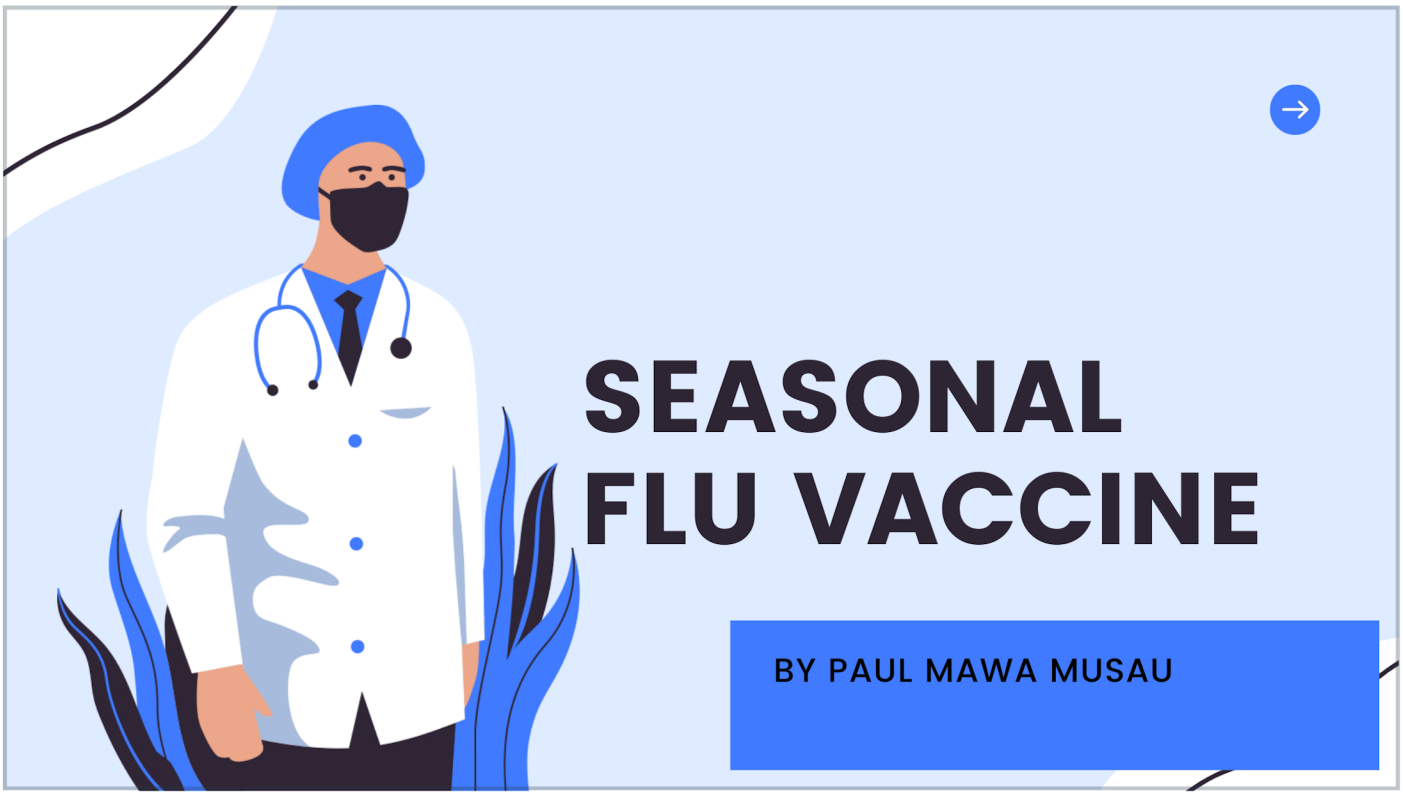

SEASONAL
FLU VACCINE
BY PAUL MAWA MUSAU

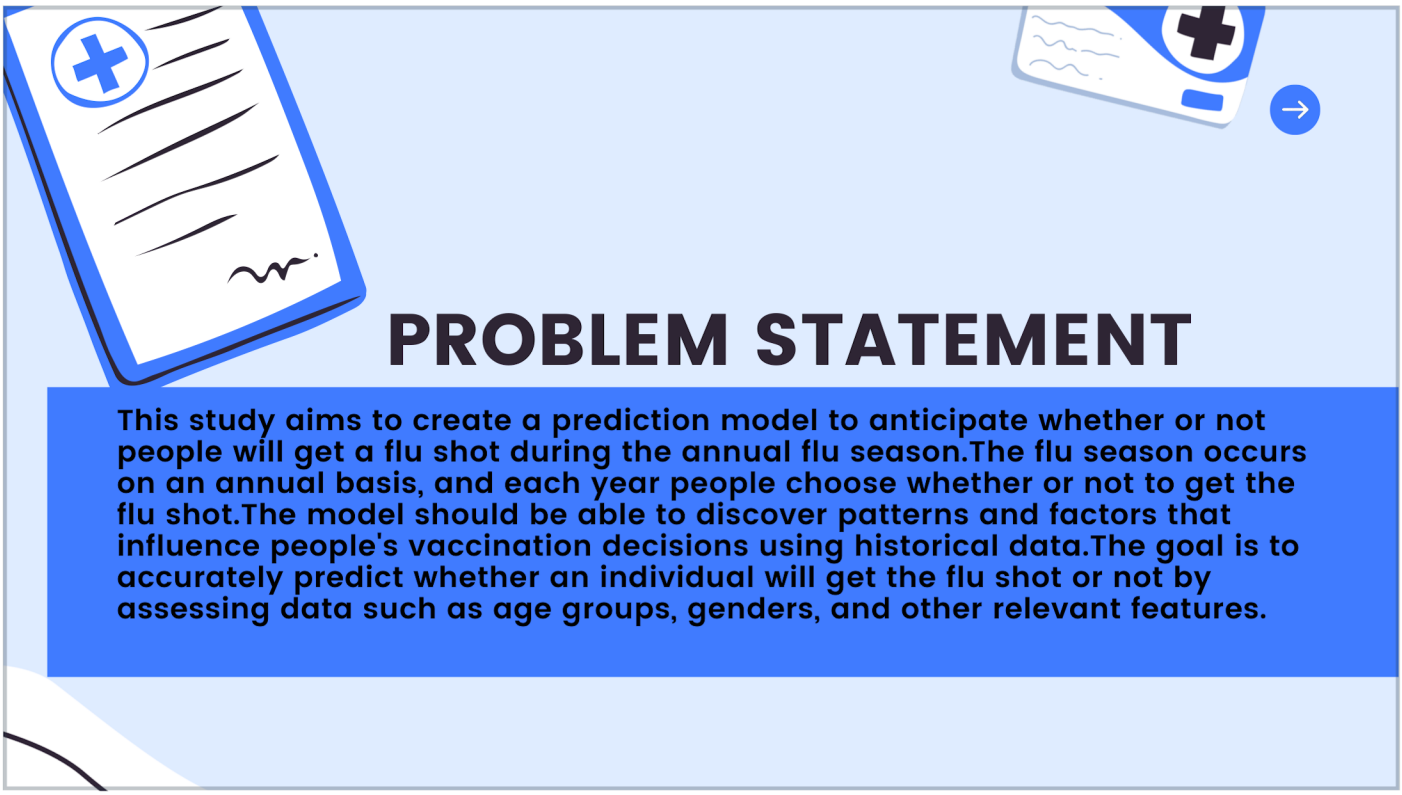
INTRODUCTION

* This study aims to create a prediction model to anticipate whether or not people will get a flu shot during the annual flu season.

* The flu season occurs on an annual basis, and each year people choose whether or not to get the flu shot.

* The model should leverage historical data to identify patterns and factors that influence people's vaccination choices.

