Covid-19 Patient Risk Prediction System

Contributors: Suriya, Swanand Kulkarni, Kasyap Velampalli

Abstract

Personalized medicine plays an important role in treatment optimization for COVID-19 patient management. Early treatment in patients at high risk of severe complications is vital to prevent death and ventilator use. Predicting COVID-19 clinical outcomes using machine learning may provide a fast and data-driven solution for optimizing patient care by estimating the need for early treatment. To overcome this existing limitation, we generate a robust machine learning model to predict patient-specific risk of death or ventilator use in COVID-19 positive patients using features available at the time of diagnosis. It establish the value of solution across patient demographics, including gender and race for COVID-19 patient's symptoms, status, and medical history. Therefore, the primary aim of this study is to develop and validate a population-based prediction model, using a large, rich dataset and a selective, clinically informed approach, which dynamically estimates the COVID-19 mortality risk in confirmed diagnoses.

1. Problem Statement

As of March 2022, different strains of the SARS-CoV-2 virus have caused five global surges in the number of cases and deaths from COVID-19. It is critical to potentiate the health system struggling with managing the resources during disease surge. Early identification of risk factors and clinical outcomes might help in identifying critically ill patients, providing appropriate treatment, and preventing mortality.

2. Market / Customer/ Business Need Assessment

Timely risk assessment of COVID-19 patients can significantly improve the quality of patient care and in-hospital resource allocation. Recent studies have leveraged machine learning to derive and validate risk prediction algorithms via electronic health records (EHRs) in order to estimate the risk of COVID-19-related adverse events, such as ICU readmission and mortality. According to COVID-19 inequity and disparities studies, different gender groups have different levels of risk associated with COVID-19 infection and mortality, with males having a higher COVID-19 death rate across all age groups.

3. Target Specification and characterization

Developing a Covid-19 Patient Risk Prediction System entails the creation of a predictive model designed to evaluate the risk levels of individuals based on pertinent features and historical data. The system will utilize diverse factors such as age, gender, comorbidities, symptoms, vital signs, lab results, and travel and contact history to generate a risk assessment for Covid-19 patients. The primary target variable is the patient's risk level, categorized, for instance, as low, medium, or high risk. Data collection involves gathering comprehensive historical information from Covid-19 cases, ensuring data quality, and addressing privacy and ethical considerations.

Why Covid-19 Risk Prediction System Needed?

- Transform the current Covid-19 risk assessment process into a faster and more accurate system.
- Mitigate patient frustration and reduce mortality rates by minimizing delays in the prognosis process.
- Utilize a predefined dataset comprising Covid-19 patient records and distinguish between confirmed cases and non-infected individuals for predictive analysis.
- Understand the expectations and concerns of individuals undergoing Covid-19 risk assessment.
- Evaluate the strengths and weaknesses of current Covid-19 risk assessment methodologies.
- Investigate methods for accurately identifying and providing early-stage treatment for Covid-19 cases.
- Aim to provide patients with Covid-19 risk results within minutes, accompanied by clear guidance on the next steps if a high risk is detected.

4. External Search (information sources/references)

I use the Covid-19 dataset for this project Dataset can be found here:

Kaggle Link: https://www.kaggle.com/datasets/meirnizri/covid19-dataset

About the Dataset

Contains a vast number of anonymized patient-related information including pre-conditions. The raw dataset consists of 21 different features and 1,048,576 unique patients. In the Boolean features, 1 means "yes" and 2 means "no". values as 97 and 99 are missing data.

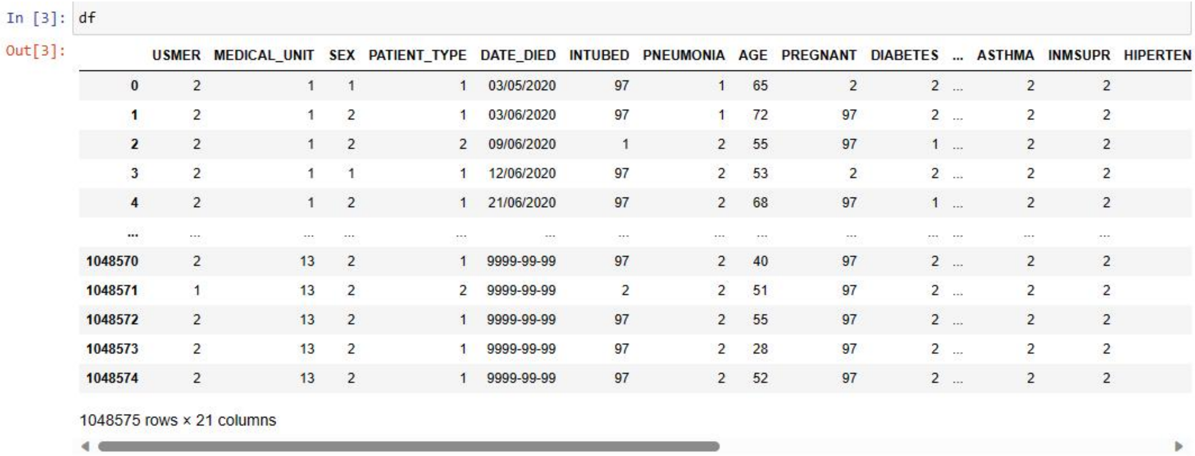
Dataset Origin: Datos Abiertos de México - Información referente a casos COVID-19 en México

Dataset Description

The dataset was provided by the Mexican government (link). This dataset contains an enormous number of anonymized patient-related information including pre-conditions. The raw dataset consists of 21 unique features and 1,048,576 unique patients. **In the Boolean features, 1 means "yes" and 2 means "no". values as 97 and 99 are missing data**.

- 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.
- pregnancy: whether the patient is pregnant or not.
- 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.
- intubed: whether the patient was connected to the ventilator.
- icu: Indicates whether the patient had been admitted to an Intensive Care Unit.

