**ML Assignment - Decision Trees**

**Sai Anirudh Yellanki(18XJ1A0548)**

**The Code can generate single and bagged decision trees with the accuracy and misclassification, with being able to take in parameters like, max\_depth and no of bagging splits and the code related to plotting the visualization is commented so, it doesn’t generate many dot files and images**

**The code is initialized with bagging set to False, bagging\_splits set to 9 and max\_depth set to 100**

**It will consider max\_depth, 0.001 times the total training set size and 0 entropy for stopping at a terminal node**

**BNA - Banknote Authentication Data set**

**F0 - Feature 0 and likewise with other features (In the graph)**

**Hyper parameters**

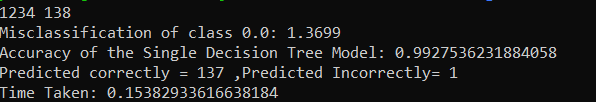
Min no of samples = 0.001 times the dataset size

Entropy = 0 (there is only one class that is predicted at a node)

Depth = 5

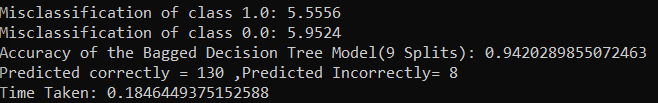
|  |  |
| --- | --- |
| Class | Percent Misclassification |
| 0 | 1.39% |
| 1 | 0% |

