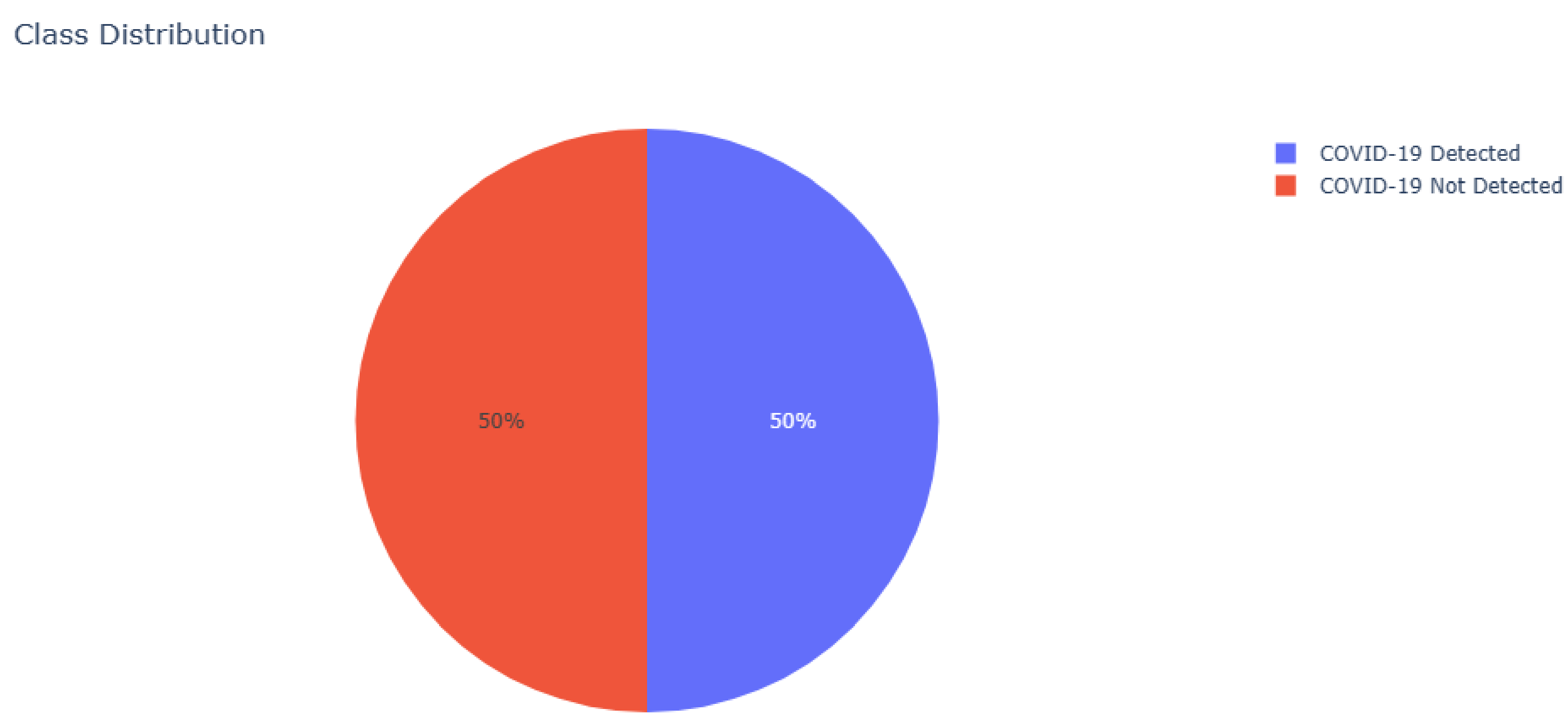
Clinical AI-ML Case Study

Coronavirus disease (COVID-19) is an infectious disease caused by a newly discovered coronavirus. Most people infected with COVID-19 virus will experience mild to moderate respiratory illness and recover without requiring special treatment. Older people, and those with underlying medical problems like cardiovascular disease, diabetes, chronic respiratory disease, and cancer are more likely to develop serious illness.

