



**PREDICTING THE LIKELIHOOD OF INDIVIDUALS TAKING H1N1 FLU AND SEASONAL FLU
VACCINES USING LOGISTIC REGRESSION MODEL AND DECISION TREE CLASSIFIER (ID3)
MODELS**

**PRESENTED BY:
CHEBII PAMELA JEPKORIR**

CONTENTS

INTRODUCTION

- Background
- Overview
- Objectives

MODELLING

- Linear Regression
- Vanilla DT
- Tuned DT
- Complex DT

EVALUATION

- Model Comparison Table

conclusion

- Findings
- Recommendations

Introduction

Business Understanding

- Vaccination is a key public health measure used to fight infectious diseases.
- It is a simple, safe, and effective way of protecting a person against harmful diseases, before they come into contact with them (WHO, 2024).
- Vaccines provide immunization for individuals, and enough immunization in a community can further reduce the spread of diseases through "herd immunity."

Overview

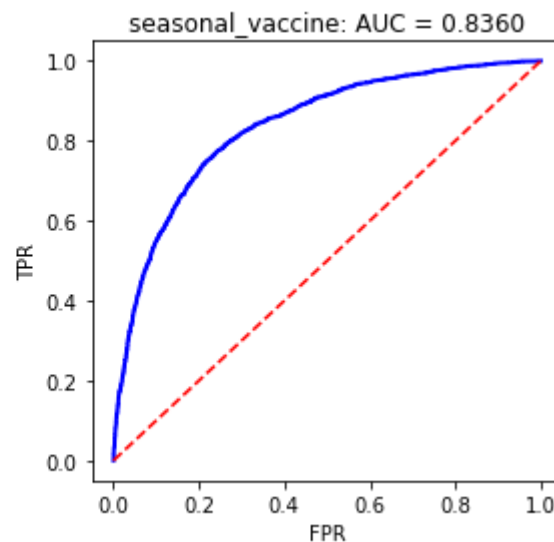
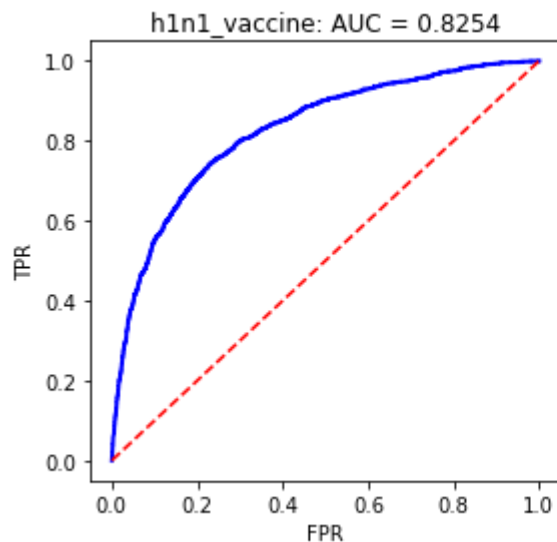
- The data for this research on uptake of vaccines based on opinions, demographics and doctor's recommendation is from United States National 2009 H1N1 Flu Survey collected to monitor vaccination rates during the US Government Vaccination Campaign that began in October 2009.
- A phone survey was used to ask people whether they had received H1N1 and seasonal flu vaccines, in conjunction with information they shared about their lives, opinions, and behaviors.

Objectives

- To predict whether people got H1N1 and Seasonal Flue Vaccines using information shared about their backgrounds, opinions and Health Behaviour.
- To determine the best performing model in predicting Vaccines

