Result of Ensemble Learning

TABLE I. The proportion of each class in different datasets

Number of Instances	Undersampling Data		Oversampling Data		Whole Dataset	
	Train	Test	Train	Test	Train	Test
Fraud	341	151	199023	151	341	151
Normal	341	85292	199023	85292	199023	85292
Total	682	85443	398046	85443	199364	85443

TABLE II. Optimized hyper-parameters for logistic regression classifier, decision tree classifier, random forest classifier given three dataset options

Recall on class 1	Undersampling Data	Oversampling Data	Whole Dataset
LR	C=1 penalty='l2'	C= 100 penalty='l2'	C= 10 penalty='l1'
DT	criterion='entropy' max_depth=3 min_samples_leaf=1 min_samples_split=4	criterion='gini' max_depth=4 min_samples_leaf=1 min_samples_split=2	criterion='gini' max_depth=4 min_samples_leaf=4 min_samples_split=2
RF	criterion='gini' max_depth=5 min_samples_leaf=9 min_samples_split=2	criterion='gini' max_depth=4 min_samples_leaf=1 min_samples_split=4	criterion='gini' max_depth=4 min_samples_leaf=1 min_samples_split=2

Table III. The performances (recall score on class 1 - fraud) from logistic regression classifier, decision tree classifier, random forest classifier and bagging method given three dataset options

Recall on class 1	Undersampling Data		Oversampling Data		Whole Dataset	
	Train	Test	Train	Test	Train	Test
LR	0.93	0.89	0.95	0.88	0.67	0.62
DT	0.87	0.82	0.94	0.88	0.80	0.73
RF	0.91	0.87	0.92	0.85	0.70	0.60
Bagging	0.94	0.91	0.95	0.88	0.94	0.73