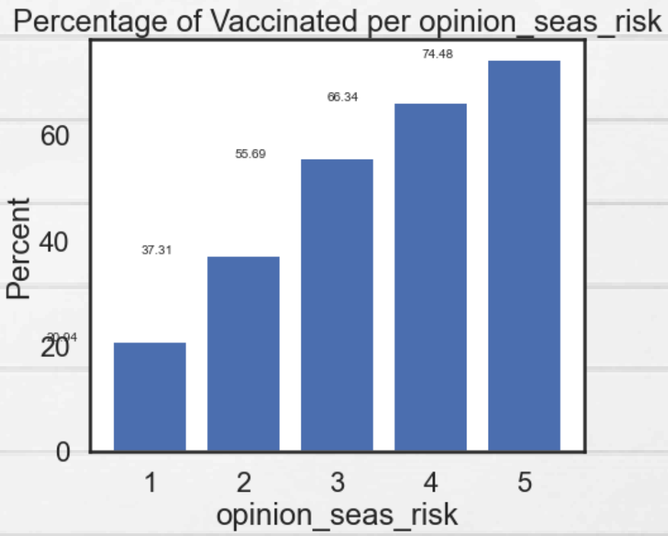


# EXPLORATORY DATA ANALYSIS

## opinion on risk

opinions definitely affect the vaccine responses both for hana and seasonal

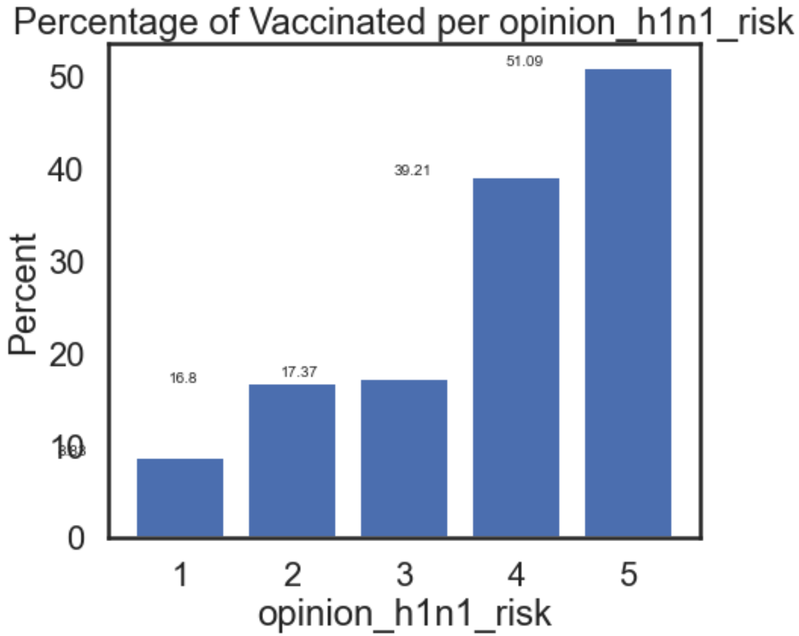


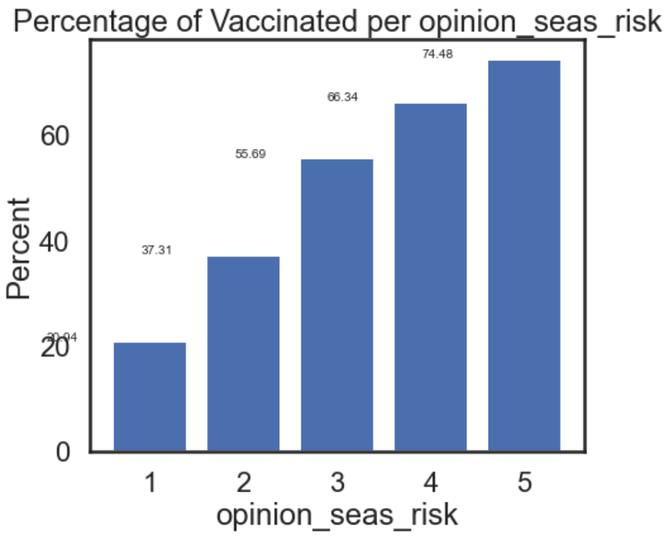



# EXPLORATORY DATA ANALYSIS

## opinion on risk

opinions definitely affect the vaccine responses both for hana and seasonal

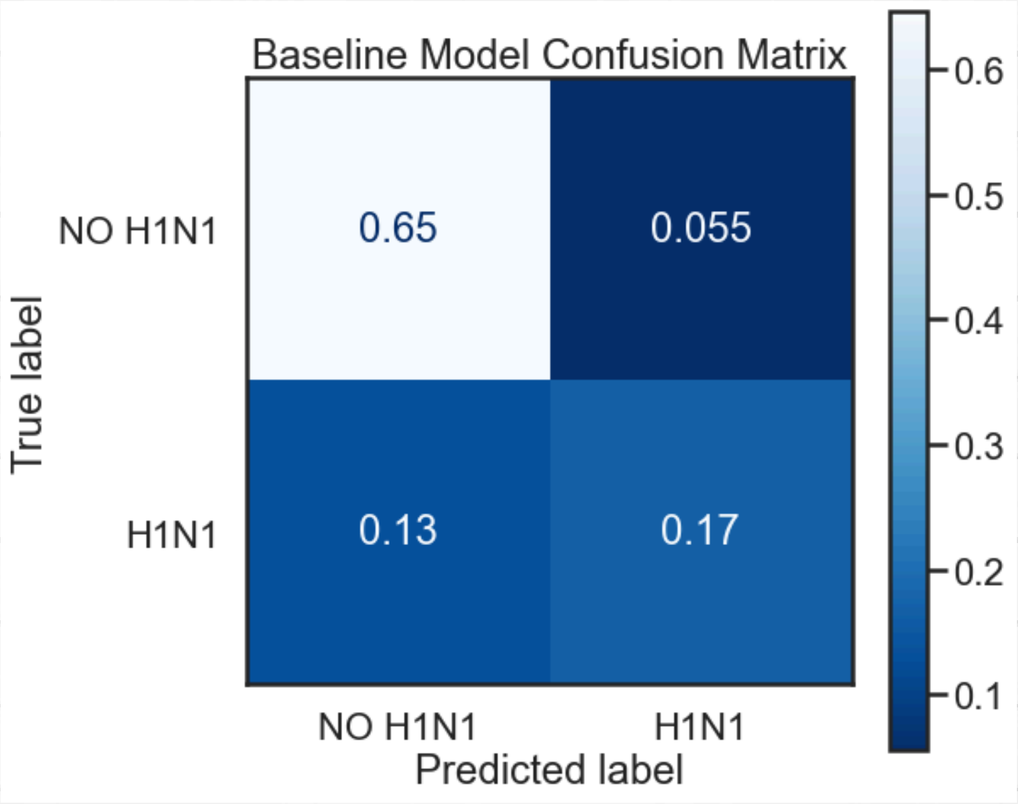


# MODEL 1

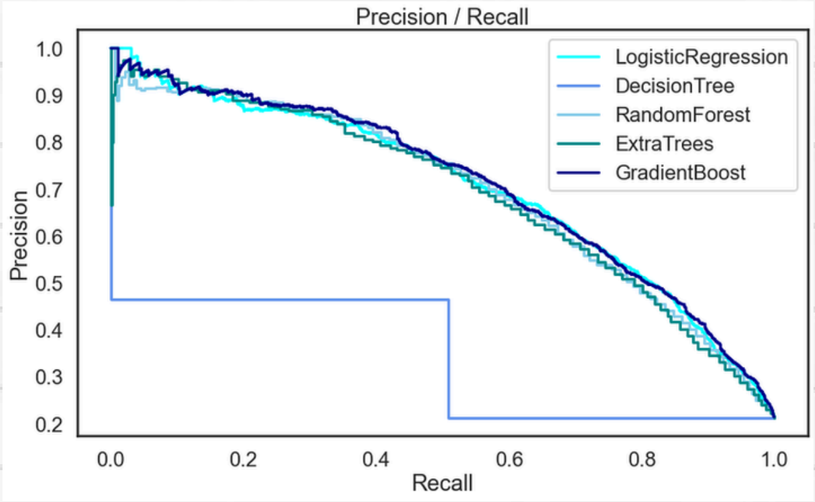
## logistic regression model:

Tbaseline model exhibits solid predictive capability, there is room for improvement, particularly in capturing more positive instances without significantly increasing false positives.

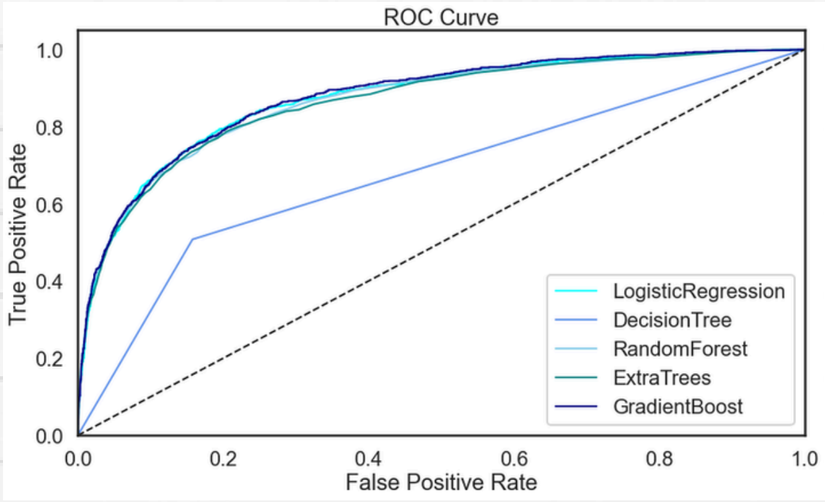


# Different models

- different models such as decision tree,random forest and extra trees show inmprovement in the data :Logistic Regression Model:
- Cross Validation Score: 0.8491
- Test Accuracy Score: 0.8588
- DecisionTree Model:
- Cross Validation Score: 0.7765
- Test Accuracy Score: 0.7712
- RandomForest Model:
- Cross Validation Score: 0.845
- Test Accuracy Score: 0.8576
- ExtraTrees Model:
- Cross Validation Score: 0.8467
- Test Accuracy Score: 0.8559

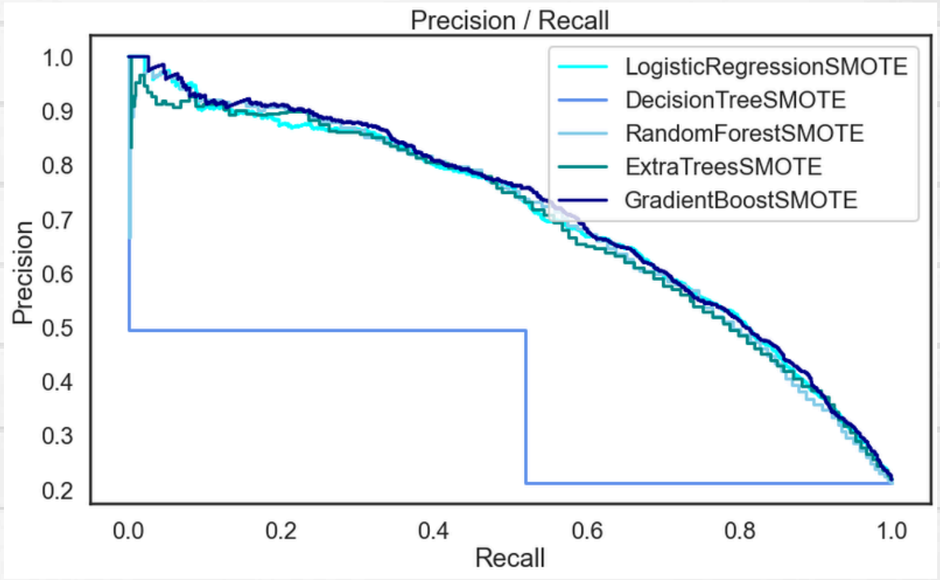


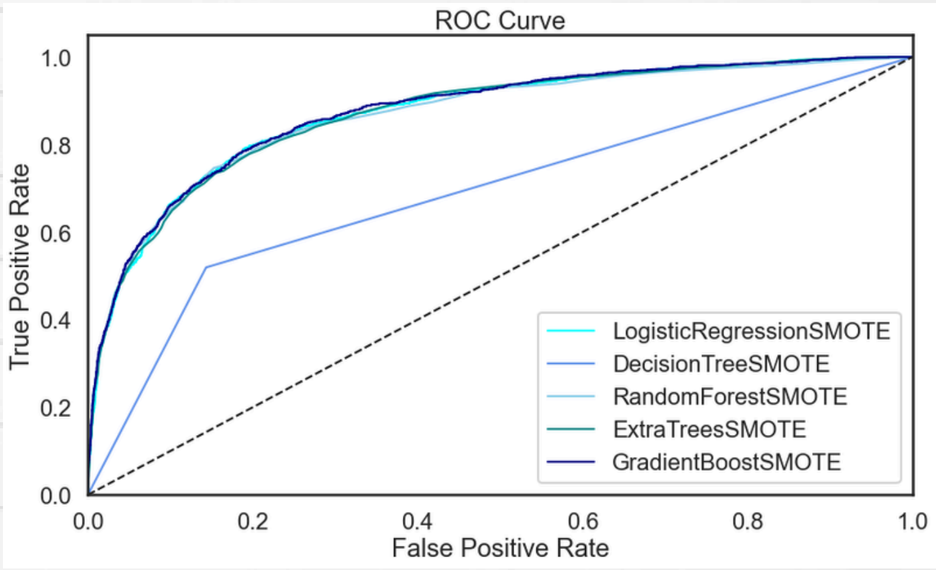

different models exhibit improvement in the model of the data both dislayed by the roc and auc curves .



# smote models

- Logistic Regression Model:
- Cross Validation Score: 0.7929
- Test Accuracy Score: 0.8068
- Decision Tree Model:
- Cross Validation Score: 0.7711
- Test Accuracy Score: 0.7849
- Random Forest Model:
- Cross Validation Score: 0.8472
- Test Accuracy Score: 0.8603
- Extra Trees Model:
- Cross Validation Score: 0.8480
- Test Accuracy Score: 0.8589
- Gradient Boosting Model:
- Cross Validation Score: 0.8507
- Test Accuracy Score: 0.8609
-




**smote models exhibit improvement in the model of the data both dislayed by the roc and auc curves .**

# Conclusion

- Strong negative correlation observed between behavioral-related questions and the likelihood of receiving the vaccine.
- Opinion_seas_sick_from_vaccine and opinion_h1n1_sick_from_vacc demonstrate a strong negative correlation with the H1N1 target variable.
- Baseline model performs well on various test metrics:
- Accuracy: Approximately 81.49%, indicating correct prediction for a significant portion of instances.
- Recall: Around 56.55%, suggesting potential for missed positive instances.
- Precision: Approximately 75.3%, indicating low false positive rate among positive predictions.
- F1 Score: About 64.59%, reflecting a balance between precision and recall.

# *Recommendation*

- utilize more recent data for efficacy and recent vaccinations
- Educate people on the need of vaccines
-
- work on minimizing risks of vaccine

# THANK YOU VERY MUCH!