COVID-19 Patient Survival Prediction

USING LOGISTIC REGRESSION, RANDOM FOREST, SGD CLASSIFIER, AND XGBOOST

AZADEH ANSARI

JULY 2, 2024

Introduction

- Overview of the COVID-19 pandemic
- Significant challenges to healthcare systems worldwide posed by COVID-19

- Importance of predicting patient outcomes
- Predicting patient outcomes based on their symptoms and medical history can be crucial

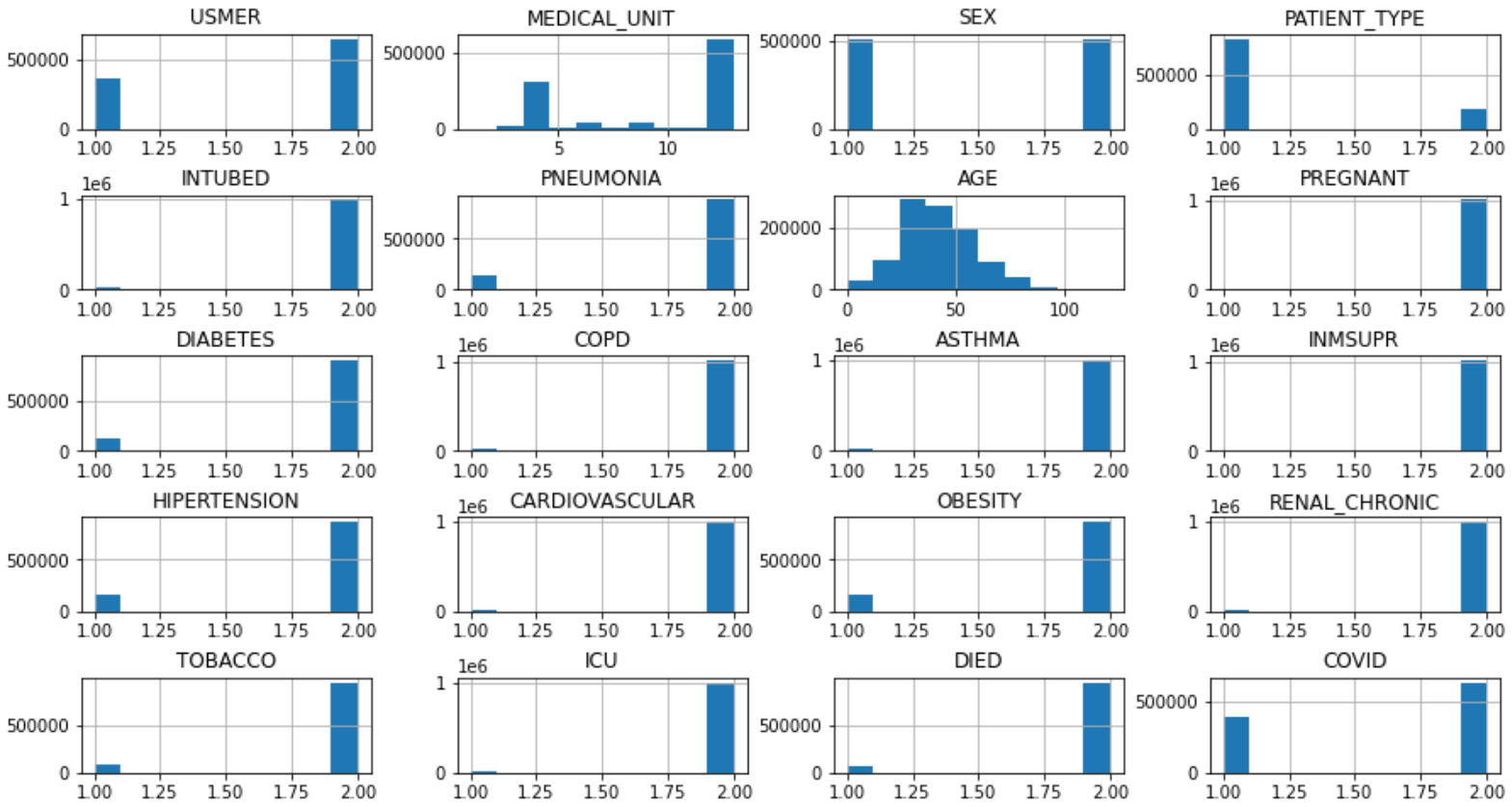
Data Description

- Description of the dataset
  - COVID-19 patients' medical history
  - Patient symptoms.

