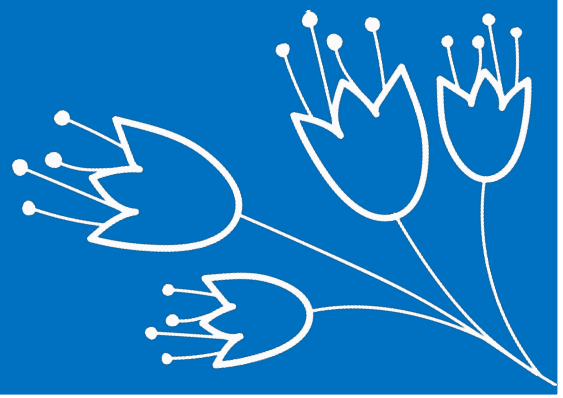
Predicting vaccination rate

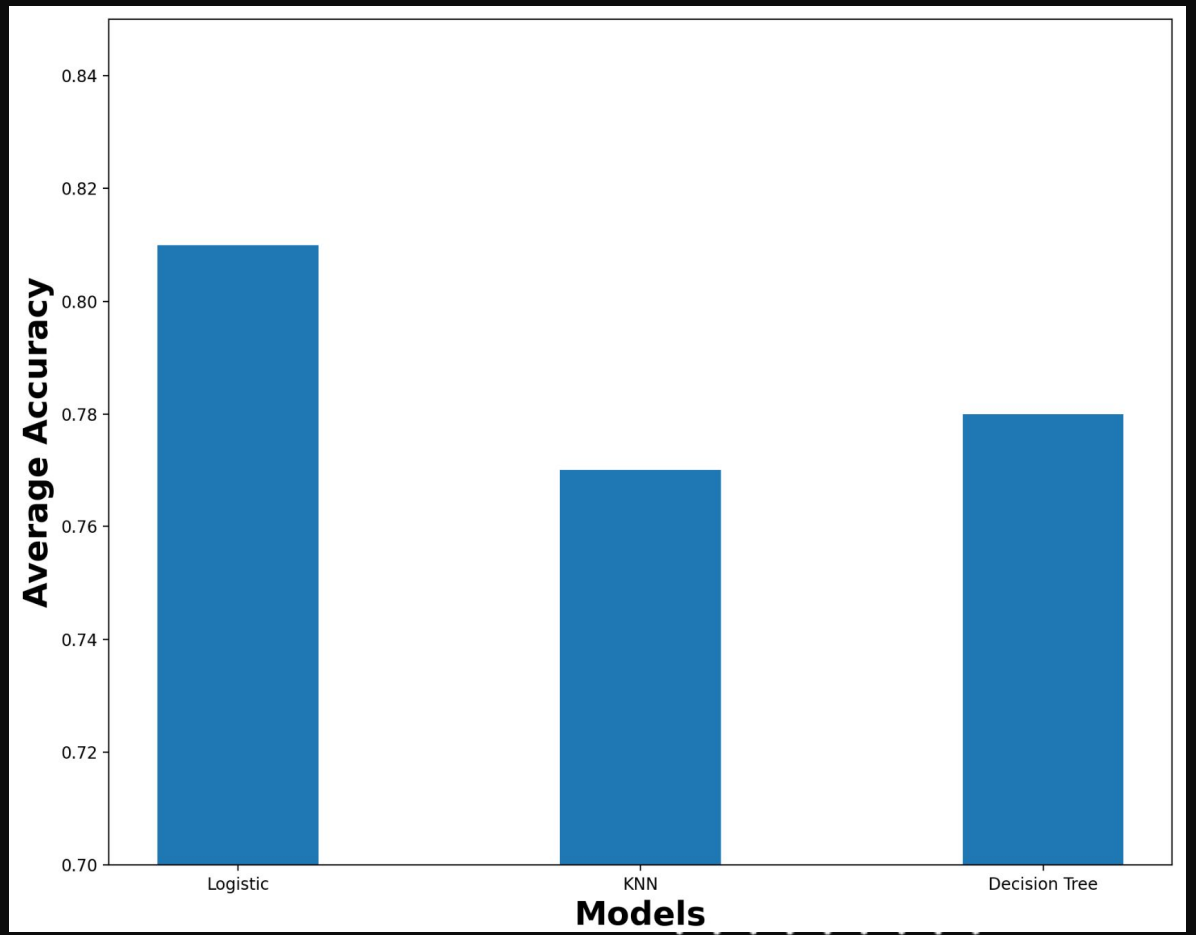
Joshua Ko

About the data

- ✔ Provided by Drivendata
- ✔ Features
  - Behavioral
  - Economic/Social status
  - Medical history
  - etc.
- ✔ Targets
  - H1N1 Vaccination
