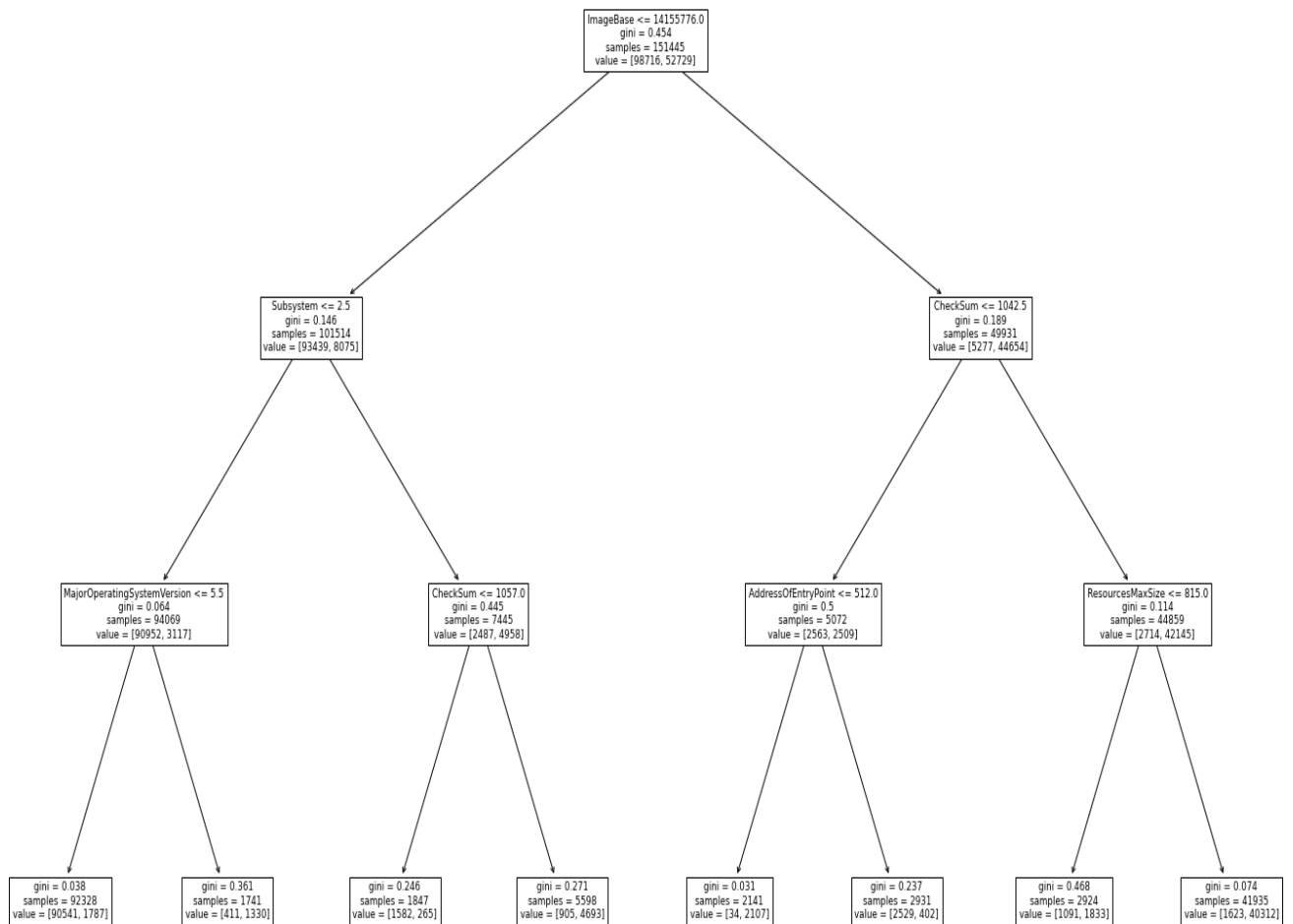


INDEPENDENT STUDY WEEK 2 REPORT

Top 6 feature detection from Decision Tree (Max Depth =3)

