CSE 556: Natural Language Processing

Assignment 2 Report

Group 16

Data samples prepared in Part 1A:

1005: {'text': 'dr. s. k. kaushik pw 3 deposed 1.40 a. m. night intervening 22nd 23rd july 1986 recorded dying declaration ex p/104 injured jetha ram bhartiya general hospital churu presence dr. rakesh bhargava got thumb impression jetha ram place x',

'labels': ['O', 'B_WITNESS', 'I_WITNESS', 'I_WITNESS', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'B_DATE', 'I_DATE', 'I_DATE', 'I_DATE', 'O', 'O', 'O', 'O', 'O', 'O', 'B_OTHER_PERSON', 'I_OTHER_PERSON', 'B_ORG', 'I_ORG', 'I_ORG', 'I_ORG', 'O', 'O', 'B_OTHER_PERSON', 'I_OTHER_PERSON', 'O', 'O', 'O', 'B_OTHER_PERSON', 'I_OTHER_PERSON', 'O', 'O']},

999: {'text': 'september 10 1965 legislature state maha- rashtra enacted maharashtra municipalities act repealed bombay district municipal act 3 1901',
'labels': ['B_DATE', 'I_DATE', 'I_DATE', 'B_ORG', 'I_ORG', 'I_ORG', 'I_ORG', 'O', 'B_STATUTE', 'I_STATUTE', 'I_STATUTE', 'O', 'B_STATUTE', 'I_STATUTE', 'I_STATUTE', 'I_STATUTE', 'I_STATUTE', 'I_STATUTE']},

Data samples prepared in Part 1B:

{'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']}

{'text': ['However', ',', 'the', 'multi-touch', 'gestures', 'and', 'large', 'tracking', 'area', 'make', 'having', 'an', 'external', 'mouse', 'unnecessary', '(', 'unless', 'you', "'re", 'gaming', ')', '.'],
'labels': ['O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'B', 'I', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O']}

Additional preprocessing performed on the text:

For dataset 1:
We defined the preprocess_text function that takes an input text and prepares it for analysis by first converting it to lowercase to ensure consistency, then tokenizing it into individual words. It removes common stopwords and punctuation marks that typically do not contribute significantly to the meaning of the text. Additionally, it filters out any remaining empty tokens.

For dataset 2:
No preprocessing was necessary for Dataset 2 since it was a relatively simpler dataset.

We pad the text sequences in our dataset to a specified maximum length. If a text sequence exceeds or equals the maximum length, the sample is removed from the dataset during training. For sequences shorter than the maximum length, the function pads them with a padding token until they match the specified length. The corresponding labels are also padded with 'O' tags accordingly.

Task-1

Model Type	**Embedding Used**	**Accuracy**	**F1-Score**
RNN	Word2Vec	0.964	0.38
RNN	Glove	0.968	0.42
RNN	Fast-Text	0.968	0.39
LSTM	Word2Vec	0.964	0.44
LSTM	Glove	0.968	0.48
LSTM	Fast-Text	0.971	0.45
GRU	Word2Vec	0.964	0.44
GRU	Glove	0.968	0.47
GRU	Fast-Text	0.968	0.47
Bilstm-Crf	Word2Vec	0.968	0.55
Bilstm-Crf	Glove	0.971	0.61
Bilstm-Crf	Fast-Text	0.9704	0.57

