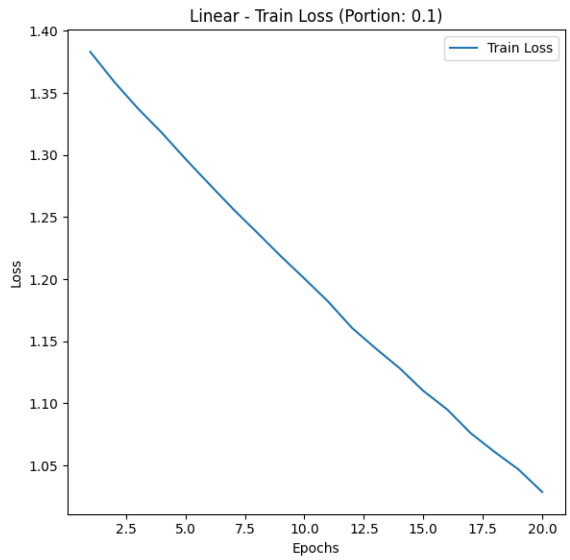
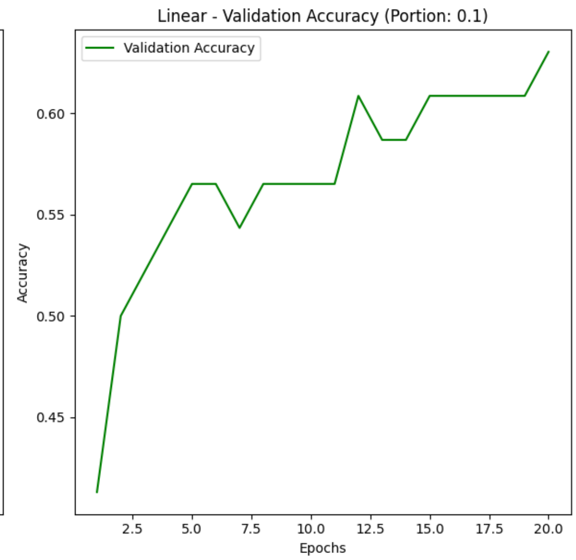
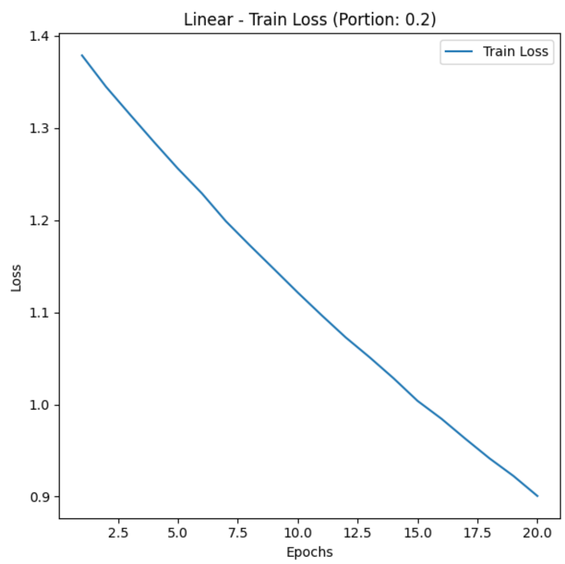