Let’s view our dataset



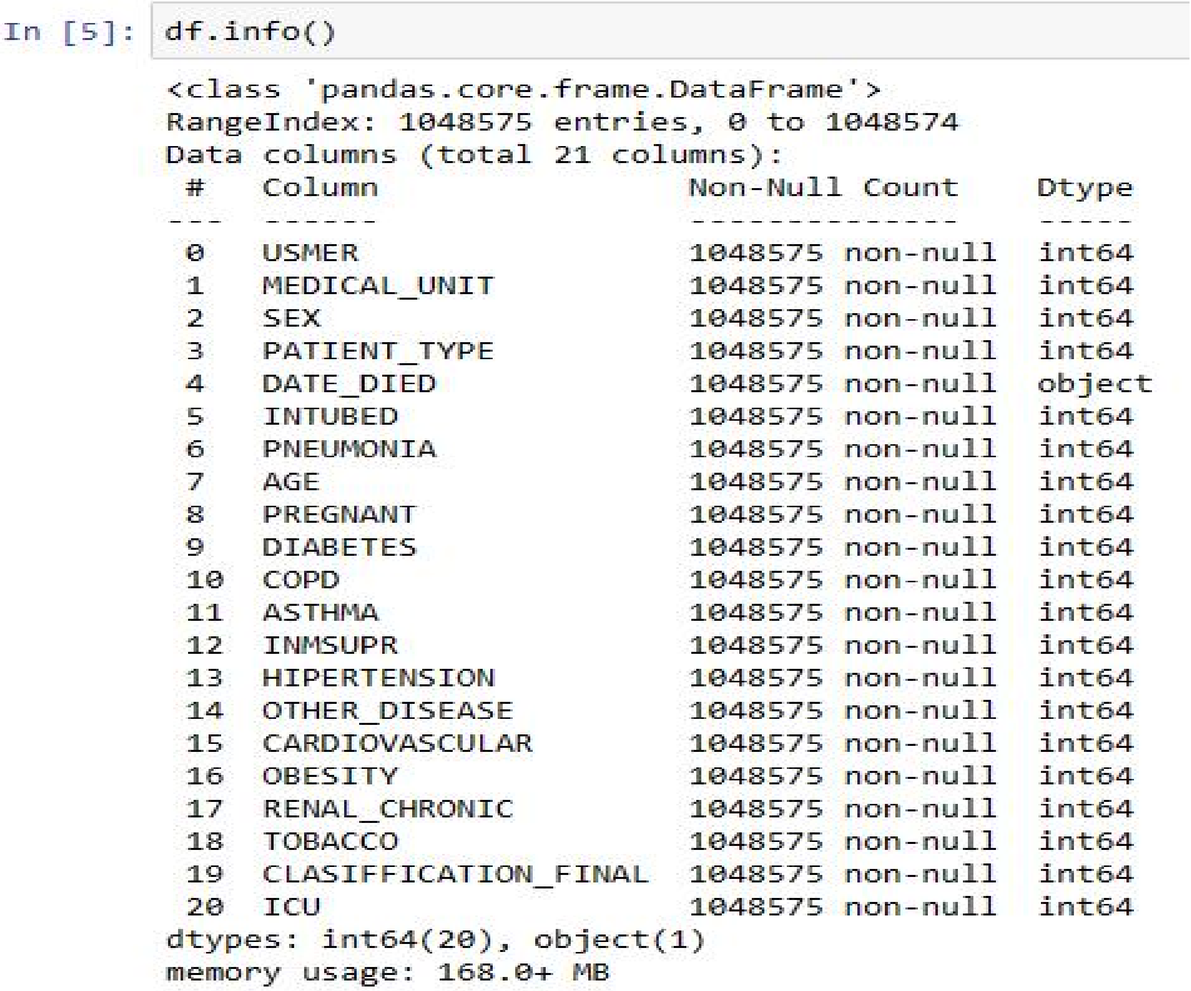
```
In [3]: df
```

Out[3]:

	USMER	MEDICAL_UNIT	SEX	PATIENT_TYPE	DATE_DIED	INTUBED	PNEUMONIA	AGE	PREGNANT	DIABETES	...	ASTHMA	INMSUPR	HIPERTEN
0	2	1	1	1	03/05/2020	97	1	65	2	2	...	2	2	
1	2	1	2	1	03/06/2020	97	1	72	97	2	...	2	2	
2	2	1	2	2	09/06/2020	1	2	55	97	1	...	2	2	
3	2	1	1	1	12/06/2020	97	2	53	2	2	...	2	2	
4	2	1	2	1	21/06/2020	97	2	68	97	1	...	2	2	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	
1048570	2	13	2	1	9999-99-99	97	2	40	97	2	...	2	2	
1048571	1	13	2	2	9999-99-99	2	2	51	97	2	...	2	2	
1048572	2	13	2	1	9999-99-99	97	2	55	97	2	...	2	2	
1048573	2	13	2	1	9999-99-99	97	2	28	97	2	...	2	2	
1048574	2	13	2	1	9999-99-99	97	2	52	97	2	...	2	2	

1048575 rows × 21 columns

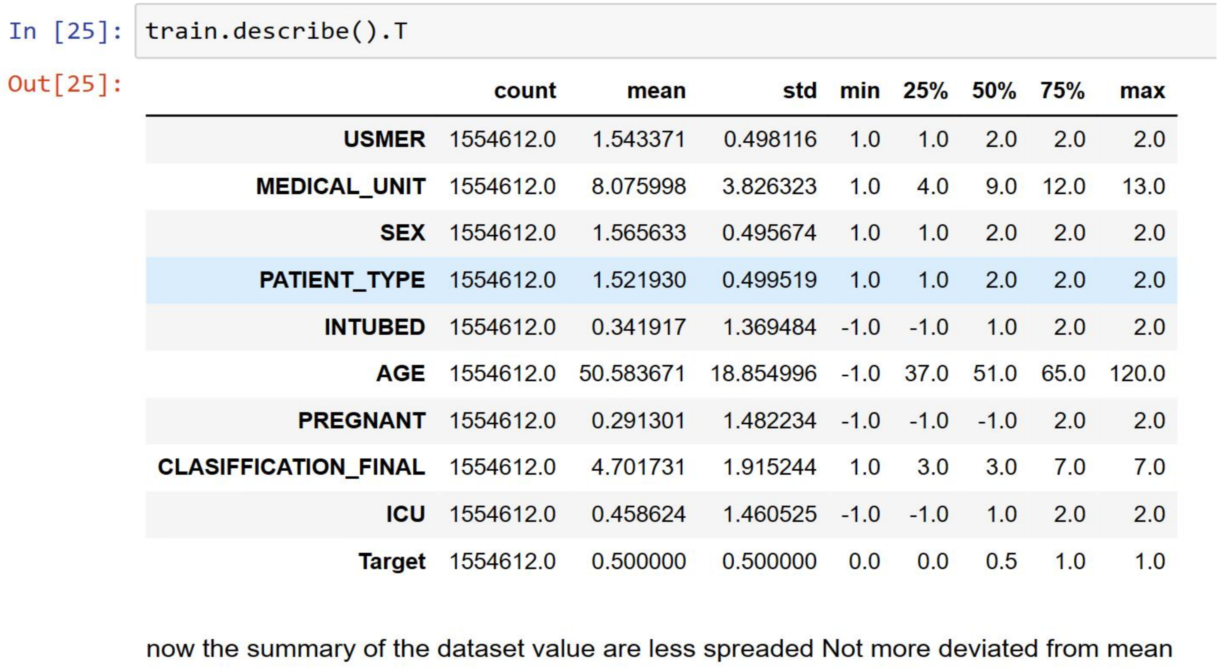
Look into the Dataset Information



```
In [5]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1048575 entries, 0 to 1048574
Data columns (total 21 columns):
 #   Column                Non-Null Count    Dtype 
---  ------                --------------    ----- 
 0   USMER                 1048575 non-null  int64 
 1   MEDICAL_UNIT          1048575 non-null  int64 
 2   SEX                   1048575 non-null  int64 
 3   PATIENT_TYPE          1048575 non-null  int64 
 4   DATE_DIED             1048575 non-null  object
 5   INTUBED               1048575 non-null  int64 
 6   PNEUMONIA             1048575 non-null  int64 
 7   AGE                   1048575 non-null  int64 
 8   PREGNANT              1048575 non-null  int64 
 9   DIABETES              1048575 non-null  int64 
 10  COPD                  1048575 non-null  int64 
 11  ASTHMA                1048575 non-null  int64 
 12  INMSUPR               1048575 non-null  int64 
 13  HIPERTENSION          1048575 non-null  int64 
 14  OTHER_DISEASE         1048575 non-null  int64 
 15  CARDIOVASCULAR        1048575 non-null  int64 
 16  OBESITY               1048575 non-null  int64 
 17  RENAL_CHRONIC         1048575 non-null  int64 
 18  TOBACCO               1048575 non-null  int64 
 19  CLASIFFICATION_FINAL  1048575 non-null  int64 
 20  ICU                   1048575 non-null  int64 
dtypes: int64(20), object(1)
memory usage: 168.0+ MB
```

