Foundations of machine learning

Assignment 2

Roll No	Name
CS18BTECH11039	Raj Patil
CS18BTECH11021	Karan Bhukar

We are submitting the following two approaches

1. an Ensemble of Ensembles:
   - probability based voting from a Bagger(Random Forest) and Booster(XGboost)
   - with minimal feature engineering i.e. we have only dropped features 16,17,18 due to large amount of missing data
   - kaggle submission comment as 'Submission 1 : Ensemble of Ensembles'
   - Public score : 0.93896
   - Private Score : 0.93427
2. Stratified Cross Validation with oversampling : base model being XGBoost
   - oversampled the rare cases until they met a minimum number
   - minimal feature engineering here as well : only dropping 17,18,19
   - because normalization/ standardization did not improve the model given our algorithm
   - Public score : 0.91549
   - Private score : 0.92488

for both the models , the 'test_input.csv' file is expected to be in their directories. They also need the original training set(already placed) to be in their directories.

Running instructions :

```
python Submission1.py
```

and

```
python Submission2.py
```

will output test_output.csv in their directory