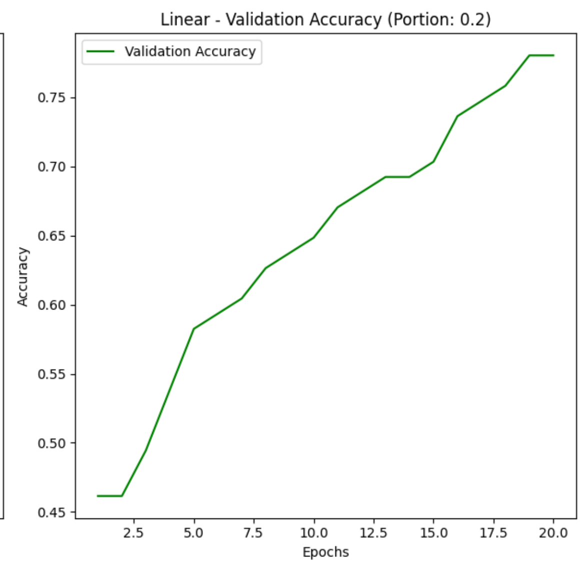
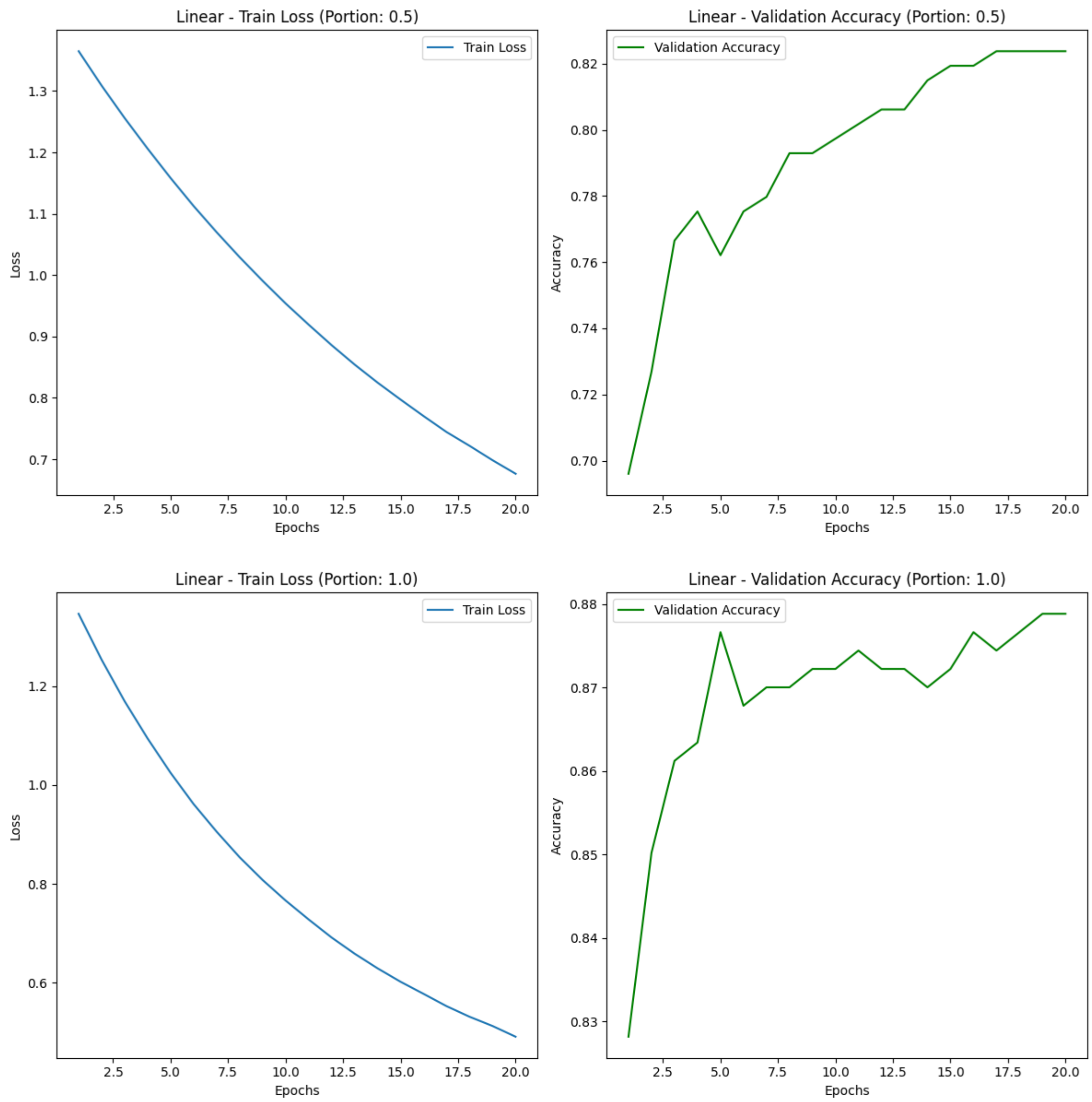
# Question 1


