Report [Group 36: Assignment 2]

- ## Snippets of Data Samples

1. Task-1 Dataset (Named Entity Recognition)


```
"4bbb0629e66146edaf4ac7bde47062fb": {
    "text": "The factum of accident, allegation of rash and negligent driving causing death of Sukendra Pal Singh were denied.",
    "labels": [
        "O",
        "O",
        "O",
        "O",
        "O",
        "O",
        "O",
        "O",
        "O",
        "O",
        "O",
        "O",
        "O",
        "B_OTHER_PERSON",
        "I_OTHER_PERSON",
        "I_OTHER_PERSON",
        "O",
        "O"
    ]
},
```

```
"14359a64c9dd42a0b9394ee14348b31f": {
    "text": "Thus offence u/s 307 IPC is not made out in respect of injuries caused to ASI Giriraj Singh.",
    "labels": [
        "O",
        "O",
        "B_PROVISION",
        "I_PROVISION",
        "B_STATUTE",
        "O",
        "O",
        "O",
        "O",
        "O",
        "O",
        "O",
        "O",
        "O",
        "O",
        "O",
        "B_OTHER_PERSON",
        "I_OTHER_PERSON"
    ]
},
```


Data preparation

- Given the guidelines and the sample BIO labeled data in Part 1 A, we tokenized on space and ignored all other whitespace characters.

- Punctuation was left unremoved during data preparation since it would affect the spans of the labeled data and we did not want to modify that.
- During model training, we clean up words by removing punctuation, symbols and other whitespace characters to ensure we can get valid word-embeddings for most of the tokens.
- Care has to be taken to ensure the number of tokens matches the labels.

2. Task-2 Dataset (Aspect Term Extraction)


```
"0": {
    "text": "I charge it at night and skip taking the cord with me because of the good battery life .",
    "labels": [
        "O",
        "O",
        "O",
        "O",
        "O",
        "O",
        "O",
        "O",
        "O",
        "O",
        "O",
        "O",
        "O",
        "O",
        "O",
        "O",
        "B",
        "I",
        "O"
    ]
},
```

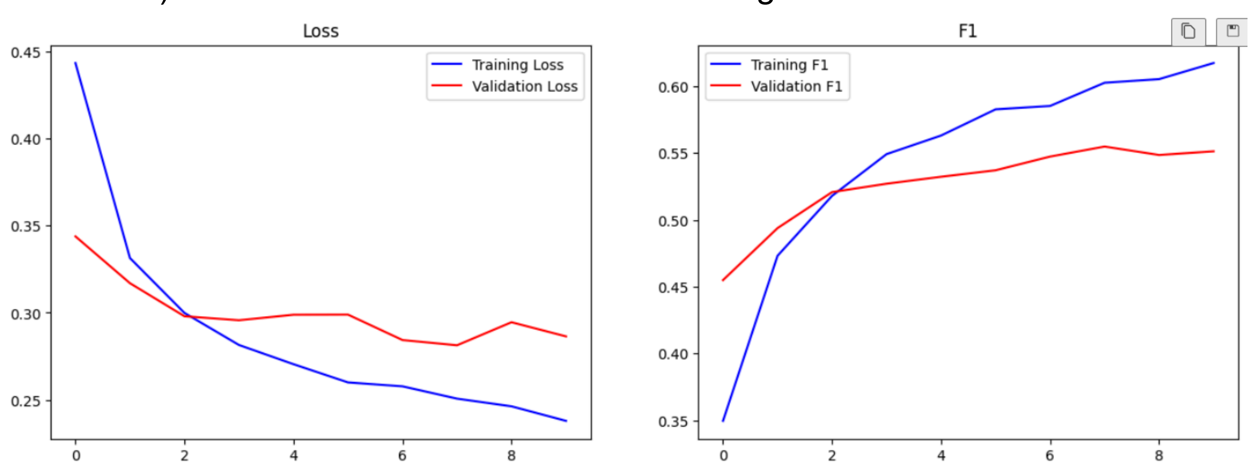

No special data pre-processing was necessary.

```
"2": {
    "text": "Easy to start up and does not overheat as much as other laptops .",
    "labels": [
        "O",
        "O",
        "B",
        "I",
        "O",
        "O",
        "O",
        "O",
        "O",
        "O",
        "O",
        "O",
        "O",
        "O"
    ]
},
```


- **Plots of Models on different Embeddings**

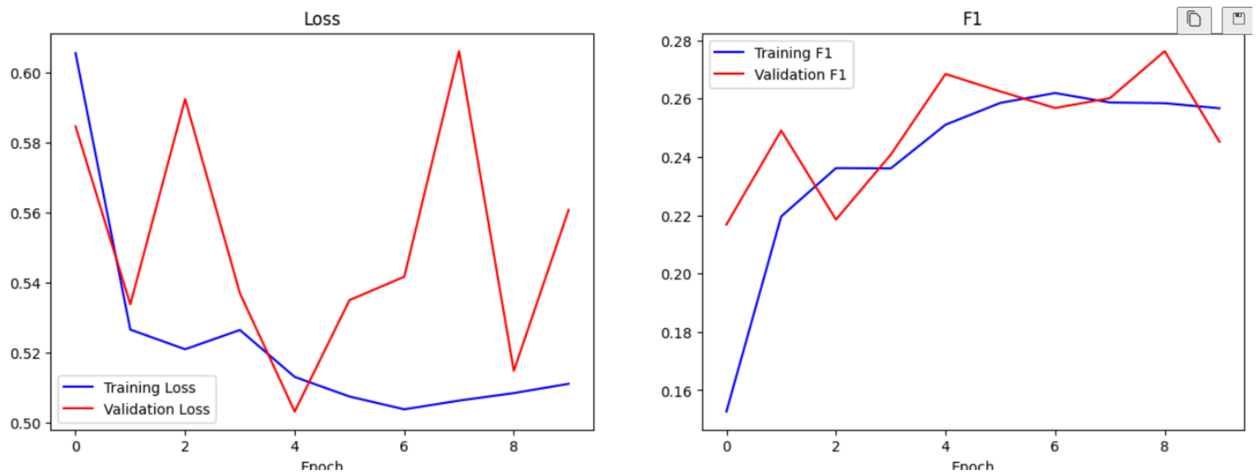
1. **Task-1 (Named Entity Recognition)**

    a) Vanilla RNN on Word2Vec Embeddings



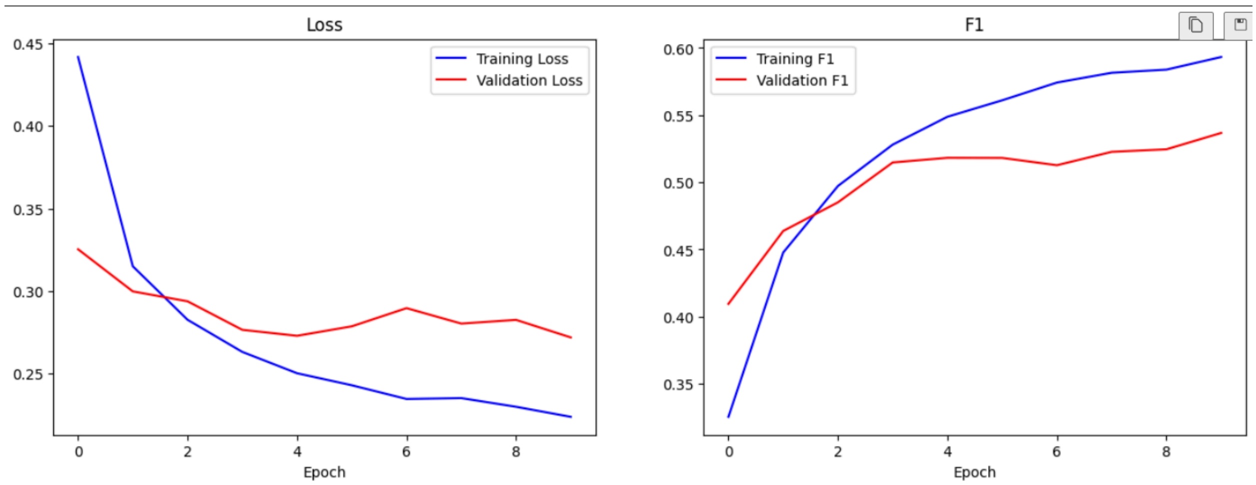
The optimizer used was Adam, so the loss curve is somewhat smooth.