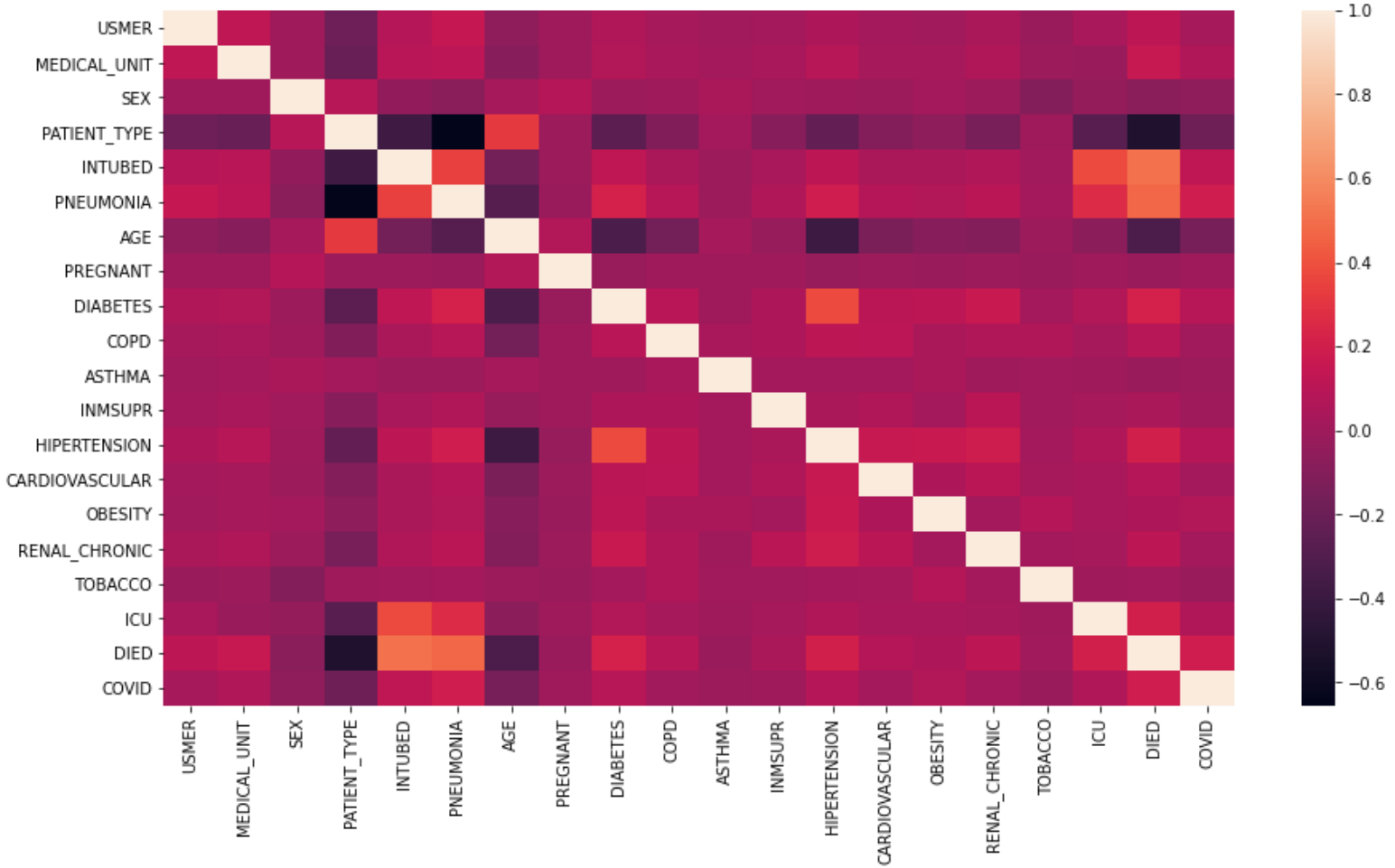
This project is from Kaggle

- Features: Age, Gender, Comorbidities, Symptoms, medical history (Diabetes, Athma …), etc.

