STATUS REPORT

FINAL SOLUTION FEATURES AND UPDATES

19/02/2022

1.Backward elimination :- Using Backward elimination for feature engineering to select the most relevant features from the original data and then feeding it to the autoencoder.
From 1033 columns ,277 important columns were selected to feed it to the autoencoder

2.Autoencoder Improvement : -By feeding only relevant columns, the performance of the autoencoder was slightly increased i.e rounded accuracy was nearly 90%.

3.Random Sampling :- Applied Random sampling to balance out the classes. In the previous code , classes were imbalanced i.e no of classes of 2 were greater than 0, but now both the classes are balanced .
Earlier **1000** instances for 2 and approx **350** instances for 0
Now **607** instances for 2 and **500** instances for 0.

4.Hyperparameters Tuning :- Hyperparameter tuning for both the models using GridSearchCV and RandomizedSearchCV and the results are as follows:

1. **Random Forest** : Random Forest had no improvement even after performing Parameter tuning and was taking a good amount of time(~15 mins) to execute. Therefore the model is executed by simple parameters only.
2. **XGBoost :** A highly enhanced model is obtained using RandomizedSearch CV on XGBoost and the accuracy is boosted to more than 95%.