b) Vanilla RNN on Glove Embeddings



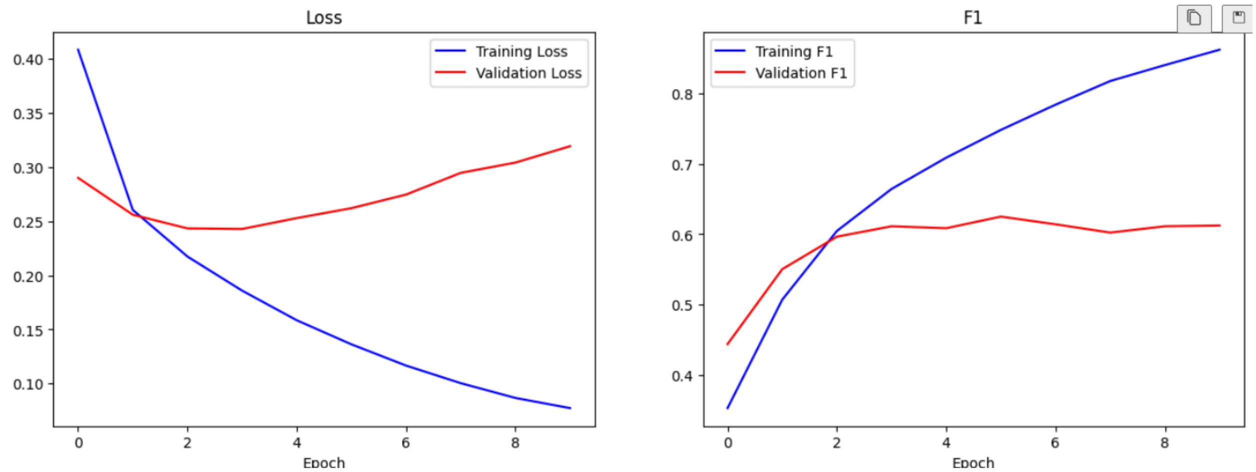
The loss plot has a lot of spikes in it, because an SGD optimizer was used.

c) Vanilla RNN on Fasttext Embeddings



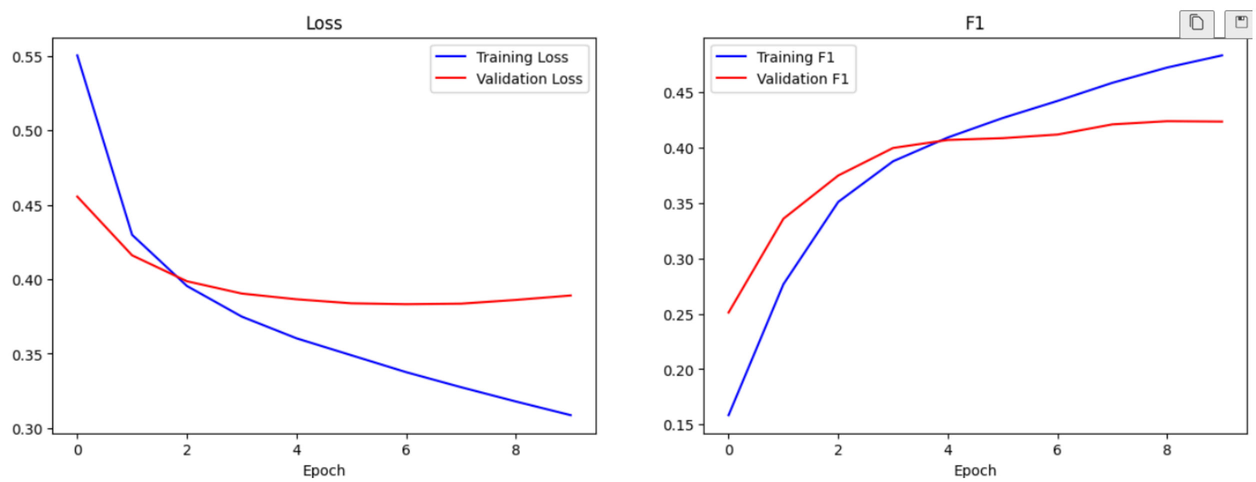
The optimizer used was Adam, so the loss curve is somewhat smooth and the loss converges too for 10 epochs.

d) LSTM on Word2Vec Embeddings



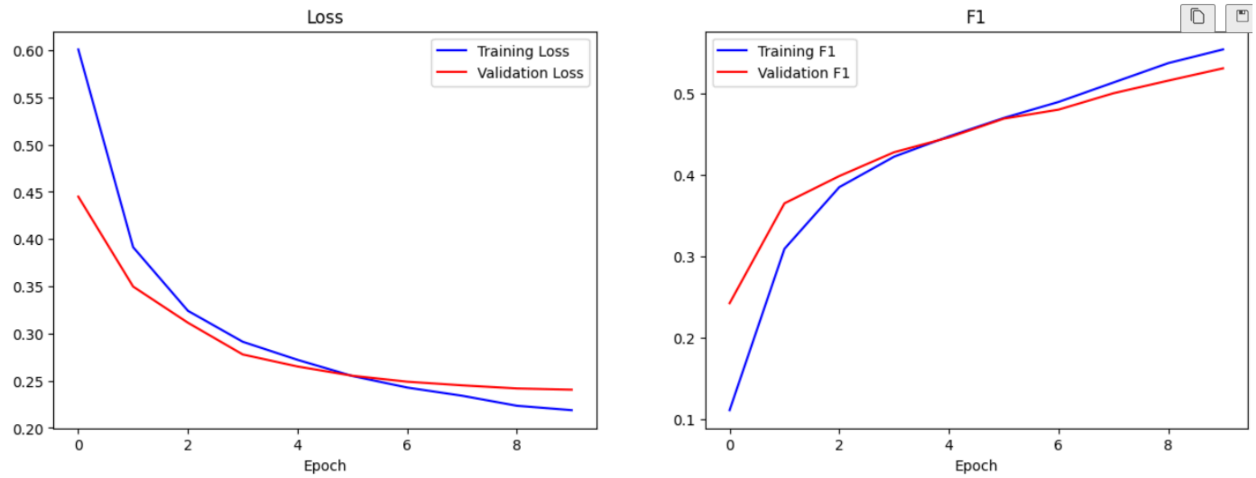
The optimizer used was Adam, so the loss curve is somewhat smooth and after a few epochs overfitting starts.

e) LSTM on Glove Embeddings



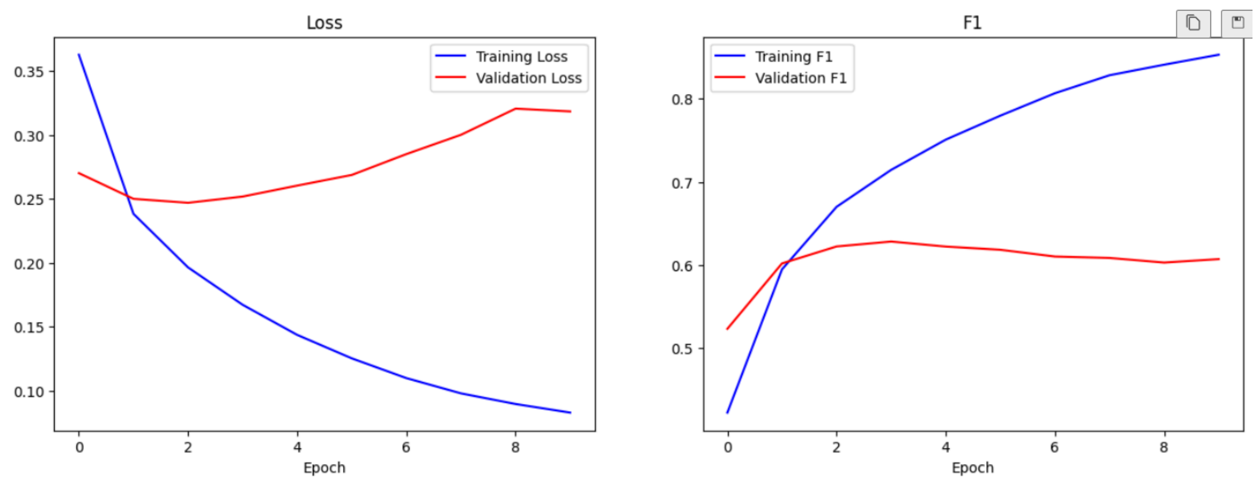
The optimizer used was Adam, so the loss curve is somewhat smooth and the loss curve saturates.