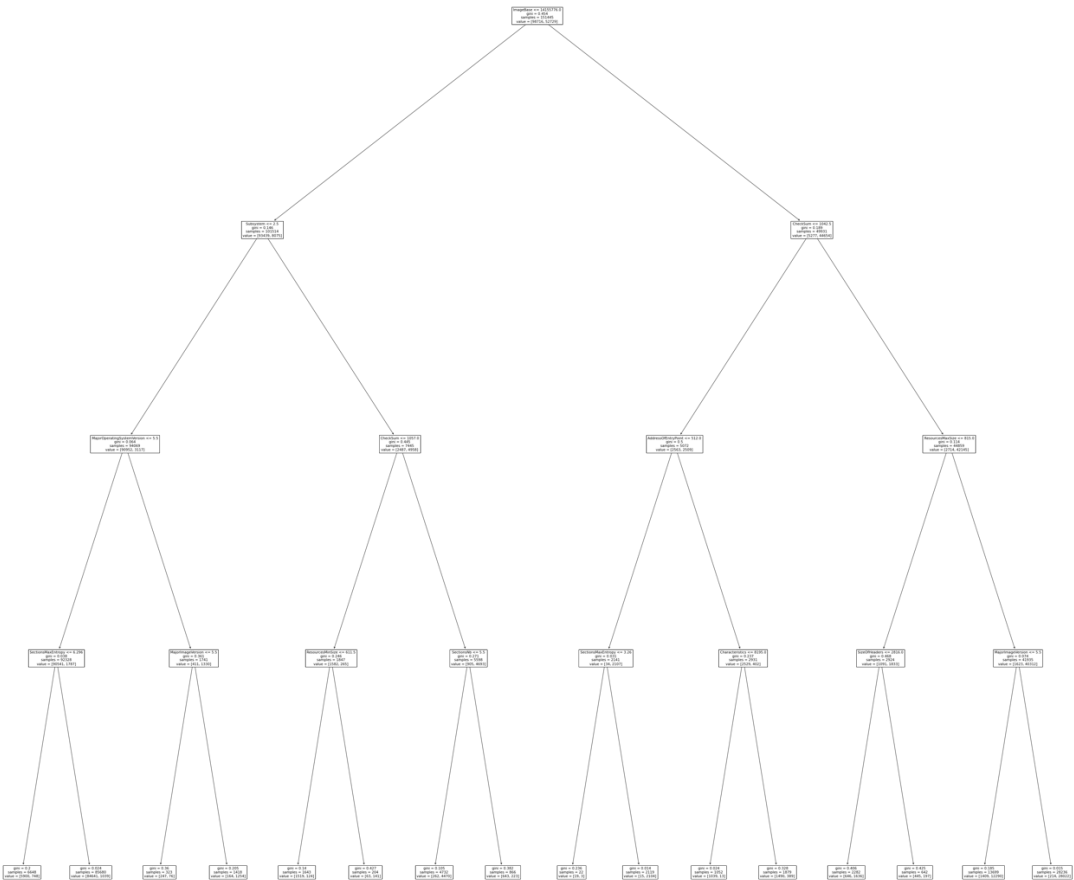
Features	Weightage
ImageBase	0.774
SubSystem	0.096
Checksum	0.054
Major OperatingSystemVersion	0.033
AddressOfEntryPoint	0.030
ResourceMaxSize	0.010

Decision Tree for Malware Detection using Gini Index



Top 10 feature detection from Decision Tree (Max Depth =4)