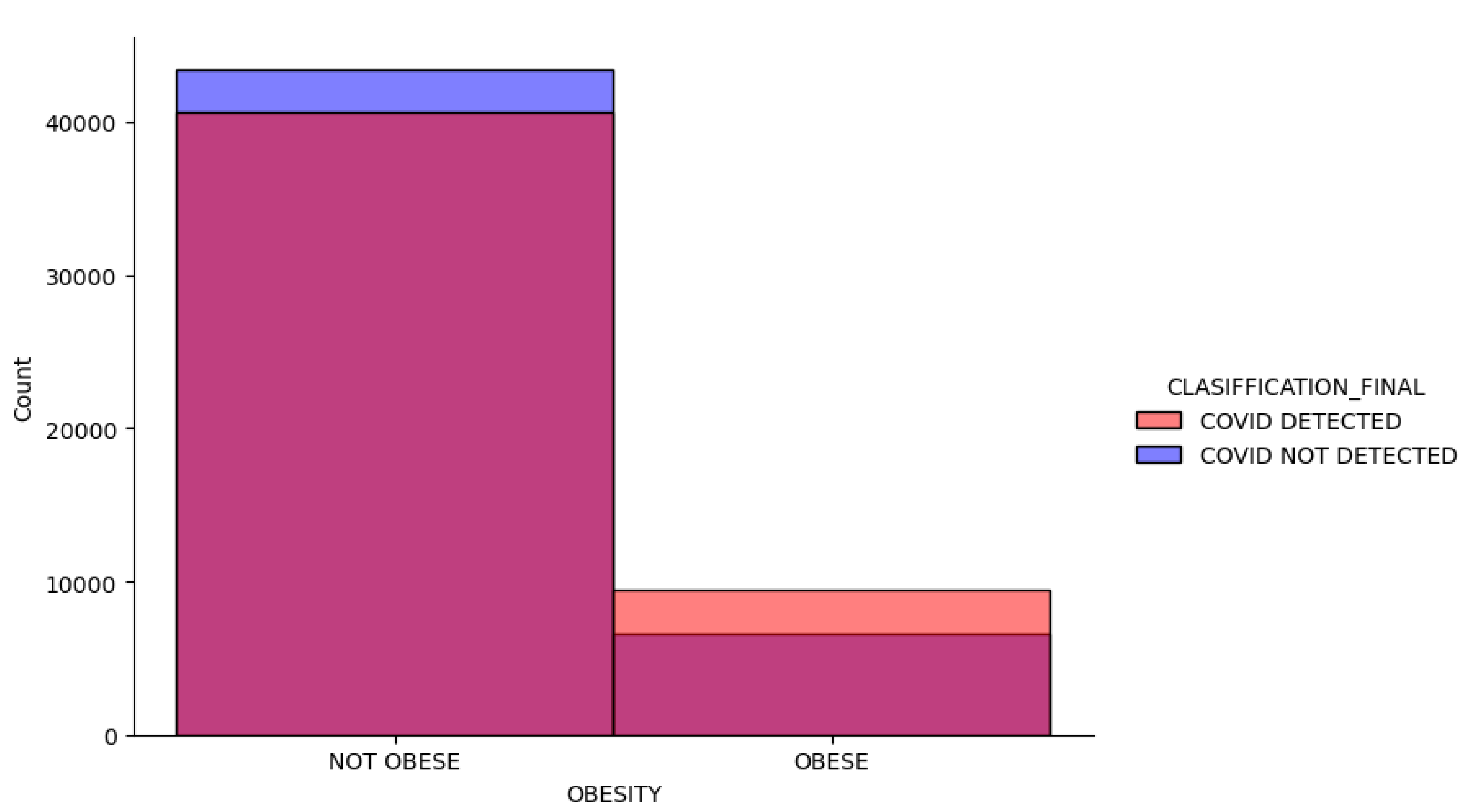
During the entire course of the pandemic, one of the main problems that healthcare providers have faced is the shortage of medical resources and a proper plan to efficiently distribute them. In these tough times, being able to predict what kind of resource an individual might require at the time of being tested positive or even before that will be of immense help to the authorities as they would be able to procure and arrange for the resources necessary to save the life of that patient.

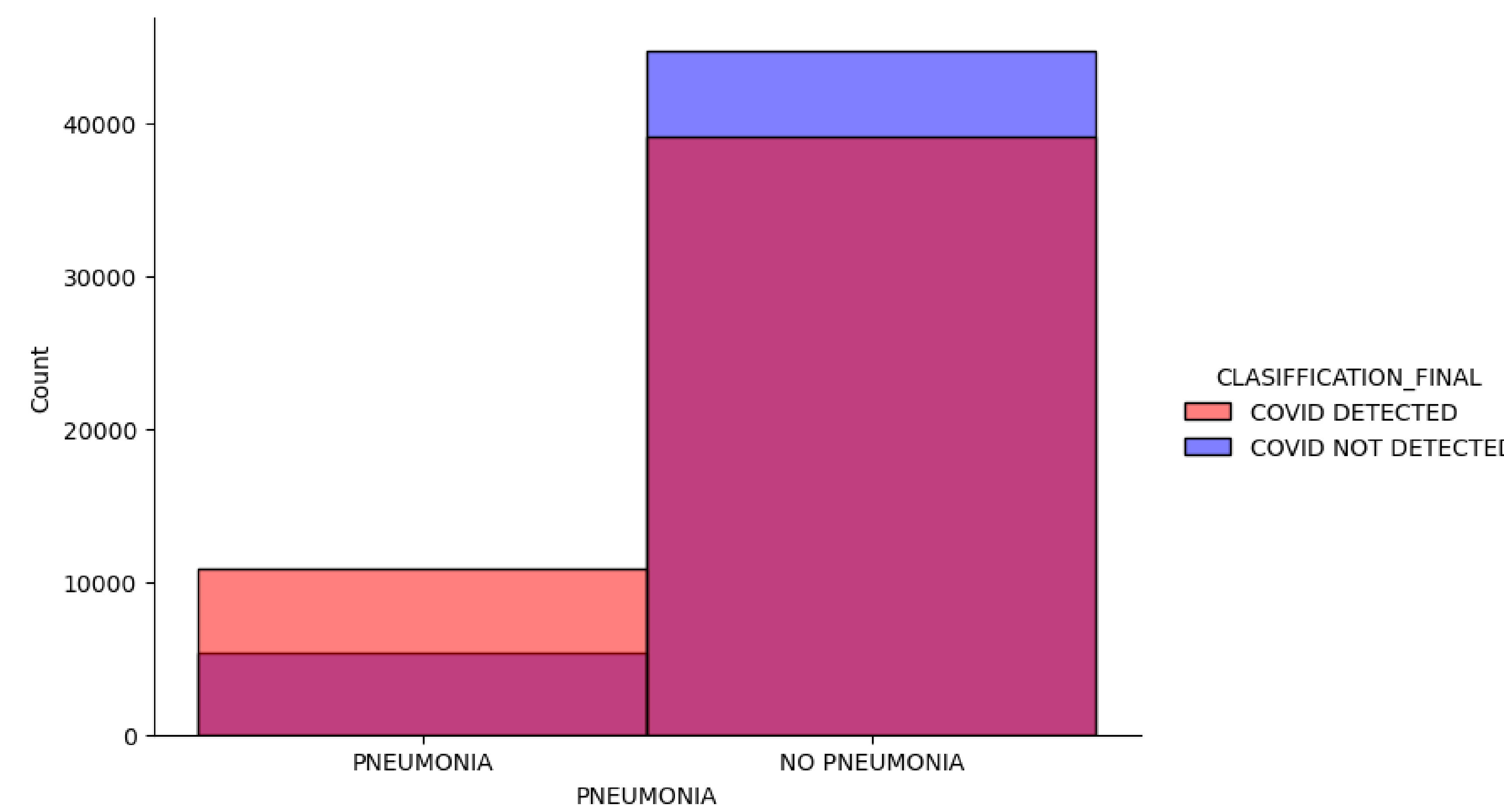
The main goal of this project is to build a machine learning model that, given a COVID-19 patient's current symptom, status, and medical history, will predict whether the patient is in high risk or not.