f) LSTM on Fasttext Embeddings



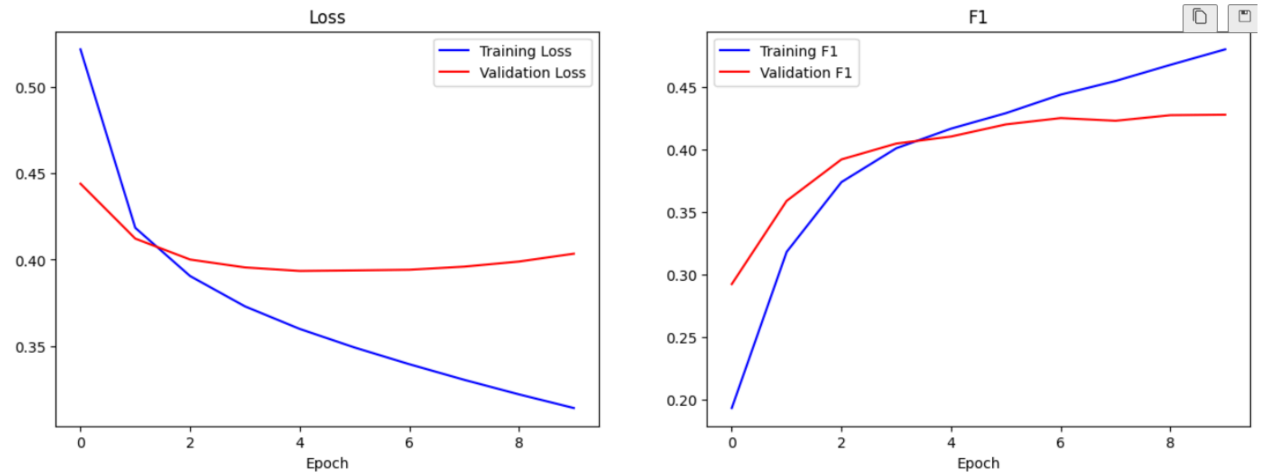
One of the perfect loss curves obtained, in terms of shapes and performance.

g) GRU on Word2Vec Embeddings



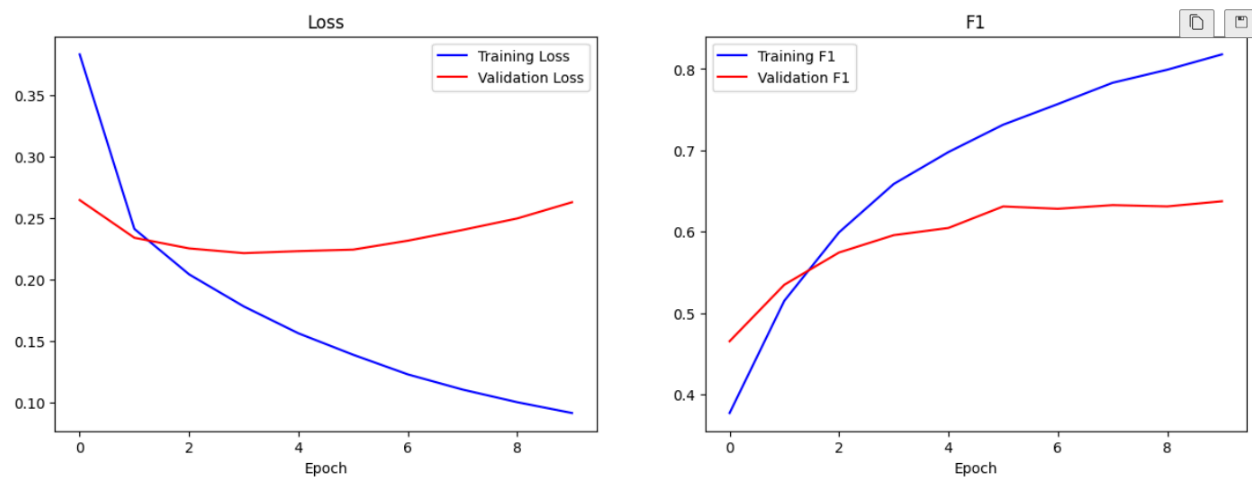
The optimizer used was Adam, so the loss curve is somewhat smooth and after a few epochs overfitting starts.

h) GRU on Glove Embeddings



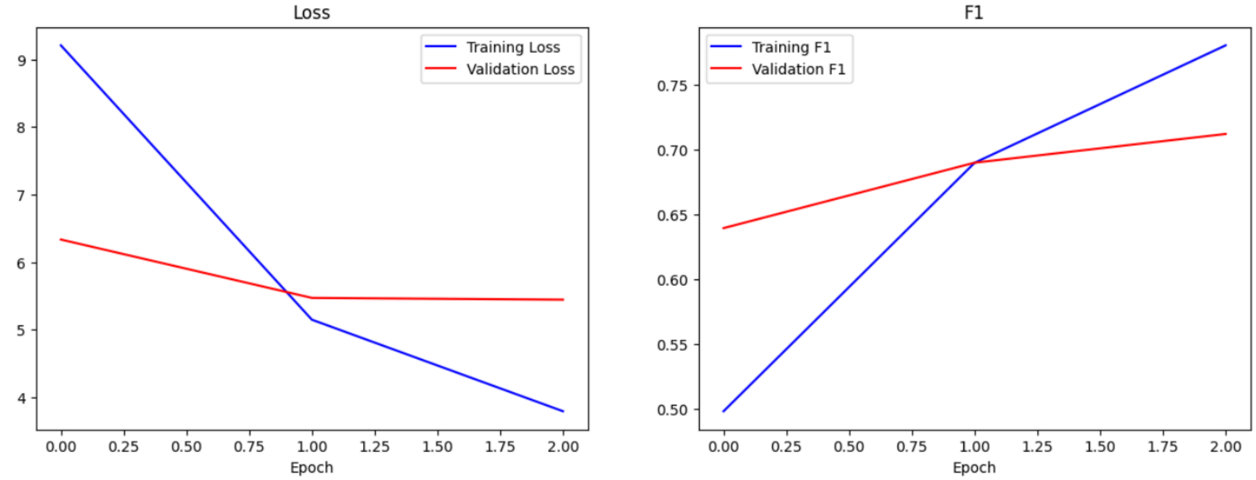
The optimizer used was Adam, so the loss curve is somewhat smooth and the curve saturates after a few epochs.

