

H1N1_VACCINE PREDICTION





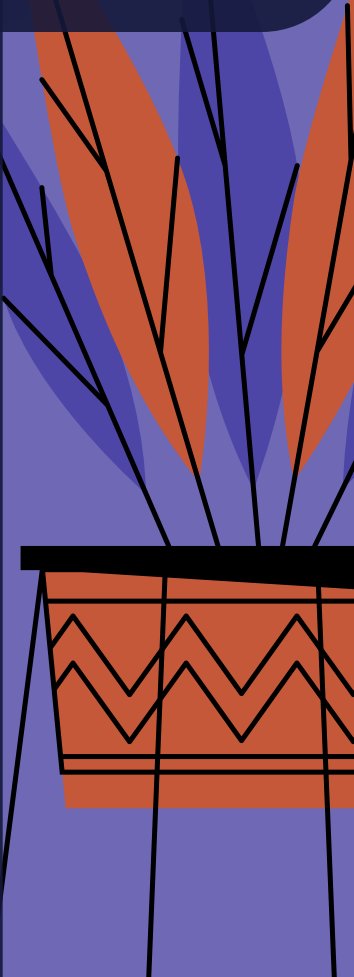
BUSINESS OVERVIEW

The Ministry of Health wants to administer Covid-19 vaccines but they do not know how people will respond. The ministry revisited past data with the information of vaccination cases in late 2009 and early 2010 conducted by the United States. I have been given this data and tasked to predict whether people received H1N1 vaccines according to their backgrounds, opinions and health behaviors.

DATA UNDERSTANDING

For this project, i will be using information from datasets:

- **training_set_features.csv**- A dataset containing personal information of individuals who responded to the survey about h1n1_vaccination.
- **training_set_labels.csv** - A dataset containing the label features 0 and 1. 0 for those who received vaccines and 1 for those who didn't.