Linear - Train Loss (Portion: 0.1)
Train Loss
Loss
Epochs
1.40
1.35
1.30
1.25
1.20
1.15
1.10
1.05
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0


Linear - Validation Accuracy (Portion: 0.1)
Validation Accuracy
Accuracy
Epochs
0.60
0.55
0.50
0.45
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0

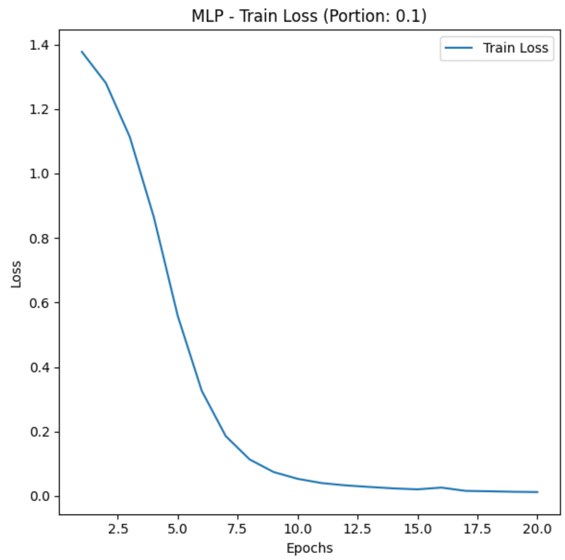

Linear - Train Loss (Portion: 0.2)
Train Loss
Loss
Epochs
1.4
1.3
1.2
1.1
1.0
0.9
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0

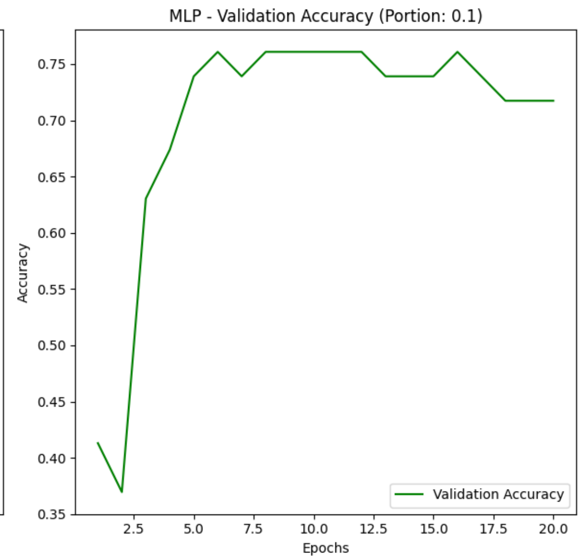

Linear - Validation Accuracy (Portion: 0.2)
Validation Accuracy
Accuracy
Epochs
0.75
0.70
0.65
0.60
0.55
0.50
0.45
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0