Patient Status	No. of Patients
COVID-19 Detected	50000
COVID-19 Not Detected	50000
COVID-19 Patient Rate	50 %
No COVID-19 Patient Rate	50 %

Features Responsible : Auto-ML

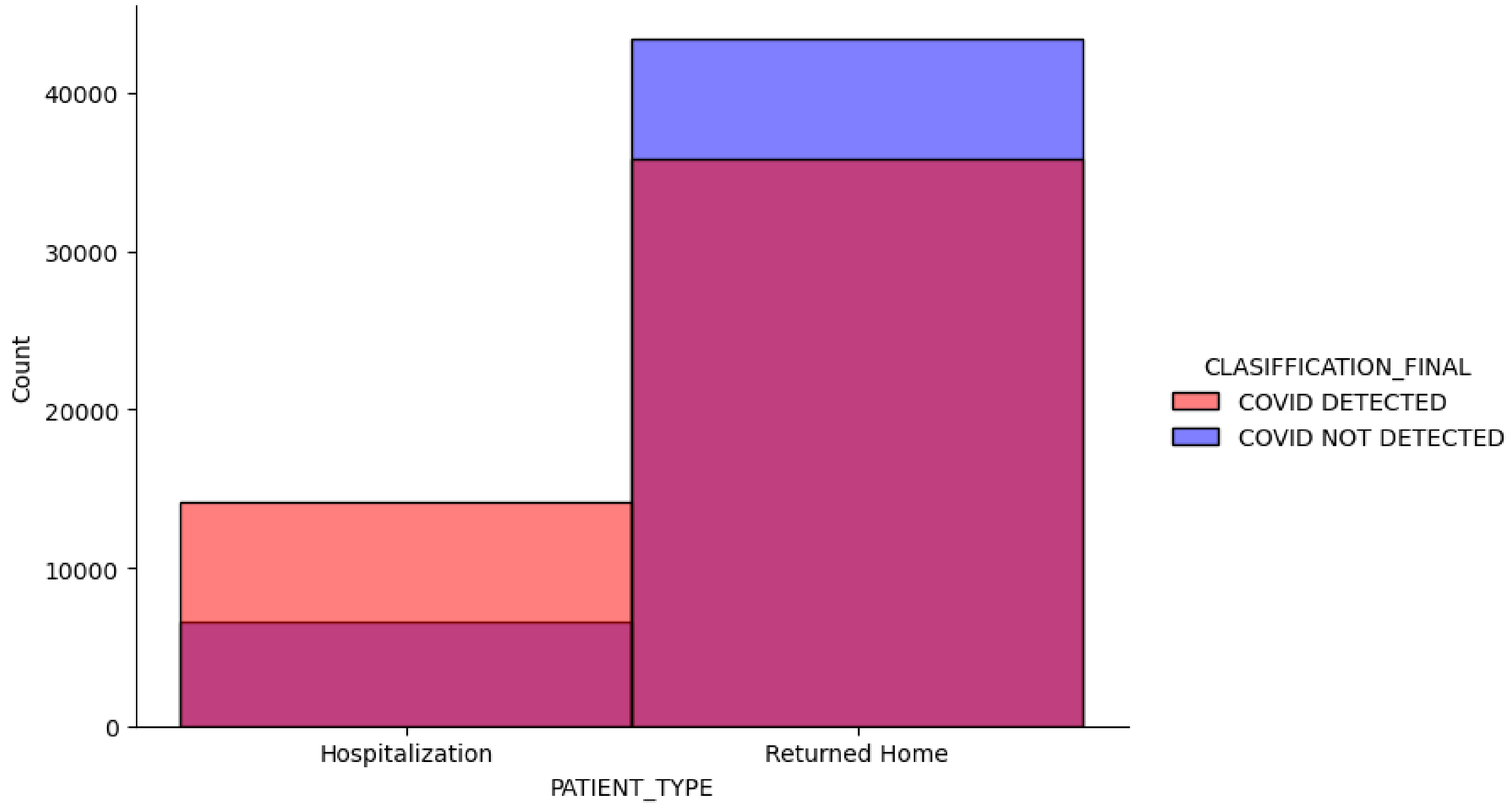


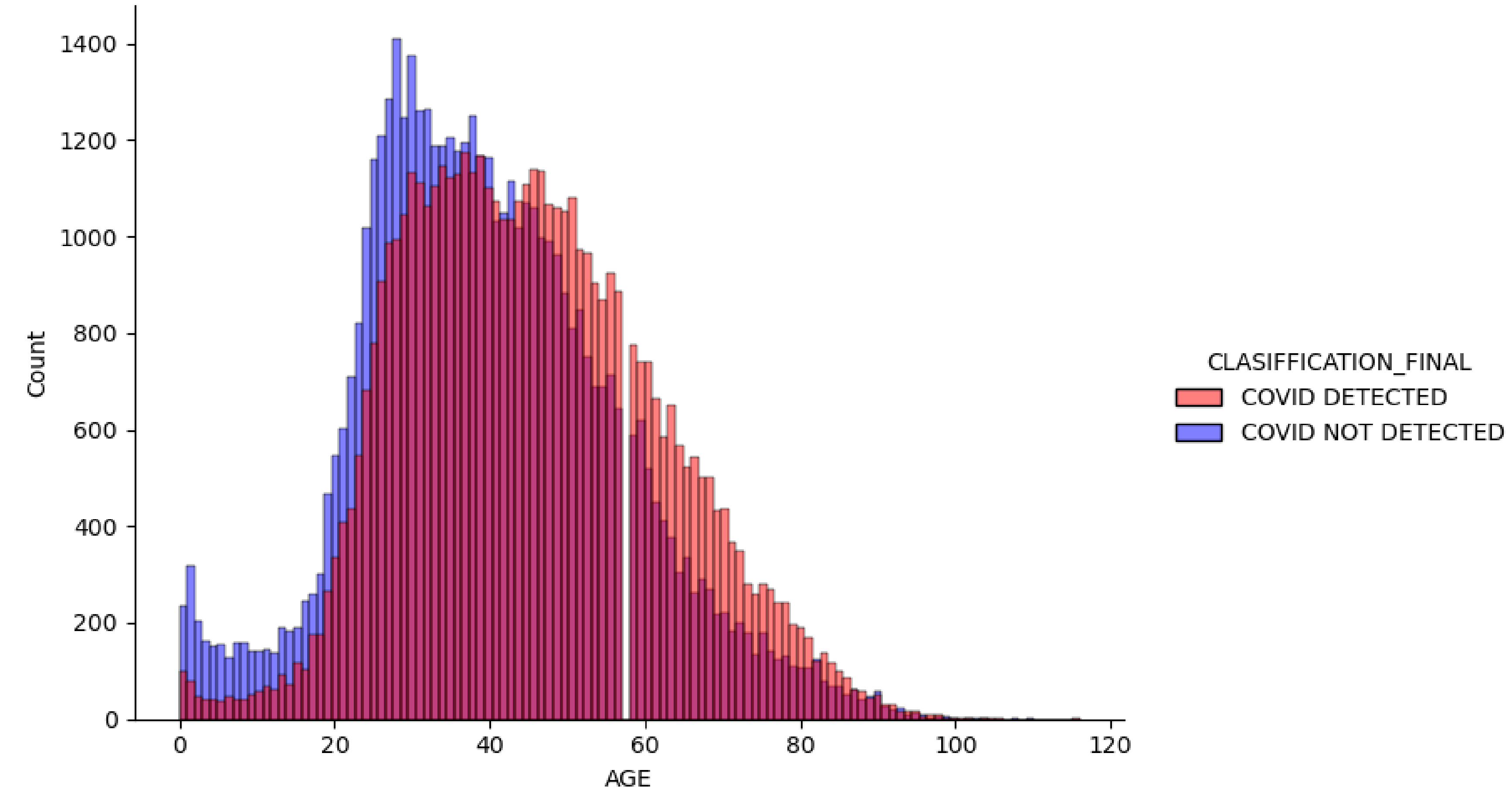


- **OBESITY** : whether the patient is obese or not
- Obesity can increase the risk of severe illness and complications from COVID-19. Obesity is defined as having a body mass index (BMI) of 30 or higher, and it can impair the immune system's ability to fight infections.

- **PNEUMONIA** : whether the patient already have air sacs inflammation or not
- Pneumonia is a medical condition characterized by inflammation and infection of the lungs, which can be caused by various pathogens such as bacteria, viruses, and fungi. Pneumonia is a common complication of COVID-19, especially in severe cases.

Features Responsible : Auto-ML





- **PATIENT_TYPE** : type of care the patient received in the unit
- Whether the patient has discharged or hospitalised

- **AGE** : Age of the patient
- Age is one of the factors that can impact a person's risk of being infected with COVID-19 and experiencing severe symptoms. Older adults, particularly those over 65 years of age, are at higher risk of developing severe COVID-19 symptoms compared to younger individuals.

Auto-ML Methodology Results

Algorithms	Test Accuracy (25 percentile)	Test Accuracy (50 percentile)	Test Accuracy (75 percentile)	Test Accuracy (90 percentile)
Decision Tree	61.4	61.3	59.34	59.4
Random Forest	61.39	61.37	60.13	59.76
XGBoost	61.64	62.13	61.77	61.53
MLP	62.2	62.73	63.09	63.29
RNN	58.78	58.99	59.19	59.35