i) GRU on Fasttext Embeddings



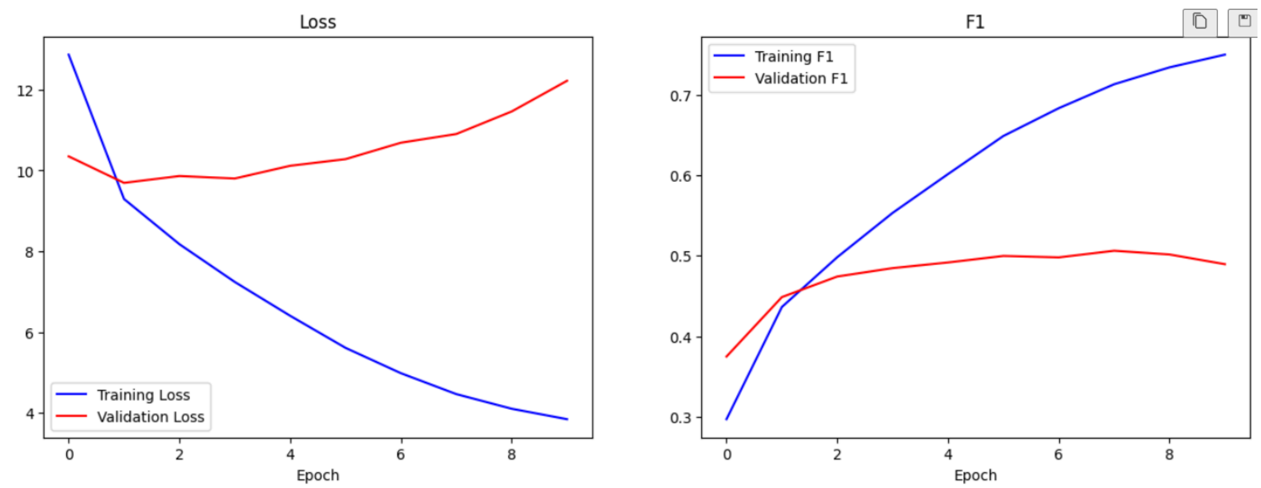
The optimizer used was Adam, so the loss curve is somewhat smooth and after a few epochs overfitting starts.

j) BiLSTM CFR on Word2Vec Embeddings



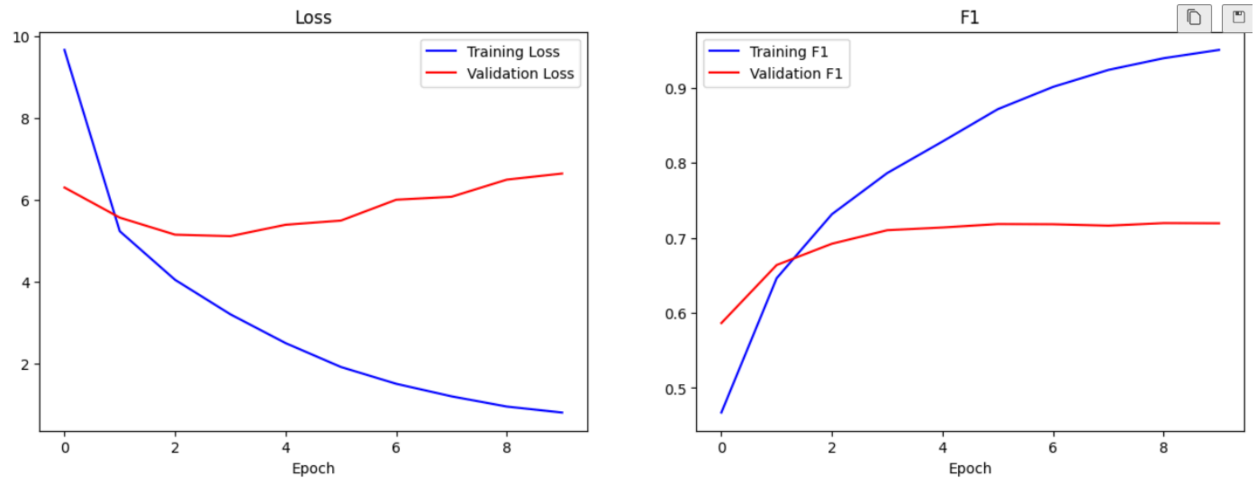
Ran the model only for a few epochs as training time was high and performance was excellent, curve seems linear because of only three epochs.

k) BiLSTM CFR on Glove Embeddings



The optimizer used was Adam, so the loss curve is somewhat smooth and after a few epochs, overfitting starts.

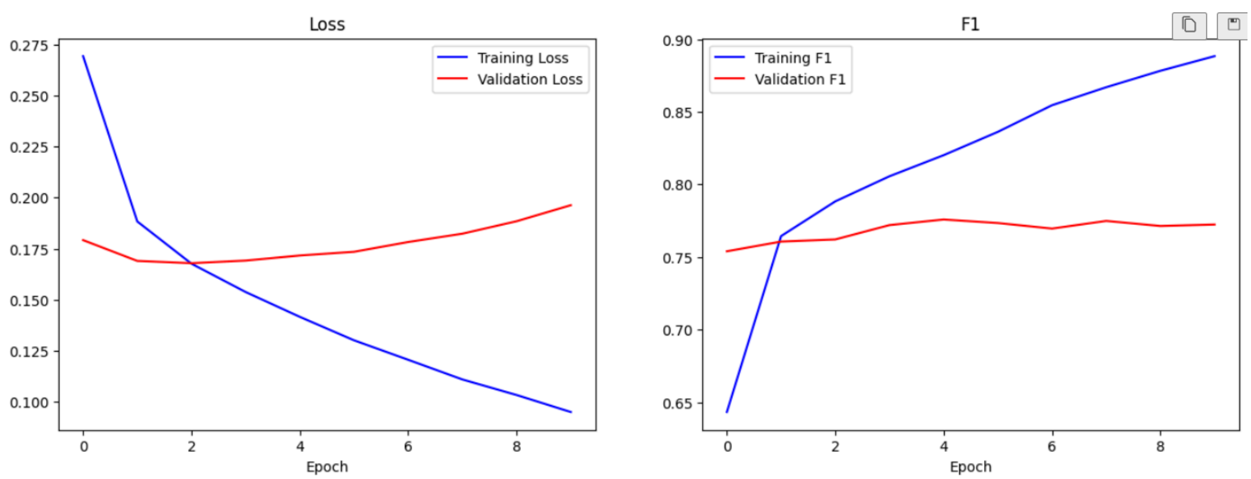
l) BiLSTM CFR on Fasttext Embeddings



Although overfitting starts to happen after few epochs, this was the best model in terms of performance