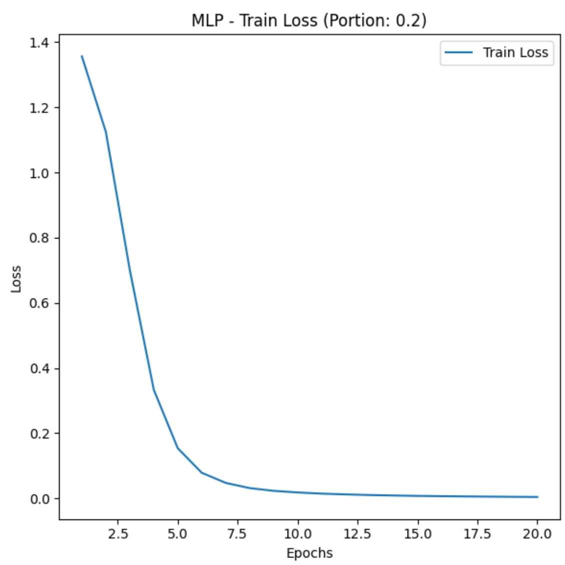


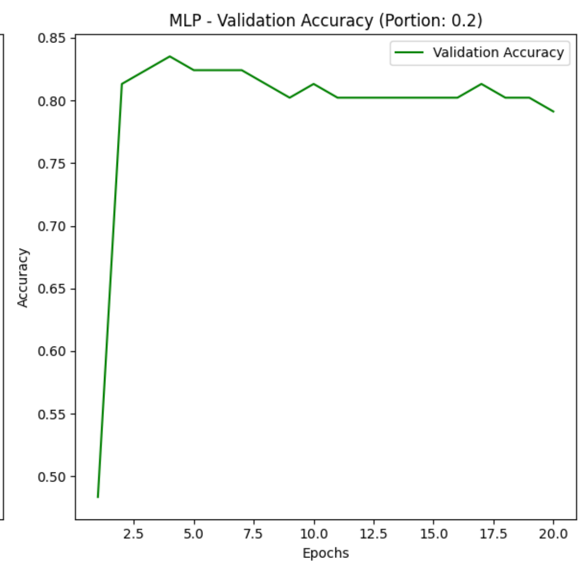
Linear - Train Loss (Portion: 0.5)
Train Loss
Loss
Epochs
Linear - Validation Accuracy (Portion: 0.5)
Validation Accuracy
Accuracy
Epochs
Linear - Train Loss (Portion: 1.0)
Train Loss
Loss
Epochs
Linear - Validation Accuracy (Portion: 1.0)
Validation Accuracy
Accuracy
Epochs