  - Seasonal Flu Vaccination



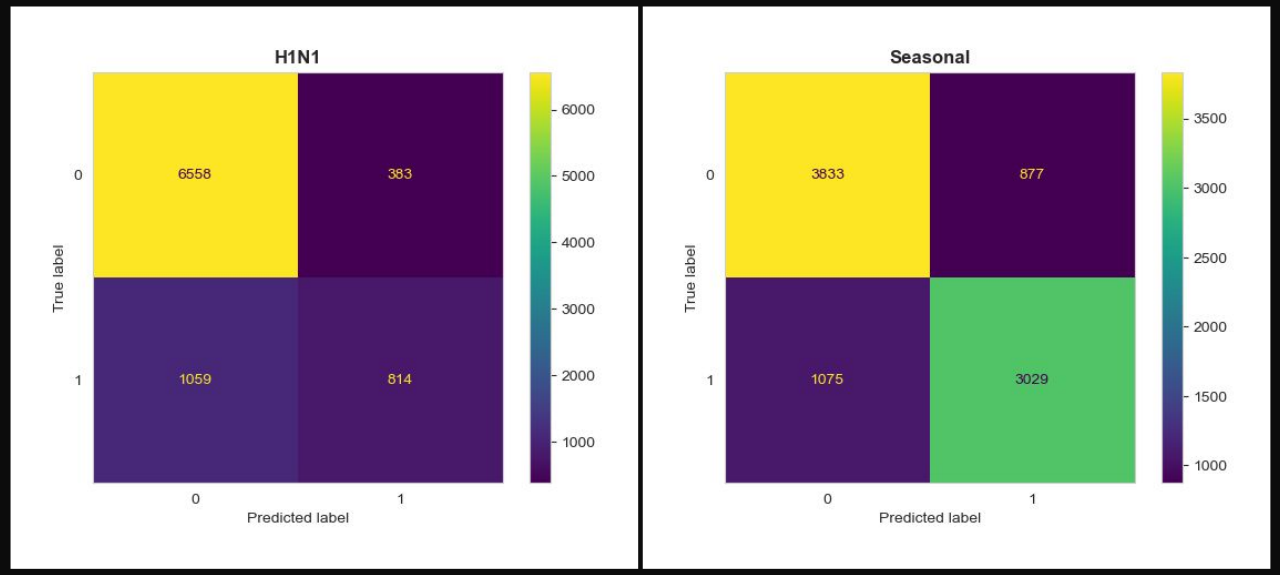
Accuracy of each model



Performance of the model

- Confusion Matrix
- Positive and negative predictive values
  - Precision
  - Recall
  - Accuracy
  - F1
- Receiver operator characteristic curve (ROC)