Dataset Summary

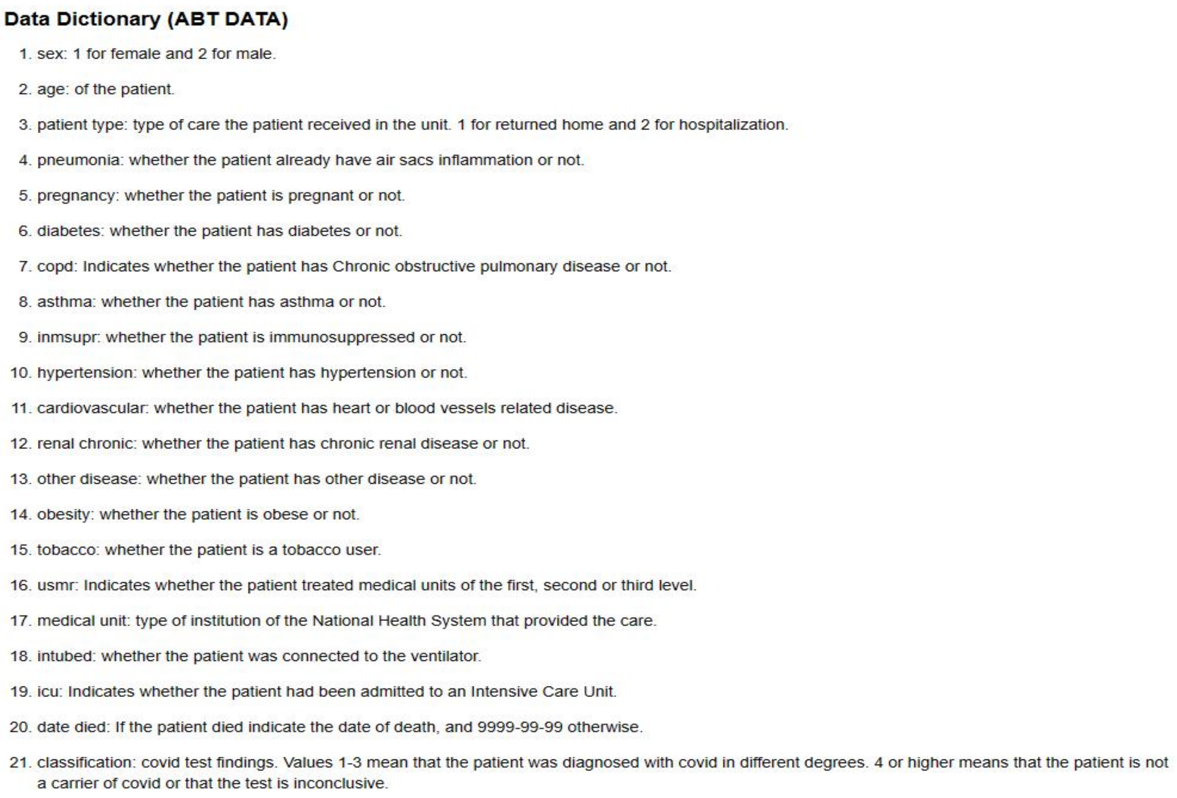


```
In [25]: train.describe().T
```

```
Out[25]:
                          count       mean        std   min   25%   50%   75%    max
               USMER  1554612.0   1.543371   0.498116   1.0   1.0   2.0   2.0    2.0
        MEDICAL_UNIT  1554612.0   8.075998   3.826323   1.0   4.0   9.0  12.0   13.0
                 SEX  1554612.0   1.565633   0.495674   1.0   1.0   2.0   2.0    2.0
        PATIENT_TYPE  1554612.0   1.521930   0.499519   1.0   1.0   2.0   2.0    2.0
             INTUBED  1554612.0   0.341917   1.369484  -1.0  -1.0   1.0   2.0    2.0
                 AGE  1554612.0  50.583671  18.854996  -1.0  37.0  51.0  65.0  120.0
            PREGNANT  1554612.0   0.291301   1.482234  -1.0  -1.0  -1.0   2.0    2.0
CLASIFFICATION_FINAL  1554612.0   4.701731   1.915244   1.0   3.0   3.0   7.0    7.0
                 ICU  1554612.0   0.458624   1.460525  -1.0  -1.0   1.0   2.0    2.0
              Target  1554612.0   0.500000   0.500000   0.0   0.0   0.5   1.0    1.0
```

now the summary of the dataset value are less spreaded Not more deviated from mean

Data Dictionary



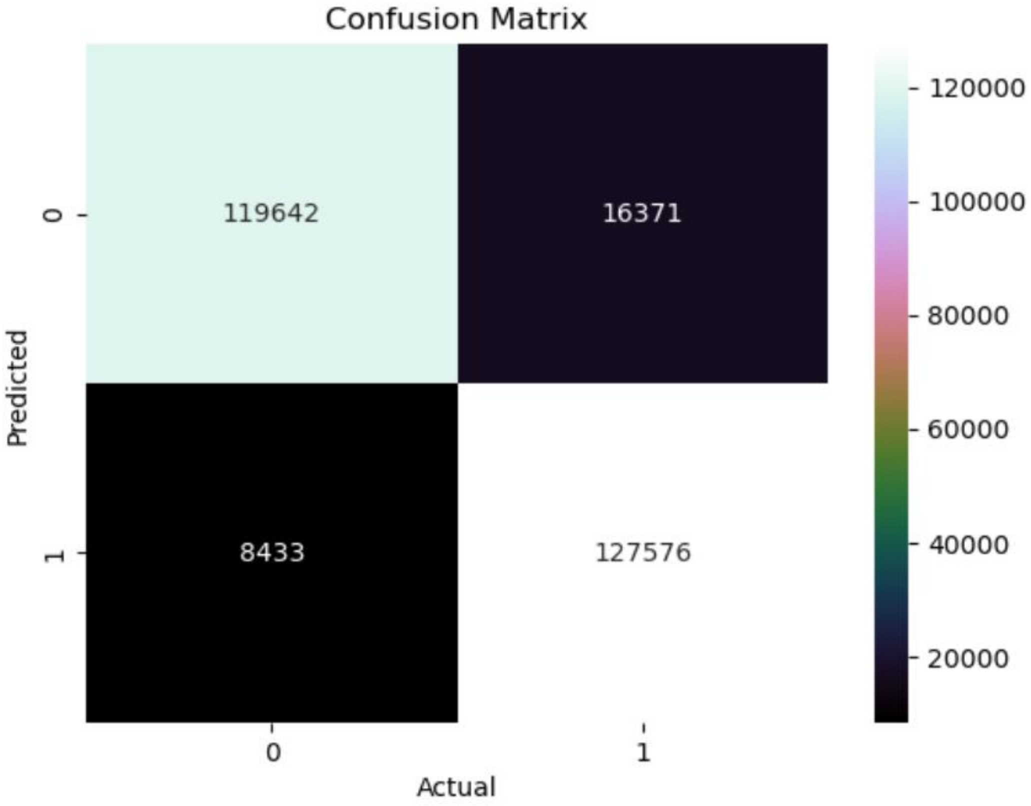
Data Dictionary (ABT DATA)

1. sex: 1 for female and 2 for male.
2. age: of the patient.
3. patient type: type of care the patient received in the unit. 1 for returned home and 2 for hospitalization.
4. pneumonia: whether the patient already have air sacs inflammation or not.
5. pregnancy: whether the patient is pregnant or not.
6. diabetes: whether the patient has diabetes or not.
7. copd: Indicates whether the patient has Chronic obstructive pulmonary disease or not.
8. asthma: whether the patient has asthma or not.
9. inmsupr: whether the patient is immunosuppressed or not.
10. hypertension: whether the patient has hypertension or not.
11. cardiovascular: whether the patient has heart or blood vessels related disease.
12. renal chronic: whether the patient has chronic renal disease or not.
13. other disease: whether the patient has other disease or not.
14. obesity: whether the patient is obese or not.
15. tobacco: whether the patient is a tobacco user.
16. usmr: Indicates whether the patient treated medical units of the first, second or third level.
17. medical unit: type of institution of the National Health System that provided the care.
18. intubed: whether the patient was connected to the ventilator.
19. icu: Indicates whether the patient had been admitted to an Intensive Care Unit.
20. date died: If the patient died indicate the date of death, and 9999-99-99 otherwise.
21. 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.