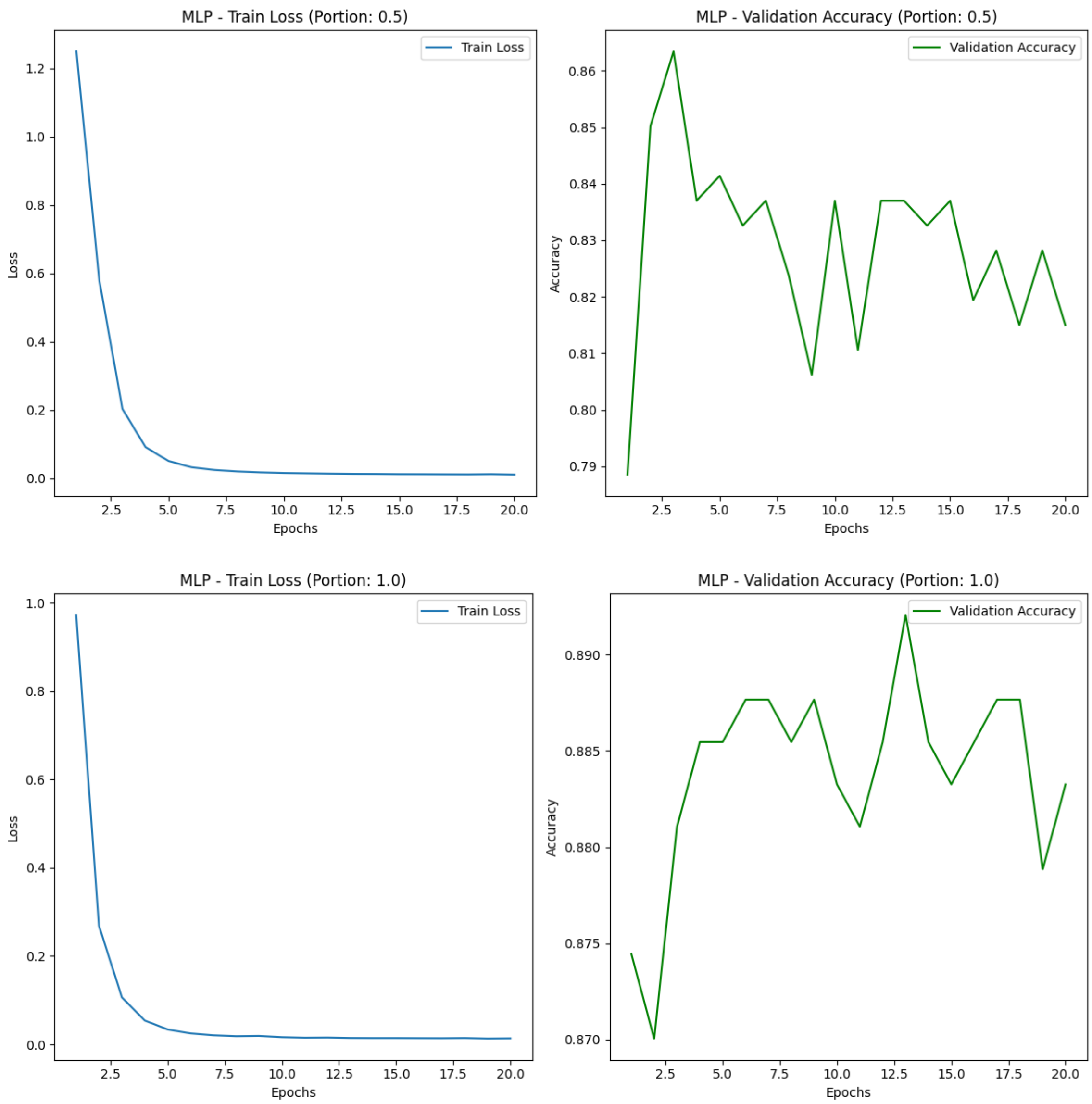
# Question 2


MLP - Train Loss (Portion: 0.1)
Train Loss
Loss
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0
Epochs

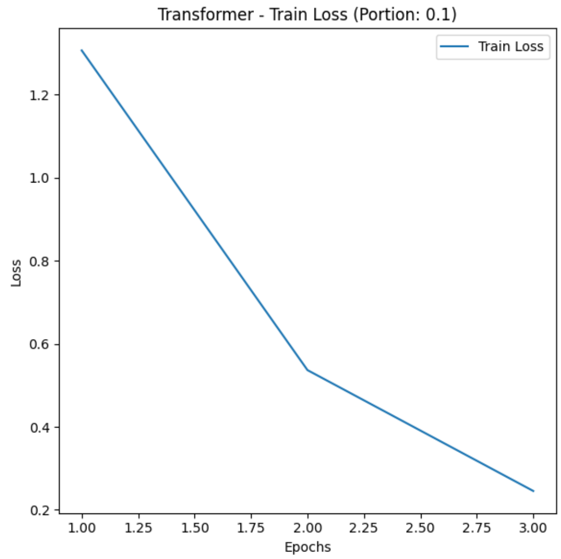

MLP - Validation Accuracy (Portion: 0.1)
Validation Accuracy
Accuracy
0.35
0.40
0.45
0.50
0.55
0.60
0.65
0.70
0.75
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0
Epochs

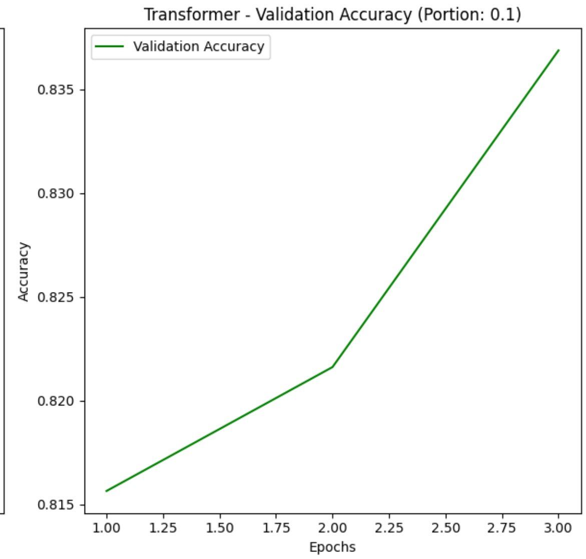

MLP - Train Loss (Portion: 0.2)
Train Loss
Loss
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0
Epochs

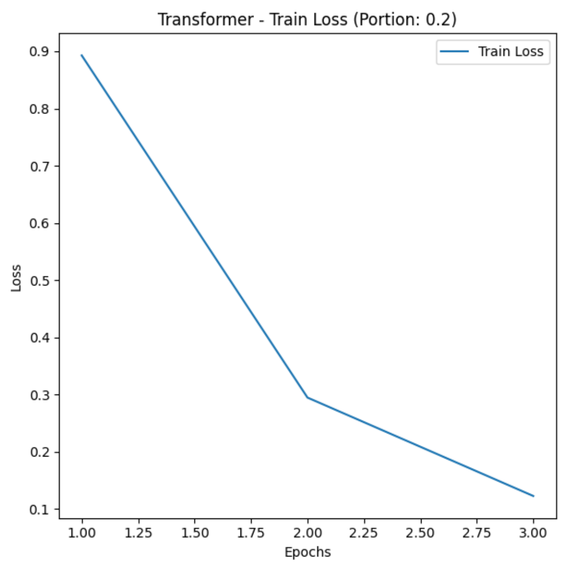

MLP - Validation Accuracy (Portion: 0.2)
Validation Accuracy
Accuracy
0.50
0.55
0.60
0.65
0.70
0.75
0.80
0.85
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0
Epochs

