Problem 1

1.trying different test\_split sizes (randomness disabled)

25% training accuracy = 0.9591836734693877

Qr code

Description automatically generated

10% training accuracy = 0.9360323886639677

Timeline

Description automatically generated

50% training accuracy = 0.9693877551020408

Timeline

Description automatically generated

75% training accuracy = 0.9883381924198251

A picture containing qr code

Description automatically generated

Conclusion: with increasing the training size the accuracy increased but so did the model size and the tree complexity and the tree became considerably more overfitting

2.a. accuracy for each train size

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 |
| 1 | 0.98231009 | 0.98422 | 0.97959183 | 0.97959 | 0.99029126 |
| 2 | 0.98439125 | 0.98301 | 0.98104956 | 0.98105 | 0.98300970 |
| 3 | 0.98126951 | 0.97937 | 0.98688046 | 0.98688 | 0.98300970 |
| 4 | 0.98126951 | 0.98422 | 0.98542274 | 0.98542 | 0.98786407 |
| 5 | 0.98231009 | 0.98301 | 0.98688046 | 0.98688 | 0.99029126 |
| Mean | 0.98231009 | 0.982767 | 0.982507 | 0.983965 | 0.986893 |
| max | 0.986472 | 0.984223 | 0.98688 | 0.989071 | 0.990291 |
| min | 0.98127 | 0.970874 | 0.979592 | 0.984429 | 0.98301 |

Chart, line chart

Description automatically generated

Number of nodes for each train size

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 |
| 1 | 31 | 37 | 33 | 35 | 51 |
| 2 | 31 | 37 | 33 | 35 | 51 |
| 3 | 31 | 37 | 33 | 35 | 51 |
| 4 | 31 | 37 | 33 | 35 | 51 |
| 5 | 31 | 37 | 33 | 35 | 51 |
| Mean | 31 | 37 | 33 | 35 | 51 |
| max | 31 | 37 | 33 | 35 | 51 |
| min | 31 | 37 | 33 | 35 | 51 |

Chart, line chart

Description automatically generated