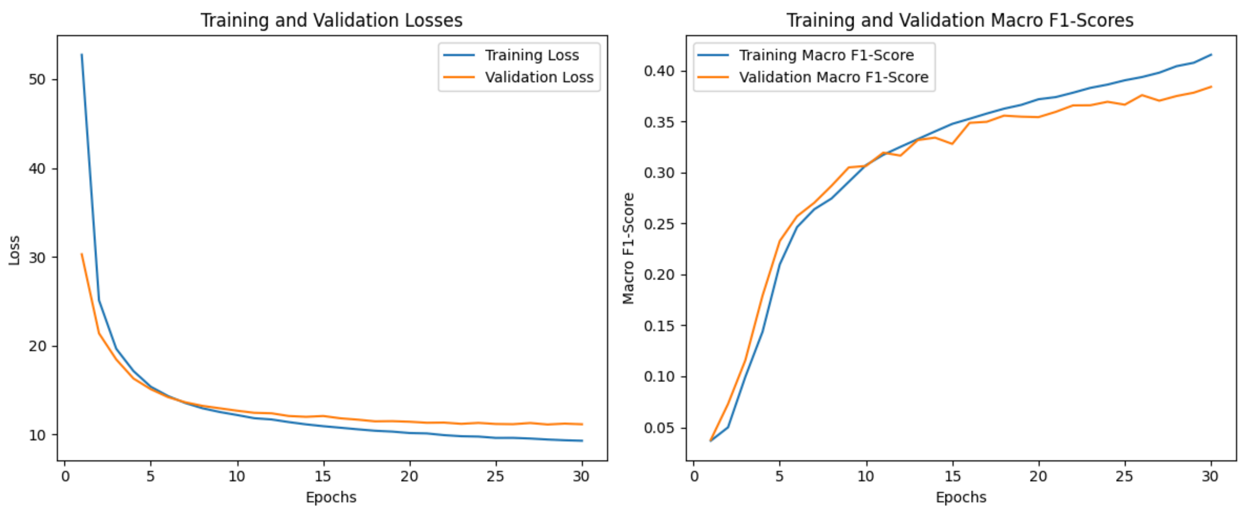
Task-2

Model Type	Embedding Used	Accuracy	F1-Score
RNN	Word2Vec	0.977	0.49
RNN	Glove	0.977	0.49
RNN	Fast-Text	0.975	0.38
LSTM	Word2Vec	0.981	0.64
LSTM	Glove	0.982	0.58
LSTM	Fast-Text	0.974	0.39
GRU	Word2Vec	0.982	0.64
GRU	Glove	0.976	0.46
GRU	Fast-Text	0.982	0.66
Bilstm-Crf	Word2Vec	0.975	0.73
Bilstm-Crf	Glove	0.977	0.73
Bilstm-Crf	Fast-Text	0.984	0.73