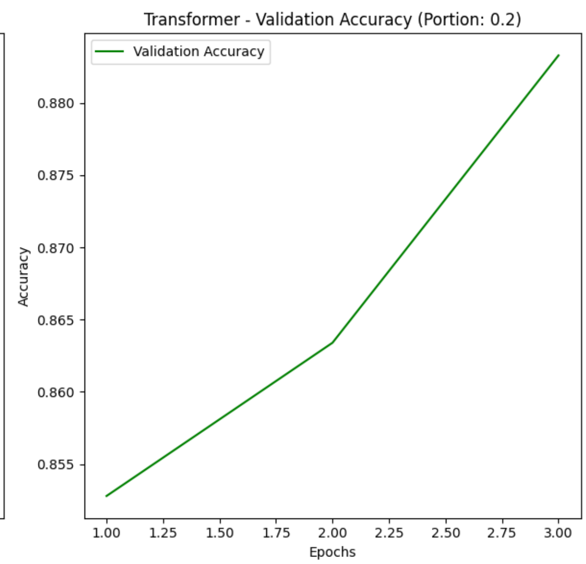

MLP - Train Loss (Portion: 0.5)
Train Loss
Loss
Epochs
MLP - Validation Accuracy (Portion: 0.5)
Validation Accuracy
Accuracy
Epochs
MLP - Train Loss (Portion: 1.0)
Train Loss
Loss
Epochs
MLP - Validation Accuracy (Portion: 1.0)
Validation Accuracy
Accuracy
Epochs

# Question 3


Transformer - Train Loss (Portion: 0.1)
Train Loss
Loss
Epochs
1.2
1.0
0.8
0.6
0.4
0.2
1.00
1.25
1.50
1.75
2.00
2.25
2.50
2.75
3.00


Transformer - Validation Accuracy (Portion: 0.1)
Validation Accuracy
Accuracy
Epochs
0.835
0.830
0.825
0.820
0.815
1.00
1.25
1.50
1.75
2.00
2.25
2.50
2.75
3.00


Transformer - Train Loss (Portion: 0.2)
Train Loss
Loss
Epochs
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
1.00
1.25
1.50
1.75
2.00
2.25
2.50
2.75
3.00


Transformer - Validation Accuracy (Portion: 0.2)
Validation Accuracy
Accuracy
Epochs
0.880
0.875
0.870
0.865
0.860
0.855
1.00
1.25
1.50
1.75
2.00
2.25
2.50
2.75
3.00

# Question 4

(a) Of all the models trained on 10%-20% of the data, the transformer with 20% had the highest accuracy with >88%. Of all models trained, regardless of percentage of training data, MLP had a brief >89% peak while training on all data at epoch 12

(b) Both in terms of loss and accuracy, the Log-linear classifier had the biggest changes when the size of the training set was changed.

(c) The number of trainable parameters for each model is as follows:

- **Log-linear model**: 8,004
- **MLP model**: 1,002,504
- **Transformer model (distilroberta-base)**: 82,121,476

  The Log-linear model, while simple and efficient, was the most sensitive to changes in the size of the training set. However, this simplicity also restricts its performance compared to the other models.

  The MLP model introduces significantly more parameters through its hidden layer, which improves its capacity to model non-linear relationships in the data, leading to better performance as the training set grows.

  The Transformer model, with the highest number of parameters, leverages pretraining and a sophisticated architecture to excel, particularly on larger datasets. However, it is also computationally expensive and more prone to overfitting on smaller datasets.

  In general, additional parameters help when they enhance the model's ability to capture meaningful information. Simply increasing the size of a model (e.g., adding parameters to a linear model) does not guarantee better predictions unless the architecture and data support the increased complexity.