

Predicting ICU Visits From Covid-19 Patients

By Shahzab

Overview of Dataset

- <https://www.kaggle.com/datasets/meirnizri/covid19-dataset>
- 21 variables which includes age, pneumonia, diabetes, and cardiovascular
- 1,048,575 individuals
- Data originates from the Government of Mexico
- Reduced to 1,019,666 individuals from data cleaning
- Whether they end up in ICU or not is the indicator variable

Overview of Variables

- Sex: 1 for female and 2 for male
- Age: of the patient
- Classification: covid test findings. Values 1-3 mean that the patient was diagnosed with covid in different Degrees. 4 or higher means that the patient is not a carrier of covid or that the test is inconclusive
- Patient type: type of care the patient received in the unit. 1 for returned home and 2 for hospitalization
- Pneumonia: whether the patient already have air sacs inflammation 1 or not 2
- Pregnancy: whether the patient is pregnant 1 or not 2 (2 for males)
- Diabetes: whether the patient has diabetes 1 or not 2

Overview of Variables

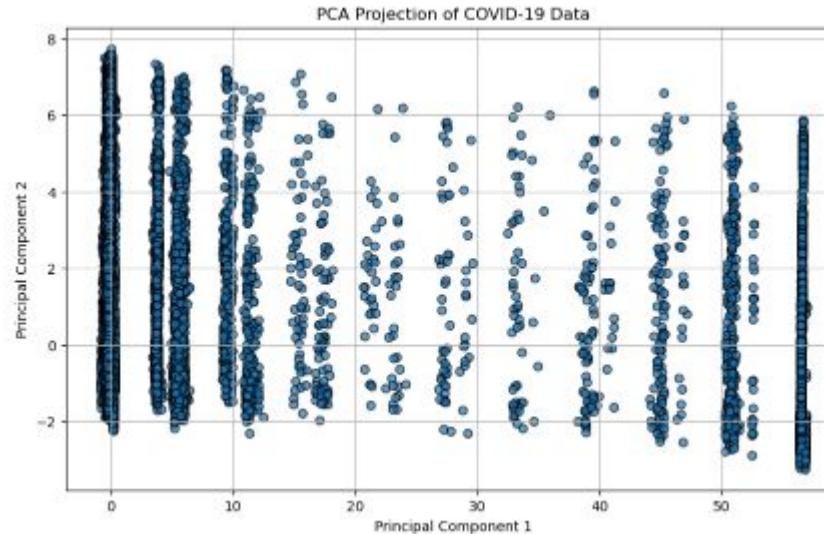
- Copd: Indicates whether the patient has Chronic obstructive pulmonary disease 1 or not 2
- Asthma: whether the patient has asthma 1 or not 2
- Inmsupr: whether the patient is immunosuppressed 1 or not 2
- Hypertension: whether the patient has hypertension 1 or not 2
- Cardiovascular: whether the patient has heart or blood vessels related disease 1 or not 2
- Renal chronic: whether the patient has chronic renal disease 1 or not 2
- Other disease: whether the patient has other disease 1 or not 2

Overview of Variables

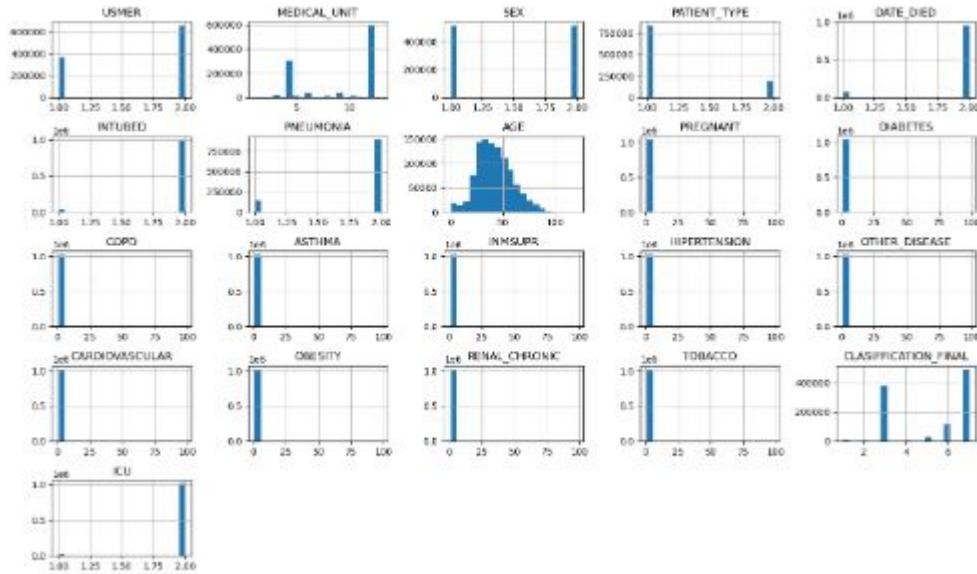
- Obesity: whether the patient is obese 1 or not 2
- Tobacco: whether the patient is a tobacco user 1 or not 2
- Usmr: Indicates whether the patient treated medical units of the first, second or third level
- Medical unit: type of institution of the National Health System that provided the care
- Intubated: whether the patient was connected to the ventilator 1 or not 2 (2 for patients who returned home)
- ICU: Indicates whether the patient had been admitted to an Intensive Care Unit 1 or not 2 (2 for patients who returned home)
- Date died: If the patient died indicate the date of death, and 9999-99-99 otherwise (1 for died and 2 for survived)