Task 1 Plots

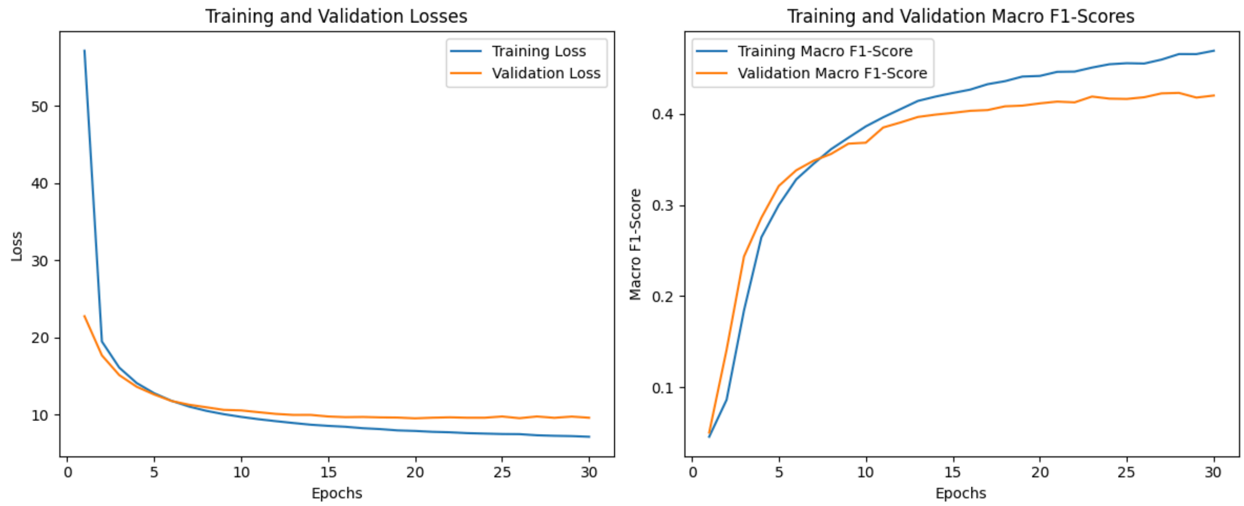
a) Vanilla RNN

1) Word2Vec



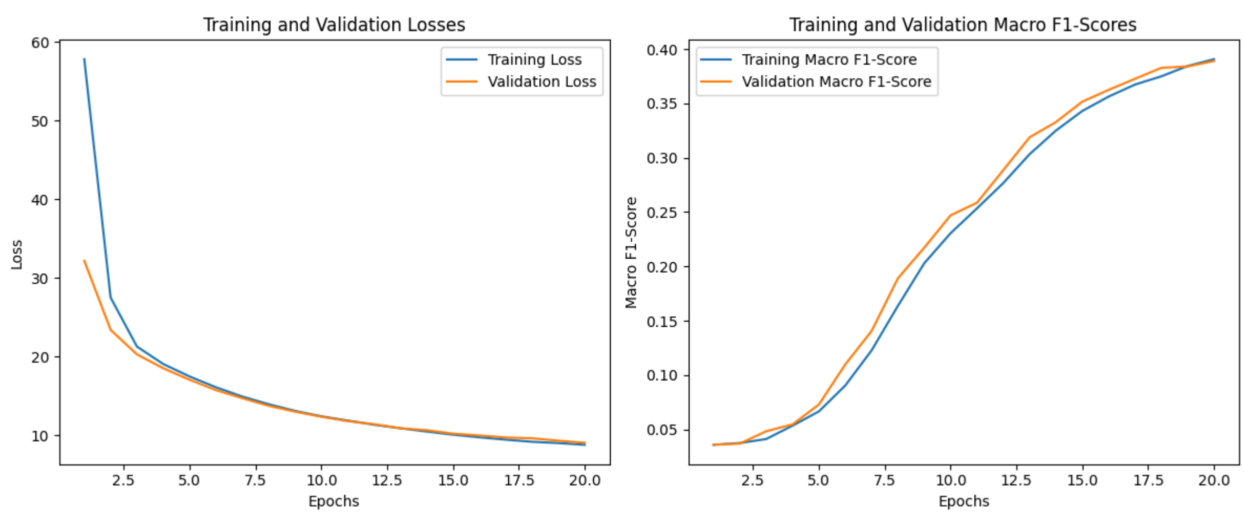
The left plot displays training and validation losses over 30 epochs. The training loss starts high but rapidly decreases as the model learns, while the validation loss remains higher in the end, indicating the model's generalization performance on unseen data. The right plot shows training and validation macro F1-scores, both initially low but steadily increasing as the model improves. The validation score lags behind training, mirroring the behavior seen in the loss plot. Overall, the plots suggest the model effectively learns from the training data, demonstrated by decreasing losses and increasing F1-scores.

2) Glove



The left plot displays training and validation losses over 30 epochs. The training loss starts high but rapidly decreases as the model learns, while the validation loss remains higher in the end, indicating the model's generalization performance on unseen data. The right plot shows training and validation macro F1-scores, both initially low but steadily increasing as the model improves. The validation score lags behind training, mirroring the behavior seen in the loss plot. Overall, the plots suggest the model effectively learns from the training data, demonstrated by decreasing losses and increasing F1-scores.

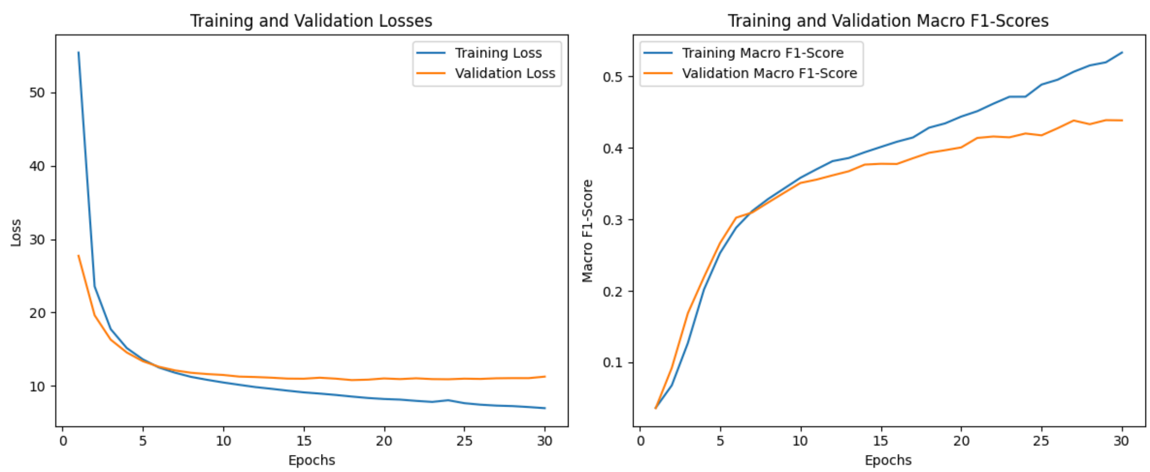
3) Fasttext



The left plot displays training and validation losses over 20 epochs. The training loss starts high but rapidly decreases as the model learns, while the validation loss remains somewhat equal or we can say slightly higher than the training loss in the end , indicating the model's generalization performance on unseen data. The right plot shows training and validation macro F1-scores, both initially low but steadily increasing as the model improves. In this case , the validation score remains higher than the training score. This implies that the model is generalizing well to unseen data. Usually, the training F1 score is higher than or equal to the validation F1 score because the model is trained to minimize the training loss. However, in this case the validation F1 score consistently outperforms the training F1 score, it suggests that the model is not overfitting to the training data and is effectively capturing the underlying patterns in the data.