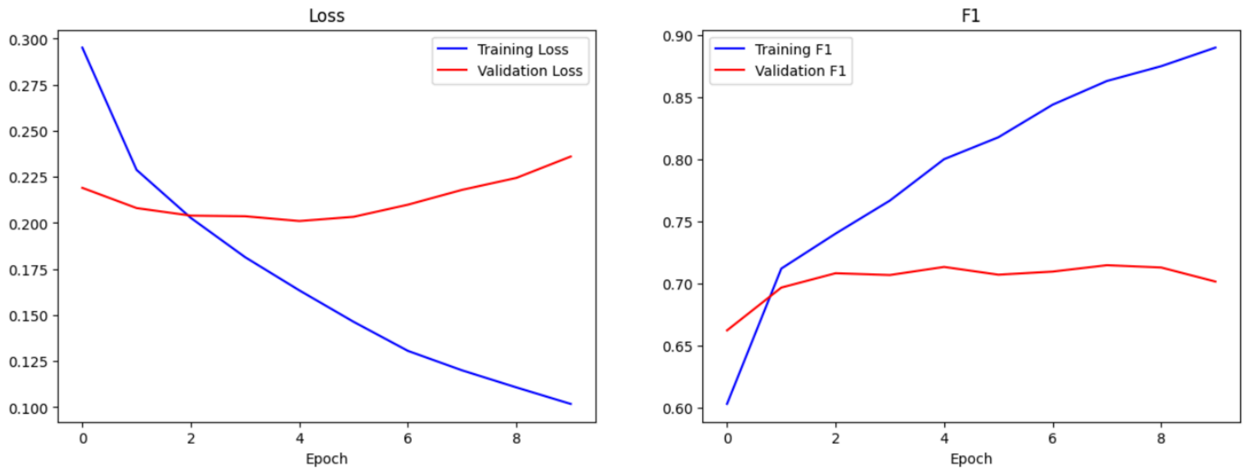
2. **Task-2 (Aspect Term Extraction)**

m) Vanilla RNN on Word2Vec Embeddings



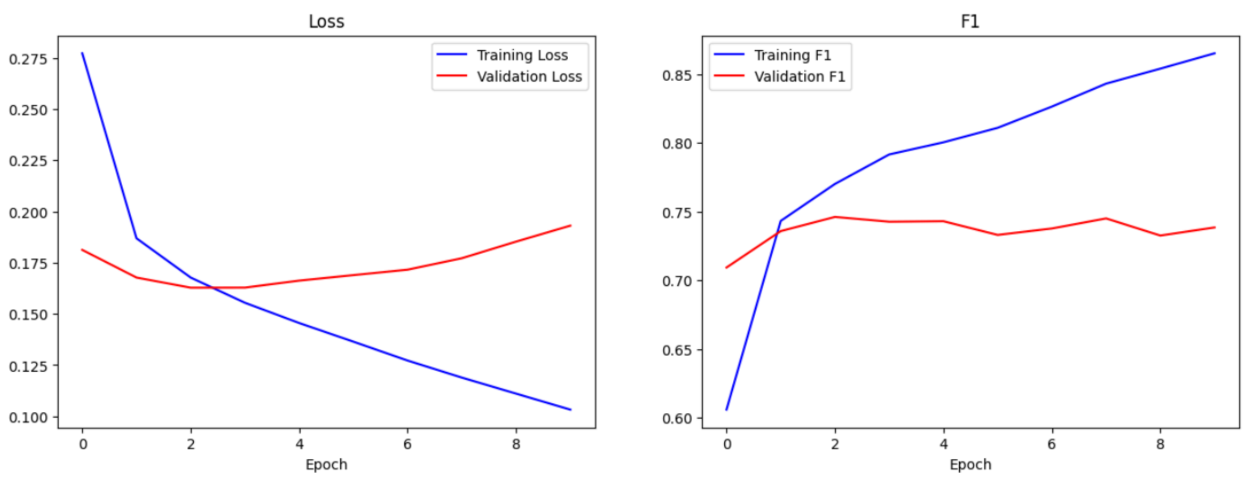
The optimizer used was Adam, so the loss curve is somewhat smooth and overfits.

n) Vanilla RNN on Glove Embeddings



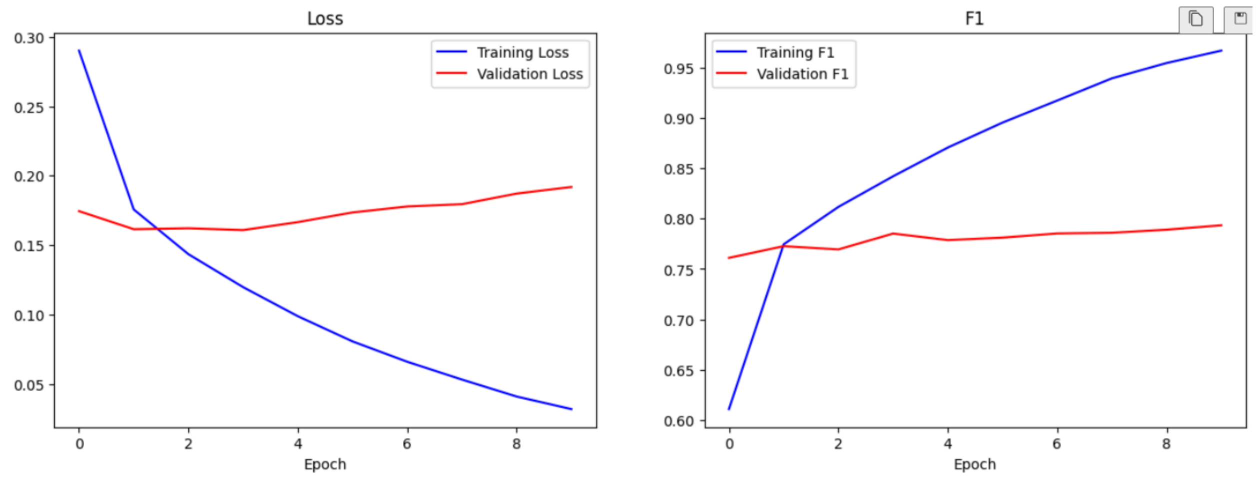
The optimizer was Adam, so the loss curve is smooth and overfit.

o) Vanilla RNN on Fasttext Embeddings



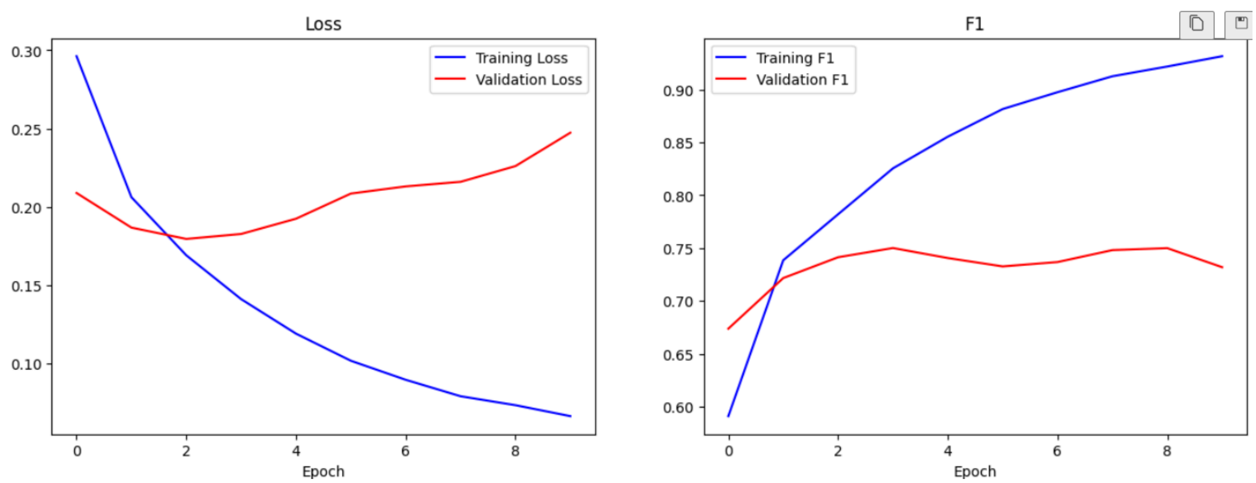
The optimizer was Adam, so the loss curve is smooth and overfit.

p) LSTM on Word2Vec Embeddings



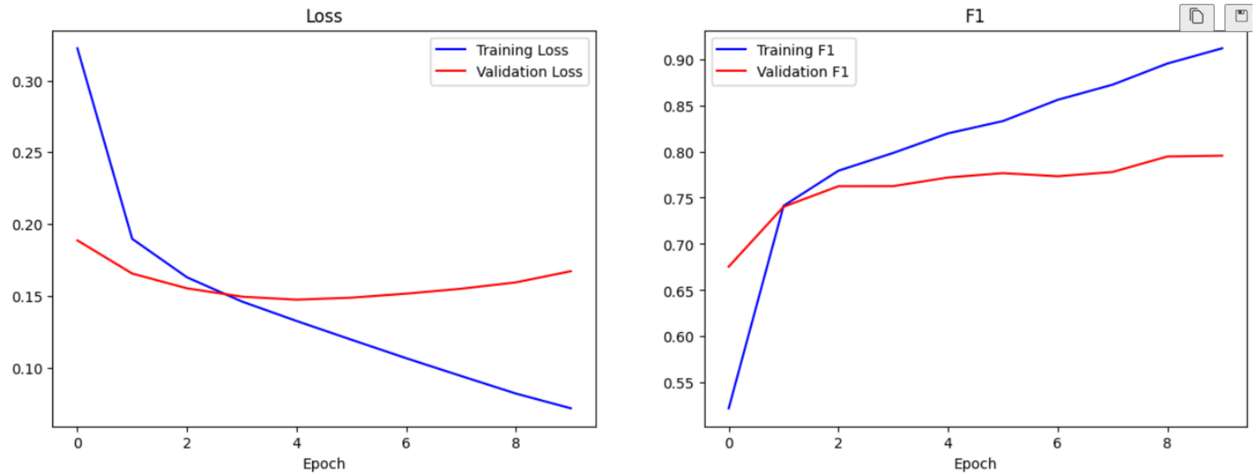
The optimizer was Adam, so the loss curve is smooth and overfit.

