# Report

Naimeesh Narayan Tiwari

2020101074

# Scores

1. **After training for 10 iterations:**
    a. **Scores for Train Set**

        Training Loss: 0.05500851929187775
        Dev Set Loss: 0.05423680506646633
        Accuracy: 0.9685430463576159
        Recall: 0.7956121295452249
        Precision: 0.8370360184519238
        F1 Score: 0.813210046413094

    b. **Scores for Test Set**

        Accuracy: 0.9729969604863222
        Recall: 0.8578627245705487
        Precision: 0.8989357760563247
        F1 score: 0.8741218270148622

2. **After training for 40 iterations:**
    a. **Scores for Train Set**

        Training Loss: 0.05220522746443748
        Dev Set Loss: 0.052186185494065285
        Accuracy: 0.9685430463576159
        Recall: 0.7956121295452249
        Precision: 0.8370360184519238
        F1 Score: 0.813210046413094

b. **Scores for Test Set**

Accuracy: 0.977
Recall: 0.889
Precision: 0.899
F1 score: 0.893

# Hyperparameters

1. embedding size
   - Originally taken 200
   - on reducing to 20 → accuracy decreases to 83%
   - on making it to 2000 → accuracy remains ~ 97% (overfitting occurs at this range)
2. hidden size
   - originally taken 200
   - drops to 94 on decreasing hidden size by 10
3. learning rate
   - currently 0.5
   - accuracy shoots down to 30% in lr = 0.001
   - accuracy at 60% for lr = 0.01
4. dropout rate - not much change
5. number of epochs ( mentioned in point 1 - SCORES)
6. batch size
   - currently set as 100
   - accuracy drops to 90% for batch size = 10
7. bidirectional nature
   - on making a unidirectional GRU → accuracy drops
8. complexity of the network (talked in analysis)

# Analysis

1. The scores for NOUN POS tag are low. This is because NOUNs don’t often get repeated and instead we see newer NOUNs in probably every new sentence - making it hard to predict.
2. High complexity leads to overfitting and hence lesser accuracy, hence hyperparameters were chosen so that everything remains balanced
3. The scores look good since the dataset is not that large