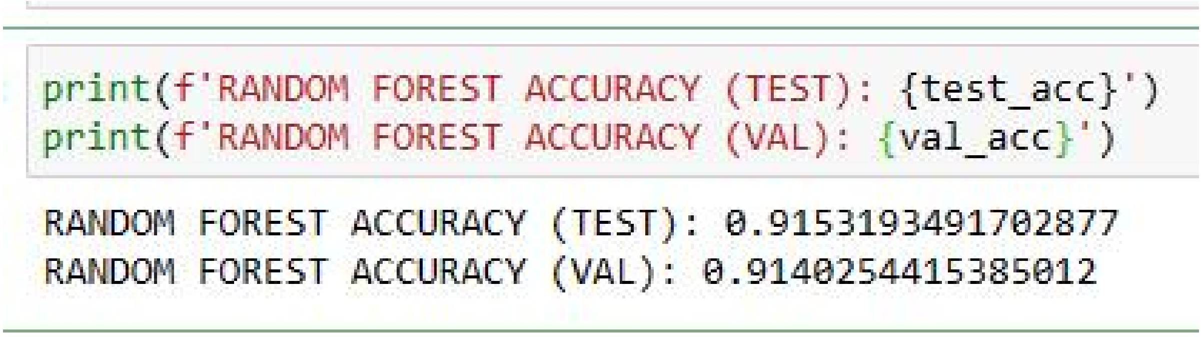
5. Benchmarking

5.1 Performance Metrics

Confusion Matrix



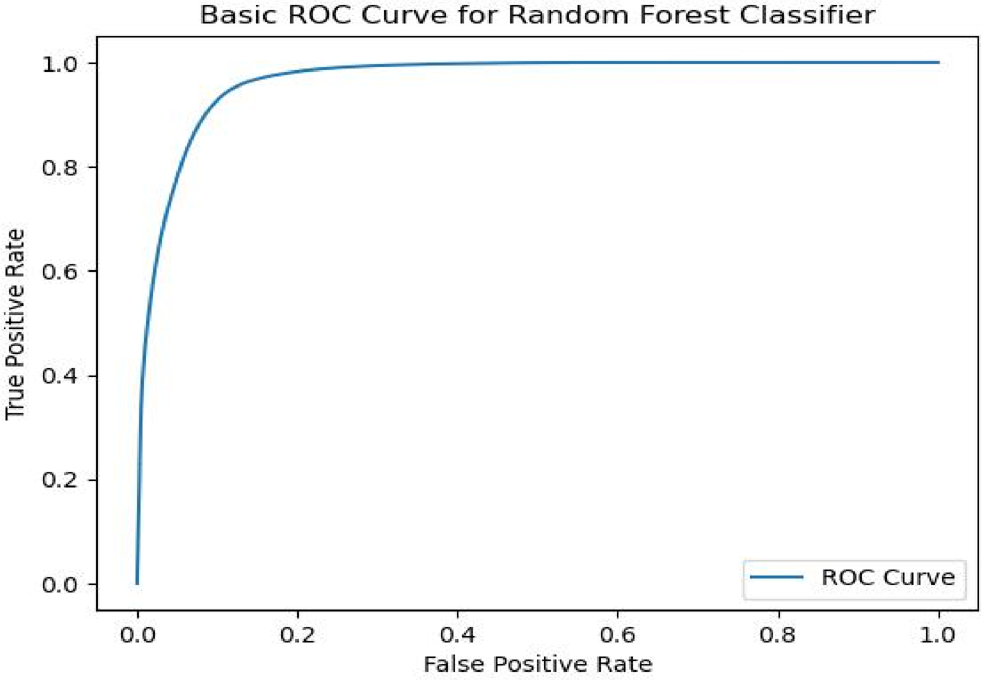
Accuracy Score

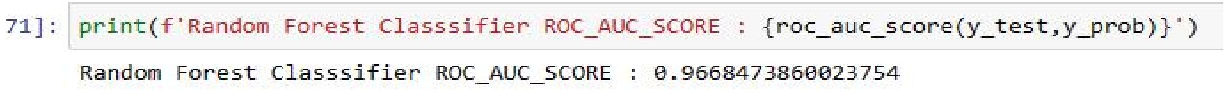


```
print(f'RANDOM FOREST ACCURACY (TEST): {test_acc}')
print(f'RANDOM FOREST ACCURACY (VAL): {val_acc}')
```

```
RANDOM FOREST ACCURACY (TEST): 0.9153193491702877
RANDOM FOREST ACCURACY (VAL): 0.9140254415385012
```

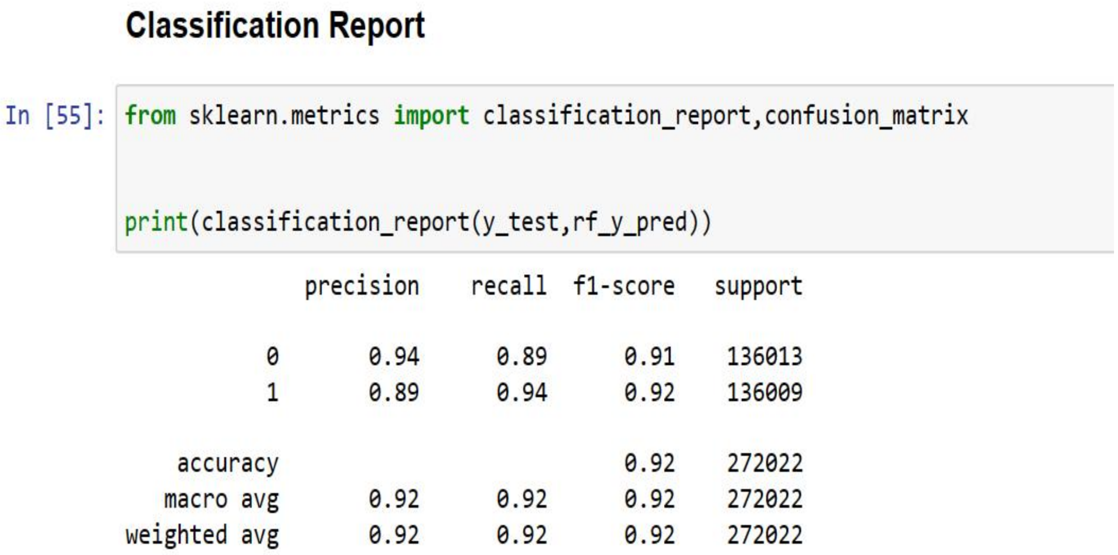
Receiver Operating Curve





```
71]: print(f'Random Forest Classsifier ROC_AUC_SCORE : {roc_auc_score(y_test,y_prob)}')
     Random Forest Classsifier ROC_AUC_SCORE : 0.9668473860023754
```

Classification Report

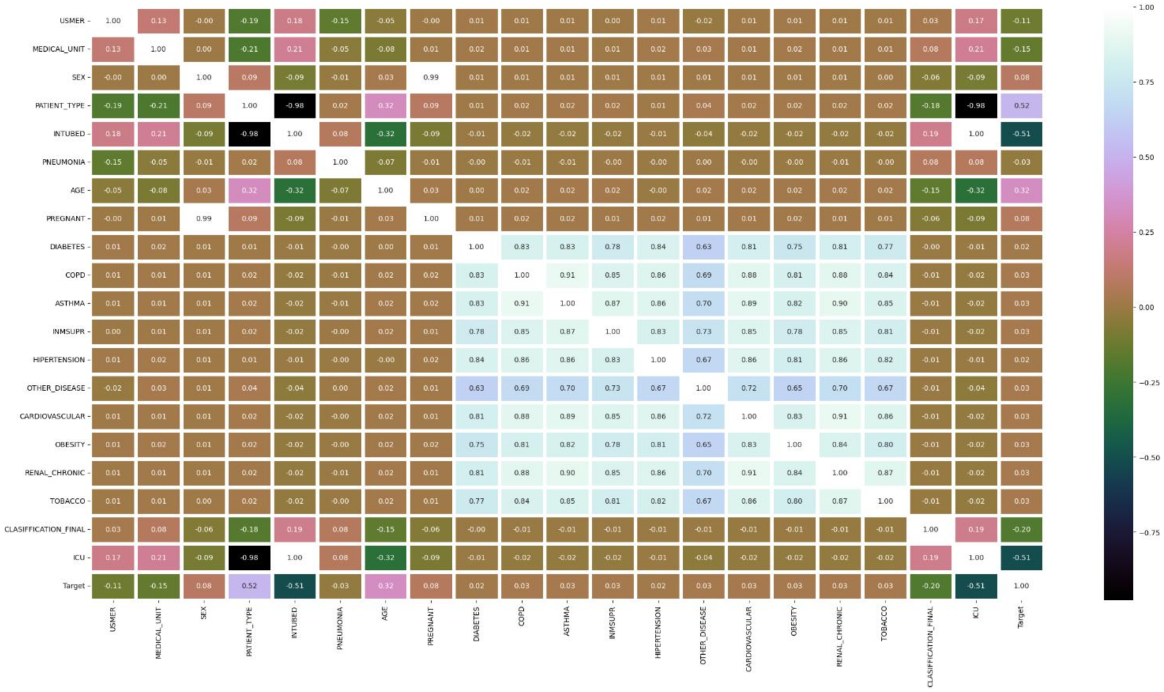


Classification Report

```
In [55]: from sklearn.metrics import classification_report,confusion_matrix


         print(classification_report(y_test,rf_y_pred))
```

```
              precision    recall  f1-score   support

           0       0.94      0.89      0.91    136013
           1       0.89      0.94      0.92    136009

    accuracy                           0.92    272022
   macro avg       0.92      0.92      0.92    272022
weighted avg       0.92      0.92      0.92    272022
```

Correlation Matrix (COVID-19 Dataset)