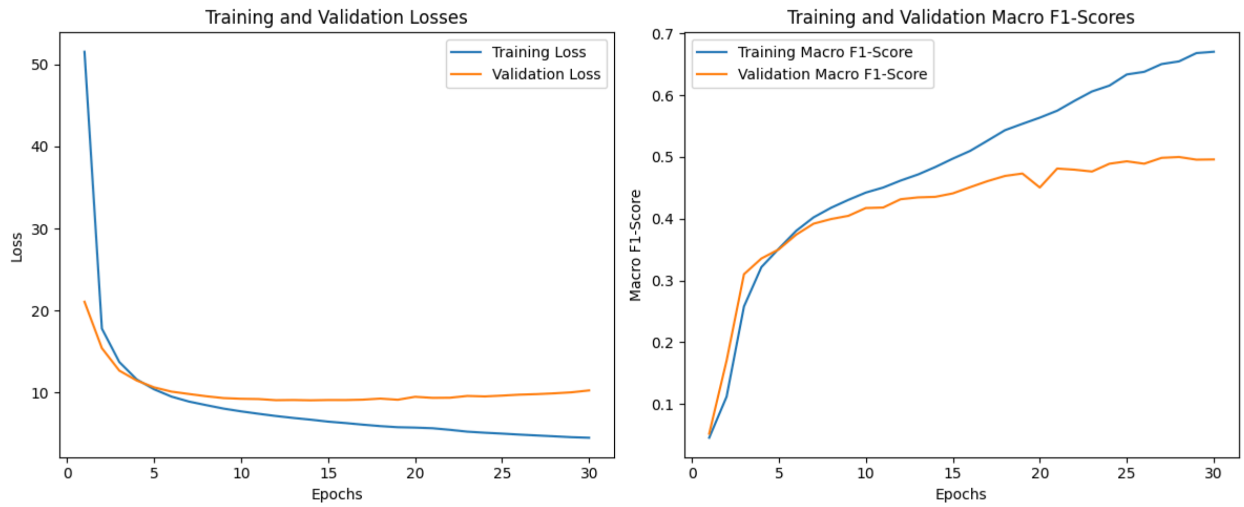
b) LSTM

1) Word2Vec



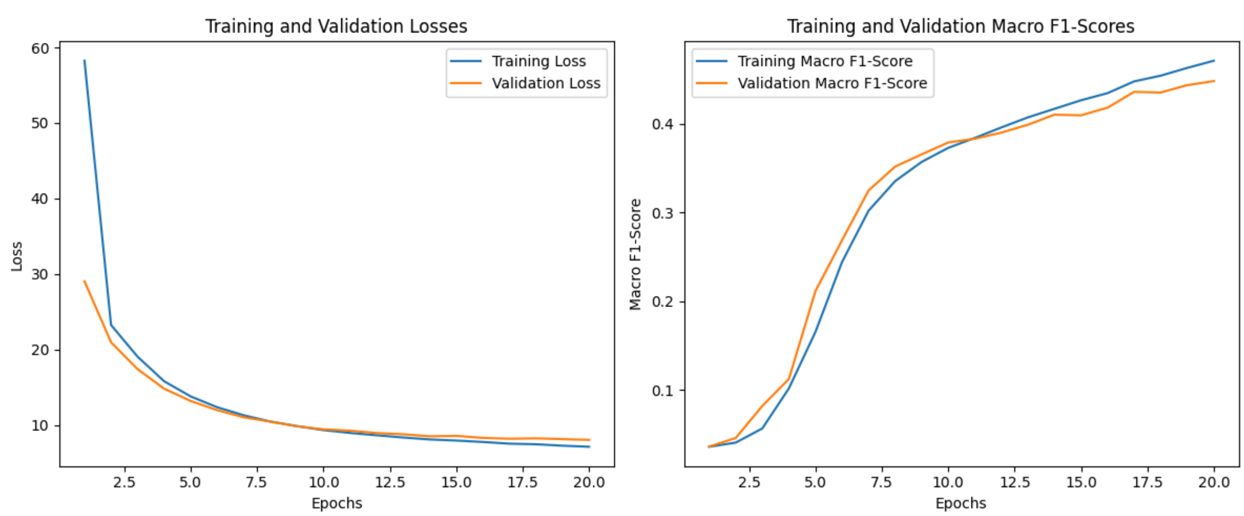
The left plot displays training and validation losses over 30 epochs. The training loss starts high but rapidly decreases as the model learns, while the validation loss remains higher in the end, indicating the model's generalization performance on unseen data. The right plot shows training and validation macro F1-scores, both initially low but steadily increasing as the model improves. The validation score lags behind training, mirroring the behavior seen in the loss plot. Overall, the plots suggest the model effectively learns from the training data, demonstrated by decreasing losses and increasing F1-scores.