Total Features	4	8	12	15
Avg. Accuracy	61.082	61.304	60.7	60.66

- **Based on our observation from the standard ML algorithms, 50 percentile has the best average accuracy**
- **MLP was the best performing algorithm with 63.29% accuracy in 90 percentile.**

Auto-ML Methodology Conclusion

- Hospitals and Clinics are leveraging the power of AI to aid the relatively expensive and competitive drug discovery process. AI solutions can successfully identify disease patterns in large datasets and help understand which drug compositions would be best suited for treating different diseases and the symptoms of various diseases. The dataset contains 1,048,575 records with 2 Categorical and 19 Numerical Features.
- For classification, models were created with algorithms using Auto-ML techniques like: Decision tree, Multi-layer Perceptron, Random forest , XGBoost and RNN. With these models, performance measurement values were obtained for feature sets of 4, 8, 12 and 15.
- The Auto-ML algorithms were able to predict whether the patient had COVID-19 or did not have COVID-19 based on their features with an average accuracy between 60% – 62% and helped to identify features that determine the symptoms of COVID-19 patients and predict whether the patient is in high risk or not to catch COVID-19. The major factors which affected the patients risk of catching COVID-19 are the features Obesity, Pneumonia, Patient_Type and Age.

- Overall, the application of Auto-ML in predicting patient is risk or not for COVID-19 and its dependent diseases can help in saving the patients life by prescribing medicine quickly and more efficiently.