MODELING

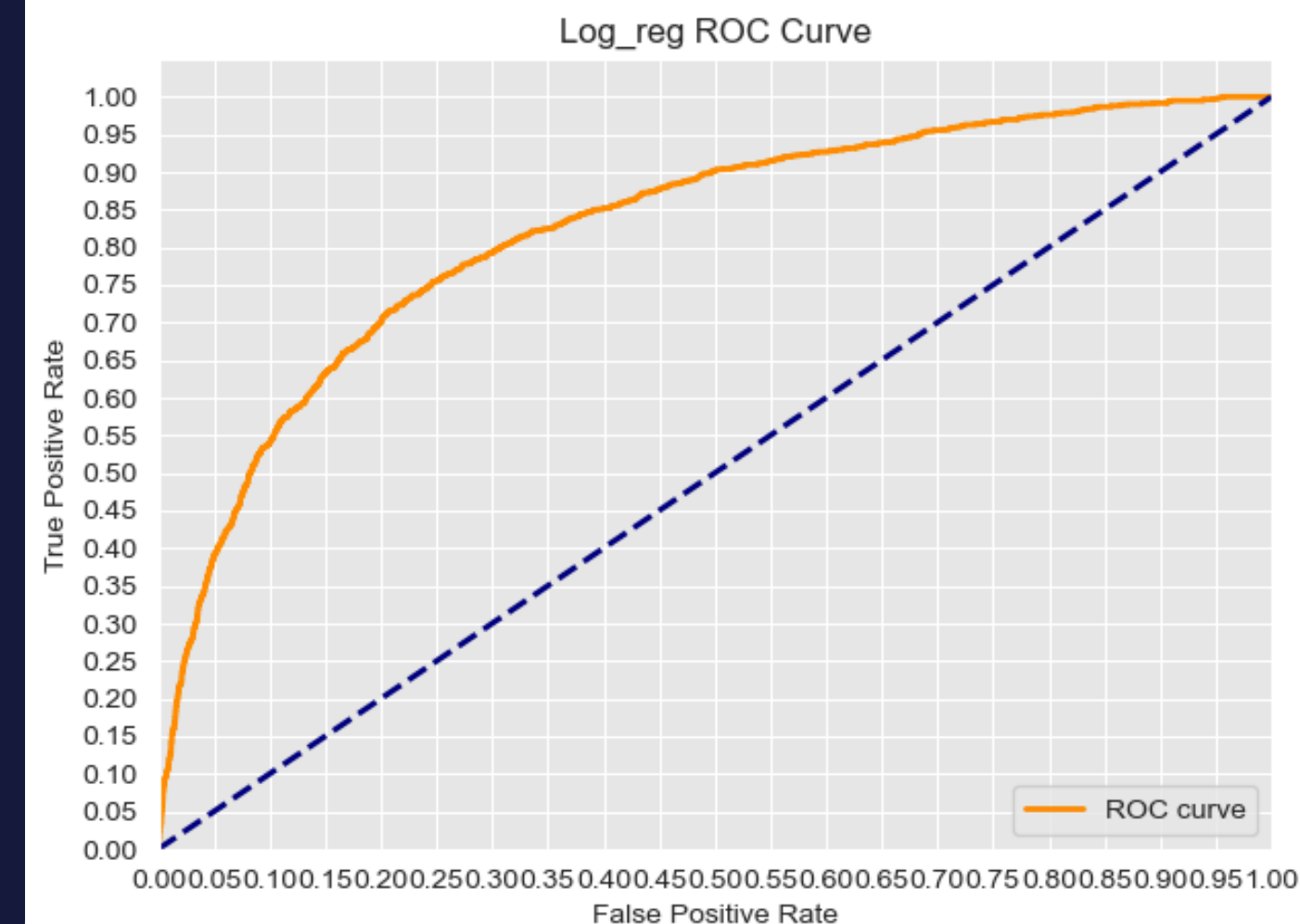
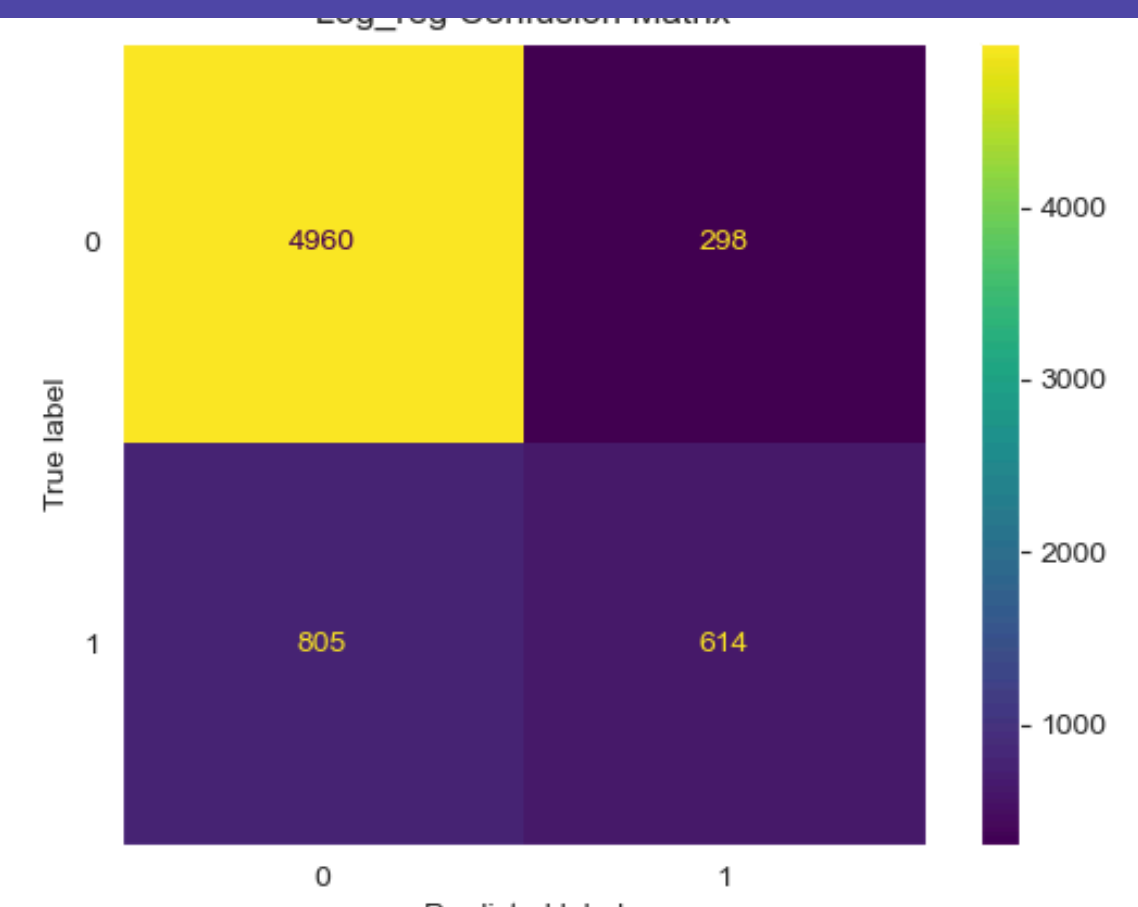
The case modeling is a Binary classification problem. I made two models,