**Misclassification, Accuracy and Time taken for Single Decision Tree**

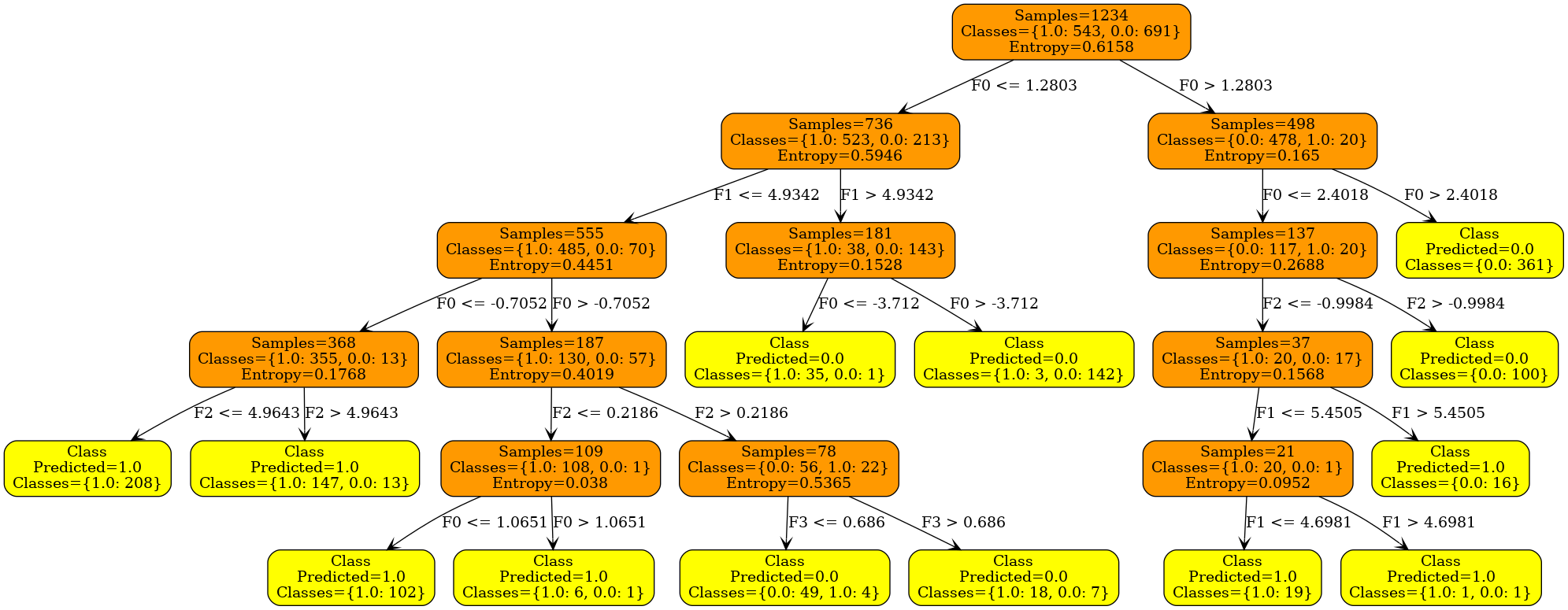


|  |  |
| --- | --- |
| Class | Percent Misclassification |
| 0 | 5.95% |
| 1 | 5.55% |

**Misclassification, Accuracy and Time taken for Bagged Decision Tree (9 Splits)**



**Single Decision Tree generated by the code for the tree with depth = 5**



**SDD - Sensorless Drive Diagnostics Data set**

**Hyper parameters**

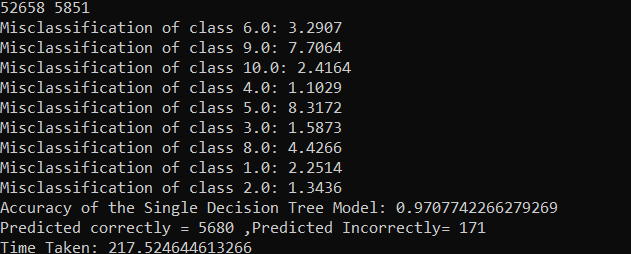
Min no of samples = 0.001 times the dataset size

Entropy = 0 (there is only one class that is predicted at a node)

Depth = 5 and no set depth

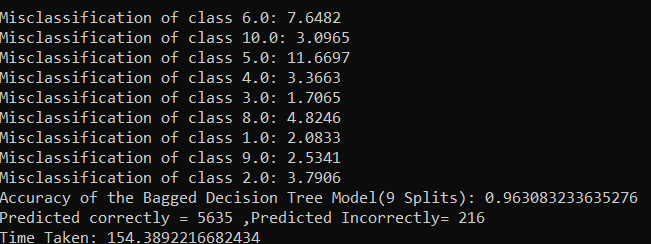
|  |  |
| --- | --- |
| Class | Percent Misclassification |
| 1 | 2.25% |
| 2 | 1.34% |
| 3 | 1.58% |
| 4 | 1.10% |
| 5 | 8.31% |
| 6 | 3.29% |
| 7 | 0.00% |
| 8 | 4.42% |
| 9 | 7.70% |
| 10 | 2.41% |
| 11 | 0.00% |

**Misclassification, Accuracy and Time taken for Single Decision Tree**



|  |  |
| --- | --- |
| Class | Percent Misclassification |
| 1 | 2.08% |
| 2 | 3.79% |
| 3 | 1.70% |
| 4 | 3.36% |
| 5 | 11.66% |
| 6 | 7.64% |
| 7 | 0.00% |
| 8 | 4.82% |
| 9 | 2.53% |
| 10 | 3.09% |
| 11 | 0.00% |