Sensitivity Analysis

Actual Data Values

MEDICAL_UNIT	PNEUMONIA	HIPERTENSION	DIABETES	OBESITY	PATIENT STATUS
12.0 -90 %	1.0 -50 %	1.0 0 %	2.0 -50 %	2.0 -50 %	**0**

Patient Status -0 : COVID-19 Detected

Adjusted Data Values

MEDICAL_UNIT	PNEUMONIA	HIPERTENSION	DIABETES	OBESITY	PATIENT STATUS
1.2 ↓ (-10.8)	0.0 ↓ (-1.0)	1.0 ↑ (0.0)	1.0 ↓ (-1.0)	1.0 ↓ (-1.0)	**1**

Patient Status -1: COVID-19 Not Detected

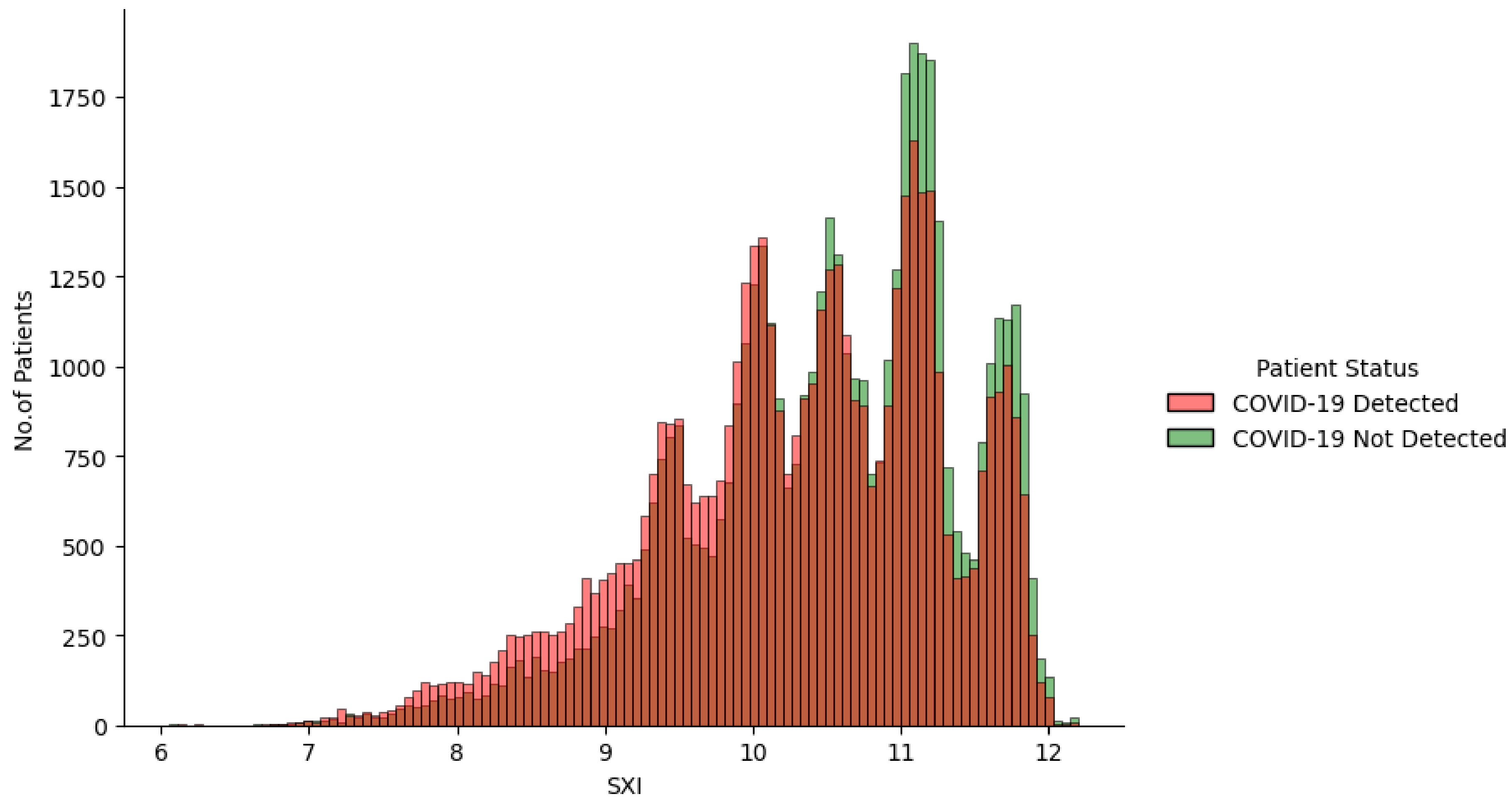
- The top 5 features :
  - MEDICAL_UNIT : Type of institution of the National Health System that provided the care.
  - PNEUMONIA: If the patient already have air sacs inflammation or not.
  - HIPERTENSION: If the patient has hypertension or not.
  - DIABETES: If the patient has diabetes or not.
  - OBESITY: If the patient is obese or not

For this patient we can see:

- If MEDICAL_UNIT with 90 % decrease from the actual .
- If PNEUMONIA 50 % decrease from the actual . (0 – No Pneumonia)
- If HIPERTENSION 0% change from the actual. (1 –No Hypertension)
- If DIABETES 50 % decrease from the actual . (1 –No Diabetics)
- If DIABETES 50 % decrease from the actual. (1 –No Obesity)