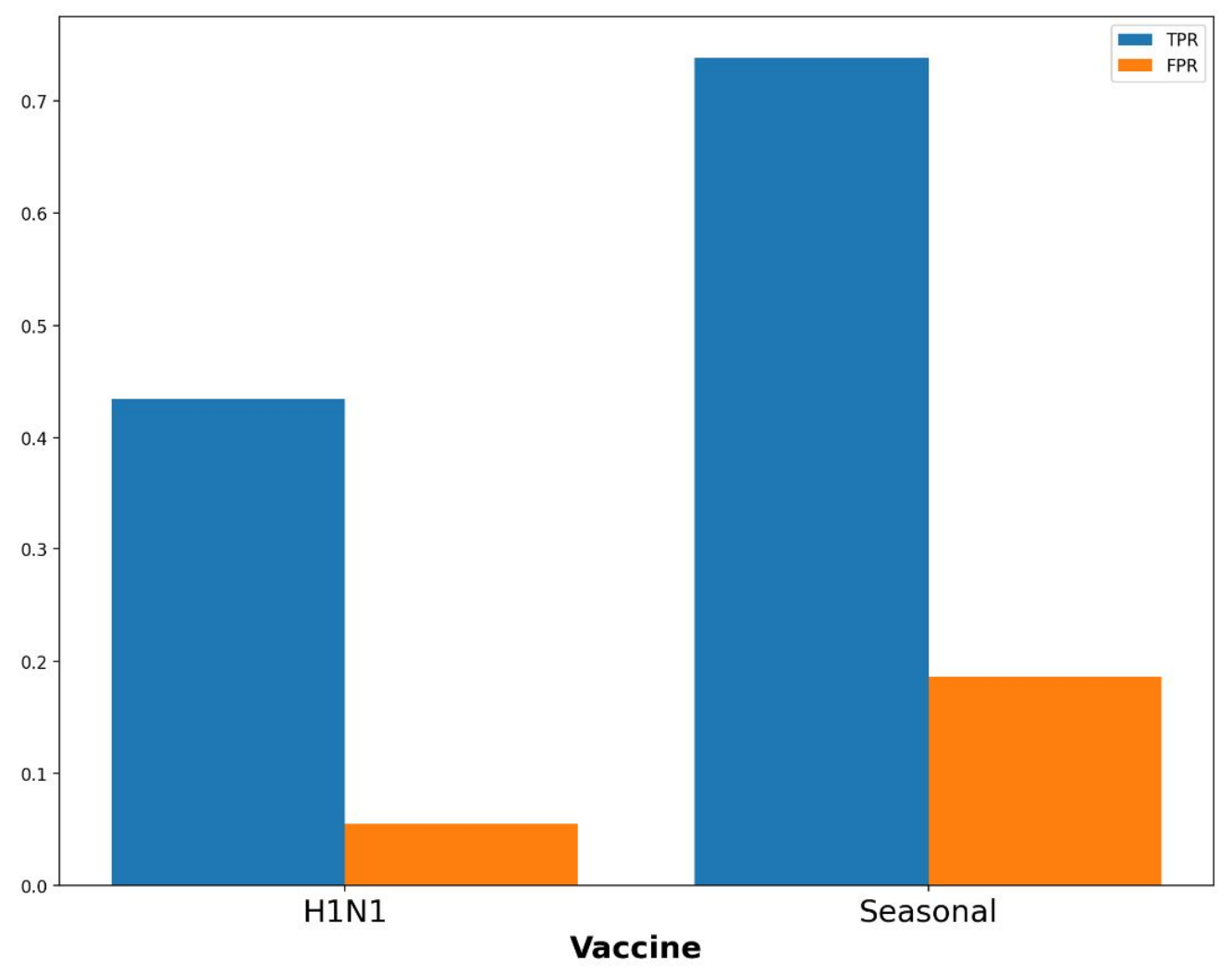
Confusion Matrix



- True positives (bottom left) and false positives (top right) are most important
- True positive indicates a correct prediction in receiving the vaccine
- False positive indicates an incorrect prediction