	USMER	MEDICAL_UNIT	SEX	PATIENT_TYPE	INTUBED	PNEUMONIA	AGE	PREGNANT	DIABETES	COPD	ASTHMA	INMSUPR	HIPERTENSION	OTHER_DISEASE	CARDIOVASCULAR	OBESITY	RENAL_CHRONIC	TOBACCO	CLASIFFICATION_FINAL	ICU	Target
USMER	1.00	0.13	-0.00	-0.19	0.18	-0.15	-0.05	-0.00	0.01	0.01	0.01	0.00	0.01	-0.02	0.01	0.01	0.01	0.01	0.03	0.17	-0.11
MEDICAL_UNIT	0.13	1.00	0.00	-0.21	0.21	-0.05	-0.08	0.01	0.02	0.01	0.01	0.01	0.02	0.03	0.01	0.02	0.01	0.01	0.08	0.21	-0.15
SEX	-0.00	0.00	1.00	0.09	-0.09	-0.01	0.03	0.99	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.00	-0.06	-0.09	0.08
PATIENT_TYPE	-0.19	-0.21	0.09	1.00	-0.98	0.02	0.32	0.09	0.01	0.02	0.02	0.02	0.01	0.04	0.02	0.02	0.02	0.02	-0.18	-0.98	0.52
INTUBED	0.18	0.21	-0.09	-0.98	1.00	0.08	-0.32	-0.09	-0.01	-0.02	-0.02	-0.02	-0.01	-0.04	-0.02	-0.02	-0.02	-0.02	0.19	1.00	-0.51
PNEUMONIA	-0.15	-0.05	-0.01	0.02	0.08	1.00	-0.07	-0.01	-0.00	-0.01	-0.01	-0.00	-0.00	0.00	-0.00	-0.00	-0.01	-0.00	0.08	0.08	-0.03
AGE	-0.05	-0.08	0.03	0.32	-0.32	-0.07	1.00	0.03	0.00	0.02	0.02	0.02	-0.00	0.02	0.02	0.02	0.02	0.02	-0.15	-0.32	0.32
PREGNANT	-0.00	0.01	0.99	0.09	-0.09	-0.01	0.03	1.00	0.01	0.02	0.02	0.01	0.02	0.01	0.01	0.02	0.01	0.01	-0.06	-0.09	0.08
DIABETES	0.01	0.02	0.01	0.01	-0.01	-0.00	0.00	0.01	1.00	0.83	0.83	0.78	0.84	0.63	0.81	0.75	0.81	0.77	-0.00	-0.01	0.02
COPD	0.01	0.01	0.01	0.02	-0.02	-0.01	0.02	0.02	0.83	1.00	0.91	0.85	0.86	0.69	0.88	0.81	0.88	0.84	-0.01	-0.02	0.03
ASTHMA	0.01	0.01	0.01	0.02	-0.02	-0.01	0.02	0.02	0.83	0.91	1.00	0.87	0.86	0.70	0.89	0.82	0.90	0.85	-0.01	-0.02	0.03
INMSUPR	0.00	0.01	0.01	0.02	-0.02	-0.00	0.02	0.01	0.78	0.85	0.87	1.00	0.83	0.73	0.85	0.78	0.85	0.81	-0.01	-0.02	0.03
HIPERTENSION	0.01	0.02	0.01	0.01	-0.01	-0.00	-0.00	0.02	0.84	0.86	0.86	0.83	1.00	0.67	0.86	0.81	0.86	0.82	-0.01	-0.01	0.02
OTHER_DISEASE	-0.02	0.03	0.01	0.04	-0.04	0.00	0.02	0.01	0.63	0.69	0.70	0.73	0.67	1.00	0.72	0.65	0.70	0.67	-0.01	-0.04	0.03
CARDIOVASCULAR	0.01	0.01	0.01	0.02	-0.02	-0.00	0.02	0.01	0.81	0.88	0.89	0.85	0.86	0.72	1.00	0.83	0.91	0.86	-0.01	-0.02	0.03
OBESITY	0.01	0.02	0.01	0.02	-0.02	-0.00	0.02	0.02	0.75	0.81	0.82	0.78	0.81	0.65	0.83	1.00	0.84	0.80	-0.01	-0.02	0.03
RENAL_CHRONIC	0.01	0.01	0.01	0.02	-0.02	-0.01	0.02	0.01	0.81	0.88	0.90	0.85	0.86	0.70	0.91	0.84	1.00	0.87	-0.01	-0.02	0.03
TOBACCO	0.01	0.01	0.00	0.02	-0.02	-0.00	0.02	0.01	0.77	0.84	0.85	0.81	0.82	0.67	0.86	0.80	0.87	1.00	-0.01	-0.02	0.03
CLASIFFICATION_FINAL	0.03	0.08	-0.06	-0.18	0.19	0.08	-0.15	-0.06	-0.00	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	1.00	0.19	-0.20
ICU	0.17	0.21	-0.09	-0.98	1.00	0.08	-0.32	-0.09	-0.01	-0.02	-0.02	-0.02	-0.01	-0.04	-0.02	-0.02	-0.02	-0.02	0.19	1.00	-0.51
Target	-0.11	-0.15	0.08	0.52	-0.51	-0.03	0.32	0.08	0.02	0.03	0.03	0.03	0.02	0.03	0.03	0.03	0.03	0.03	-0.20	-0.51	1.00

6. Applicable Patents

- ➢ Patent-1 COVID-19 prediction models: a systematic literature review
  Ref Link -> [COVID-19 prediction models: a systematic literature review - PubMed (nih.gov)]

- ➢ Patent-2 Explainable Machine Learning for Early Assessment of COVID-19 Risk Prediction in Emergency Departments
  Ref Link-> [Explainable Machine Learning for Early Assessment of COVID-19 Risk Prediction in Emergency Departments | IEEE Journals & Magazine | IEEE Xplore]

In the realm of COVID-19 patient risk prediction systems, two pivotal patents serve as noteworthy inspirations, both centered around the application of early diagnosis of risk factor infected patients. The first patent intricately outlines the methodology of Association Rule Mining and its application in delineating customer purchase behavior within a dataset. This innovative approach involves the generation of item baskets based on predetermined attributes, accompanied by the formulation of rules that effectively group items together. This foundational patent provides valuable insights into adapting various machine learning prediction models for the nuanced intricacies of COVID-19 patient data.

Complementing this, the second patent introduces an enhanced model tailored in the context of COVID-19 risk prediction in emergency departments. This model delves into historical patient data, extracting causal and predictive patterns to generate associated risk factors. The output is augmented by a quantifiable score, providing a nuanced characterization of the likelihood of specific risk predictions manifesting in patient outcomes. These patented techniques, designed for in-depth analysis and prediction, present a robust foundation for the development of a sophisticated COVID-19 patient risk prediction system.

7. Applicable Regulations

In the forthcoming development and implementation of a similar system tailored for small-scale businesses, these two patents offer not only methodological insights but also practical considerations in adapting prediction models to the unique challenges posed by the pandemic. COVID-19 patient risk prediction system aims to provide a comprehensive and nuanced understanding of risk factors, ultimately facilitating informed decision-making for healthcare professionals and contributing to the ongoing efforts to mitigate the impact of the virus.

When pursuing these patent ideas, it is imperative to collaborate with a patent professional to ensure that the claims are specific, novel, and non-obvious, meeting all legal requirements and maximizing the protection of the inventive concepts outlined in the patent application.

8. Applicable Constraints

- Limited Access to Patient Health Records
- Real-Time Data Collection Challenges
- Healthcare Professionals Technical Proficiency
- Limited Availability of High-Quality Data

9. Business Opportunity

Developing a Covid-19 patient risk prediction system can be a valuable and impactful business opportunity, especially given the ongoing global pandemic. Here are some key aspects to consider when exploring this opportunity:

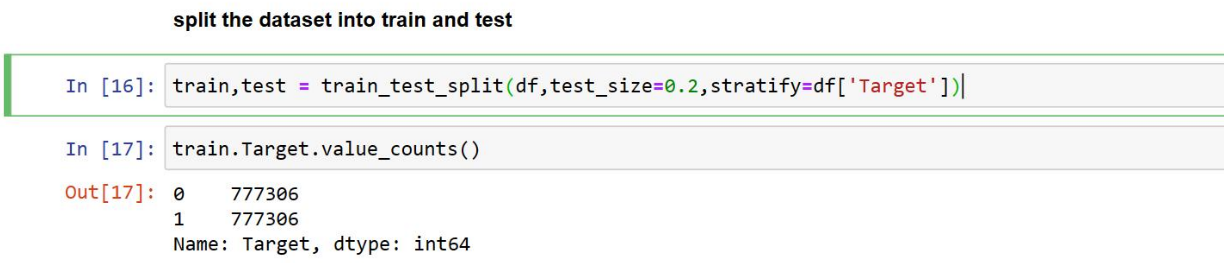
Why early risk prediction is important in business level?

