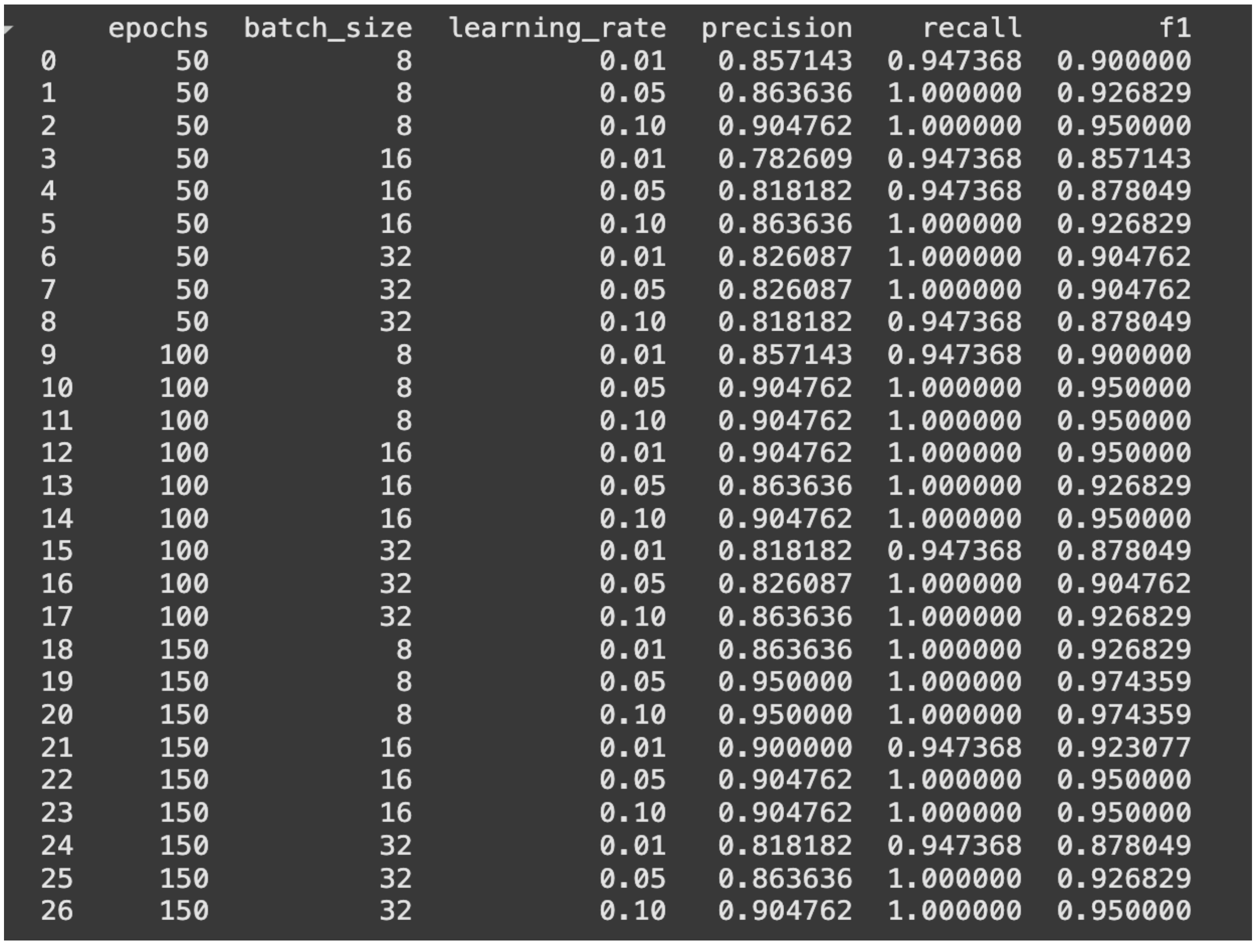
NLP Assignment 2:
Colab Link:
https://colab.research.google.com/drive/1AQajfMWI4p6WonZ3RdrK0NYz3m6sIjhj?usp=sharing

1. Describe the details for your best performing model: parameters such as number of epochs, batch size, and learning rate. Compare and contrast with other parameter settings you used—feel free to use graphs or tables to compare. Why do you think it performed better than the other models?

Solution:
Did hyperparameter tuning for different values of number of epochs, batch size and learning rate.