Data Description



Data Description



Methodology

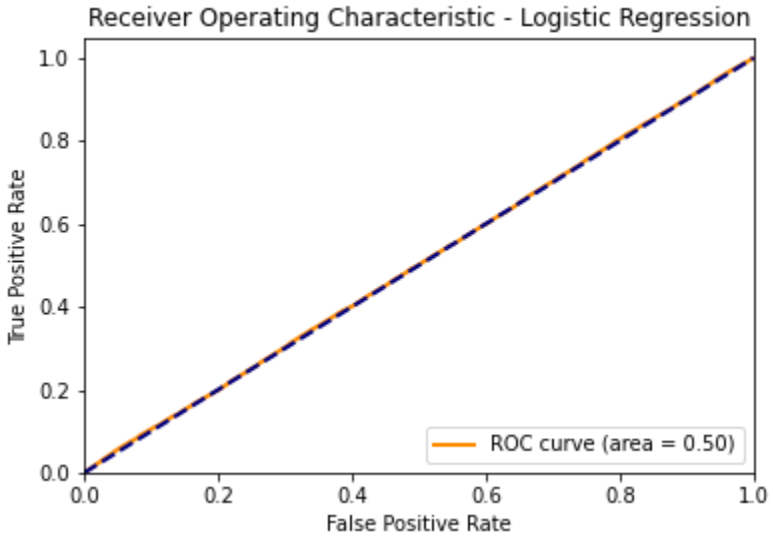
- Data Preprocessing:
  - Cleaning,
  - Feature Engineering,
  - Normalization
- Models Used:
  - Logistic Regression,
  - Random Forest,
  - SGD Classifier,
  - XGBoost