q) LSTM on Glove Embeddings



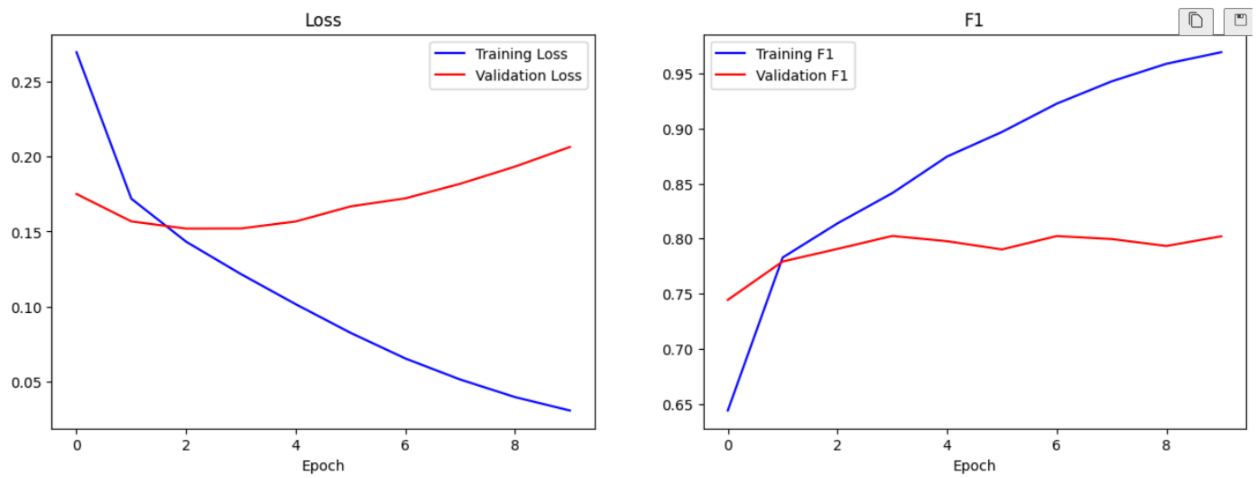
The optimizer was Adam, so the loss curve is smooth and overfit.

r) LSTM on Fasttext Embeddings



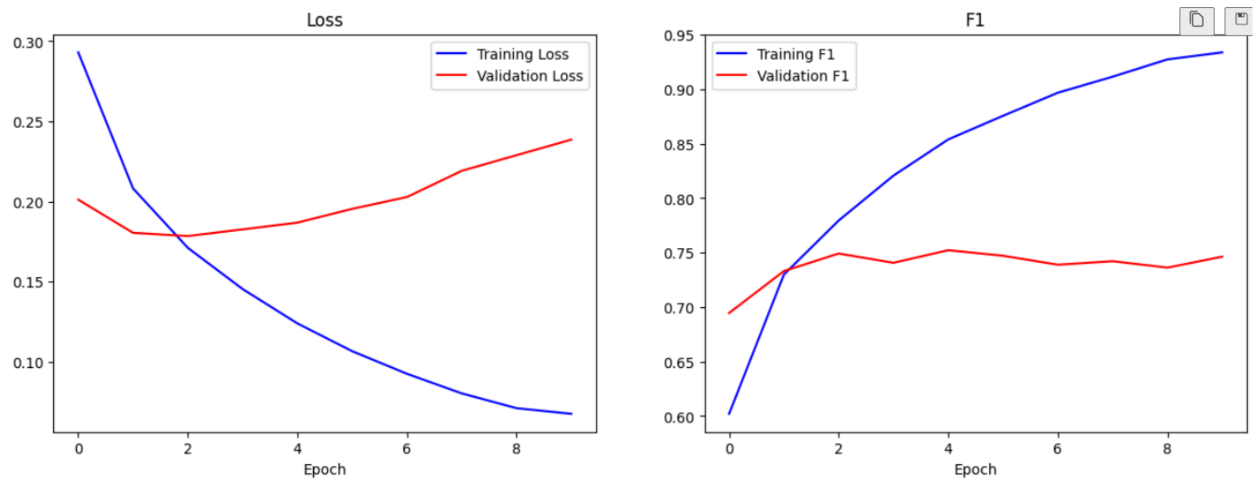
The optimizer was Adam, so the loss curve is smooth and saturates.

s) GRU on Word2Vec Embeddings



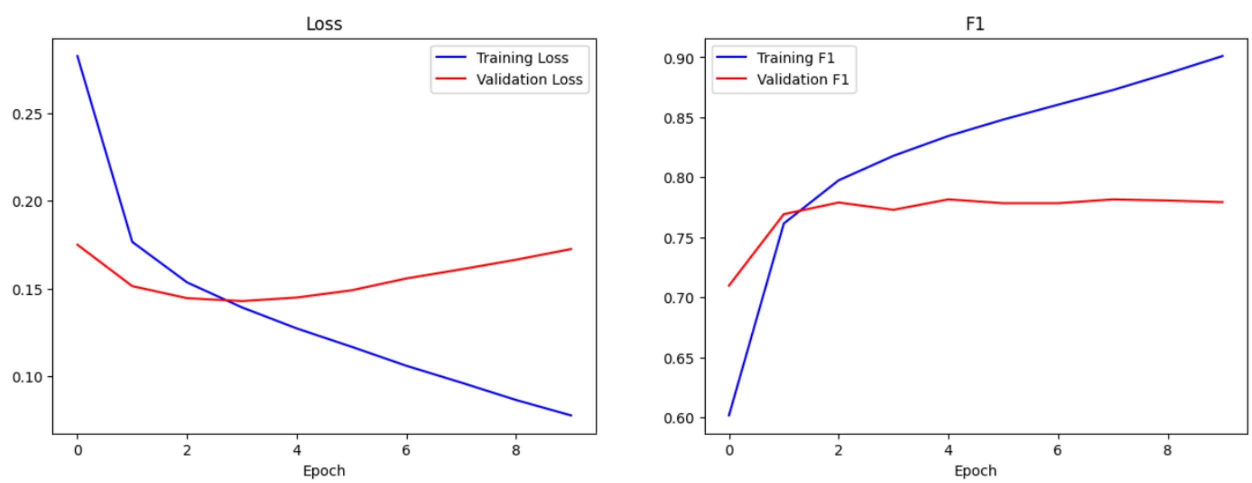
The optimizer was Adam, so the loss curve was smooth and overfit, but the performance was good.

t) GRU on Glove Embeddings



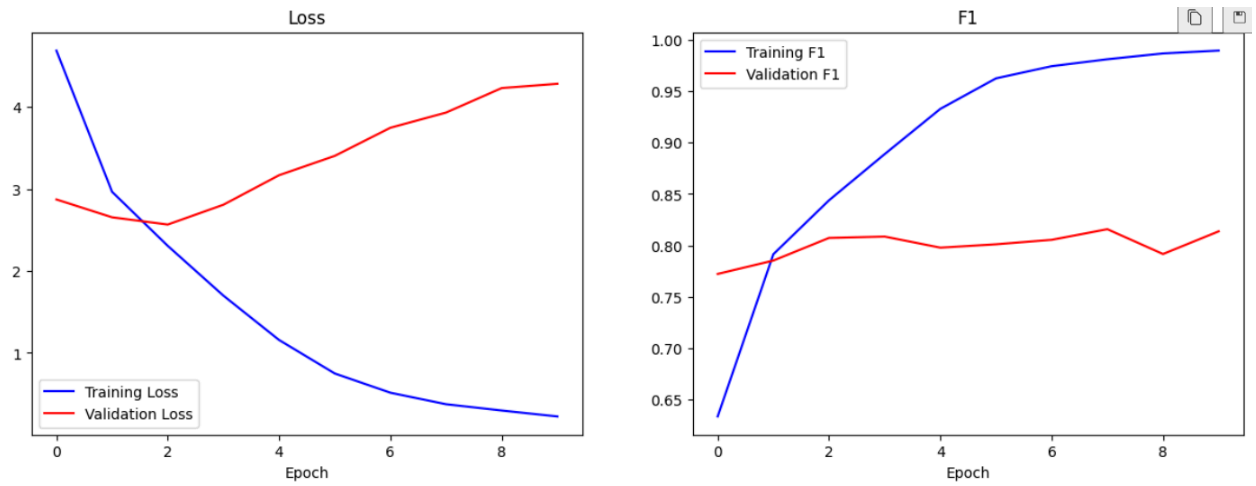
The optimizer was Adam, so the loss curve was smooth and overfit, and the performance was poor.