2) Glove



The left plot displays training and validation losses over 30 epochs. The training loss starts high but rapidly decreases as the model learns, while the validation loss remains higher in the end, indicating the model's generalization performance on unseen data. The right plot shows training and validation macro F1-scores, both initially low but steadily increasing as the model improves. The validation score lags behind training, mirroring the behavior seen in the loss plot. Overall, the plots suggest the model effectively learns from the training data, demonstrated by decreasing losses and increasing F1-scores.

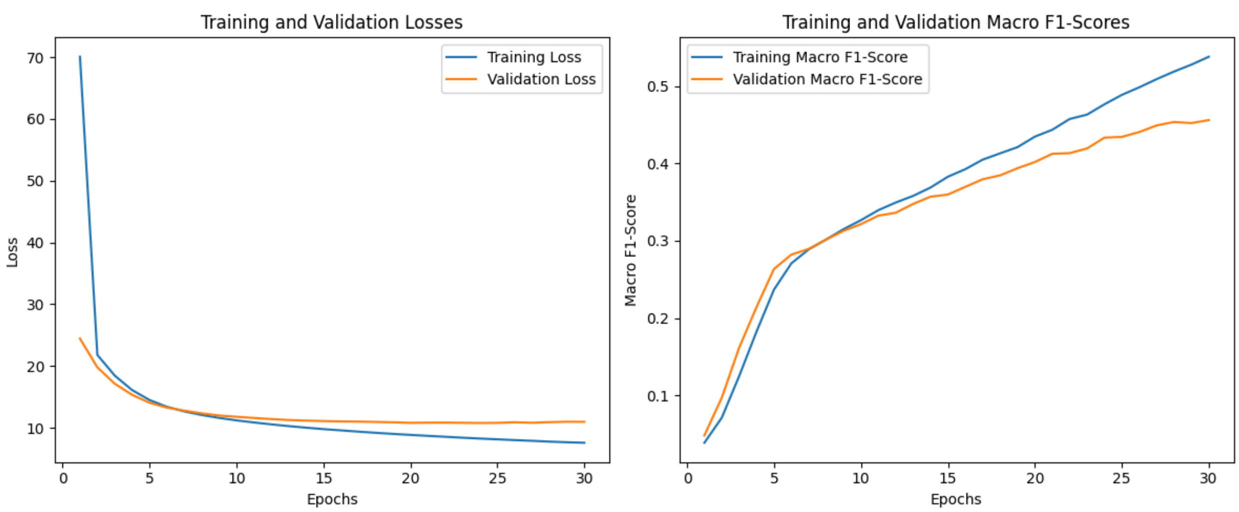
3) Fasttext



The left plot displays training and validation losses over 20 epochs. The training loss starts high but rapidly decreases as the model learns, while the validation loss remains higher in the end, indicating the model's generalization performance on unseen data. The right plot shows training and validation macro F1-scores, both initially low but steadily increasing as the model improves. The validation score is higher than the training score initially, but it eventually lags behind the training score. Overall, the plots suggest the model effectively learns from the training data, demonstrated by decreasing losses and increasing F1-scores.