- Identify key features that contribute to accurate predictions. This may involve analyzing the importance of different factors in determining the risk level of a Covid-19 patient.
- Ensure seamless integration with existing healthcare systems and electronic health records (EHR) to provide healthcare professionals with easy access to the prediction results. This could involve developing APIs or partnerships with healthcare IT providers.
- Implement a real-time monitoring system that continuously updates risk predictions based on the latest patient data. This can help healthcare providers make timely and informed decisions.
- Design an intuitive and user-friendly interface for healthcare professionals to interact with the system. The interface should provide clear and actionable insights to aid in patient management.
- Consider different business models, such as licensing the system to healthcare institutions, charging subscription fees, or partnering with healthcare providers for a revenue-sharing model.

Remember to consult with healthcare professionals, data scientists, and legal experts to ensure the accuracy, reliability, and ethical use of your Covid-19 patient risk prediction system.

10. Concept Generation

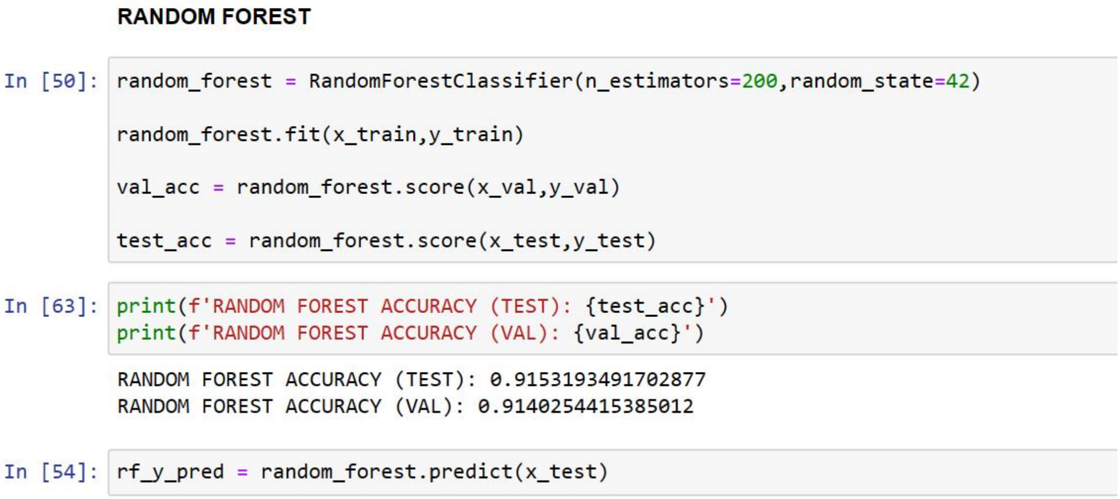
Splitting the dataset into train and test set



split the dataset into train and test

```
In [16]: train,test = train_test_split(df,test_size=0.2,stratify=df['Target'])

In [17]: train.Target.value_counts()

Out[17]: 0    777306
         1    777306
         Name: Target, dtype: int64
```

Random Forest Classifier with 0.91 ACC



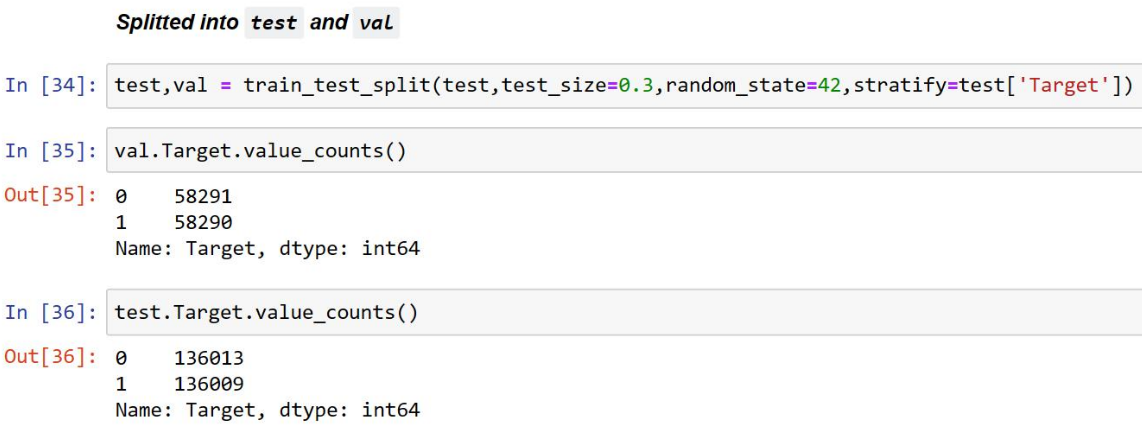
RANDOM FOREST

```
In [50]: random_forest = RandomForestClassifier(n_estimators=200,random_state=42)

         random_forest.fit(x_train,y_train)

         val_acc = random_forest.score(x_val,y_val)

         test_acc = random_forest.score(x_test,y_test)
```

```
In [63]: print(f'RANDOM FOREST ACCURACY (TEST): {test_acc}')
         print(f'RANDOM FOREST ACCURACY (VAL): {val_acc}')

         RANDOM FOREST ACCURACY (TEST): 0.9153193491702877
         RANDOM FOREST ACCURACY (VAL): 0.9140254415385012
```

```
In [54]: rf_y_pred = random_forest.predict(x_test)
```

Splitting into test and validation set



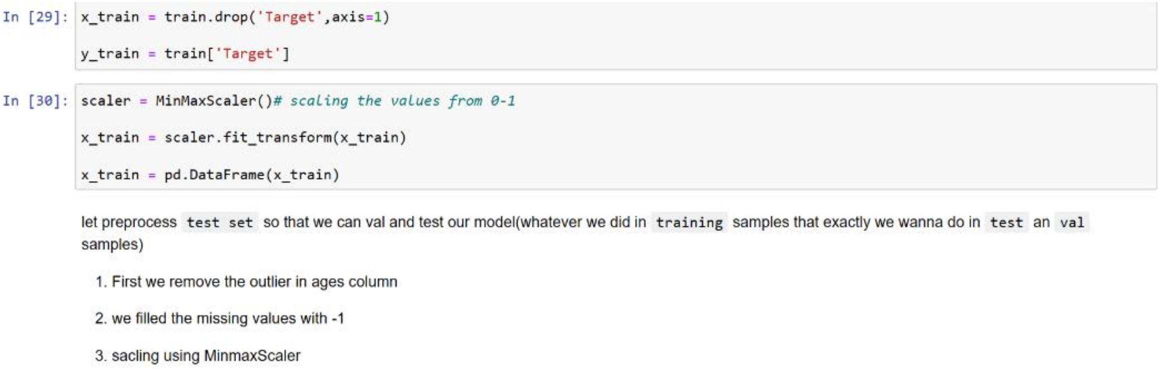
Splitted into `test` and `val`

```
In [34]: test,val = train_test_split(test,test_size=0.3,random_state=42,stratify=test['Target'])
```

```
In [35]: val.Target.value_counts()

Out[35]: 0    58291
         1    58290
         Name: Target, dtype: int64
```

```
In [36]: test.Target.value_counts()

Out[36]: 0    136013
         1    136009
         Name: Target, dtype: int64
```

Feature Scaling



```
In [29]: x_train = train.drop('Target',axis=1)

         y_train = train['Target']
```

```
In [30]: scaler = MinMaxScaler()# scaling the values from 0-1

         x_train = scaler.fit_transform(x_train)

         x_train = pd.DataFrame(x_train)
```

let preprocess `test set` so that we can val and test our model(whatever we did in `training` samples that exactly we wanna do in `test` an `val` samples)

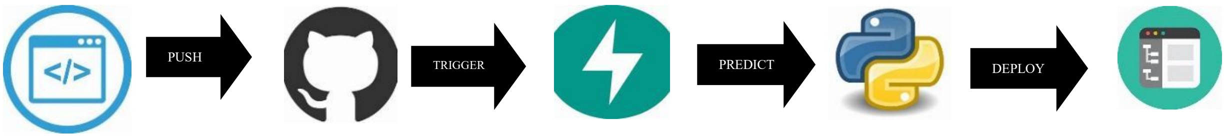
1. First we remove the outlier in ages column

2. we filled the missing values with -1

3. sacling using MinmaxScaler

12. Concept Development

The COVID-19 risk prediction system has been successfully developed, employing FastAPI for the backend, a Random Forest machine learning algorithm for predictions, and a frontend implemented with HTML, CSS, and JavaScript.