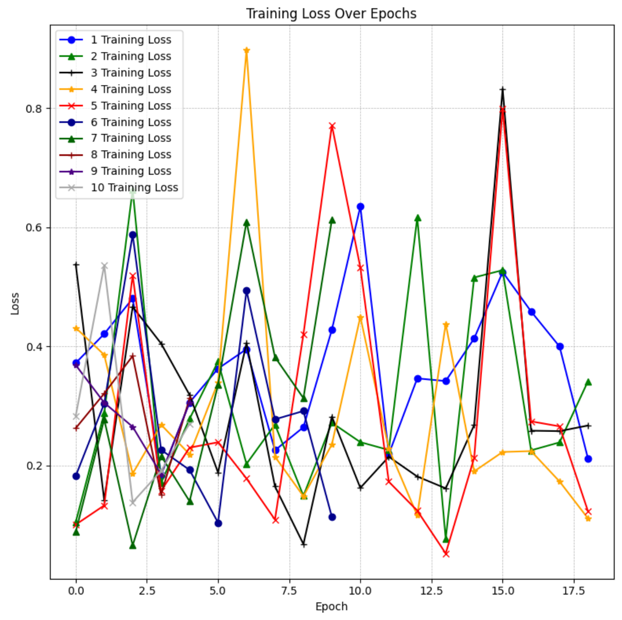
Results:

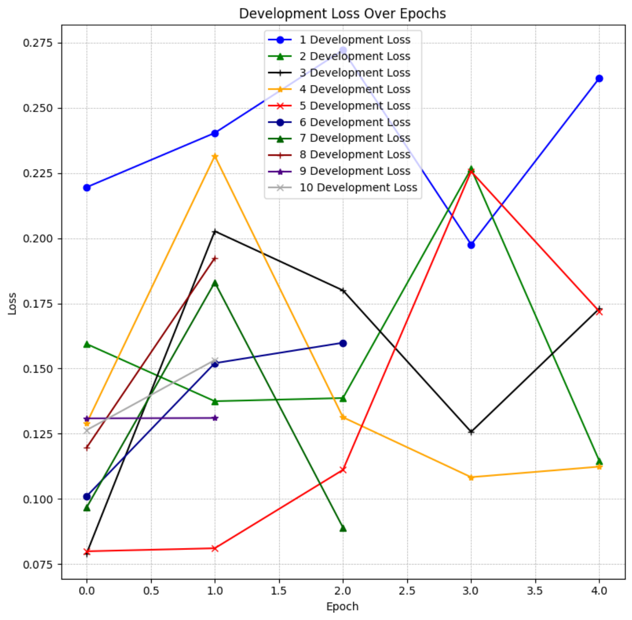


```
    epochs  batch_size  learning_rate  precision    recall        f1
0       50           8           0.01   0.857143  0.947368  0.900000
1       50           8           0.05   0.863636  1.000000  0.926829
2       50           8           0.10   0.904762  1.000000  0.950000
3       50          16           0.01   0.782609  0.947368  0.857143
4       50          16           0.05   0.818182  0.947368  0.878049
5       50          16           0.10   0.863636  1.000000  0.926829
6       50          32           0.01   0.826087  1.000000  0.904762
7       50          32           0.05   0.826087  1.000000  0.904762
8       50          32           0.10   0.818182  0.947368  0.878049
9      100           8           0.01   0.857143  0.947368  0.900000
10     100           8           0.05   0.904762  1.000000  0.950000
11     100           8           0.10   0.904762  1.000000  0.950000
12     100          16           0.01   0.904762  1.000000  0.950000
13     100          16           0.05   0.863636  1.000000  0.926829
14     100          16           0.10   0.904762  1.000000  0.950000
15     100          32           0.01   0.818182  0.947368  0.878049
16     100          32           0.05   0.826087  1.000000  0.904762
17     100          32           0.10   0.863636  1.000000  0.926829
18     150           8           0.01   0.863636  1.000000  0.926829
19     150           8           0.05   0.950000  1.000000  0.974359
20     150           8           0.10   0.950000  1.000000  0.974359
21     150          16           0.01   0.900000  0.947368  0.923077
22     150          16           0.05   0.904762  1.000000  0.950000
23     150          16           0.10   0.904762  1.000000  0.950000
24     150          32           0.01   0.818182  0.947368  0.878049
25     150          32           0.05   0.863636  1.000000  0.926829
26     150          32           0.10   0.904762  1.000000  0.950000
```

Best F1 Score Row:

epochs	150.000000
batch_size	8.000000
learning_rate	0.050000

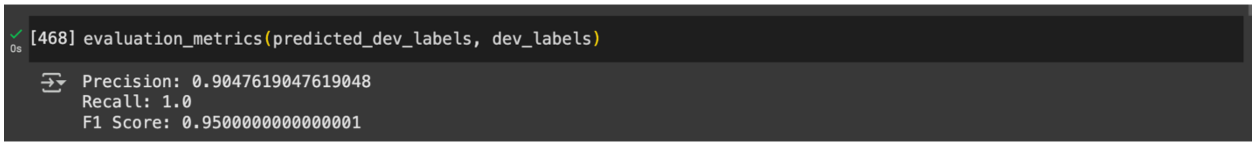




Comparison of top 5 and bottom 5 training and dev loss values.
The blue denotes the best one.
Why it performed well?