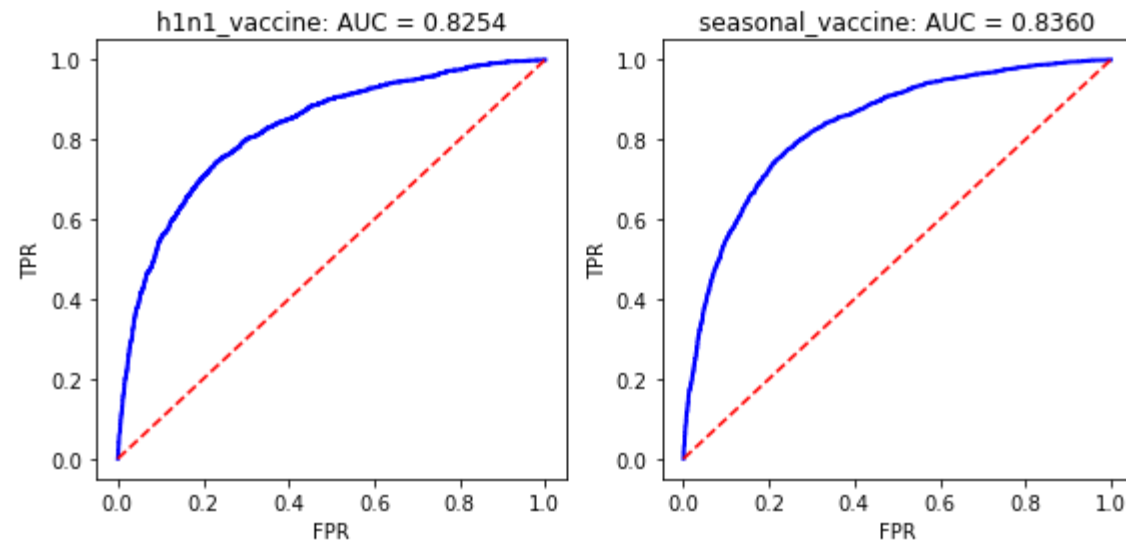
Logistic Regression Model

- It is evident that based on the information that was shared about people's backgrounds, opinions and health behaviours, 83% people took H1N1 and Seasonal Flue Vaccines.
- It therefore implies that a persons background, Opinion and Health Behaviour affect the uptake of a vaccine.
- The model performance is good but it is not perfect since it has an 83% performance out of the total 100%



Logistic Regression Model

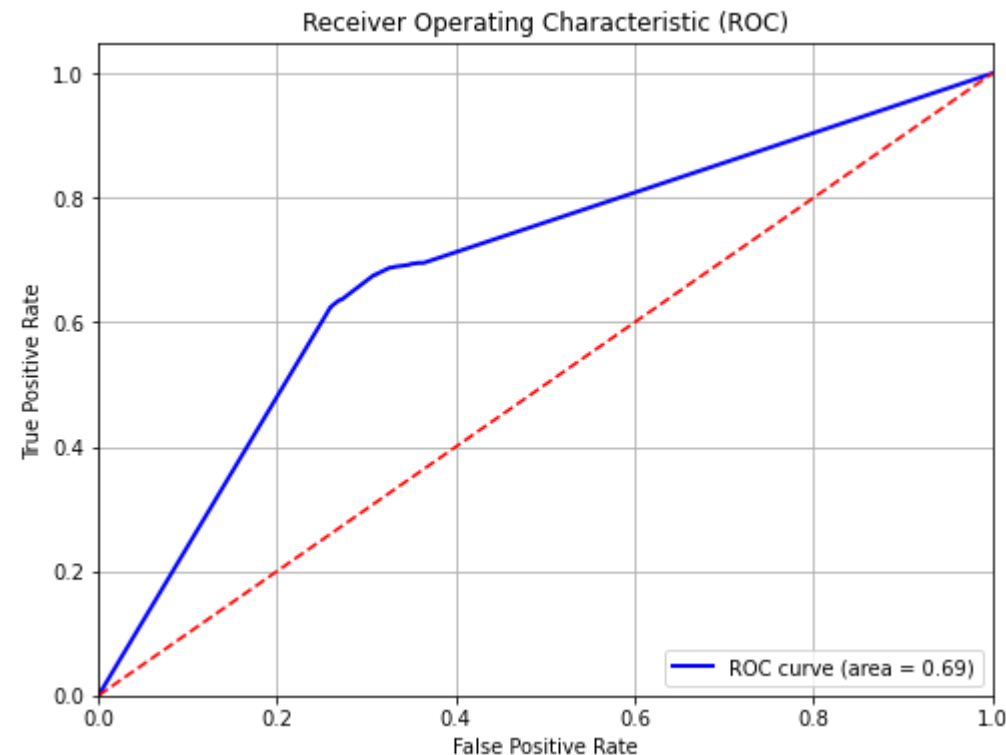
- It is evident that based on the information that was shared about people's backgrounds, opinions and health behaviours, 83% people took H1N1 and Seasonal Flue Vaccines.
- It therefore implies that a persons background, Opinion and Health Behaviour affect the uptake of a vaccine.
- The model performance is good but it is not perfect since it has an 83% performance out of the total 100%



