Assignment-1
AI20BTECH11006

Question -3

1) Without any improvement strategy, the program outputs

```
kfold: 1, accuracy: 0.8122
kfold: 2, accuracy: 0.8388
kfold: 3, accuracy: 0.8020
kfold: 4, accuracy: 0.8000
kfold: 5, accuracy: 0.8163
kfold: 6, accuracy: 0.8000
kfold: 7, accuracy: 0.8347
kfold: 8, accuracy: 0.7918
kfold: 9, accuracy: 0.7935
kfold: 10, accuracy: 0.8037
Best accuracy: 0.838776, average accuracy: 0.809306
```

2) With gini index, the program outputs

```
kfold: 1, accuracy: 0.8510
kfold: 2, accuracy: 0.8347
kfold: 3, accuracy: 0.8082
kfold: 4, accuracy: 0.8224
kfold: 5, accuracy: 0.8347
kfold: 6, accuracy: 0.7959
kfold: 7, accuracy: 0.8184
kfold: 8, accuracy: 0.7898
kfold: 9, accuracy: 0.7996
kfold: 10, accuracy: 0.7873
Best accuracy: 0.851020, average accuracy: 0.814201
```

Gini index is computationally more efficient, with regard to accuracy gini index may or may not perform any better than entropy.

3) With pre-prunning

```
kfold: 1, accuracy: 0.8265
kfold: 2, accuracy: 0.8327
kfold: 3, accuracy: 0.8000
kfold: 4, accuracy: 0.8184
kfold: 5, accuracy: 0.8245
kfold: 6, accuracy: 0.8122
kfold: 7, accuracy: 0.8204
kfold: 8, accuracy: 0.7898
kfold: 9, accuracy: 0.8078
kfold: 10, accuracy: 0.8323
Best accuracy: 0.832653, average accuracy: 0.816457
```

Pre-prunning gives better accuracy because it is used to eliminate the noise case. When the node is almost pure, it is likely that the data points which have different label may just be noise, pre-prunning removes that case. In this code, the threshold for pre-prunning was considered 0.95.