COVID-19 | 2020


PROBLEM STATEMENT
This study aims to create a prediction model to anticipate whether or not people will get a flu shot during the annual flu season.The flu season occurs on an annual basis, and each year people choose whether or not to get the flu shot.The model should be able to discover patterns and factors that influence people's vaccination decisions using historical data.The goal is to accurately predict whether an individual will get the flu shot or not by assessing data such as age groups, genders, and other relevant features.

MAIN OBJECTIVE

Create a predictive model to forecast whether or not people will receive the Vaccine based on particular characteristics or conditions.

IMPORTANT FEATURES THAT SHAPE OUR MODEL

Opinion_seas_risk
Doctor_rec_seasonal
Age_group
Employment_industry

MODELING

.ADABOOST

.GRADIENT BOOST

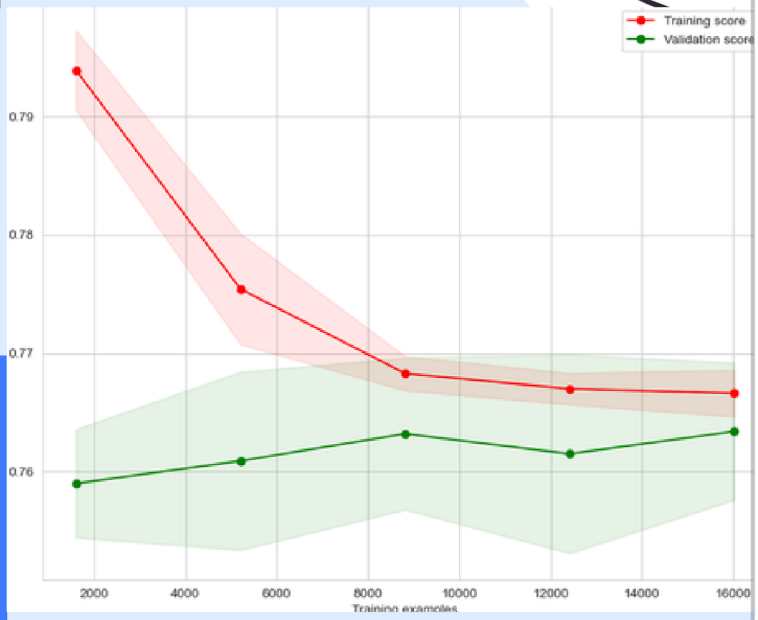
.RANDOM FOREST

.XG BOOST

RANDOM FOREST

Model Accuracy

76.9%



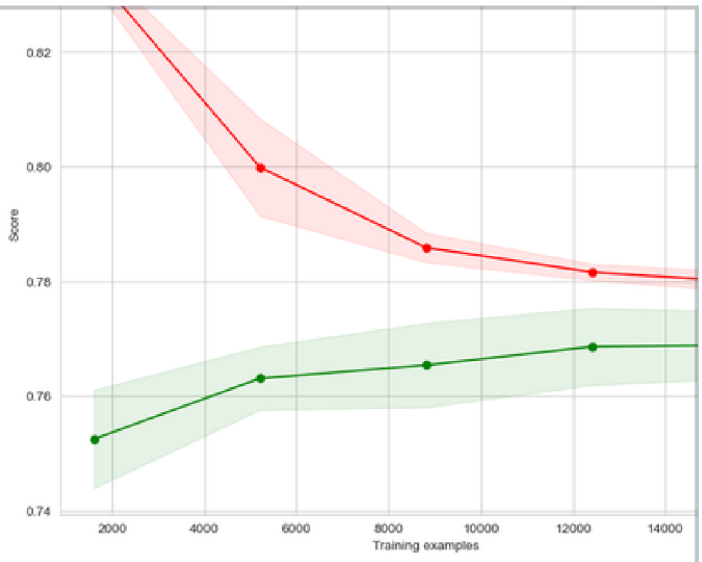
The learning curve shows that generalization is improving initially, with the test score increasing and the training score falling.

* The test score, however, gradually plateaus and begins to drop, indicating limits in catching complicated patterns and probable overfitting.

* The temporary dip suggests that there may be certain instances in the validation set where the model struggles to make accurate predictions.