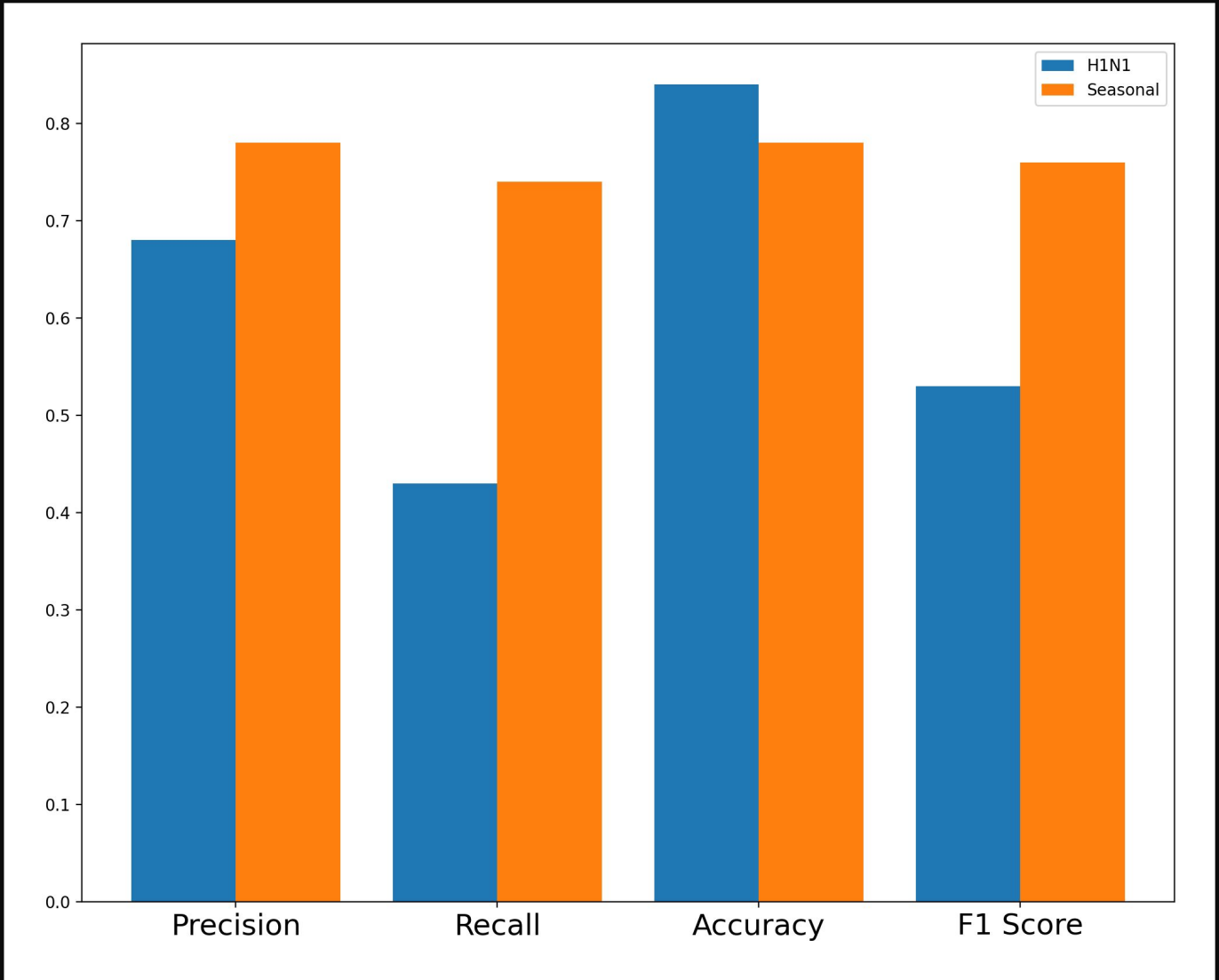
Pfizer should look for high true positive and low false positive values

True Positive Rate (TPR) and False Positive Rate (FPR)



