Problem:
Implement a deep neural network for relation extraction. The model should achieve a minimum 75% F-score or higher.

Dataset:
We use the NYT29 dataset which contains 29 relation types. The dataset is split into training (63,306 sentences with 78,973 relation tuples), development (7,033 sentences), and test (4,066 sentences and 5,859 relation tuples) sets.

Model Architecture:
The classifier model uses a pre-trained BERT model from huggingface library for encoding the sentences. A dropout layer is added to avoid overfitting followed by a linear output layer. The linear layer outputs a probability distribution over 30 classes (29 relations from the dataset and 'other' for no relation). We consider the class with the higher probability as the prediction of the model. To train the model we use cross entropy loss and Adam optimizer.

Hyperparameters:
We use the following values for the hyperparameters during training:

Adam Learning Rate	2e-5
Training batch size	16
Training epochs	4
dropout	0.3

Performance:
The trained model is evaluated on the train, test, and dev sets. We measure the micro-average F-score of the model on these sets. The results are reported below.

Dataset	**F-score**
train	0.997
test	0.919
dev	0.987

The output of the python notebook is saved as a hw2_906466769.pdf