GRADIENT BOOSTING

ACCURACY: 77%



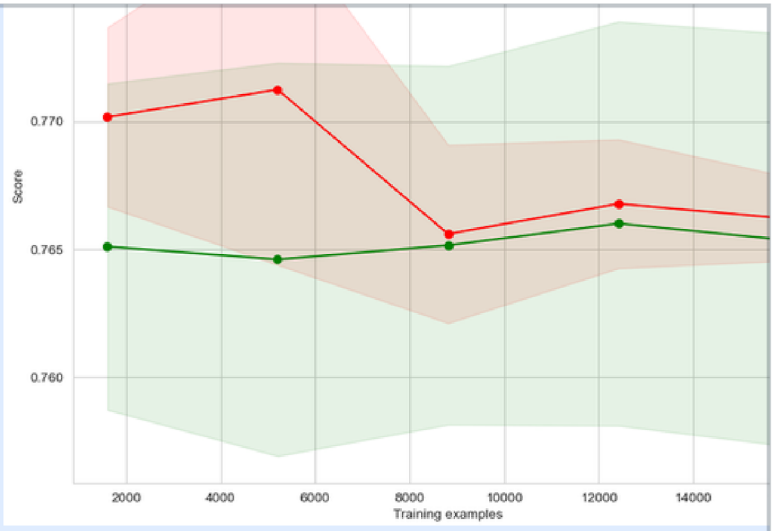
* The learning curve shows that generalization is improving, with the test score increasing and the training score falling.

* These findings show that the model performed well in terms of accuracy on both the training and test datasets, with consistent performance evaluated by cross-validation.

* The rising validation score indicates that the model generalizes well to unseen data, as it consistently improves its predictive accuracy.

ADABOOST

ACCURACY:76.9%



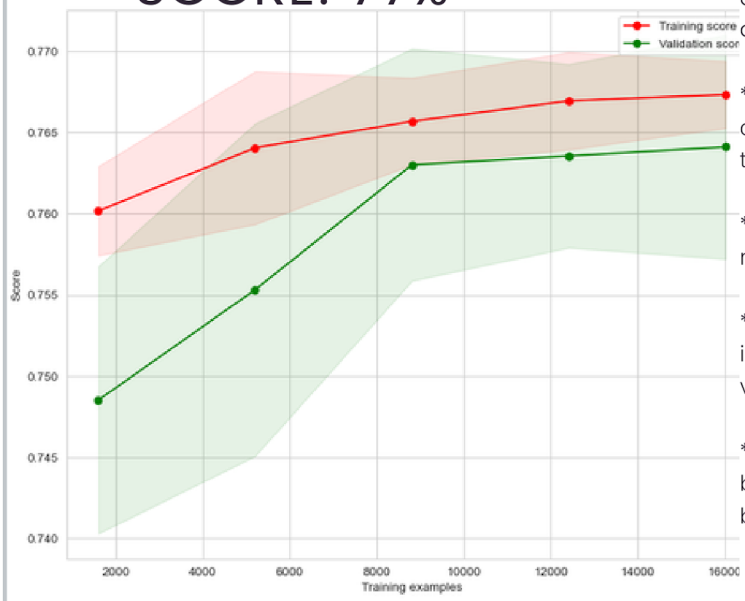
* The learning curve shows that the training score starts with a high note suggesting overfitting however, it suddenly declines.

* This indicates that the model initially struggles to fit the training data and may not capture all of the patterns available in the data.

* However, as the model receives more training data and learns from it, the training score gradually improves.

* Both the train and test set reach a point where the model start to drop gradually.

XGBOOSTING SCORE: 77%



* The learning curve shows that the training score starts with a high note suggesting overfitting however, it suddenly declines.

* These results indicate that the model performs consistently across different datasets, as the training and test accuracies are similar.

* The achieved accuracies are rather high, showing that the model predicts the outcome variable successfully.

* The model is learning from the training data and improving its performance on both the training and validation sets.

* Both scores are increasing, which implies that the model benefits from more training and has the potential to become more accurate.