1. Logistic Regression model
2. Decision Tree classification Model

These models were useful since they would help us predict whether someone received vaccine or not

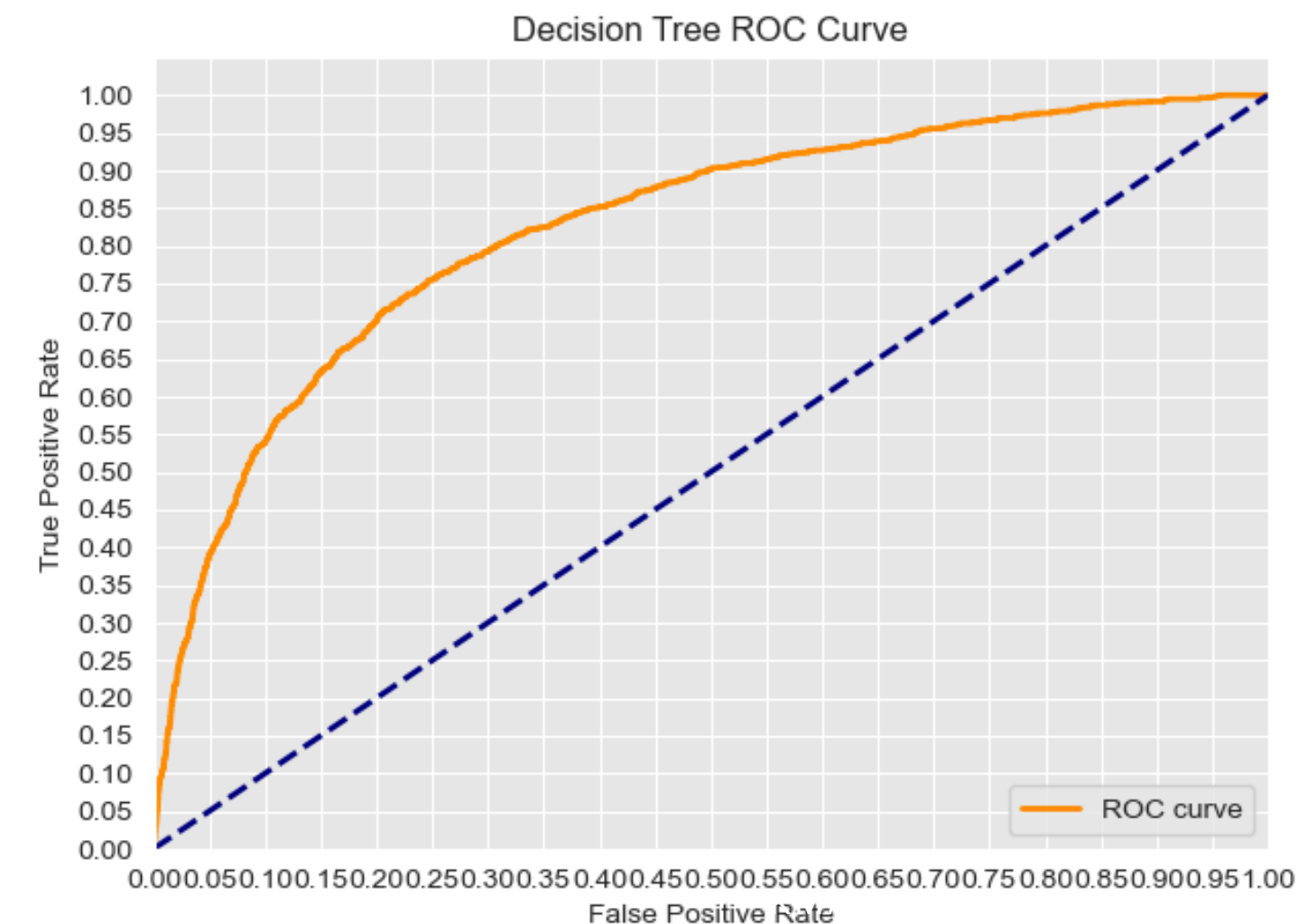
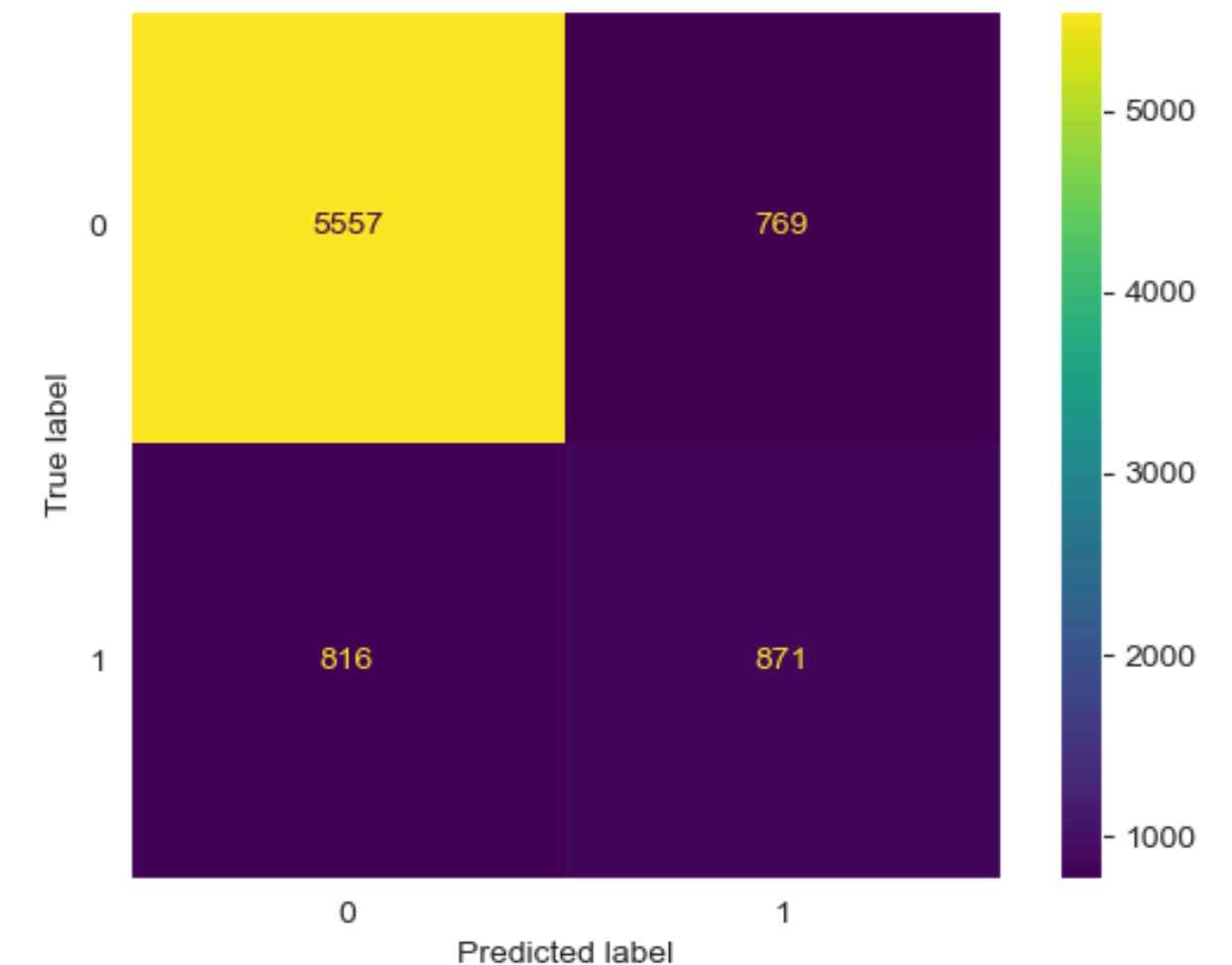
1. Logistic Regression Model