- The learning rate is not too high nor too low which helps the model to converge efficiently. I got the same result with learning rate of 0.1 as well. So I think 0.05-0.1 is a good range
- Smaller batches were performing well as compared to large
- The epoch size of 100 was good as well and 150 gave a better result

Note:
Evaluation metrics with normalization and all 6 feature vectors:

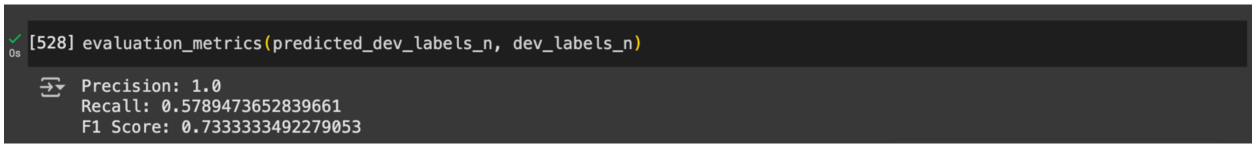

```
[468] evaluation_metrics(predicted_dev_labels, dev_labels)

Precision: 0.9047619047619048
Recall: 1.0
F1 Score: 0.9500000000000001
```

2. What do you think would happen if you didn’t normalize the feature vectors? Write down a guess for what you think would happen, and then run an experiment to test your intuitions and report back what you learned.

Solution:

- I think there should be some degrade compared to the model that used normalization.
- Normalization helps in making data to be in the same scale. For example, if one feature is of a higher scale, then it will dominate the model and may cause bias.
- The “without normalization” evaluation metrics show that there was a decrease in the F1 score. Which means our intuition was correct.

F1 score without normalization:



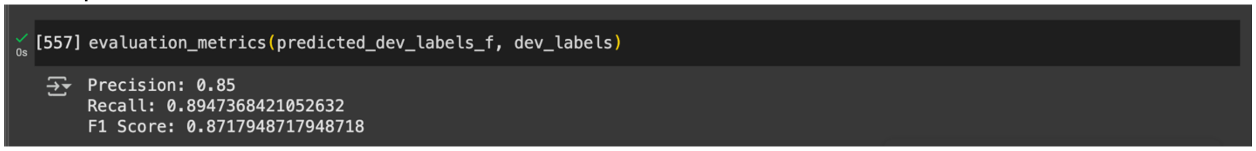
```
[528] evaluation_metrics(predicted_dev_labels_n, dev_labels_n)

      Precision: 1.0
      Recall: 0.5789473652839661
      F1 Score: 0.7333333492279053
```

3. What would happen if you removed one of the features entirely and used a 5-dimensional feature vector? Choose one feature and remove it from your vector. Then, run another experiment and see what happens. Does the test F1 go up or down? Does the model converge slower, or faster? Report which feature you removed and what you learned.

Solution:

- I think it depends on the importance of the feature vector that we are removing. If the feature vector is of more importance, it would affect the model significantly.
- The F1 score went down when a feature was removed.
- The model converges slower than before. It means the model is struggling to learn efficiently.
- I removed the negative words feature from my vector. The recall shows a decrease, as the model's seems to be confused about the positive instances. Important features like negative words help the model to learn more efficiently, giving us better results.

As expected, there was a decrease in evaluation metrics.



```
[557] evaluation_metrics(predicted_dev_labels_f, dev_labels)

      Precision: 0.85
      Recall: 0.8947368421052632
      F1 Score: 0.8717948717948718
```

4. Review Section 4.10, p. 18 of the textbook and then consider the resources we used for this task: for instance, the training data and the positive and negative lexicons. Did you notice any biases present in these resources? Can you think of any harms or unintended consequences (harmful or not) that this classifier could cause? There is no correct answer; just write a couple of sentences reflecting on this prompt.

Solution:

- If we train data on a specific training set, it may have biased data for some data that has no relation to the training set. Outliers would be the most affected in this scenario and we may observe total opposite results in that case.
- For sentiment analysis, positive and negative lexicons may have different interpretations in different settings or different cultures.
- If a decision-making software uses such a biased classifier for important decisions like healthcare or career decisions, it may cause harm to one's health or future life.