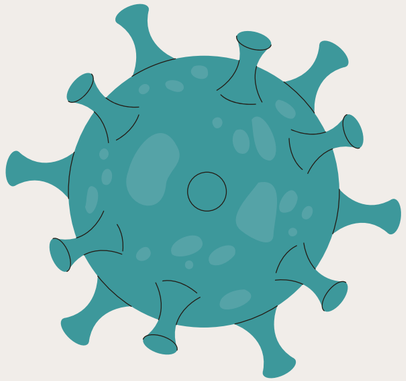
SEASONAL FLU VACCINE UPTAKE PREDICTION

BY RONALD ONGORI NYAGAKA

INTRODUCTION

The occurrence of the flu season repeats annually, and each year people make a choice to either receive the flu shot or not.

This attempts to develop a predictive model to forecast individuals' decision to receive the flu shot or not during the annual flu season.



PROBLEM STATEMENT

The aim of the project is to assist in public health planning and decision-making by providing insights into vaccination trends and helping allocate resources effectively to the United States government health Agencies

MAIN OBJECTIVE

Create a model that can predict seasonal flu vaccine uptake based on a person's background and patterns of behaviour

IMPORTANT FEATURES THAT SHAPE OUR MODEL

01

OPINION_SEAS_RISK

Respondent's opinion about risk of getting sick with seasonal flu without vaccine

02

DOCTOR_RECC_SEASONAL

Seasonal flu vaccine recommendation by daughter

03

AGE_GROUP

Age group of respondents

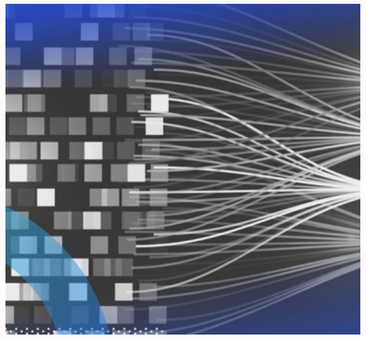
04

EMPLOYMENT_INDUSTRY

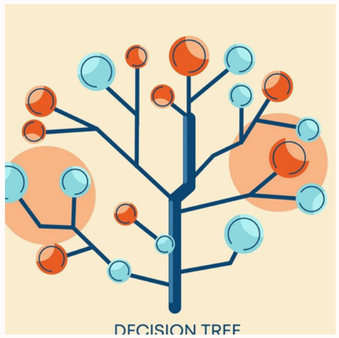
Type of industry respondent is employed in

MODELLING

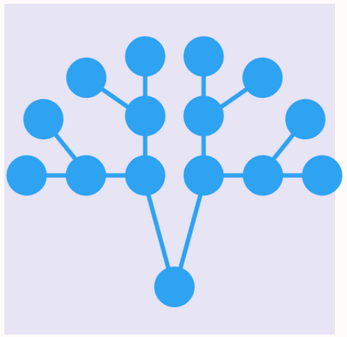
The baseline model was created using a Decision Tree Model. It had the lowest accuracy of 73.1 %



