Sepsis prediction:

Approach 1: This approach included two steps to analyse sepsis data: features selction, features engieering and then ensembling.
The features engineering schema seperated all the covariates in the data into two clusters: first cluster of covariates with the minimun number of missing values with threshold of 10%.
For the features with low missing values:

- Aggregation using Sliding windows of 5 and 11 hours frames while applying diffente methods:
    - Min , max, mean , median, and varaince.
    - Quantile of 95%,99%,5%, and 1%
- Other features to capture Long and Shot Term Dependencies like Shannon Entropy energy, mean fo the first differneces and the length of stay for a specific patient.

Morteza el. H while the second cluster grouped the remaming covariates. The total number of features is around 410 features.
In order to reduce the number of covaraites and reduce the bias and variance, a second layer of features selection was applied for two main objetives: select best performing features and five best hyperparameters.
The described procedure above was applied on only 10% of the data. One interesting point about the features selection in this appraoch is that it was perfomed using BoostARoota algorithms based on Xgboost. The procedure is that for each features we create another shadow features and use only both of them on the training. Then, the features importance will be presented. If shadow features repoted as more important tha the original features, this latter will be elimnated.
Once it's decided about which are the best performaing arameters and best performing covariates, the training start using the remianing 90% of the training data. In fact, in this approach we construct five randonmly disjoint sets and aply an undersampling technique to balance class in each sets seperately. At the end, 5Xgboost models are trained using each sets with 5-fold- cross-validation. The final otput is calculated using the geometric mean of the five outputs of the trained model.
As results, this model training structure has achieved an AUC score of 0.833 abd an accuracy of 0.8440.