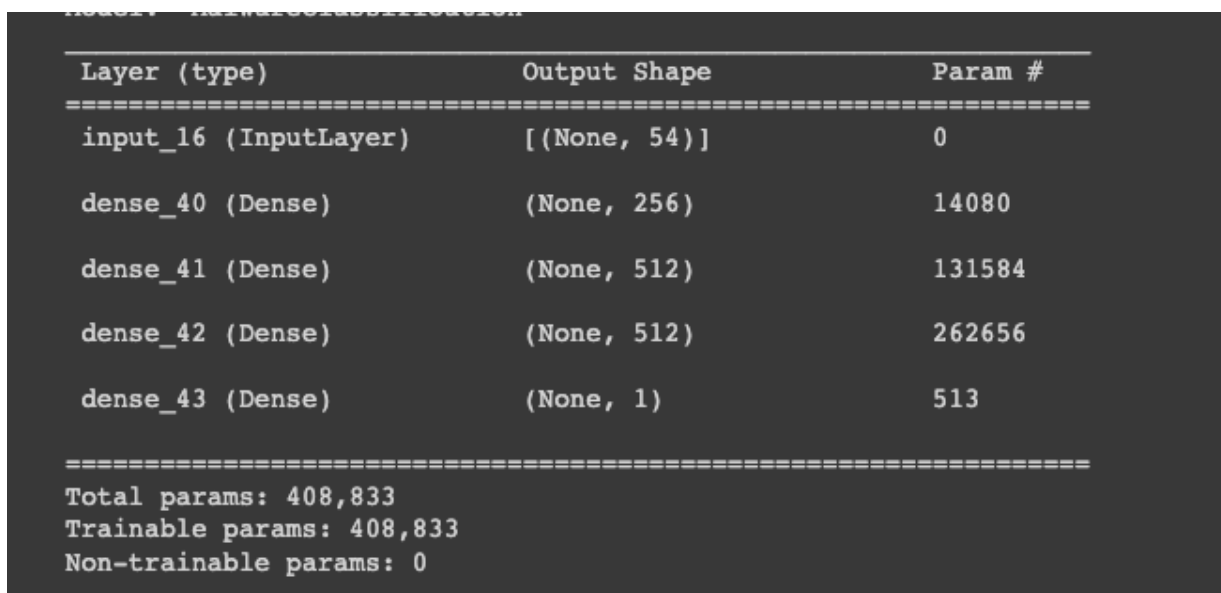
Feature	Weightage
ImageBase	0.753
Subsystem	0.093
Checksum	0.053
MajorOperatingSystemVersion	0.032
AddressOfEntrypoint	0.030
SectionsNb	0.011
ResourceMaxSize	0.010
MajorImageVersion	0.006
SizeOfHeaders	0.002



Note : Please refer to DecisionTreeDepth4.png image attached here.

Neural Network Model for Malware Classification

1) Model 1 (Complex Model)



```
Model: "MalwareClassification"
Layer (type)                 Output Shape              Param #
=====
input_16 (InputLayer)        [(None, 54)]              0
dense_40 (Dense)              (None, 256)               14080
dense_41 (Dense)              (None, 512)               131584
dense_42 (Dense)              (None, 512)               262656
dense_43 (Dense)              (None, 1)                 513
=====
Total params: 408,833
Trainable params: 408,833
Non-trainable params: 0
```

Layer (type)	Output Shape	Param #
input_16 (InputLayer)	[(None, 54)]	0
dense_40 (Dense)	(None, 256)	14080
dense_41 (Dense)	(None, 512)	131584
dense_42 (Dense)	(None, 512)	262656
dense_43 (Dense)	(None, 1)	513

=====
Total params: 408,833
Trainable params: 408,833
Non-trainable params: 0

Optimizer Used: Adam Optimizer

Loss function used: Binary Cross entropy

Model: $54 \times 256 \times 512 \times 512 \times 1$

Training Time: 2minutes 10 seconds

Total trainable parameter: 408,833

Accuracy on test data set: 92.2%

2) Model 2 (Simple Model)

Layer (type)	Output Shape	Param #
input_17 (InputLayer)	[(None, 54)]	0
dense_44 (Dense)	(None, 128)	7040
dense_45 (Dense)	(None, 128)	16512
dense_46 (Dense)	(None, 1)	129
=====		
Total params: 23,681		
Trainable params: 23,681		
Non-trainable params: 0		

Optimizer Used: Adam Optimizer
Loss function used: Binary Cross entropy
Model: $54 \times 128 \times 128 \times 1$
Training Time: 44 seconds
Total trainable parameter: 23,681

Accuracy on test data set: 90.28%

Now, we again run the model with different hyperparameters and test the accuracy.

1) Mini batch Gradient Descent

Model	Learning Rate (α)	Num of Epochs	Num of batches	Test Accuracy
$54 \times 256 \times 512 \times 512 \times 1$	1e-4	10	128	92.73
$54 \times 256 \times 512 \times 512 \times 1$	1e-3	10	128	92.23
$54 \times 256 \times 512 \times 512 \times 1$	1e-2	10	128	83.4%
$54 \times 256 \times 512 \times 512 \times 1$	1e-4	20	128	91.84%
$54 \times 128 \times 128 \times 1$	1e-4	20	64	89.88%
$54 \times 128 \times 128 \times 1$	1e-3	20	64	91.07

2) Stochastic Gradient Descent

Model	Learning Rate (α)	Num of Epochs	Num of batches	Test Accuracy
$54 \times 128 \times 128 \times 1$	1e-3	2	1	84.3%
$54 \times 128 \times 128 \times 1$	1e-4	2	1	86.2%
$54 \times 128 \times 128 \times 1$	1e-2	2	1	83.2%