ADABoost

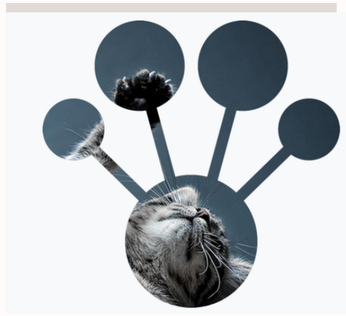


RandomForest



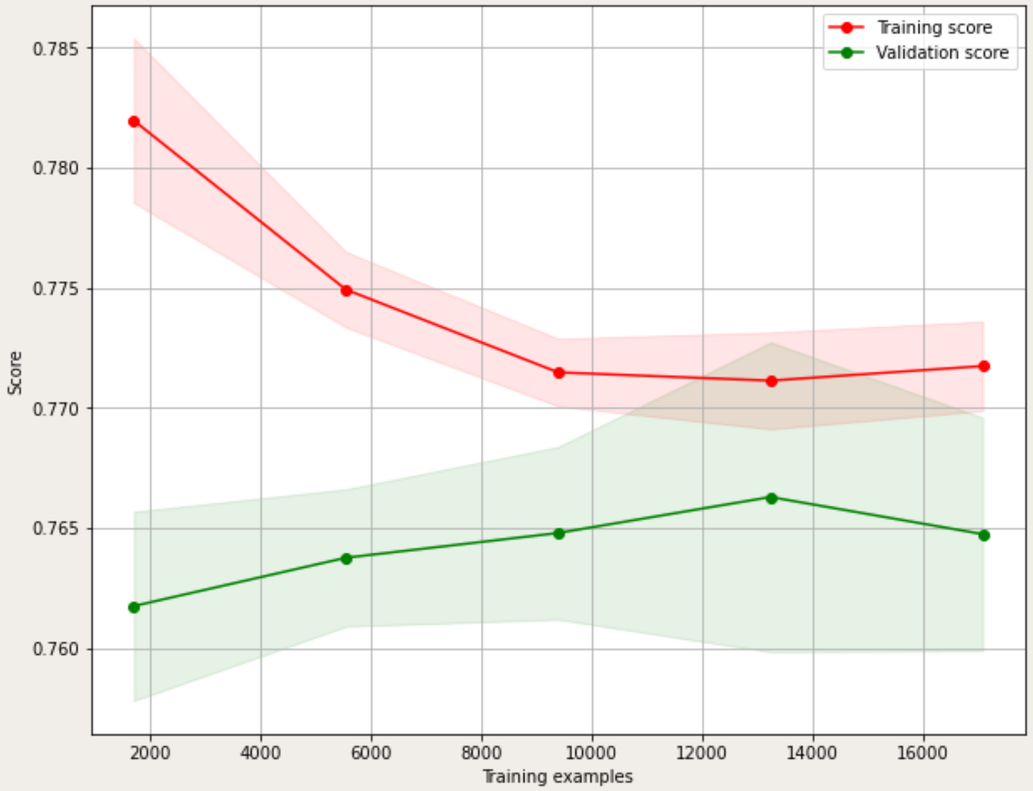
GradientBoost

XGBoost



CATboost

RANDOM FOREST



MODEL ACCURACY

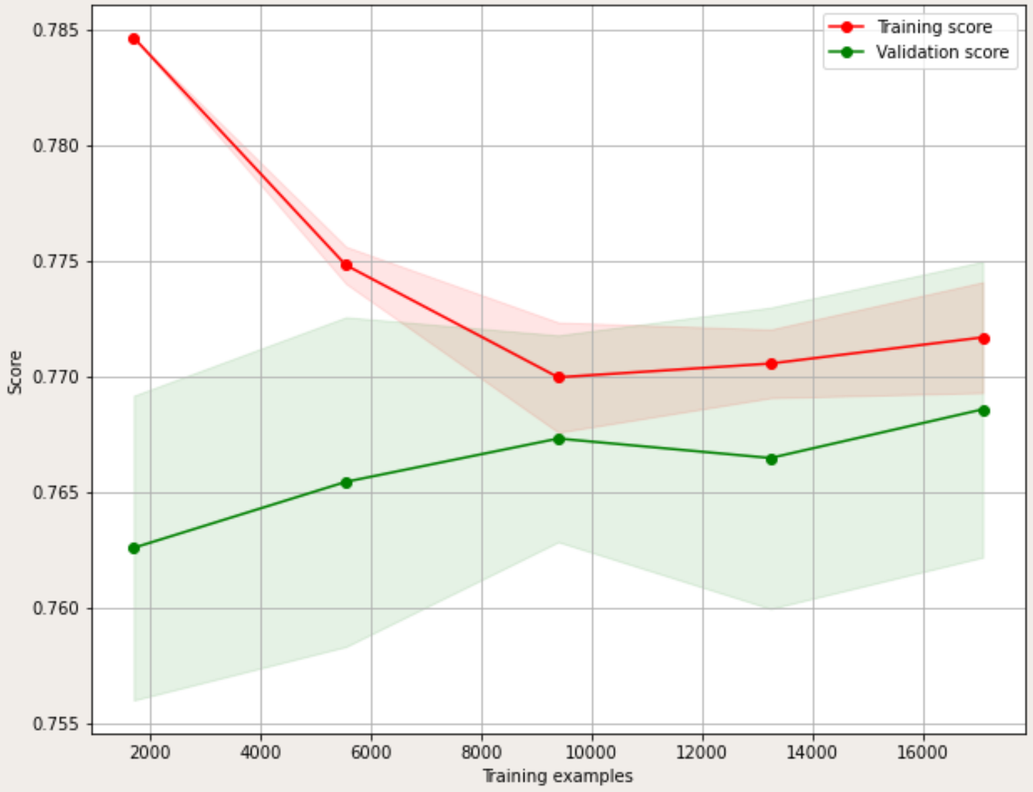
77 %

SPEED

9.48 sec

test score eventually plateaus and starts to decline, suggesting limitations in capturing complex patterns