CODE APP **GITHUB** **FAST API** **PYTHON** **WEB**

13. Final Report Prototype

COVID-19 risk prediction system implementation with good accurate results for infected and uninfected patients to check whether the person is under risk or not.

Back-end Development

The backend of the COVID-19 risk prediction system is developed using FastAPI, a modern, fast (high-performance), web framework for building APIs with Python 3.11.7 based on standard Python type hints. The backend serves as the core engine responsible for processing data, invoking the Random Forest machine learning algorithm for predictions, and handling communication with the frontend. It is optimized for performance through the use of asynchronous programming, ensuring efficient handling of concurrent requests.

- Performing EDA to realize the dependent and independent features.
- Algorithm training and optimization must be done to minimize overfitting of the model and hyperparameter tuning

Front-end Development

The frontend of the COVID-19 risk prediction system is built using a combination of HTML, CSS, and JavaScript, creating an intuitive and user-friendly interface. The user interface is designed to enhance user experience, featuring real-time updates to provide dynamic interactions. The frontend covers the following features:

- Dynamic Interactions
- User Friendly with numbering system input field
- Easy to use and Single Page Application

14. Product Details

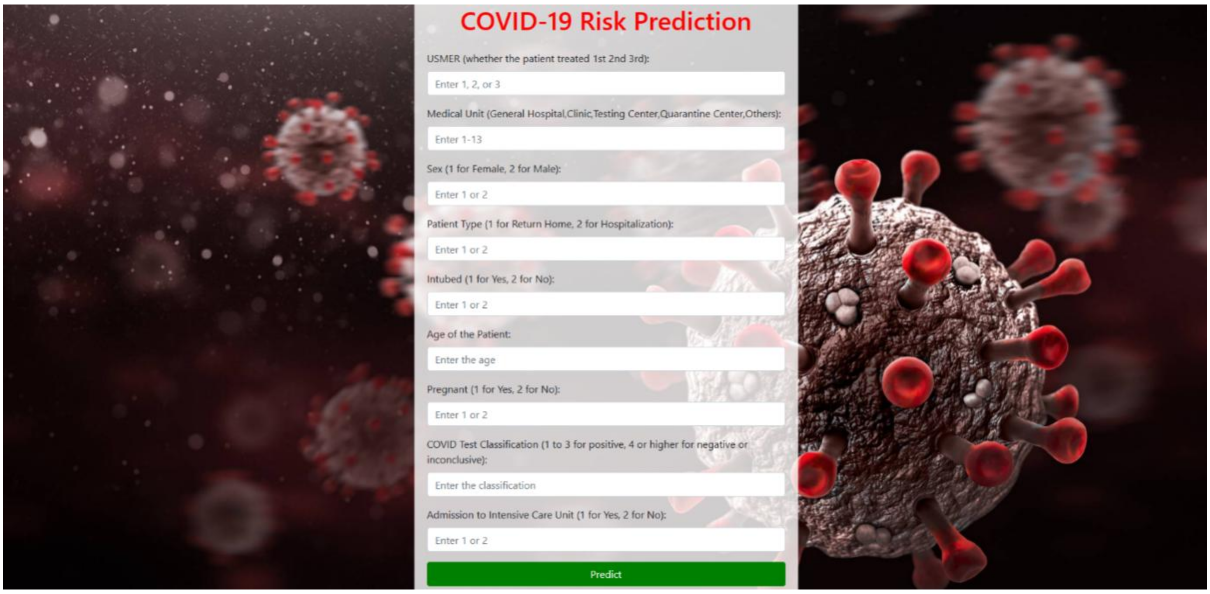
How Does it Work?

Created a web app using the Fast API framework and use random forest model for training data and testing the web app in localhost server.

For predicting whether the person is under risk or not. The following details are mandatory in order to successfully run the application

- 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.
- pregnancy: whether the patient is pregnant or not.
- 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.
- intubed: whether the patient was connected to the ventilator.
- icu: Indicates whether the patient had been admitted to an Intensive Care Unit.

Covid-19 Risk Prediction System – User Interface



COVID-19 Risk Prediction

USMER (whether the patient treated 1st 2nd 3rd):

Enter 1, 2, or 3

Medical Unit (General Hospital,Clinic,Testing Center,Quarantine Center,Others):

Enter 1-13

Sex (1 for Female, 2 for Male):

Enter 1 or 2

Patient Type (1 for Return Home, 2 for Hospitalization):

Enter 1 or 2

Intubed (1 for Yes, 2 for No):

Enter 1 or 2

Age of the Patient:

Enter the age

Pregnant (1 for Yes, 2 for No):

Enter 1 or 2

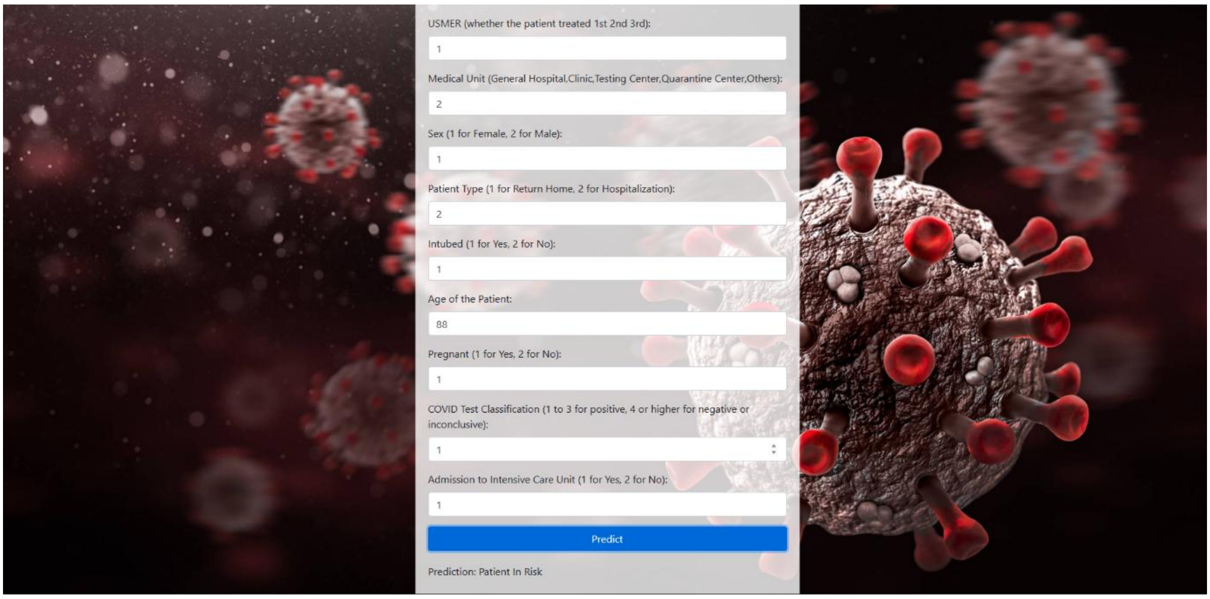
COVID Test Classification (1 to 3 for positive, 4 or higher for negative or inconclusive):

Enter the classification

Admission to Intensive Care Unit (1 for Yes, 2 for No):

Enter 1 or 2

Input field are filled with patient details – Risk



USMER (whether the patient treated 1st 2nd 3rd):

1

Medical Unit (General Hospital,Clinic,Testing Center,Quarantine Center,Others):

2

Sex (1 for Female, 2 for Male):

1

Patient Type (1 for Return Home, 2 for Hospitalization):

2

Intubed (1 for Yes, 2 for No):

1

Age of the Patient:

88

Pregnant (1 for Yes, 2 for No):

1

COVID Test Classification (1 to 3 for positive, 4 or higher for negative or inconclusive):

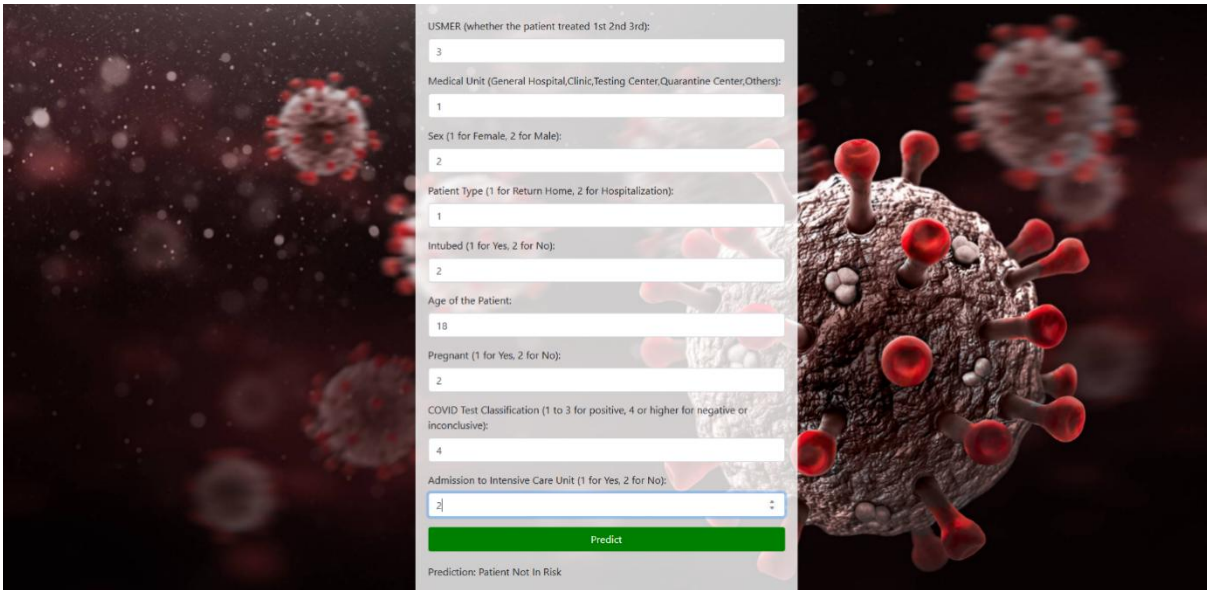
1

Admission to Intensive Care Unit (1 for Yes, 2 for No):

1

Prediction: Patient In Risk

Input field are filled with patient details – Not Risk



USMER (whether the patient treated 1st 2nd 3rd): 3

Medical Unit (General Hospital,Clinic,Testing Center,Quarantine Center,Others): 1

Sex (1 for Female, 2 for Male): 2

Patient Type (1 for Return Home, 2 for Hospitalization): 1

Intubed (1 for Yes, 2 for No): 2

Age of the Patient: 18

Pregnant (1 for Yes, 2 for No): 2

COVID Test Classification (1 to 3 for positive, 4 or higher for negative or inconclusive): 4

Admission to Intensive Care Unit (1 for Yes, 2 for No): 2

Prediction: Patient Not In Risk

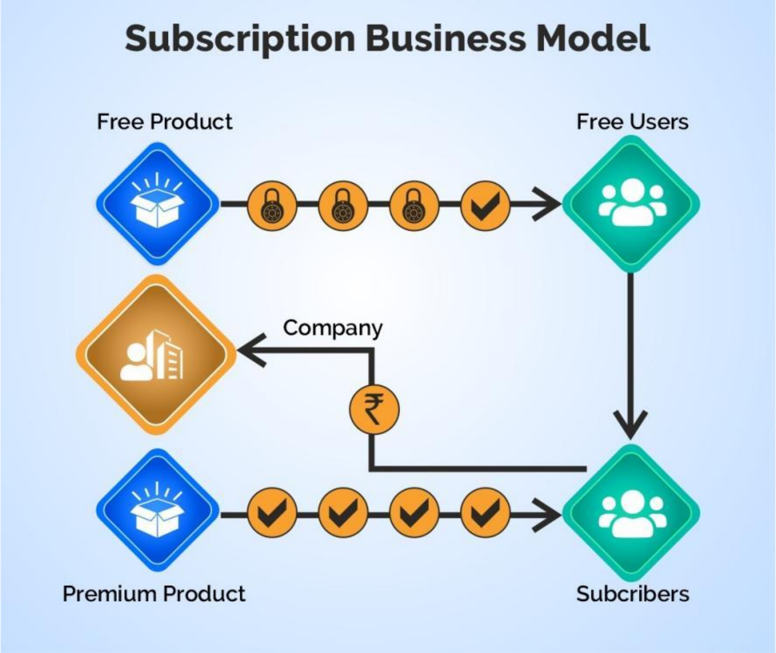
15. Code Implementation

https://github.com/theSuriya/COVID-19/tree/main/PycharmProjects/Covid-19

16. Conclusion

The primary objective of the product was to develop and validate a population-based prediction model, leveraging a large and diverse dataset. By incorporating a selective, clinically informed approach, our model dynamically estimates the COVID-19 mortality risk for confirmed diagnoses. This not only enhances the accuracy of predictions but also emphasizes the adaptability of our solution to different patient profiles. This also underscores the critical role of personalized medicine in optimizing the treatment of COVID-19 patients, particularly in the early identification of those at high risk for severe complications.

Business Model:



1. Value Proposition:

- Early Risk Prediction: Provide hospitals with a tool that enables early identification of COVID-19 risks in patients.
- Resource Optimization: Assist healthcare providers in optimizing resource allocation and prioritizing high-risk cases.
- Improved Patient Outcomes: Enhance patient outcomes by enabling proactive healthcare interventions.

2. Customer Segments:

- Hospitals and Healthcare Organizations: Primary customers for deploying the risk prediction system.
- Healthcare Professionals: Users of the system within hospitals.

3. Channels:

- Direct Sales: Engage in direct sales to hospitals and healthcare institutions.
- Online Platform: Provide an online platform for easy access and deployment.

4. Customer Relationships:

- Customer Support: Offer customer support for system integration and ongoing assistance.
- Training: Provide training programs for healthcare professionals on using and interpreting the system.

5. Revenue Streams:

- Subscription Model: Charge hospitals a subscription fee based on the number of patients or a flat rate.
- Implementation Fee: Charge an initial fee for system integration and training.

6. Key Resources:

- Technology Infrastructure: Secure and efficient server infrastructure for hosting the risk prediction model.
- Data Sources: Access to relevant and up-to-date healthcare data.
- Partnerships: Collaborate with healthcare institutions for data sharing and system implementation.

7. Key Activities:

- System Development: Ongoing development and improvement of the risk prediction model.
- Data Management: Ensure the accuracy and security of the data used by the system.
- Marketing and Sales: Promote the system to hospitals and healthcare organizations.

8. Key Partnerships:

- Healthcare Institutions: Partner with hospitals and healthcare organizations for data access and system implementation.
- Technology Partners: Collaborate with technology providers for infrastructure support.

9. Cost Structure:

- Development Costs: Invest in the continuous improvement of the risk prediction model.
- Server and Infrastructure Costs: Maintain a secure and scalable server infrastructure.
- Sales and Marketing Expenses: Promote the system to the target audience.

10. Regulatory and Ethical Considerations:

- Ensure compliance with healthcare data regulations and ethical standards in handling patient information.

Financial Equation:

The financial equation could be adapted based on the revenue streams:

$\mathbf{Y=(P\times N)+I}$

- **Y** is the financial outcome (revenue).
- **P** is the price per patient (subscription fee).
- **N** is the number of patients.
- **I** is the implementation fee.

For example:

- P = 100 (price per patient per month).
- N=1000(number of patients).
- I = 2000 (implementation fee).

Y = (1000 \times100) + 2000

GitLink:

- Suriya: https://github.com/theSuriya/Ev-Vehicle-Segmentation.
- Swanand Kulkarni: SSK007-b/EV_Segementation (github.com)
- Kasyap Velampalli: kashyyvel/EV-Market-Segmentation (github.com)