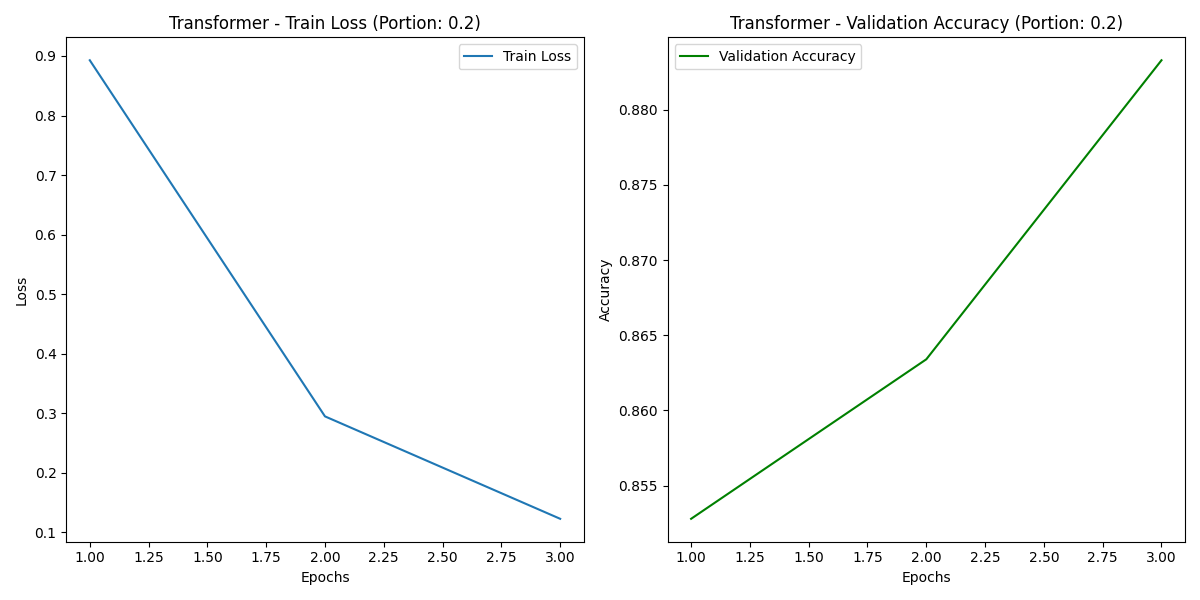
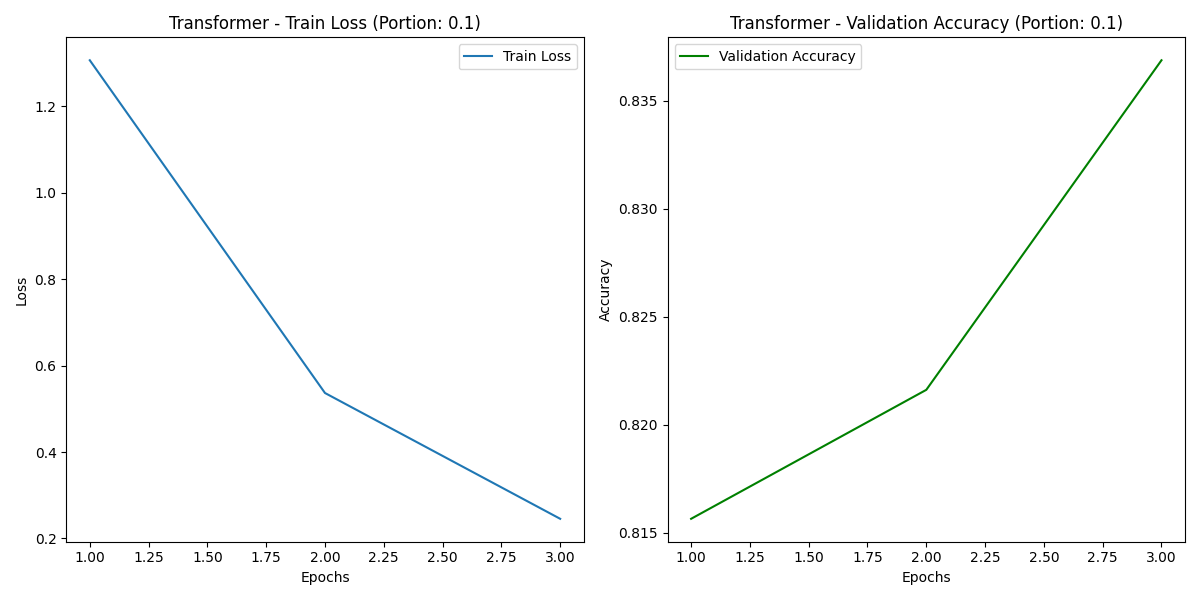
# Question 1A screenshot of a graph Description automatically generated

# Question 2A screenshot of a graph Description automatically generatedA screenshot of a graph Description automatically generatedA screenshot of a graph Description automatically generated

# Question 3



# Question 4

1. 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
2. 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.