- In The Modelling process I used Area Under the curve ,ROC and AUC as my evaluation metrics.
- I made a logistic regression model and came up with the following ROC graph and Confusion Matrix after Tuning the model.
- The logistic regression model also provided a testing accuracy score of 83% and a testing precision of 67%.
- The logistic regression has a higher precision than recall. This works better with our model since it is better to predict one as not being vaccinated while they are ,than predicting them as vaccinated while they aren't.



2.DECISION TREE CLASSIFICATION MODELING

- In the decision Tree, I used AUC ROC and Accuracy score as my evaluation metrics.
- Decision Tree Classifier Model Evaluation
- After evaluating and tuning the decision tree classifier, it had an ROC AUC of around 80%
- The Decision Tree classifier had an accuracy of about 70%.



CONCLUSIONS



From the modelling process, I made two models: logistic regression and a decision tree classification model. I Made the following conclusions:

- After evaluating and tuning the logistic regression , It had a ROC AUC score of around 83%.
- The logistic regression model also provided a testing accuracy score of 83% and a testing precession of 67%.
- The logistic regression has a higher precession than recall.
- After evaluating and tuning the decission tree classifier, it had an ROC AUC of around 80%
- The Decission Tree classifier had an accuracy of about 70%

RECOMMENDATIONS

Of the two models, it is better to utilize the regression model. The regression model has a higher accuracy score of 83% and a precision of 67%.