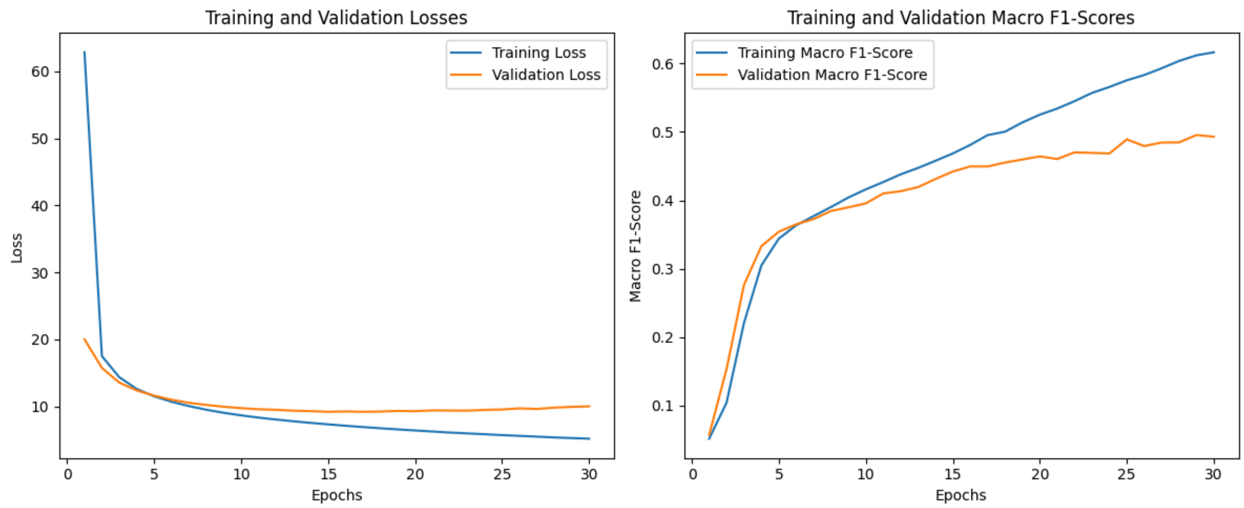
c) Model - GRU

1) Word2Vec



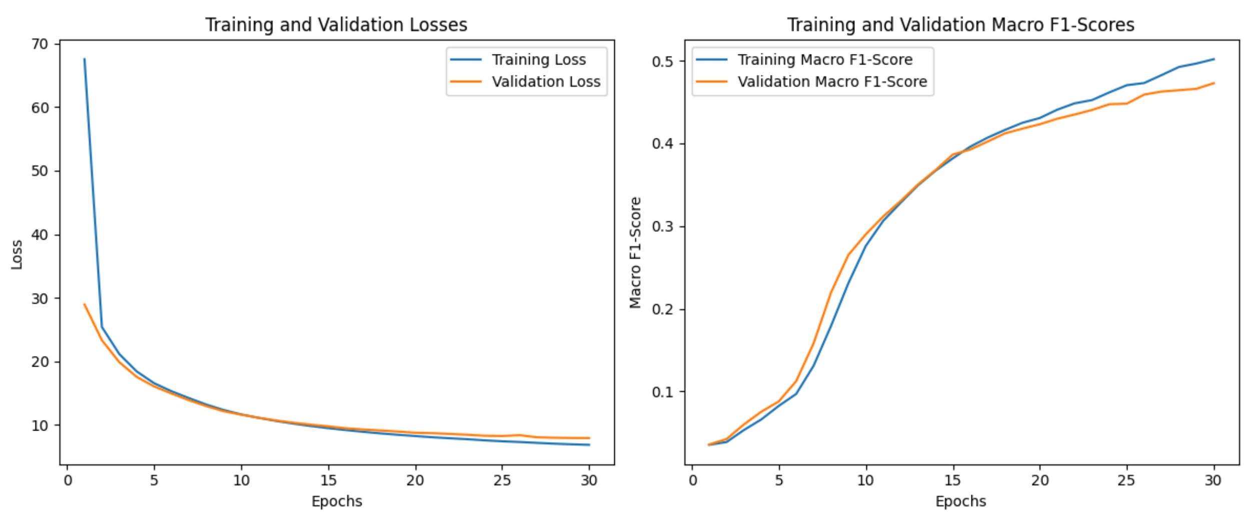
The left plot displays training and validation losses over 30 epochs. The training loss starts high but rapidly decreases as the model learns, while the validation loss remains higher in the end, indicating the model's generalization performance on unseen data. The right plot shows training and validation macro F1-scores, both initially low but steadily increasing as the model improves. The validation score is higher than the training score initially, but it eventually lags behind the training score. Overall, the plots suggest the model effectively learns from the training data, demonstrated by decreasing losses and increasing F1-scores.

2) Glove



The left plot displays training and validation losses over 30 epochs. The training loss starts high but rapidly decreases as the model learns, while the validation loss starts from a lower value and remains higher in the end, indicating the model's generalization performance on unseen data. The right plot shows training and validation macro F1-scores, both initially low but steadily increasing as the model improves. The validation score is higher than the training score initially, but it eventually lags behind the training score. Overall, the plots suggest the model effectively learns from the training data, demonstrated by decreasing losses and increasing F1-scores.