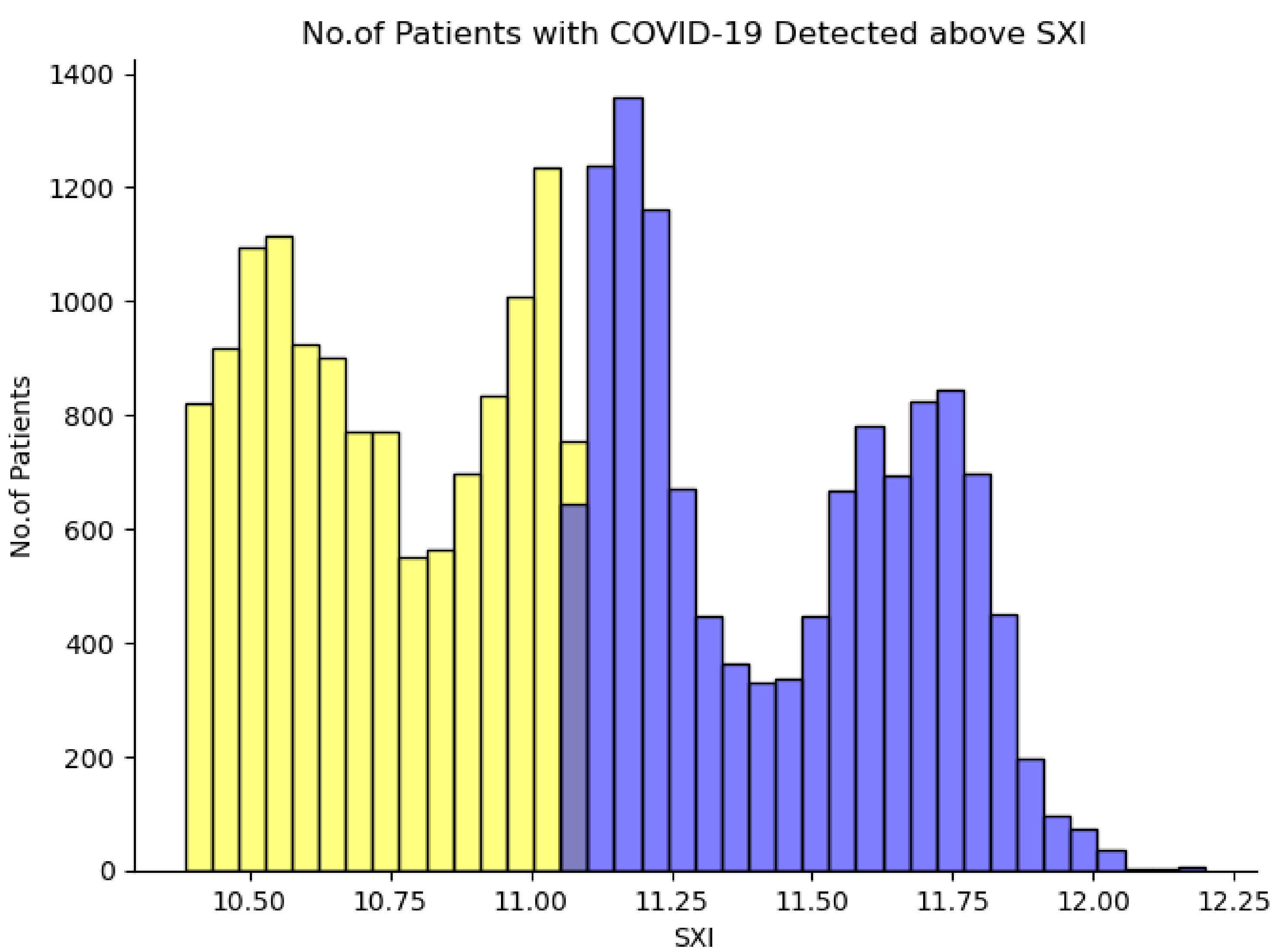
Which will lead this patient to be COVID-19 test to be not detected.

SXI Method



- Pink – COVID-19 Detected patients
- Brown– Mix of Detected and Undetected COVID-19 patients
- Green - COVID-19 Not Detected patients

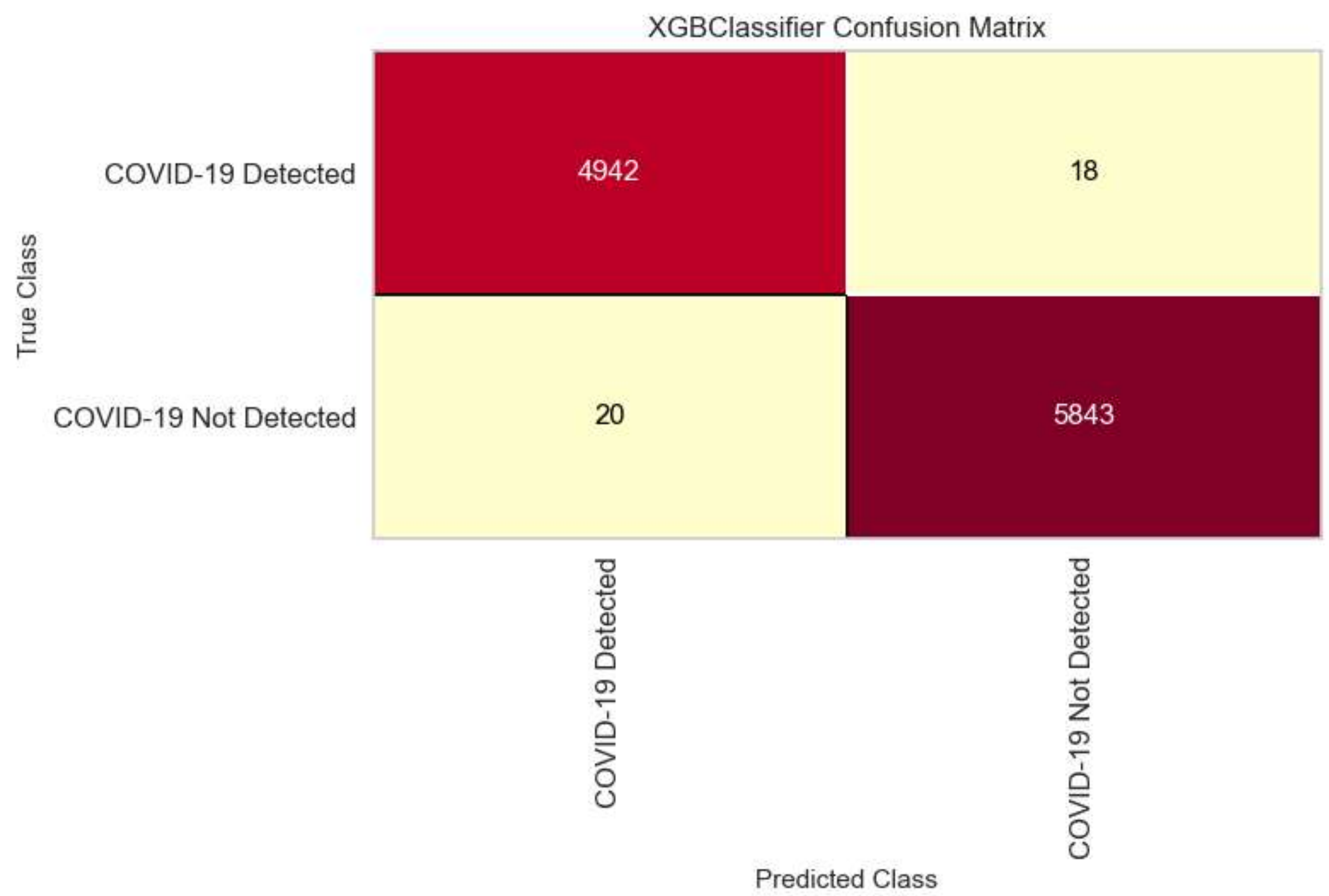
SXI	10.38
Top SXI	12.19
Minimum SXI	6.05
No. of COVID-19 patients above SXI	25312
No. of COVID-19 patients below SXI	24688
No. of COVID-19 Not Detected patients above SXI	29427
No. of COVID-19 Not Detected patients below SXI	20573
SXI Model Accuracy	99%