MODELLING

Vanilla Decision Tree Model

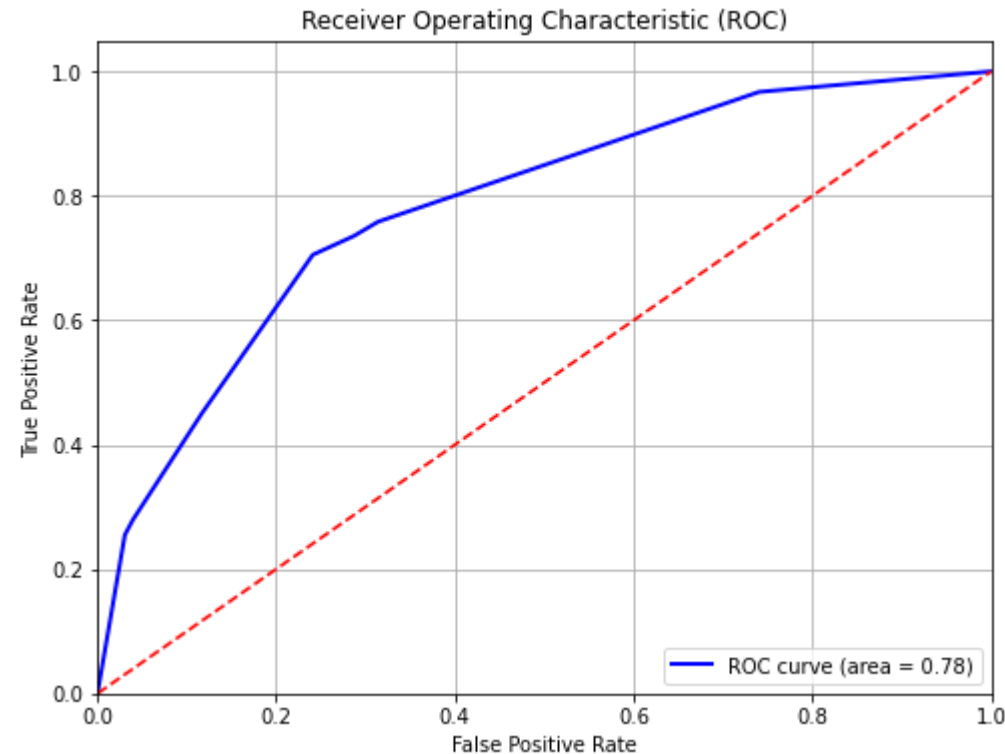
- The model performance using the decision tree is lower compared to the performance of the Logistic Regression Model.
- The score `roc_auc` is 69% whereas the score in the Logistic Regression is 83%.
- This might be because the number of features selected in the Logistic Regression Model were higher compared to the features selected for Decision Tree.
- The difference in score might also be caused by overfitting in the decision tree.
- To solve the overfitting problem, tree pruning / hyperparameter tuning will be done in the next model