- This precision is way better than recall. This is useful in our prediction cases. It is better to predict a person as having taken the vaccine than predicting a person that didn't take the vaccine as having taken the vaccine.
- This model can be used as a reference for the government to introduce vaccines for the first time.
- It will also be useful by the Ministry of Health to begin distributing vaccines due to an emerged virus.



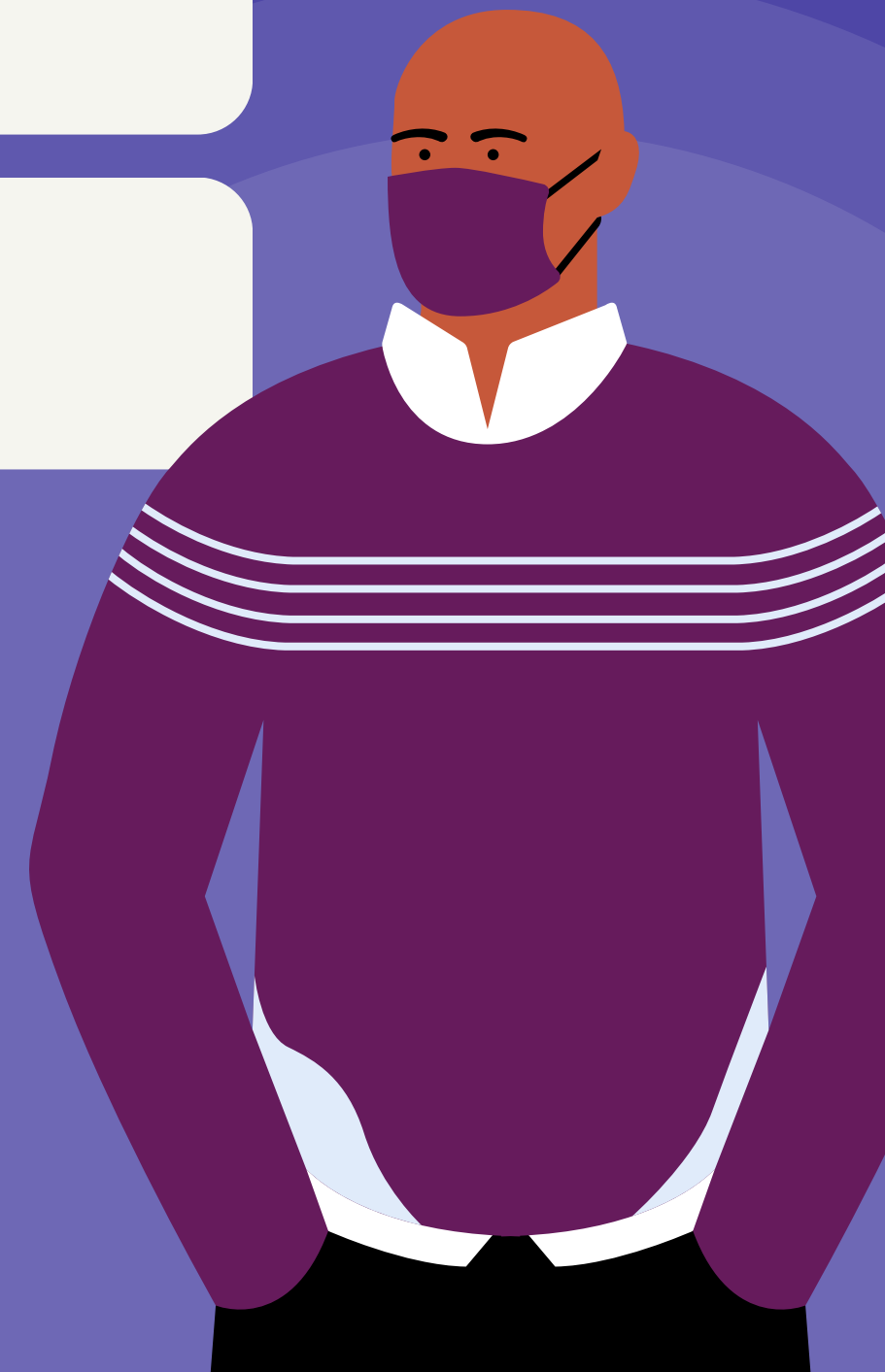
**This project was conducted
By Baraka Wambua**

Phone Number

0720011403

Email Address

baraka.nyamai@student.moringa
school.





THANK YOU !

KMA VACCINE RESEARCHERS