



**Data Glacier**

Your Deep Learning Partner

# Final Report

ABC Pharma

Julia Donato

**29-March-2023**

# Agenda

Problem Statement

Logistic Regression

Random Forest

Support Vector Machines

AdaBoost

Feature Selection

Deployment

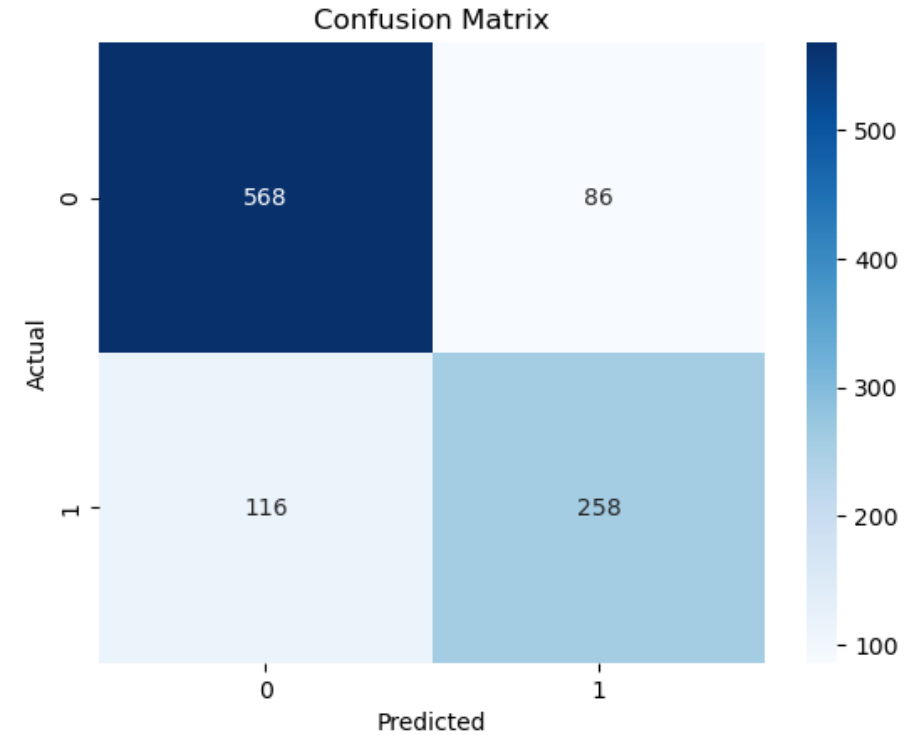
# Problem Statement

- ABC Pharma would like to gain insight on the persistency of their drug as per the physician's prescription.
- Objective: Explore model selection process and how the final model was chosen. Evaluate and deploy the model.

# Logistic Regression

- max\_iter=1000
- ROC-AUC score indicates efficiency
  - Score : 0.8723773078872916
- \*1: Persistent 0:Non-Persistent

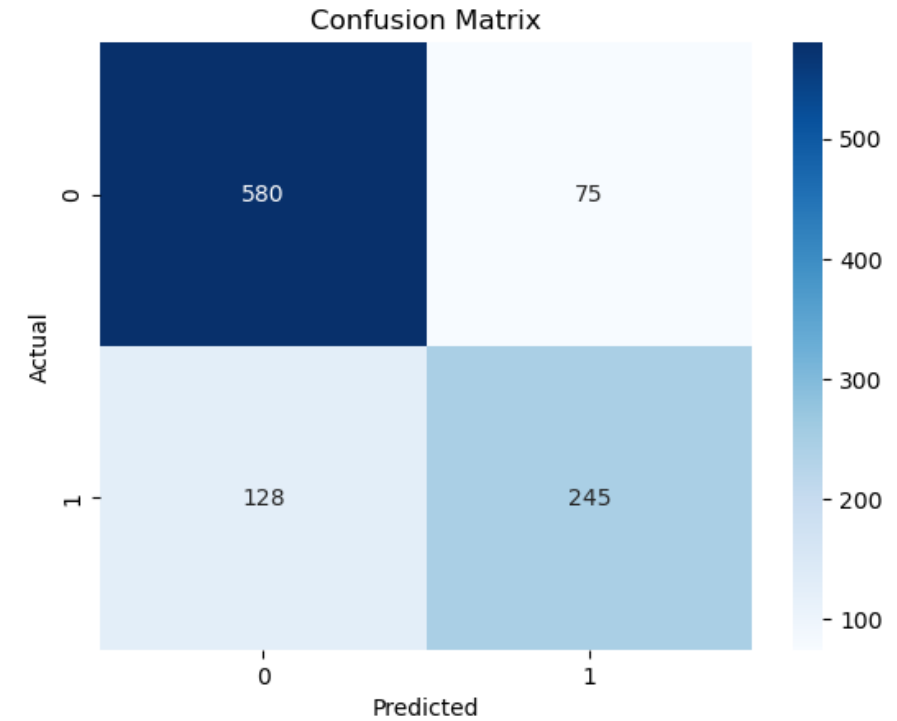
	precision	recall	f1-score	support
0	0.83	0.87	0.85	654
1	0.75	0.69	0.72	374
accuracy			0.80	1028
macro avg	0.79	0.78	0.78	1028
weighted avg	0.80	0.80	0.80	1028