GRADIENT BOOSTING



M O D E L
A C C U R A C Y

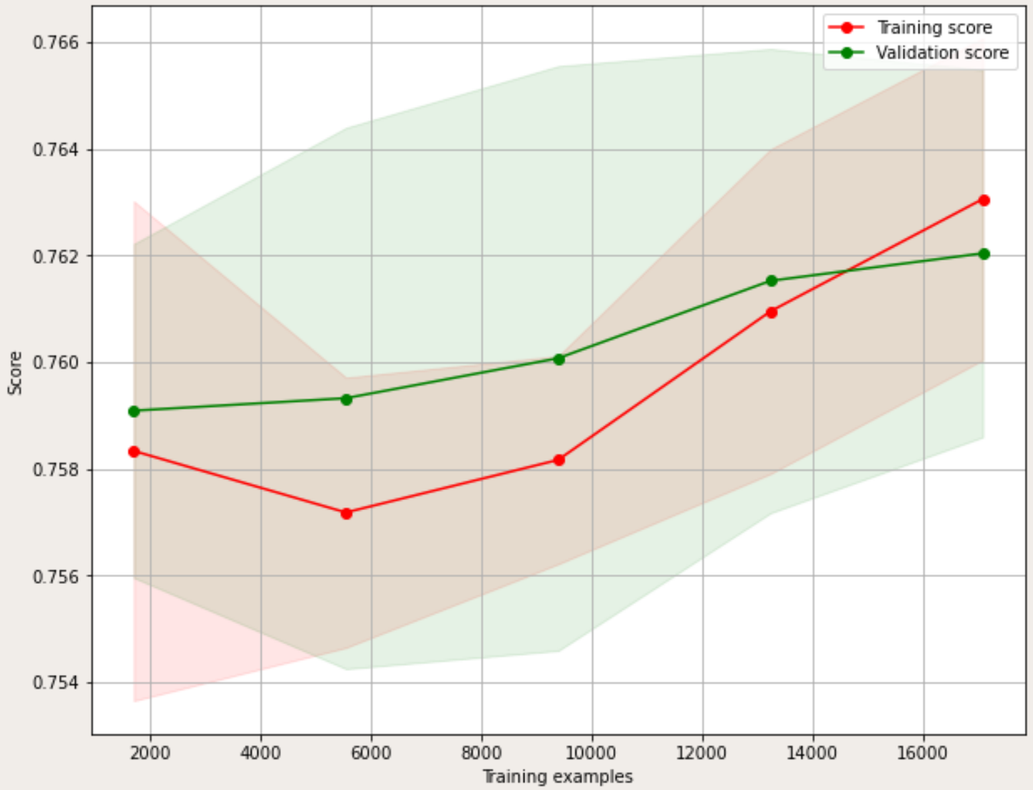
77 %

S P E E D

2.805 sec

Can adjust and imporve perfomance and generalizes well on unseen data. could improve if trained on more data