Visualization of the Dataset

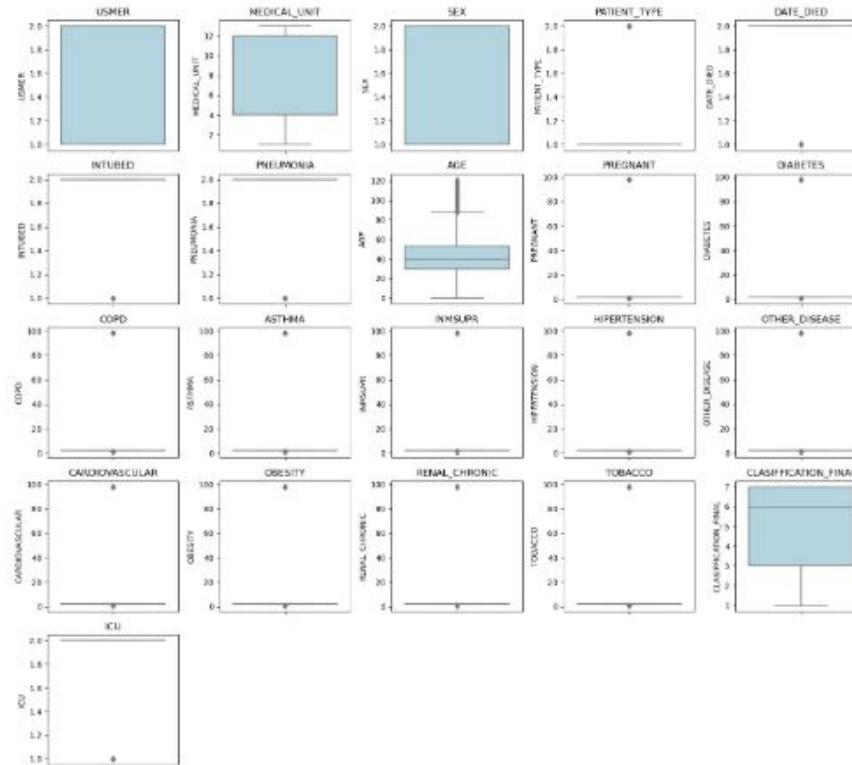
- Employed Principal Component Analysis (PCA) for dimensionality reduction
- Further analysis using machine learning models to uncover deeper insights