# Random Forest

- N=1000 estimators
- ROC-AUC score: 0.8572641876266296

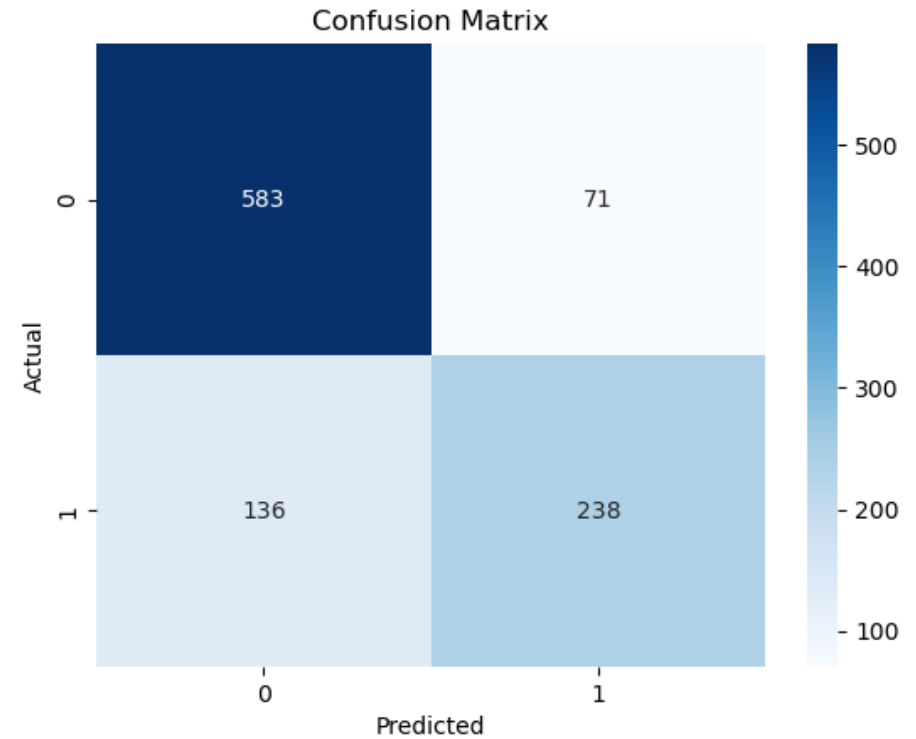
	precision	recall	f1-score	support
0	0.82	0.89	0.85	655
1	0.77	0.66	0.71	373
accuracy			0.80	1028
macro avg	0.79	0.77	0.78	1028
weighted avg	0.80	0.80	0.80	1028



# Support Vector Machines

- Probability=true
- ROC-AUC score: 0.8727125545797969

	precision	recall	f1-score	support
0	0.81	0.89	0.85	654
1	0.77	0.64	0.70	374
accuracy			0.80	1028
macro avg	0.79	0.76	0.77	1028
weighted avg	0.80	0.80	0.79	1028