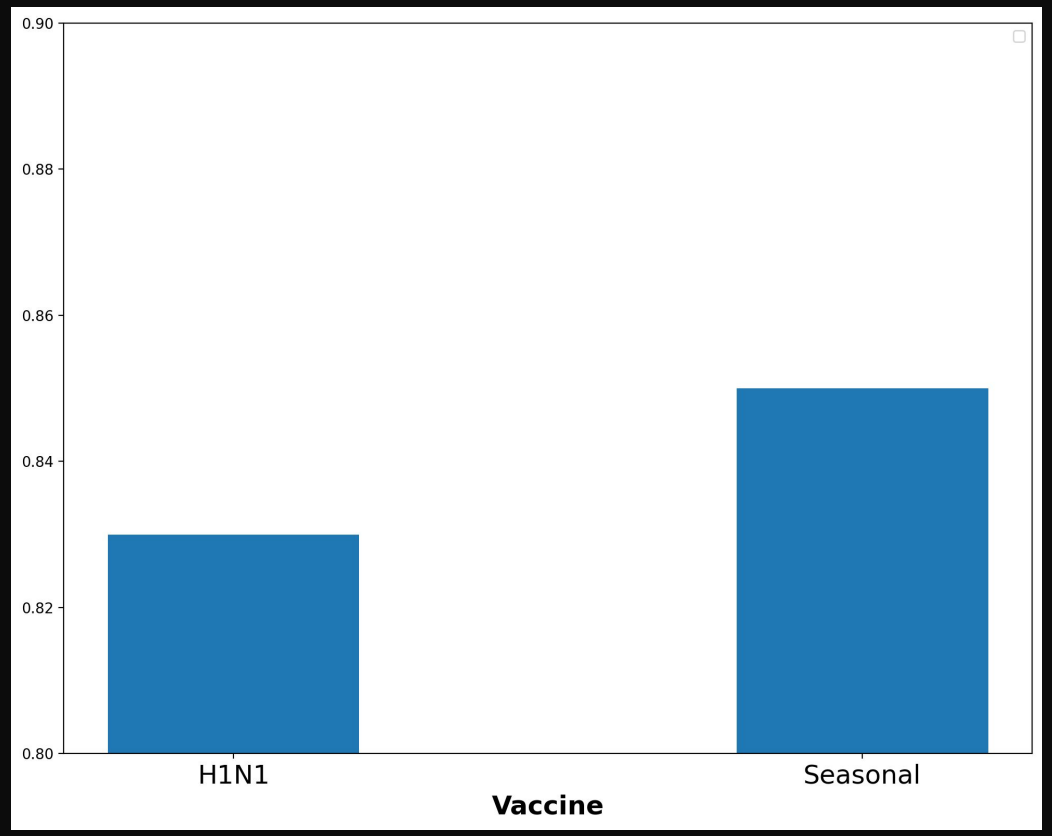
- TPR can help predict how much money Pfizer will make
- Seasonal model had the higher TPR but also the higher FPR

Prediction Scores



- H1N1 has the higher accuracy but is lower in all other metrics
- Low recall is same as TPR
- Precision is another metric for true positives
- F1 score is the average of the two (Precision & Recall)
- Higher F1 score the better

Area under ROC curve



- A higher AUC value means better efficiency of the model
- Seasonal vaccine model has a higher value
    - Able to distinguish between positive and negative classes

Final predictions & recommendations

Vaccine	Average Probability of Receiving the Seasonal Vaccine
H1N1	0.21
Seasonal	0.46

- The logistic model should be used to predict vaccination rate
  - Values of the table above was obtained from the logistic model
- More people are likely to take the seasonal vaccine
  - Pfizer should produce more seasonal vaccine
- Investigate why certain variables yield lower rates
  - Race
  - Education
  - Income
- May use this outcome to push future vaccinations

Improvements & Further developments

- **Making a more accurate model**
  - Working with less null values
  - Finding the best predictive model for each column
  - Trying different combinations of hyperparameters
- **Removing variables**
  - The dataset has multiple columns
  - Removing more irrelevant columns

Thank You!

This Presentation is Prepared by

Link:
https://github.com/Jko0425

Email:
joshuawko@gmail.com