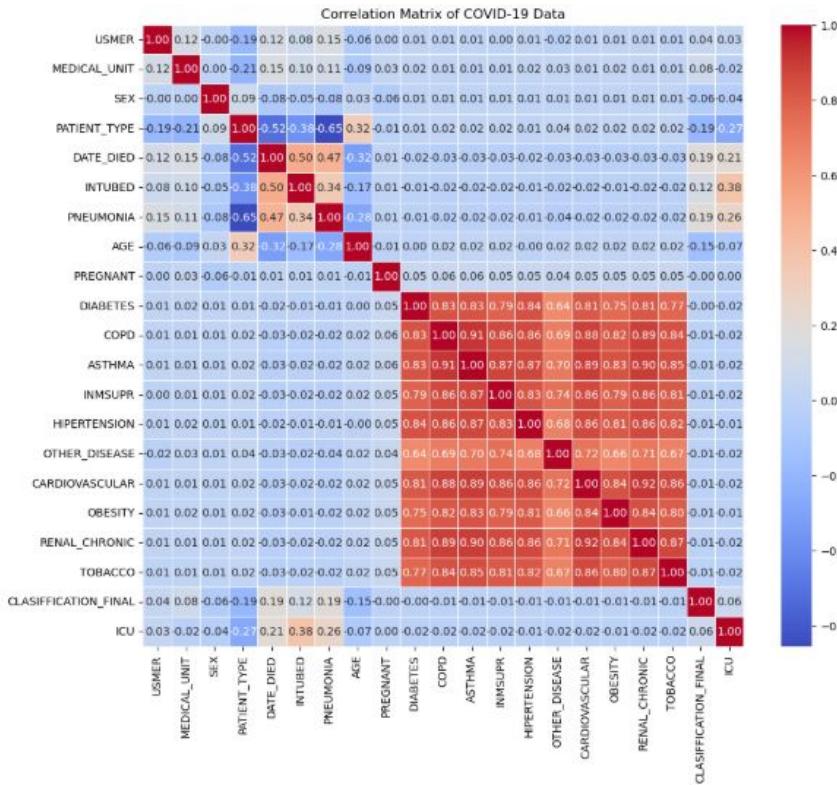
Visualization of the Dataset



Visualization of the Dataset

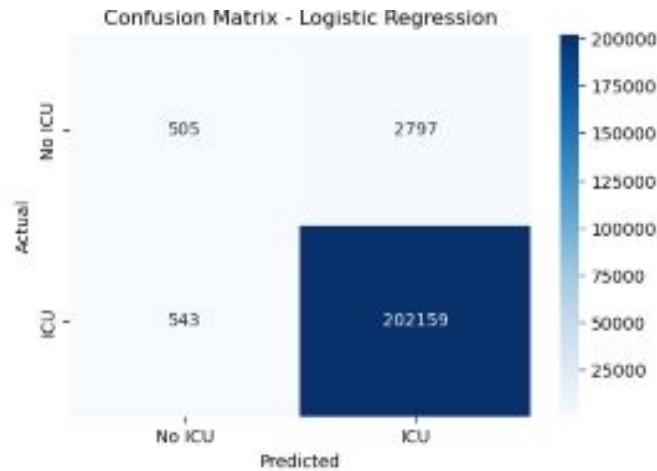


Visualization of The Dataset



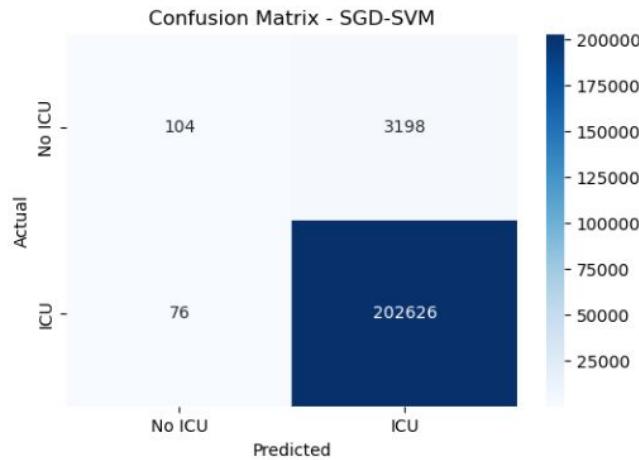
Logistic Regression Results

- 0.5751 accuracy
- 0.9973 sensitivity
- 0.1529 specificity



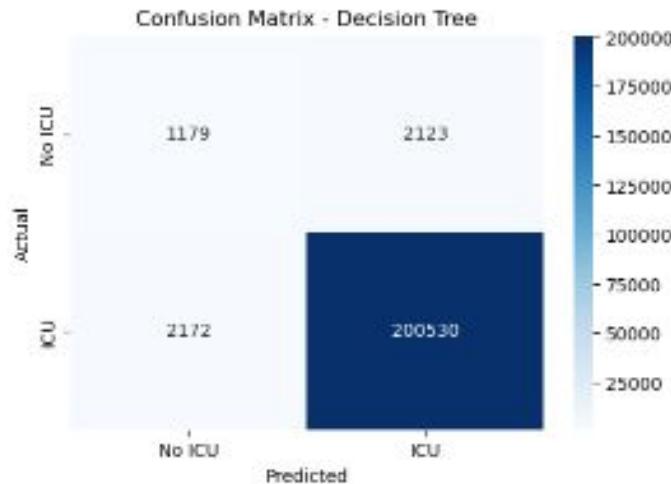
Support Vector Machine Results

- 0.5156 accuracy
- 0.9996 sensitivity
- 0.0315 specificity



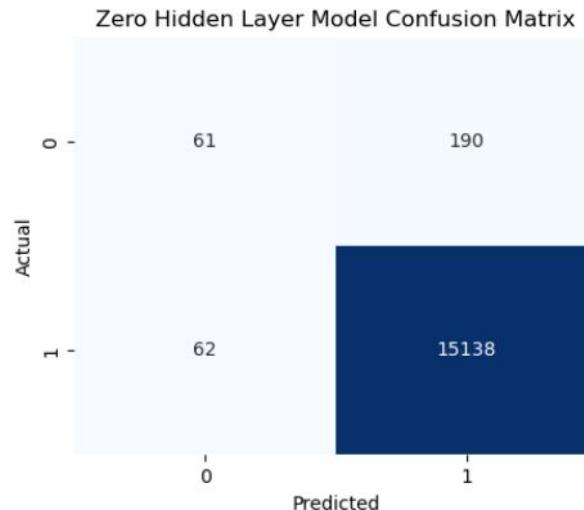
Decision Tree Results

- 0.6732 accuracy
- 0.9893 sensitivity
- 0.3571 specificity



Neural Net Results

- 0.6185 accuracy
- 0.9959 sensitivity
- 0.243 specificity



Overall Conclusions

	Accuracy	Sensitivity	Specificity
Logistic Regression	0.5751	0.9973	0.1529
SVM	0.5156	0.9996	0.0315
Decision Tree	0.6732	0.9893	0.3571
Neural Network	0.6185	0.9959	0.243

Questions?