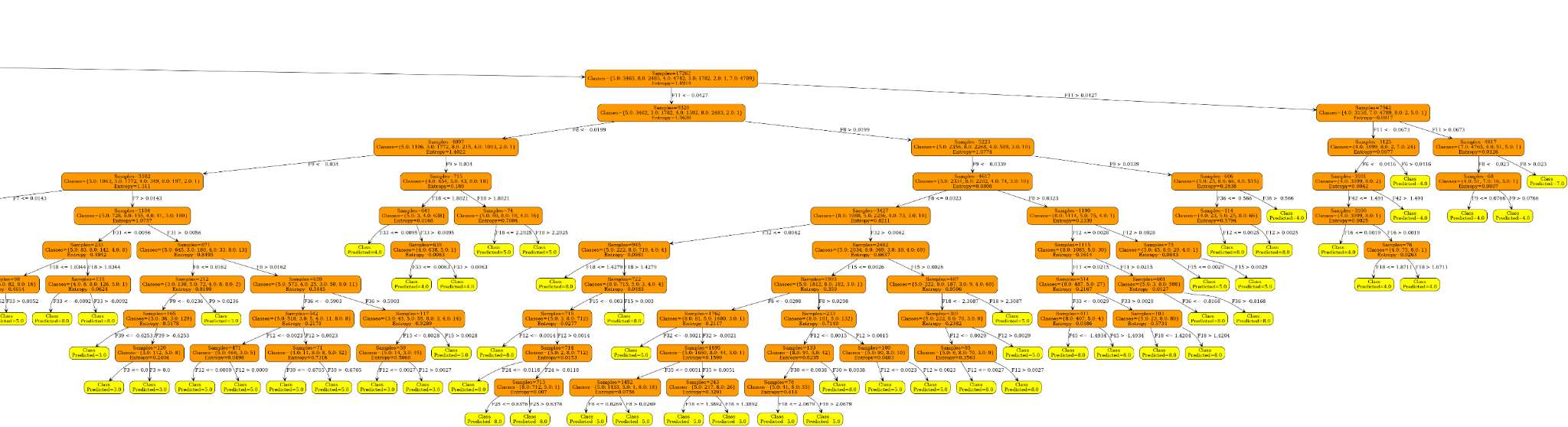
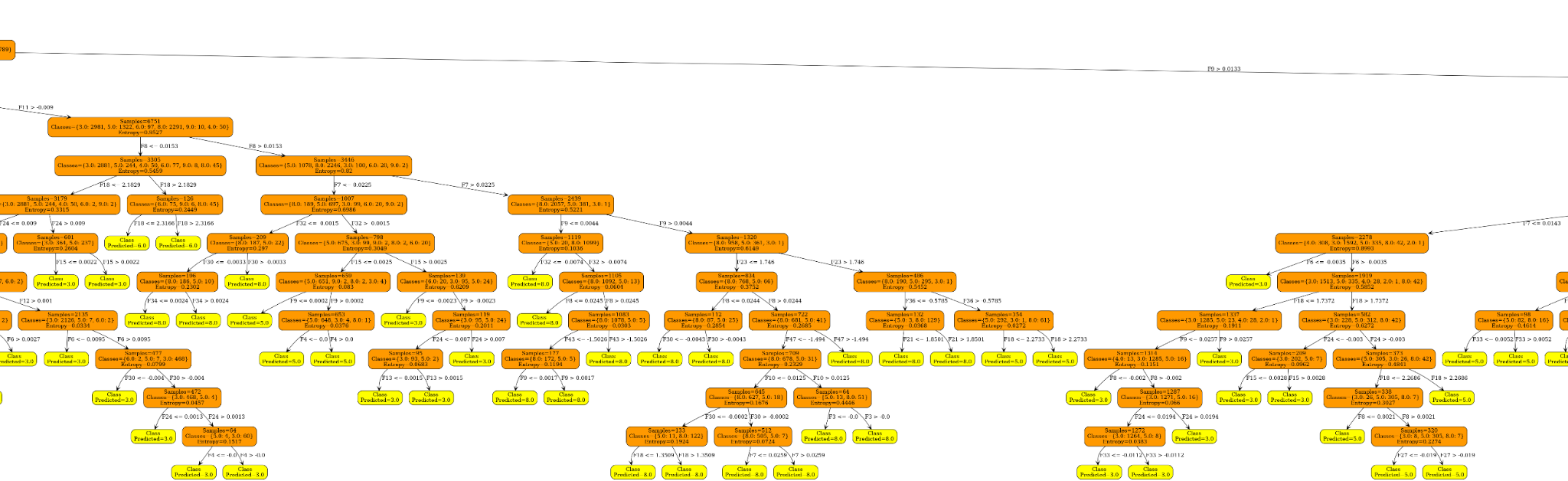
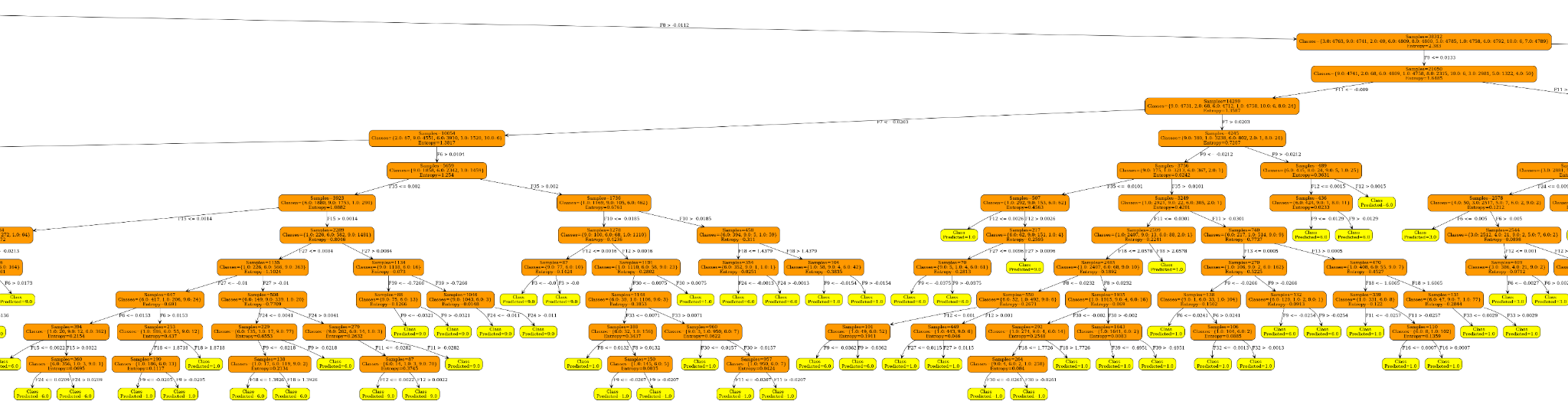
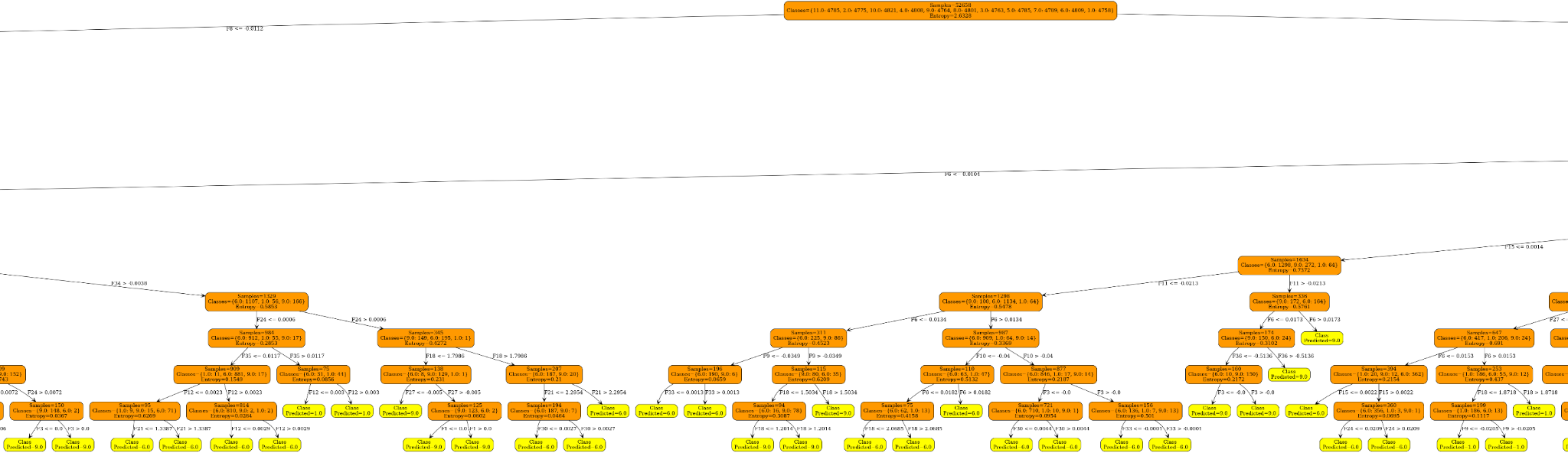
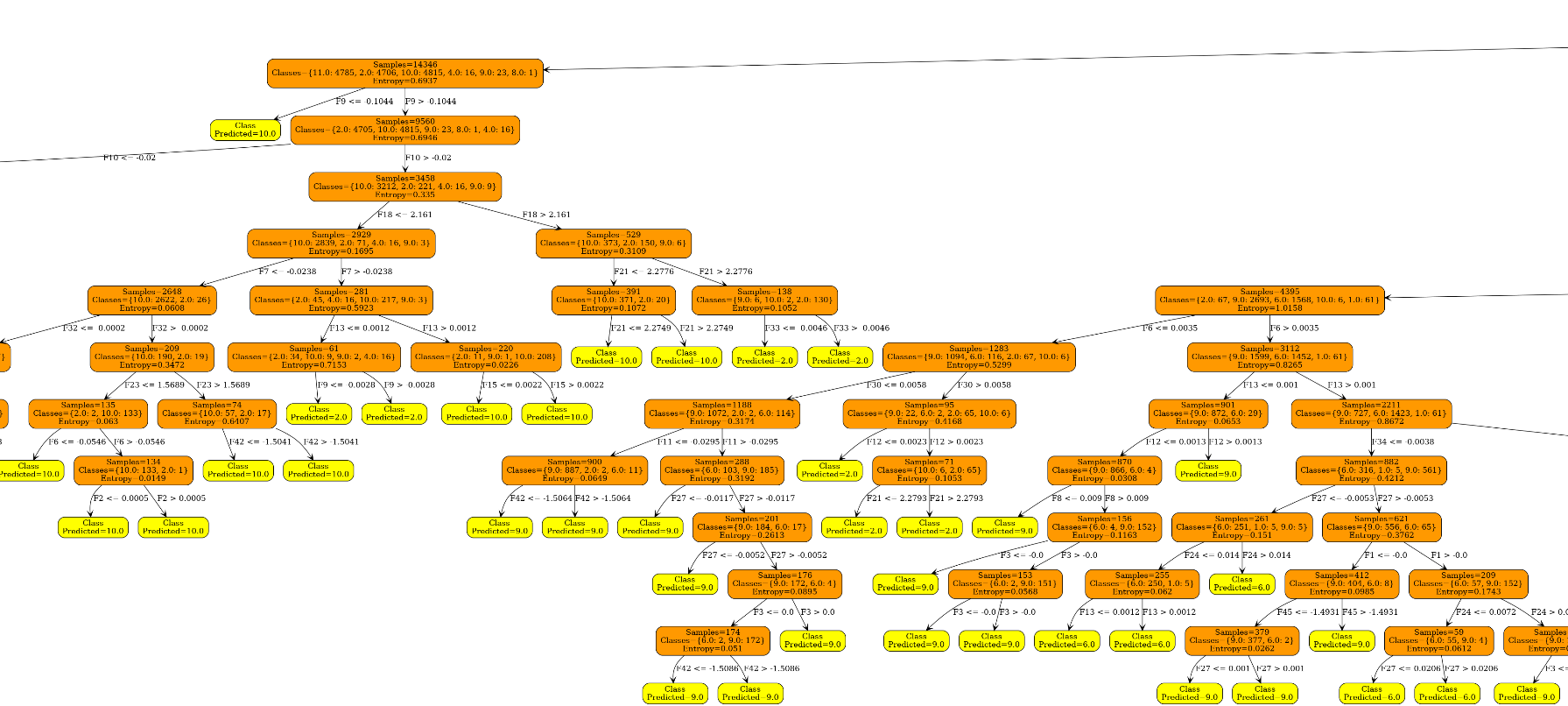
