

1) Bootstrap Aggregation Algorithm

- Please see the notebook at the end of this document.
- Please see the notebook at the end of this document.
-

Num of trees	10	50	100	150	200
Accuracy	.73	.79	.75	.78	.75

The best accuracy is for $n = 50$. Accuracy of test set for case $n=50$: 0.794

Please see the notebook at the end of this document.

2) Missing Values

-

c1	c2	c3	c4	c5
5	35	227	374	11

- Please see the notebook at the end of this document.
- Please see the notebook at the end of this document. Accuracy: 0.428
- Please see the notebook at the end of this document. Accuracy: 0.693

3) Imbalanced Data

- Please see the notebook at the end of this document. Accuracy: 0.996. If the classifier simply classifies all the samples as the class of the majority of samples, the accuracy would be high as the distribution of the test set is the same as the training set and still, most of the test samples fall into the same class. The number of misclassified samples in regard to correctly classified samples is negligible, so the accuracy will still remain high.

-

```
Accuracy: 0.9992135052386169
Recall: 0.6439393939393939
Precision: 0.9042553191489362
F1 score: 0.7522123893805309
ConfMat
  0  1
0 71061  9
1  47 85
```

-

```
Accuracy: 0.9379563584180982
Recall: 0.9023495685899778
Precision: 0.9716875491798317
F1 score: 0.9357358320096761
ConfMat
  0  1
0 69125 1871
1  6949 64213
```

-

```
Accuracy: 0.9512195121951219
Recall: 0.8983050847457628
Precision: 1.0
F1 score: 0.9464285714285715
ConfMat
  0  1
0 128  0
1  12 106
```

-

```
Accuracy: 0.9995786635206876
Recall: 0.803030303030303
Precision: 0.9636363636363636
F1 score: 0.8760330578512396
ConfMat
  0  1
0 71066  4
1   26 106
```

Respecting Recall, Precision, and F1 score measures, upsampling and downsampling results in better performances in regard to the application of random forest. However, learning a classifier on this dataset by random forest algorithm also works fine.