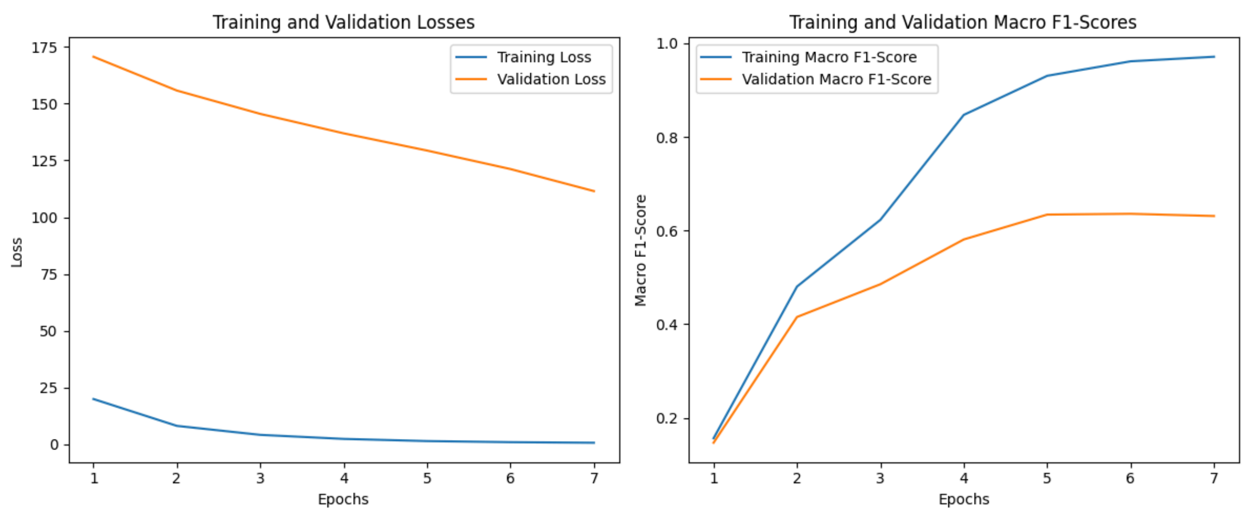
3) Fasttext



The left plot displays training and validation losses over 30 epochs. The training loss starts high but rapidly decreases as the model learns, while the validation loss remains higher in the end, indicating the model's generalization performance on unseen data. The right plot shows training and validation macro F1-scores, both initially low but steadily increasing as the model improves. The validation score is higher than the training score initially, but it eventually lags behind the training score. Overall, the plots suggest the model effectively learns from the training data, demonstrated by decreasing losses and increasing F1-scores.

d) BiLSTM - CRF

1) Word2Vec



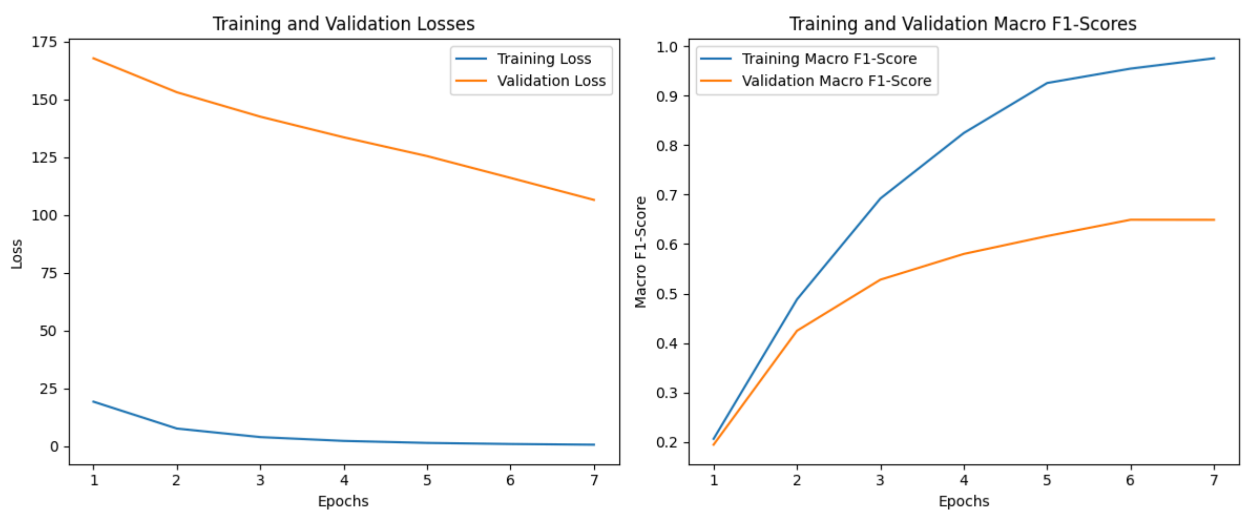
The plot on the left displays the training loss (blue line) and validation loss (orange line) over 7 epochs. The training loss starts around 25 and decreases gradually in the first few epochs as the model learns from the training data. The validation loss starts very high around 150 and decreases rapidly, but remains higher than the training loss eventually. However, the fact that the validation loss starts significantly higher and remains consistently higher than the training loss implies that the model may be overfitting to some extent.