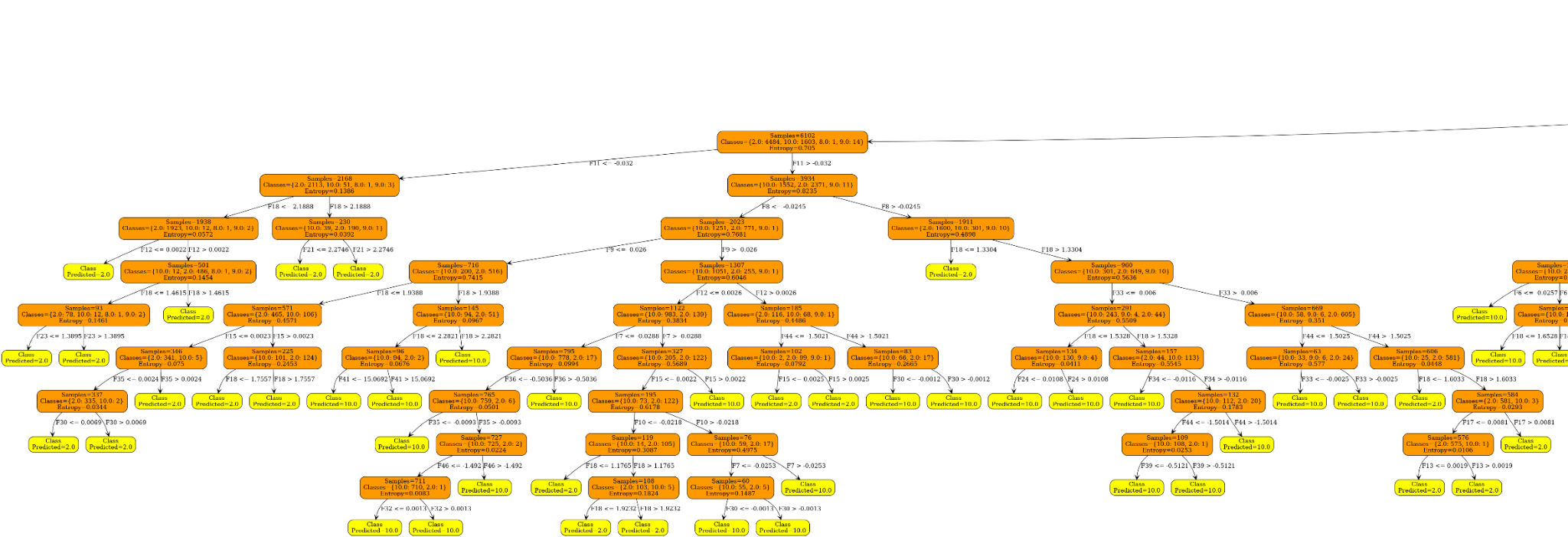
**Misclassification, Accuracy and Time taken for Bagged Decision Tree (9 Splits)**



**Single Decision Tree generated by the code for the tree with depth = 11**

****

**The graph plotted for the Single decision tree for SDD looks this big so I am cropping the image into parts so that it is visible clearly. The cropped images will be in order.**

****

**Zooming into these cropped images will give you an idea on how the decision tree is split for this dataset. It is not very clear but is readable when zoomed in.**