Predicting Health

Mining patient data to gain invaluable insights

Group 2

Anthony Le

Overview

For my data mining project, I built a predictive model using a comprehensive dataset of patient data related to COVID-19 published by the Mexican government. The goal of this project was to leverage machine learning techniques to address key questions that could significantly impact patient care and resource management in healthcare settings.

Interesting Questions

- Can I predict whether a patient will need intubation?
- Can I predict whether a patient will need to be admitted to the ICU?
- Can I predict the mortality outcome of a patient?

Tools Used

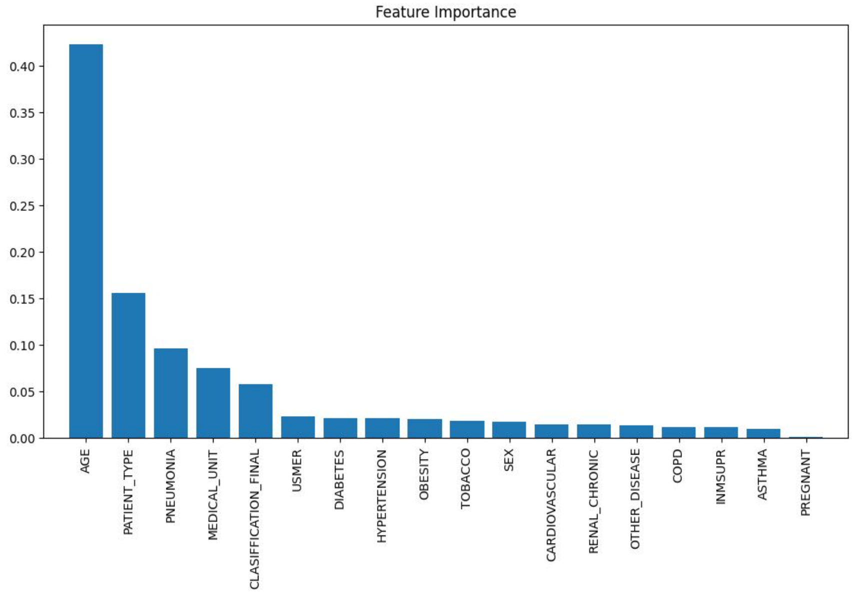
- Python
    - Pandas
    - NumPy
    - Matplotlib
    - Seaborn
    - Sklearn
- Git/GitHub
- Trello

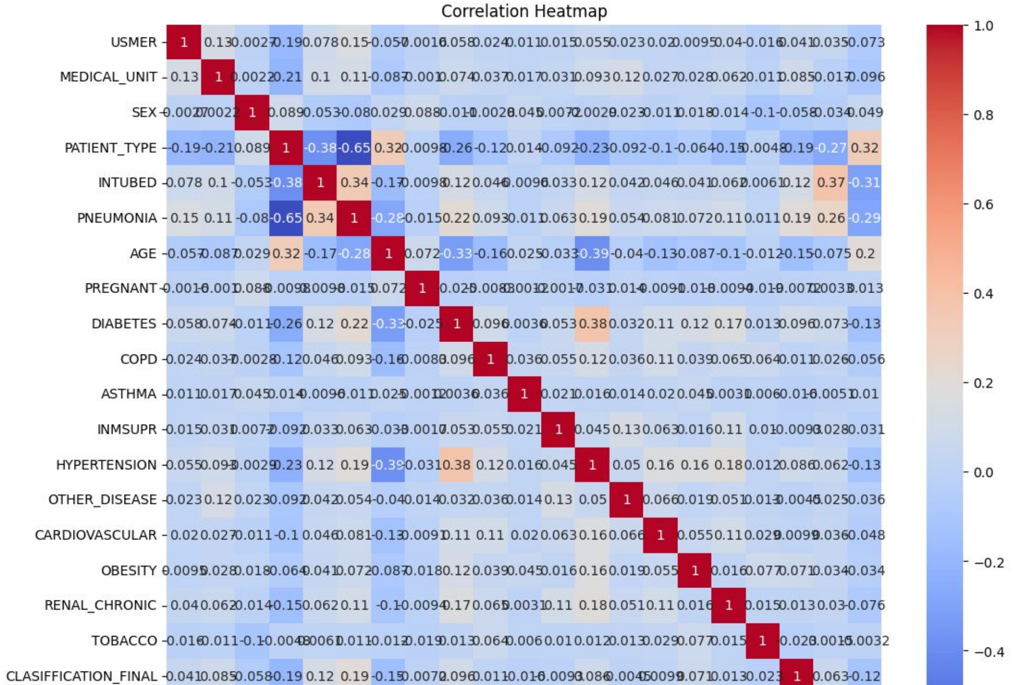
Data Preparation

- Data Loading and Initial Exploration
    - Inspect dataset and identify features and distribution of values
    - Note presence of syntactical errors and missing values
- Data Cleaning
    - Fix misspelled feature names
    - Address missing values
    - Post-cleaning verification
    - Prepare date for use of training and evaluating predictive model
- Model Development
    - Choosing an algorithm
    - Prepare data
    - Train model
    - Evaluate model

```
Dimensions of COVID Dataset: (1048575, 21)
Attributes of COVID Dataset: ['USMER' 'MEDICAL_UNIT' 'SEX' 'PATIENT_TYPE' 'DATE_DIED' 'INTUBED'
 'PNEUMONIA' 'AGE' 'PREGNANT' 'DIABETES' 'COPD' 'ASTHMA' 'INMSUPR'
 'HIPERTENSION' 'OTHER_DISEASE' 'CARDIOVASCULAR' 'OBESITY' 'RENAL_CHRONIC'
 'TOBACCO' 'CLASIFFICATION_FINAL' 'ICU']
```

Classification





Insights

- Model had good overall performance
    - Still had its own shortcomings
- Strengths of model
    - Feature importance
    - Robust predictions
- Limitations
    - Handling of rare cases

Intubation Prediction

Class	Precision	Recall	F1-Score	Support
1.0	0.27	0.11	0.15	6,582
2.0	0.97	0.99	0.98	197,352
Accuracy				0.96
Macro Avg	0.62	0.55	0.57	203,934
Weighted Avg	0.95	0.96	0.95	203,934

ICU Admission Prediction

Class	Precision	Recall	F1-Score	Support
1.0	0.25	0.09	0.13	3,252
2.0	0.99	1.00	0.99	200,682
Accuracy				0.98
Macro Avg	0.62	0.54	0.56	203,934
Weighted Avg	0.97	0.98	0.98	203,934

Mortality Prediction

Class	Precision	Recall	F1-Score	Support
1.0	0.97	0.99	0.98	198155
2.0	0.23	0.08	0.11	5779

Application

- Predictive modeling can be helpful in other aspects of healthcare
    - Prioritization management
    - Resource allocation
    - Planning and strategy
    - Policy and protocol development
- Strong use case in other industries outside of healthcare
    - Finance
    - Retail
    - Manufacturing
    - Transporation