u) GRU on Fasttext Embeddings



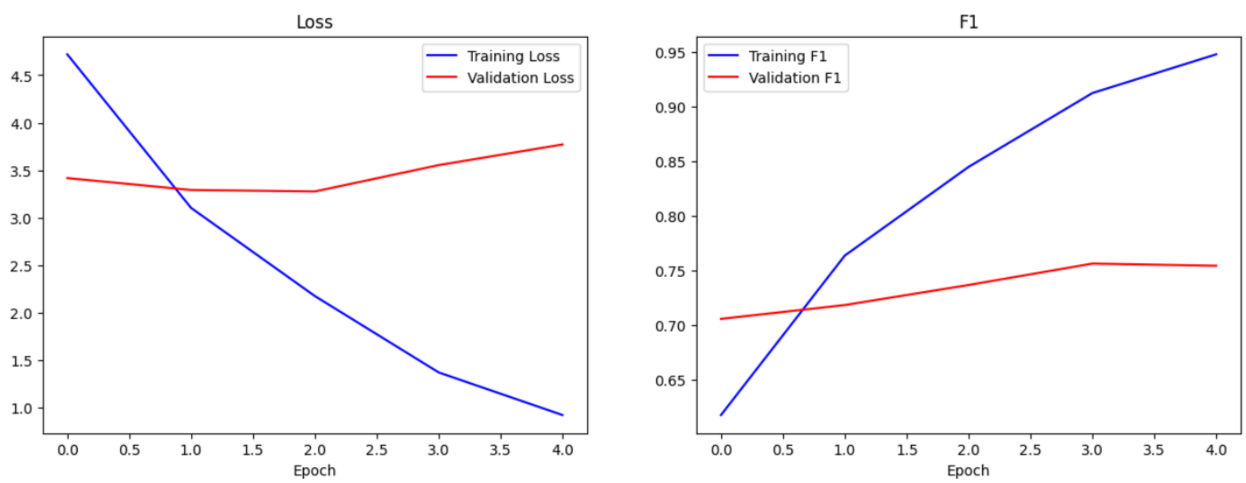
The optimizer was Adam, so the loss curve was smooth and overfit, but the performance was good.

v) BiLSTM CFR on Word2Vec Embeddings



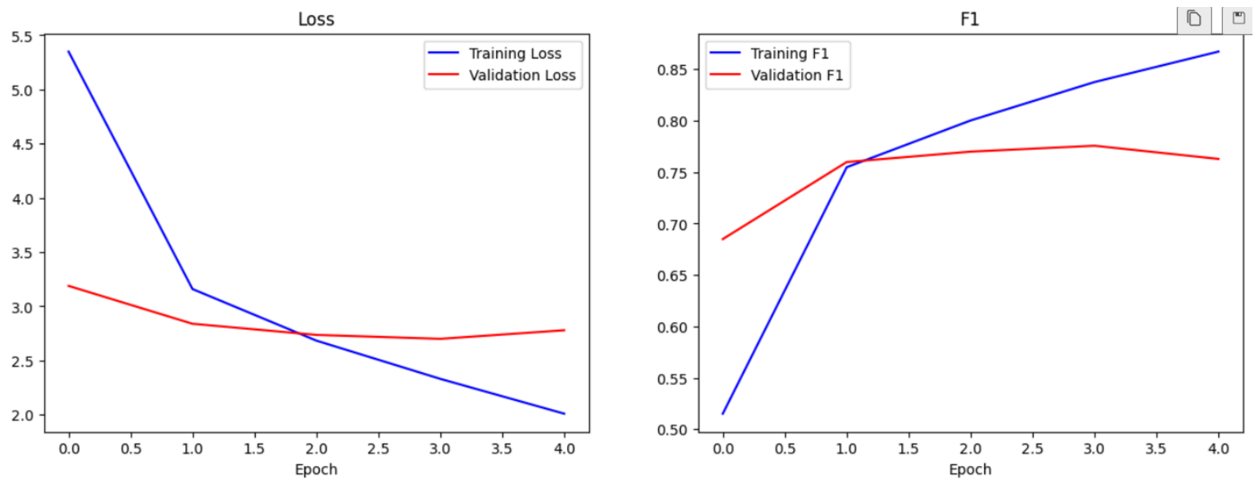
Although the curves were overfitting, the performance was good.

w) BiLSTM CFR on Glove Embeddings



Curves were overfitting and the performance was poor.

x) BiLSTM CFR on Fasttext Embeddings



Good loss saturating and converging loss curve with reasonably good f1 performance.

- **Accuracy and Macro-F1 Tables**

Task-1 (Named Entity Recognition):

Model no	Embedding_used	Accuracy	Macro F1
1 (RNN)	Word2Vec	92.26	55.12
1 (RNN)	Glove	85.71	23.93
1 (RNN)	Fast-text	92.54	50.79
2 (LSTM)	Word2Vec	93.03	59.94
2 (LSTM)	Glove	89.56	39.76
2 (LSTM)	Fast-text	93.14	52.37
3 (GRU)	Word2Vec	92.80	58.08
3 (GRU)	Glove	89.39	40.68
3 (GRU)	Fast-text	93.61	61.22
4 (BiLSTM CRF)	Word2Vec	95.21	69.67
4 (BiLSTM CRF)	Glove	90.97	43.96

4 (BiLSTM CRF)	Fast-text	95.46	70.63

Task-2 (Aspect Term Extraction):

Model no	Embedding_used	Accuracy	Macro F1
1 (RNN)	Word2Vec	92.30	75.54
1 (RNN)	Glove	90.47	67.97
1 (RNN)	Fast-text	92.13	73.12
2 (LSTM)	Word2Vec	92.77	77.58
2 (LSTM)	Glove	91.49	70.22
2 (LSTM)	Fast-text	92.48	76.36
3 (GRU)	Word2Vec	92.57	76.51
3 (GRU)	Glove	91.45	71.96
3 (GRU)	Fast-text	92.98	77.18
4 (BiLSTM CRF)	Word2Vec	93.58	80.07
4 (BiLSTM CRF)	Glove	92.26	74.18
4 (BiLSTM CRF)	Fast-text	93.56	78.84