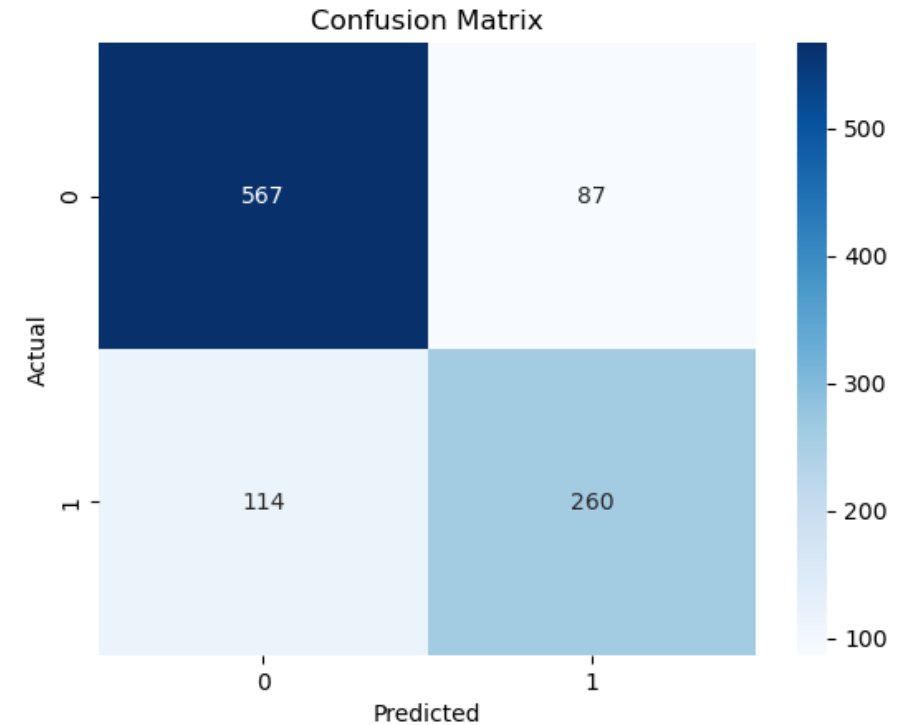


# AdaBoost

- `base_estimator=DecisionTreeClassifier(max_depth=1)`
- `n_estimators=1000`
- `learning_rate=1.0`
- ROC-AUC score: 0.8720420611947864

	precision	recall	f1-score	support
0	0.83	0.87	0.85	654
1	0.75	0.70	0.72	374
accuracy			0.80	1028
macro avg	0.79	0.78	0.79	1028
weighted avg	0.80	0.80	0.80	1028

Overall, the accuracy, confusion matrix, and ROC-AUC score point to AdaBoost being the ideal model when compared with the other models



# Feature Selection

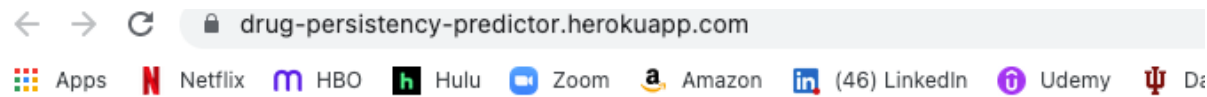
- In order to make the web app user friendly, Chi-Squared testing was used for feature selection in order to identify the top 5 features
- Scores reflect the feature's influence with Dexa\_Freq\_During\_Rx having the most influence by a significant amount

	Feature	Score
0	Dexa_Freq_During_Rx	8443.923832
1	Dexa_During_Rx	601.821735
2	Comorb_Long_Term_Current_Drug_Therapy	324.413431
3	Comorb_Encounter_For_Screening_For_Malignant_N...	196.456669
4	Comorb_Encounter_For_Immunization	189.482869



# Model Deployment

- The AdaBoost model with feature selection was deployed to Heroku
- <https://drug-persistence-predictor.herokuapp.com/>



## Model Prediction

Dexa\_Freq\_During\_Rx:

Dexa\_During\_Rx(Yes:1 No:0):

Comorb\_Long\_Term\_Current\_Drug\_Therapy(Yes:1 No:0):

Comorb\_Encounter\_For\_Screening\_For\_Malignant\_Neoplasms(Yes:1 No:0):

Comorb\_Encounter\_For\_Immunization(Yes:1 No:0):

## Model Prediction

Dexa\_Freq\_During\_Rx:

Dexa\_During\_Rx(Yes:1 No:0):

Comorb\_Long\_Term\_Current\_Drug\_Therapy(Yes:1 No:0):

Comorb\_Encounter\_For\_Screening\_For\_Malignant\_Neoplasms(Yes:1 No:0):

Comorb\_Encounter\_For\_Immunization(Yes:1 No:0):

The predicted label is Non-Persistent.

# Thank You