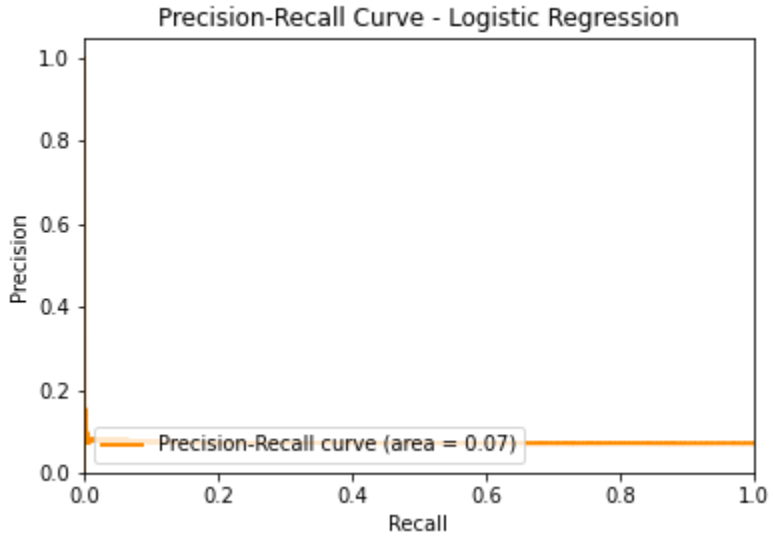
Evaluation Metrics

- Accuracy
- ROC Curve
- Precision-Recall Curve

Logistic Regression Results

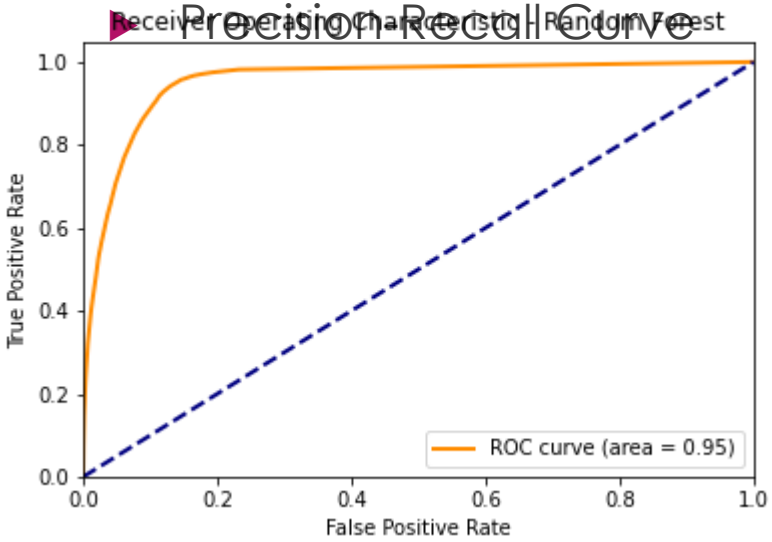
- Accuracy: [0.95]
- ROC Curve
- Precision-Recall Curve

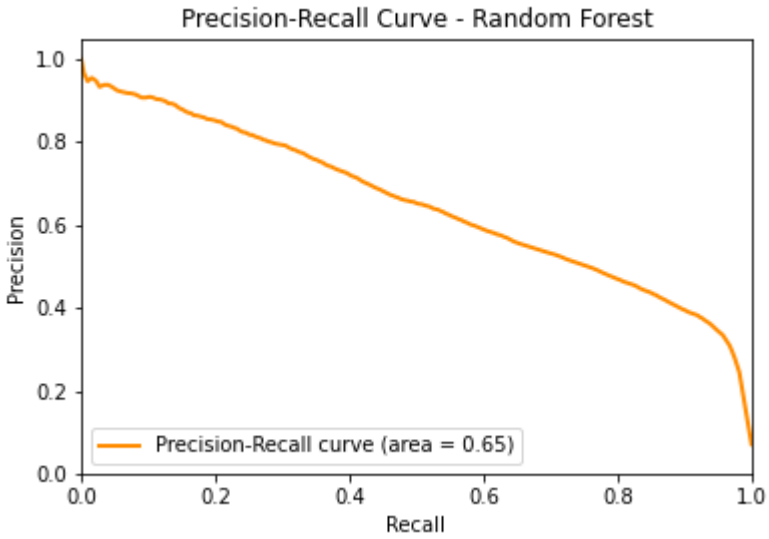




Random Forest Results

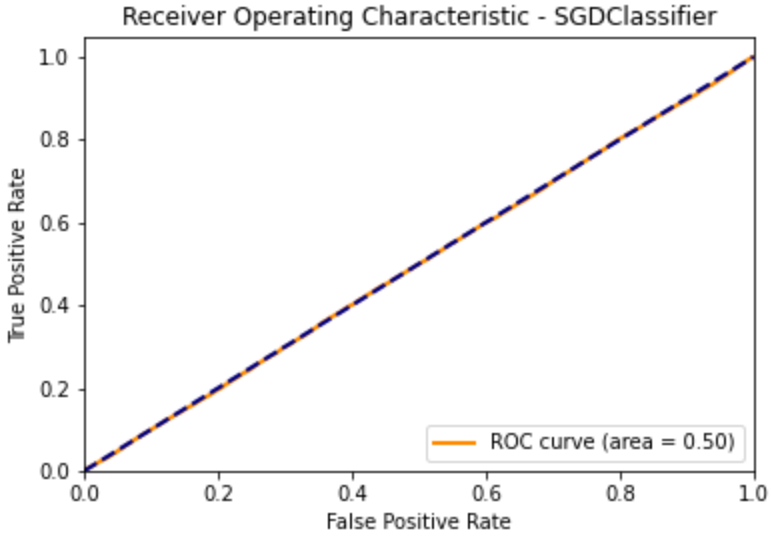
- Accuracy: [0.95]
- ROC Curve
- Precision-Recall Curve

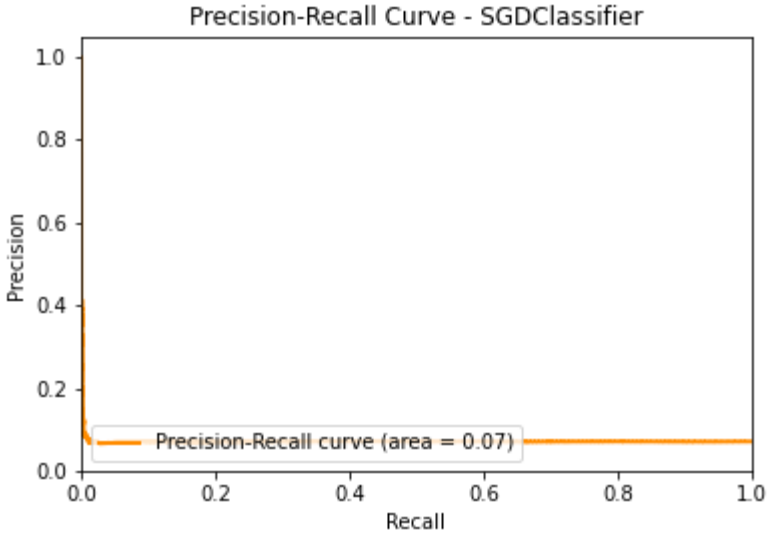




SGD Classifier Results

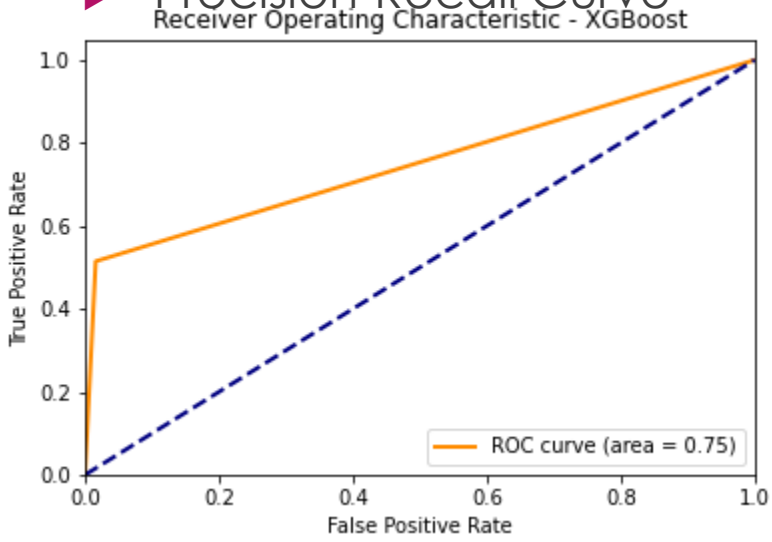
- Accuracy: [0.073]
- ROC Curve
- Precision-Recall Curve