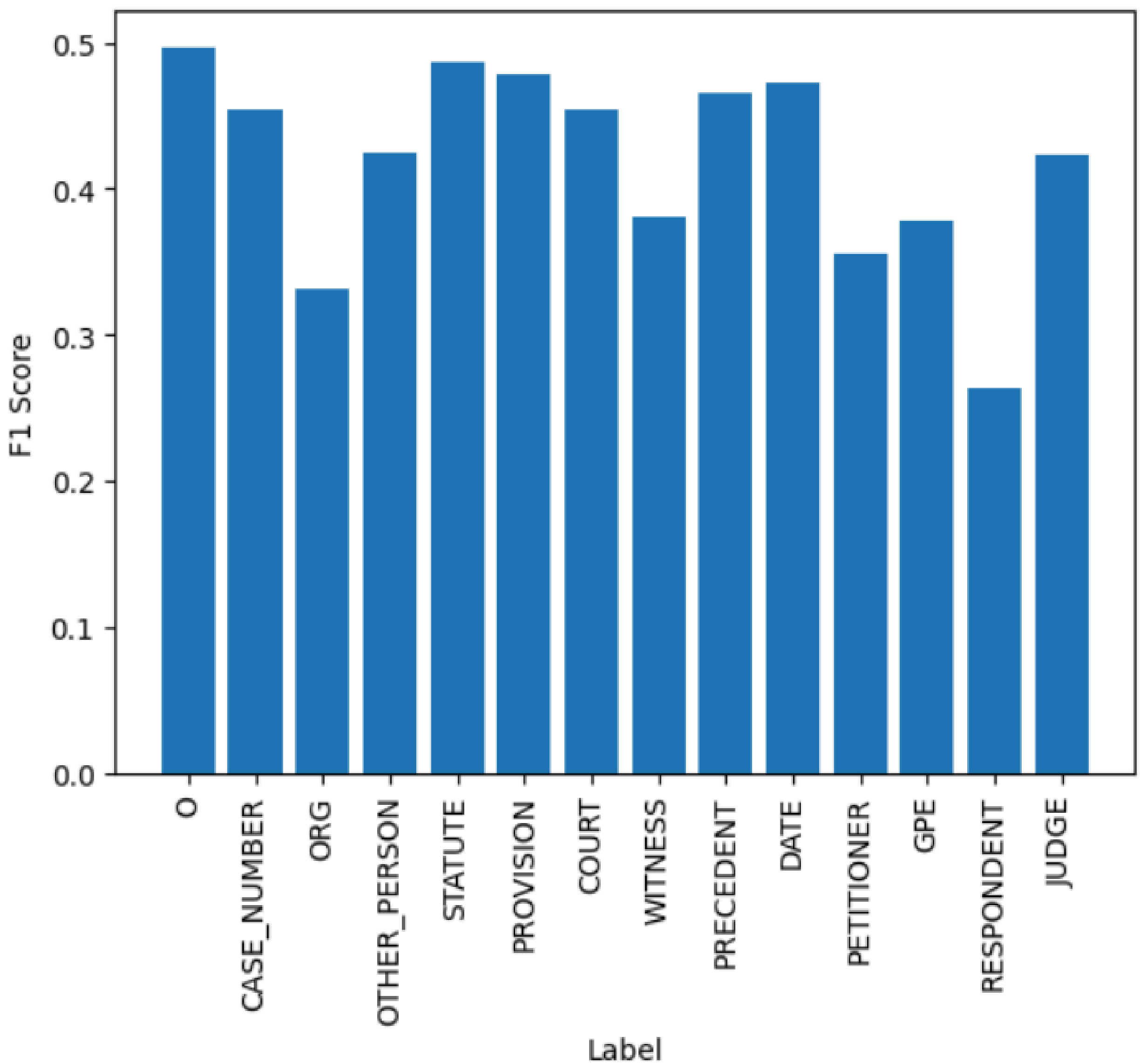
- **Overall trends:**
  - Performance of LSTM and GRU models was significantly better than vanilla RNN, and the BiLSTM CRF was giving the best performance as expected.
  - We noticed that models with fastText gave the best performance, followed by Word2Vec and Glove gave the worst performance.

- This can be explained by the fact that fastText provides embeddings for most words by looking at subword embeddings and combining them and so fewer random embeddings have to be used.

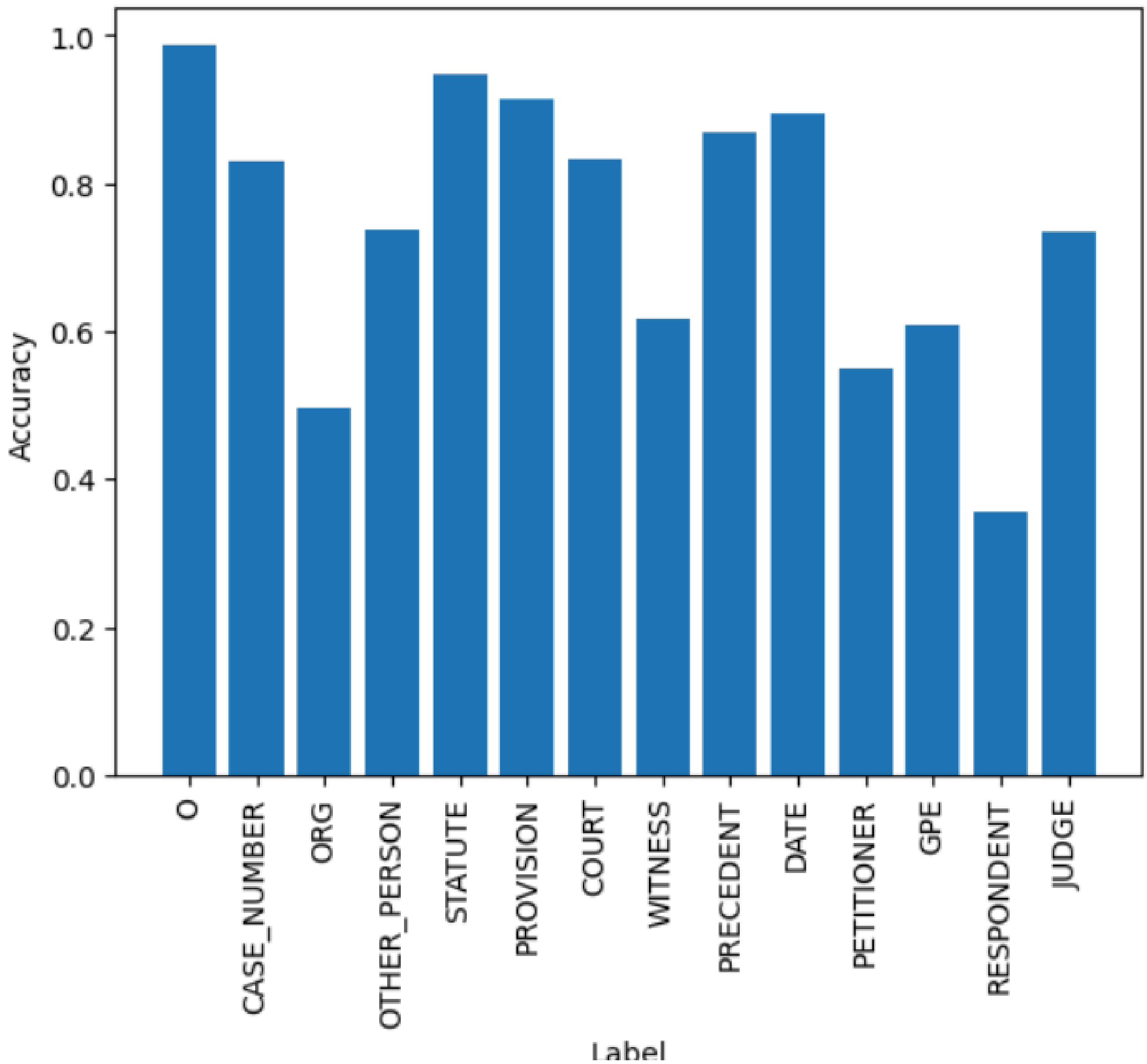
- **Label-wise F1 plot for Best Model on Task-1 (NER)**

Best Model was observed to be Bi-LSTM CRF trained with FastText Embeddings

Label-wise F1 plot on the Best Model



Label-wise Accuracy plot on the Best Model



- **Contributions:**

    1. Khushdev: Task-2 dataset preparation, Part-2 Baseline and CRF models implementation.
    2. Apurv: Task-1 data preparation, Helped with Part-2 Baseline models and improvements for Task-1 dataset performance.
    3. Pankaj:Task-2 data preparation, and helped in Task 2 BiLSTM-CRF model.
    4. Saara Moin: Tried Task 2 BiLSTM-CRF.