COVID-19 Patients above SXI

- Patients who have above SXI values and who are being detected with COVID-19 are the important patients.
- In order to decrease the COVID-19 detection rate these patients are one to focus on, because in future they have more chance to be get cured.
- The average SXI on “COVID-19 Patients above SXI” is 11.08. Here we further categorize ,so patient’s SXI above 11.08 are to be mostly given importance in future.

SXI Method Accuracy is 99%



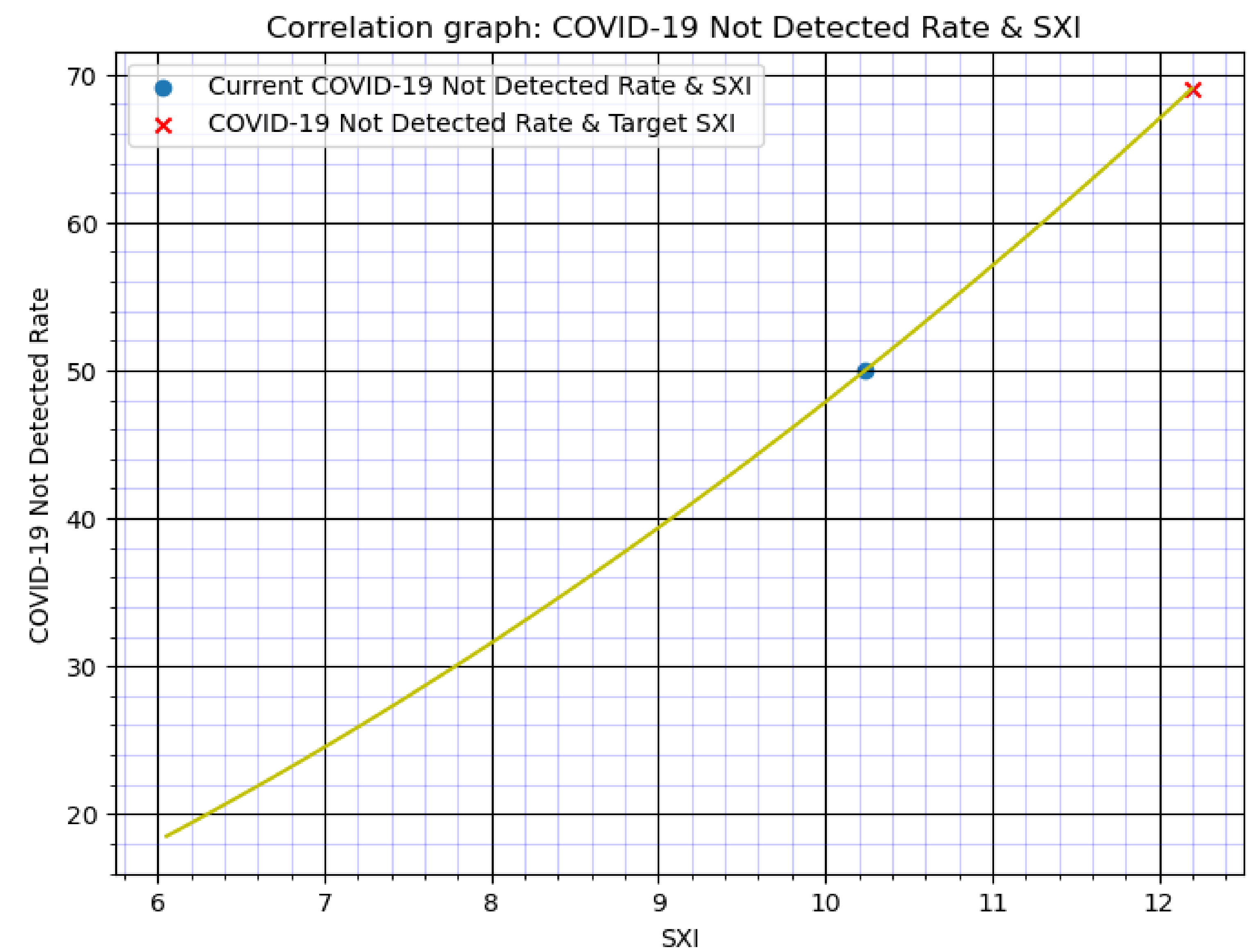
- Actual and Predicted was COVID-19 Not Detected Patients (TP) : 5843
- Actual and Predicted was COVID-19 Detected Patients (TN) : 4942
- Actual COVID-19 Not Detected Patients and Predicted COVID-19 Detected Patients (FN): 20
- Actual COVID-19 Detected Patients Predicted COVID-19 Not Detected Patients (FP) : 18

