

Shivanigowda_Ps1

September 20, 2021

```
[75]: sys.path.append("source")
      sys.path.append("data")
```

```
[77]: import titanic
      import matplotlib.pyplot as plt
      plt.rcParams["figure.figsize"] = (16, 8)
```

```
[78]: from titanic import main
```

```
[95]: main()
```

Classifying using Majority Vote...

-- training error: 0.404

Classifying using Random...

-- training error: 0.485

Classifying using Decision Tree...

-- training error: 0.014

Investigating various classifiers...

Manjority Vote Classifier train error = 0.397 and test error = 0.434

Ramdom Classifier train error = 0.517 and test error = 0.483

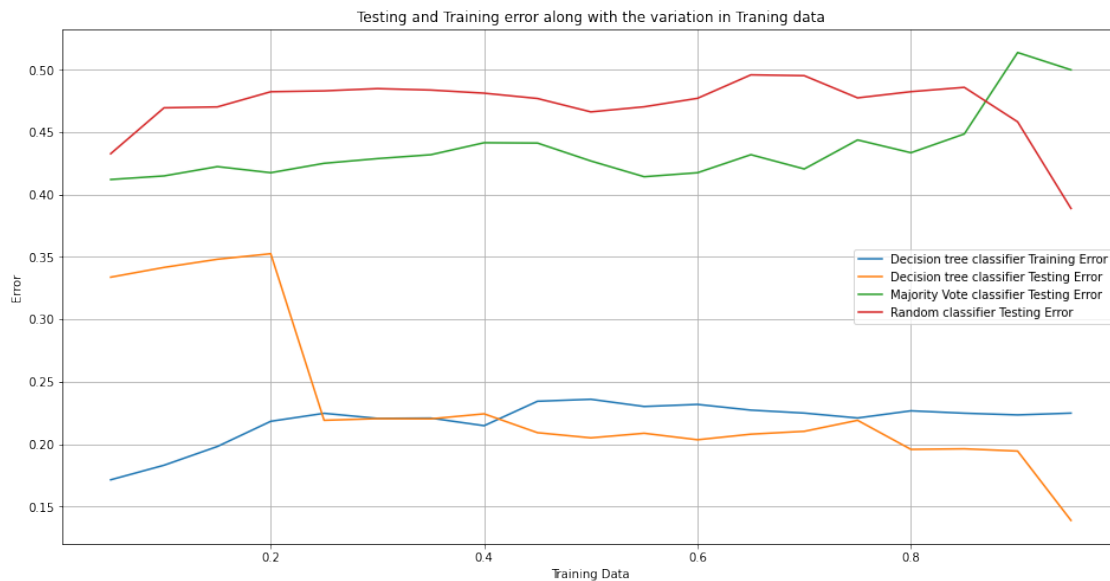
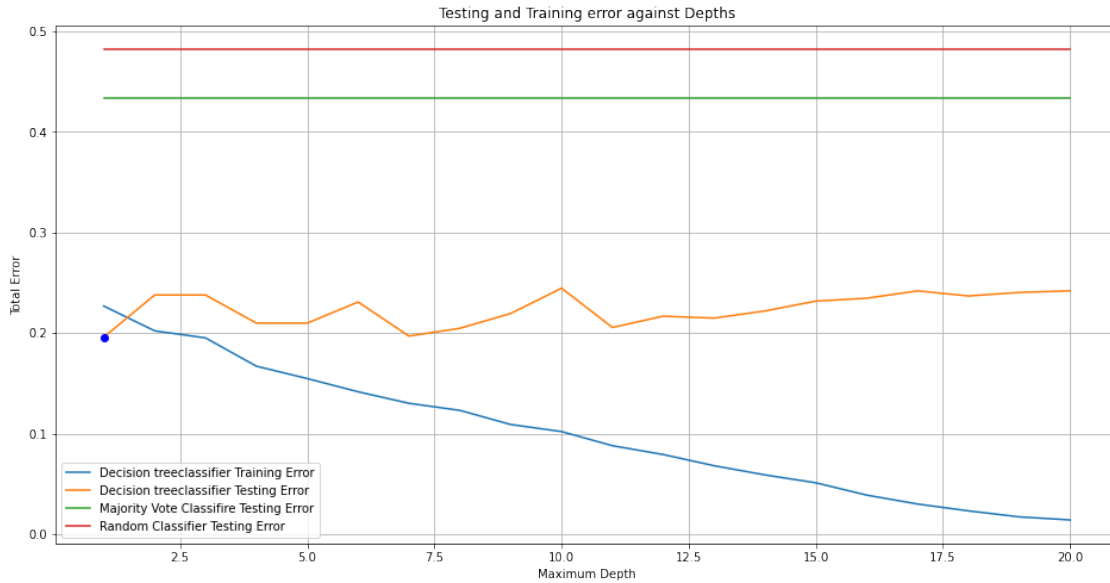
Decision treeclassifier train error = 0.012 and test error = 0.241

Investigating depths...

Investigating training set sizes...

Train error 0.221 and Test error 0.074 with the implementation of 3 Folds

Done



1 Outputs

1. b. Manjority Vote Classifier train error = 0.397 and test error = 0.434 Random Classifier train error = 0.517 and test error = 0.483 Decision tree classifier train error = 0.012 and test error = 0.241
- c. The best depth to use this data is 1. Yes, overfitting is observed. We can observe in plot that after the depth 11, the test error starts increasing and training error starts incuding

because of the overfitting of the training data. Best Maximum depth 1, because the test error at that point is comparatively lower than the rest.

- d. When training data is 20%, we can observe that the test error is high, and the training error for decision tree does not have much variations. The spike and dip in error is observed only after 80% of the training data in Majority vote classifier and Random classifier respectively.

- 2. For the decision tree classifier, the modification made is, the K Fold Cross validations.

[]: