

I wrote some comments and results on the code file, so I only post the link of my code(colab)

https://colab.research.google.com/drive/1GX_6QnnNrbk01PzOFtjFU8OuAVThoi8g?usp=sharing

Because of the notice, I add

```
train.loc[(300 < train['7']) & (train['7'] < 450), '7'] = train.loc[(300 < train['7']) & (train['7'] < 450), '7'] / 10  
  
train.loc[(3000 < train['7']) & (train['7'] < 4500), '7'] = train.loc[(3000 < train['7']) & (train['7'] < 4500), '7'] / 100
```

line at preprocessing part, so that I change some values in column 7 (it might be temperature of human body).

For the assignment, I change the values in timestamp(day) because the assignment said that -3,-4,-5 values have to be -3

And column 1 and 2 are gender and age, so I change them into 'info_target' column and change the value as 'age_gender'.

Also I want to find the end of the patients record, so I add 'end_of_record' column and if both timestamp values are zero, 'end_of_record' value would return to 1(true).

I delete some rows that 'original' value is null (Nan), and after all these things, I use SMOTE to add values that the data is null(Nan).

After these process, I found EDA values such as mean, median, standard deviation, correlation matrix (I don't know why but correlation matrix has different values or null values from part 1 code).

And I did PCA process, it takes about 16minutes.

Now the important parts, machine learning process.

When I choose many models such as MLPClassifier, SGDClassifier, KNN, RNN, etc. but I can't find the models that f1 score(macro) is more than 0.20 or 0.43. All the models get accuracies more than 0.45, but the other values, such as f1 score, recall, precision, are 0.00 values (if I print whole values they are like 0.0068, 0.00000043).

I want to find another models to get proper values but because of SGD model (what I tried the last), it takes more than 4 hours to get values, and my computer ran down, so I couldn't do this more.

```
model = SGDClassifier()

for batch in range(num_batches):
    start = batch * batch_size
    end = min(start + batch_size, len(X_train))

    X_batch = X_train[start:end]
    y_batch = y_train[start:end]

    model.partial_fit(X_batch, y_batch, classes=np.unique(y_train))

y_pred = model.predict(X_test)
```

✓ 244m 15.0s

+ 코드 + Markdown

And confusion matrix and roc curves, I can't make it and don't know why the confusion matrix is not printed and the error occurs when I run the code that make roc_curves.