Train Records	Test Records	Actual Train count for COVID-19 Not Detected Patients	Actual train count for COVID-19 Detected Patients	Actual test count for COVID-19 Not Detected Patients	Actual test count for COVID-19 Detected Patients	Predicted test count for COVID-19 Not Detected Patients	Predicted test count for COVID-19 Detected Patients	Precision rate	Recall rate	Model Accuracy
40514	10823	23564	19728	5863	4960	5861	4962	0.99	0.99	**0.99**

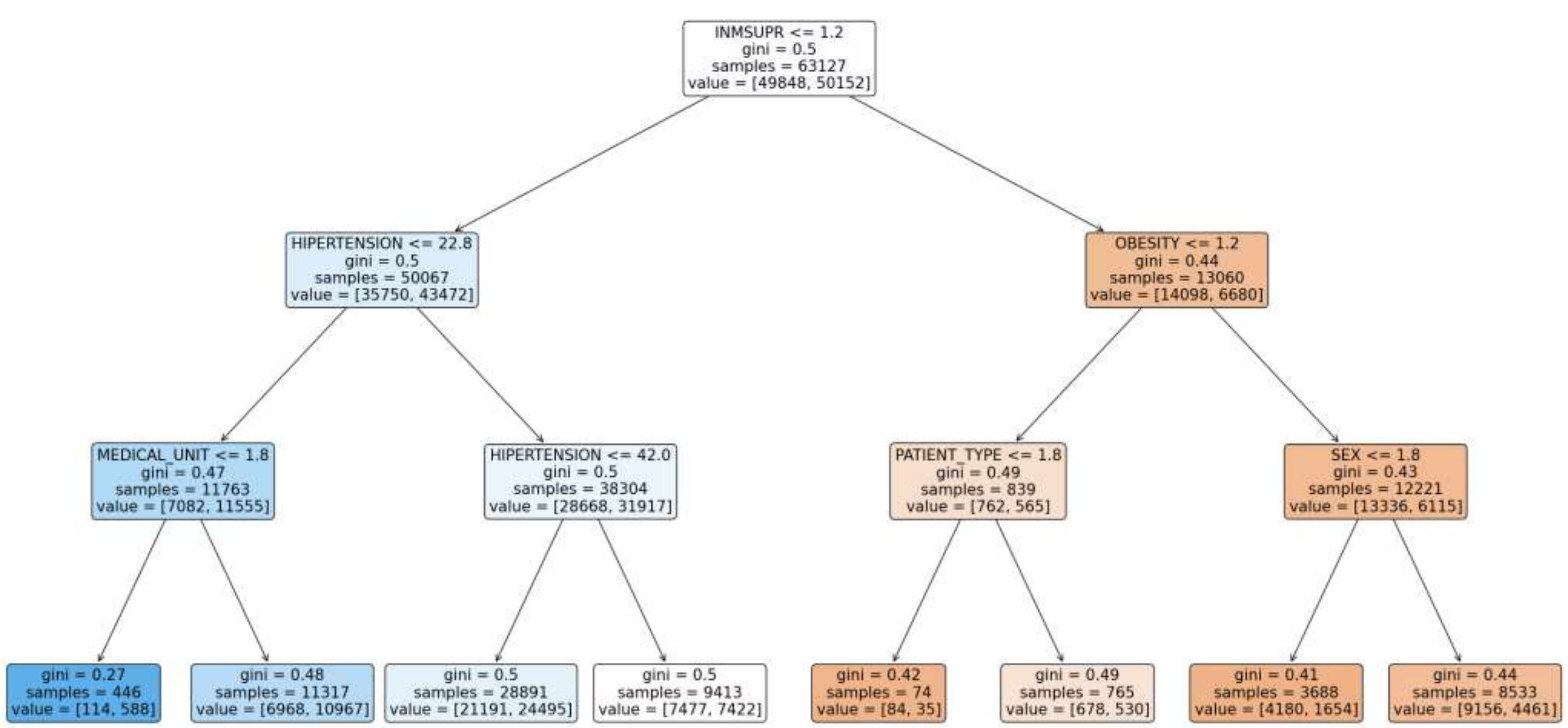
- **Precision rate**: Precision is defined as the ratio of actual COVID-19 Not Detected Patients customers (True Positive) to a total number of predicted COVID-19 Not Detected Patients customers. ***TP / (TP + FP)***
- **Recall rate**: The recall is calculated as the ratio between the Actual numbers of COVID-19 Not Detected Patients to the total number of wrongly predicted COVID-19 Not Detected Patients as COVID-19 Detected Patients plus actual number of COVID-19 Not Detected Patients. ***TP / (TP + FN)***
- **Model Accuracy**: It is the fraction of predictions where the model got right. ***(TP + TN) / (TP + FP + TN + FN)***

SXI Method - Conclusion

SXI	10.38
Target SXI	12.2
COVID-19 Not Detected rate	50%
Target COVID-19 Not Detected rate	69%



The correlation between SXI and COVID-19 Not Detected Rate is ***0.99***



Tree - Interpretation

- If the patient is **not immunosuppressed.**
- **If the patient has less Hypertension (<=** 22.8 %**)**
- **Type of institution of the National Health System that provided the care =** Type 1
- 446 samples leading to higher COVID-19 Not Detected rate.