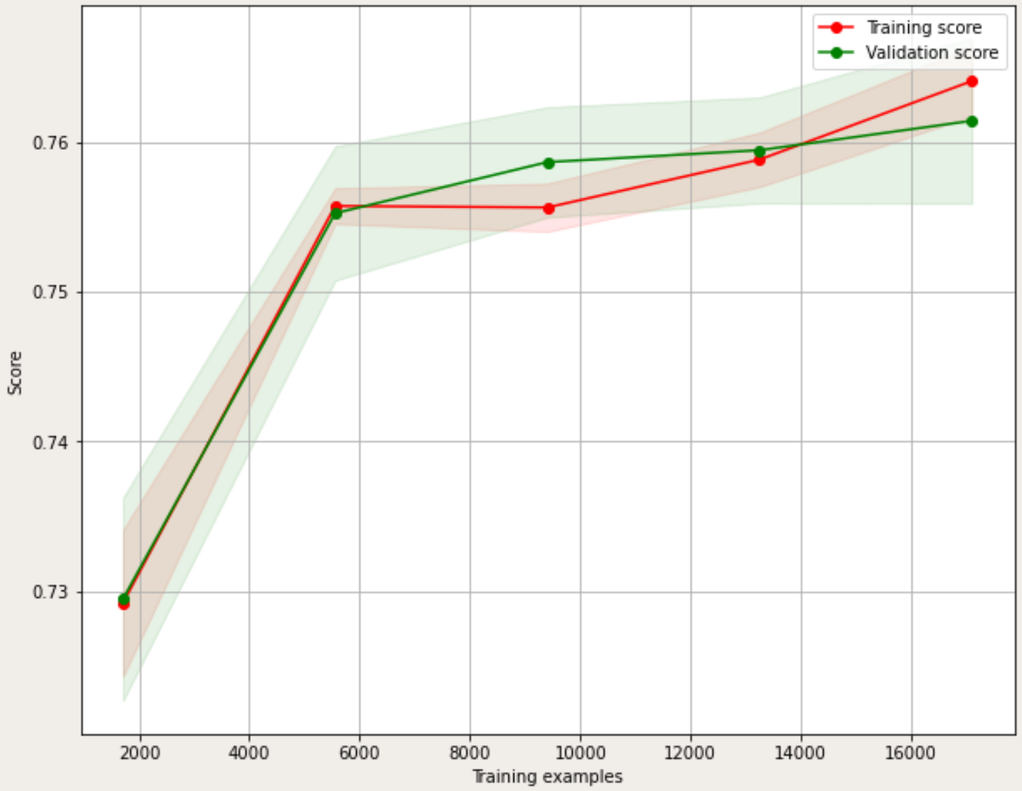
ADABOOSTING



MODEL ACCURACY	76 %
SPEED	3.50 sec

Model struggles to fit data initially but gradually improves. However its prediction ability starts to decline after some point

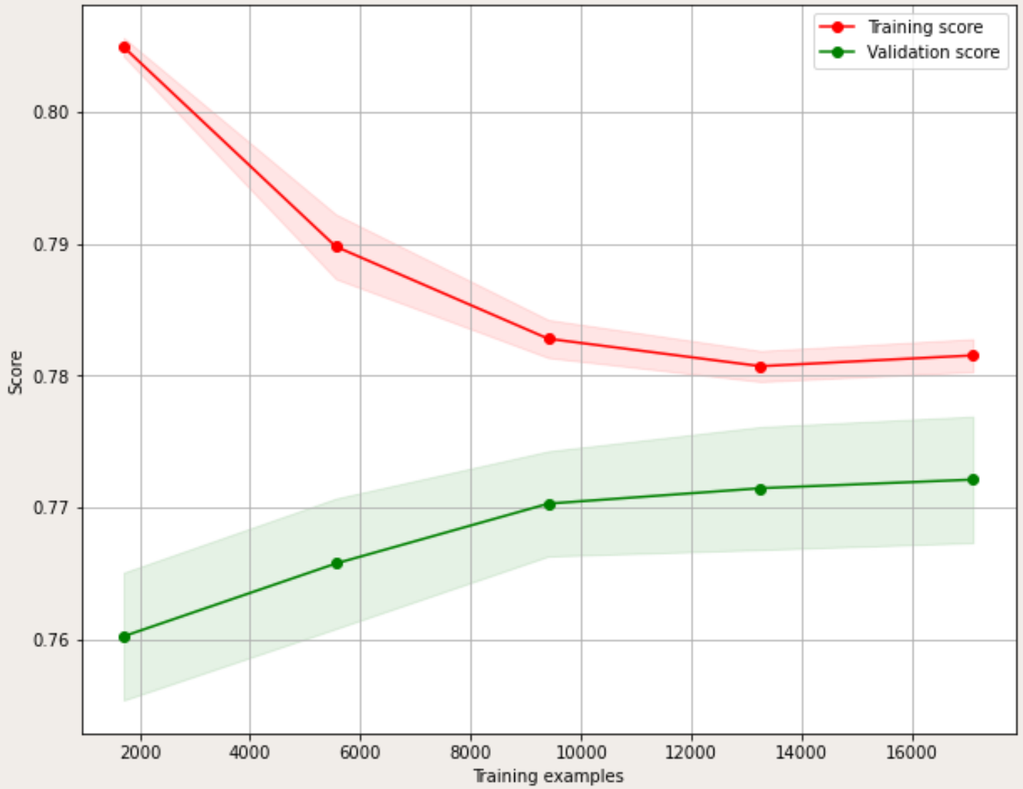
XGBOOST



MODEL ACCURACY	77 %
SPEED	8.58 sec

model benefits from additional training and has the potential to further improve its accuracy. However it starts to overfit

CATBOOST



MODEL ACCURACY	77 %
SPEED	4.15 sec

model is overfitting and shows signs of a plateauing performance as more data is trained

THE BEST MODEL: GRADIENT BOOSTING

the validation score rises steadily together with the training score as more data is trained. It is improving its predictive accuracy and it adjusts to improve performance on challenging instances.

Gradient boost has better prospects on predictive performance as more data is introduced

CONCLUSION

The data set was biased towards certain communities and classes.

This makes the models generated biased towards this group of people

Better insights would be generated if the data was more balanced.

Recommendation

- Communication of limitations, biases, and specific applicability of the models, Should be adhered to.
- Resampling techniques such as oversampling, undersampling can be employed to mitigate biase.
- Careful examination of the significance of identified predictors will help understand underlying factors and ensure fairness in decision-making
- Include a wider range of demographic groups, thereby achieving a more diverse and balanced dataset.