MODELLING

Tuned Decision Tree Model

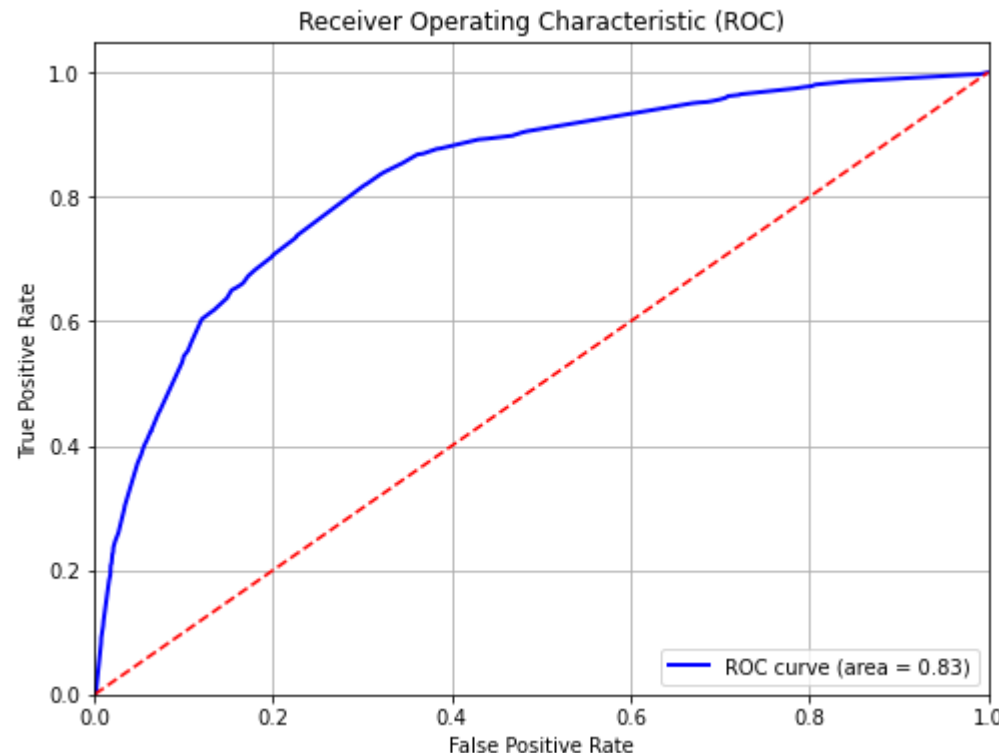
- The model has significantly improved (73% performance score) compared to the Vanilla decision tree model (69% performance score) that was not pruned.
- The Logistic Regression Model that had two target variables still performs better (83% performance score) compared to the two Decision Tree models.
- This might be attributed to the size of the Logistic Regression Model that had a higher number of Features compared to the features selected for Decision Tree.



MODELLING

Complex Decision Tree Model

- The performance of the model has significantly improved (83% auc_roc score) compared to the second and third models which were both decision tree models but had a performance of 69% and 78% respectively.
- This improvement in performance is attributed to the use of more features and the use of correct depth limit achieved through hyperparameter tuning.
- The model however, has the same performance when compared to the Logistic Regression model that had a performance score of 83% despite model four having more features than the Logistic Regression Model.
- To further improve the performance of the Decision Tree Classifier, other tree pruning methods like (minimum samples with leaf split, minimum leaf sample size, maximum leaf nodes, and maximum features) can be used instead of using maximum depth alone.



MODEL PERFORMANCE COMPARISON TABLE

CLASSIFICATION MODEL	TRAIN SCORE	TEST SCORE
Logistic Regression Model	0.8307	0.8307
Vanilla Decision Tree	0.9962	0.6902
Tuned Decision Tree	0.7856	0.7828
Complex Decision Tree	0.8491	0.8312

FINDINGS

- Logistic Regression Model has an average roc_auc score of 83.07% indicating that the model performs well in predicting H1N1 and Seasonal Flu Vaccines uptake
- The Vanilla Decision Tree Model has an roc_auc score of 69% which is a poor performance compared to the logistic regression model. This indicates that use of less variables without hyperparameter tuning results in poor model performance in Decision Trees.
- The tuned decision tree using similar features as the Vanilla model had an roc_auc score of 78% indicating that proper pruning of decision tree can lead to good performance of a model
- The tuned decision tree that was developed using an optimal depth with more than three more features (`sex`, `race` and `age_group`) compared to the logistic regression that had only 23 features performed better than all the models (roc_auc of 83.12%). This indicates that with proper tuning and increasing the number of features the model can perform better in predicting vaccines

Conclusion & Recommendations

- When performing hyperparameter tuning for the decision tree classifier, only the Maximum Depth was used. To improve the performance of the Decision Tree Classifier, other tree pruning methods like (minimum samples with leaf split, minimum leaf sample size, maximum leaf nodes, and maximum features) can be used instead of using maximum depth alone.
- The decision tree classifier used only one target variable (seasonal_vaccine). Both outcomes can be included in the decision tree model to determine whether the model performance will improve.



THANK YOU!