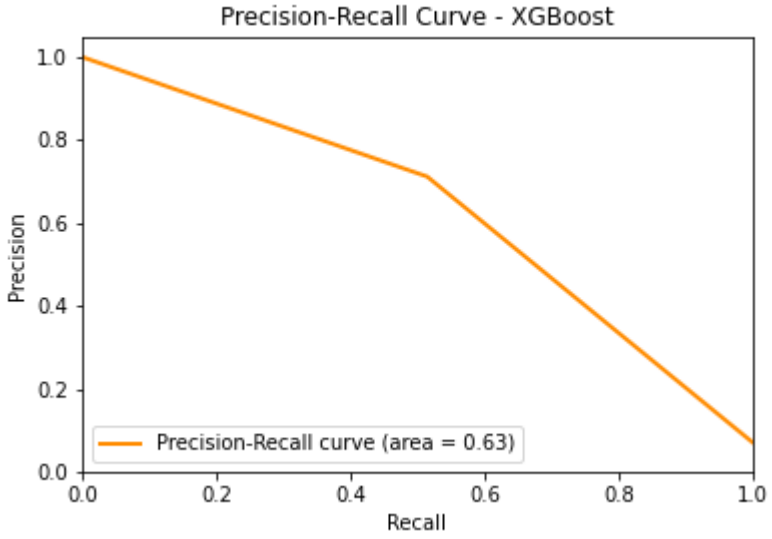




XGBoost Results

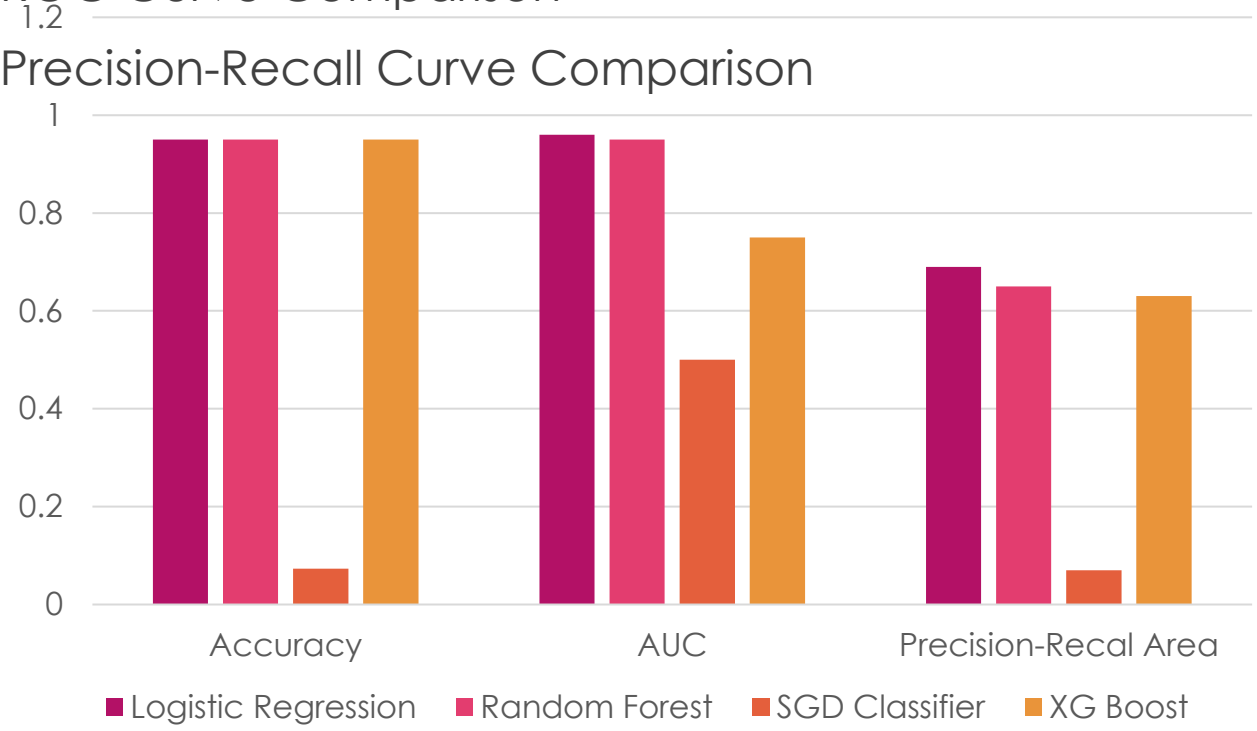
- Accuracy: [0.95]
- ROC Curve
- Precision-Recall Curve





Comparison of Models

- Accuracy Comparison
- ROC Curve Comparison
- Precision-Recall Curve Comparison



Conclusion

- Summary of findings

SGD Classifier is as good as a random model. SVM could perform much better, but the run time is large due to the size of the data.

XG Boost performs well in terms of accuracy but not so well in terms of AUC and area under precision-recall curve.

- Best performing model

Logistic Regression and Random Forest perform well in terms of the metrics we looked; however Logistic Regression performs slightly better.

Future Work

- Potential improvements

Tuning hyper parameters and model optimization

- Exploring additional models

- Incorporating more data