The plot on the right shows the training macro F1-score (blue line) and validation macro F1-score (orange line) over the same 7 epochs. Both the training and validation F1-scores start low but steadily increase as the model improves during training.

However, the validation F1-score lags behind the training score during the entire training process. Key observations from these plots are:

1. The model is learning effectively from the training data, as evidenced by the decreasing losses and increasing F1-scores.
2. There is a gap between the training and validation metrics, indicating some degree of overfitting, where the model performs better on the training data than on the unseen validation data.
3. The validation loss and F1-score appear to plateau towards the later epochs, suggesting that the model may be starting to overfit or reach its maximum performance on the given data.

2) Glove



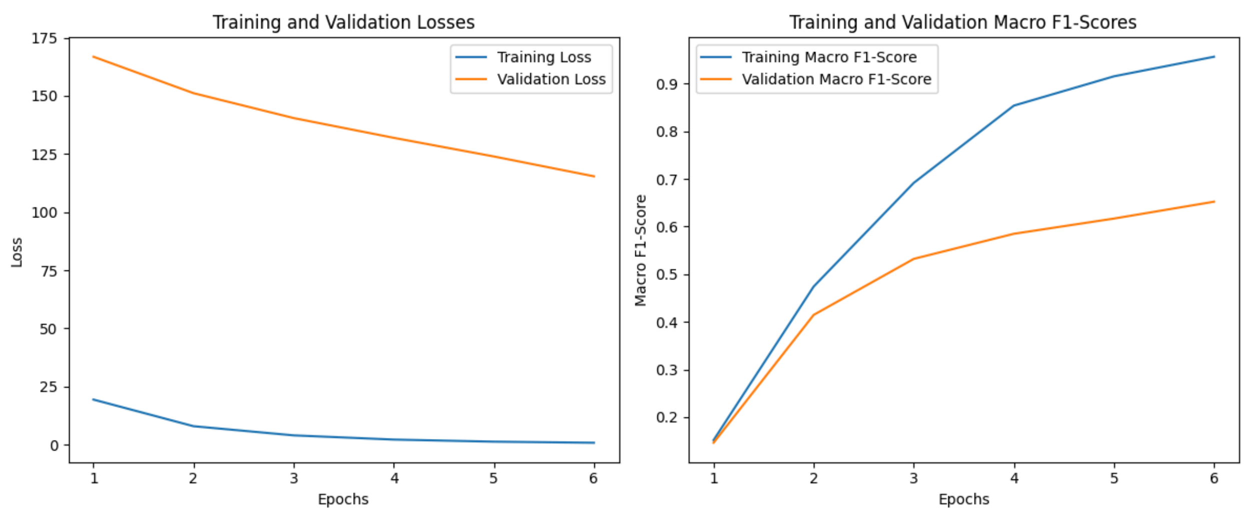
The plot on the left displays the training loss (blue line) and validation loss (orange line) over 7 epochs. The training loss starts around 25 and decreases gradually in the first few epochs as the model learns from the training data. The validation loss starts very high around 170 and decreases rapidly, but remains higher than the training loss eventually. However, the fact that the validation loss starts significantly higher and remains consistently higher than the training loss implies that the model may be overfitting to some extent.

The plot on the right shows the training macro F1-score (blue line) and validation macro F1-score (orange line) over the same 7 epochs. Both the training and validation F1-scores start low but steadily increase as the model improves during training.

However, the validation F1-score lags behind the training score during the entire training process. Key observations from these plots are:

4. The model is learning effectively from the training data, as evidenced by the decreasing losses and increasing F1-scores.
5. There is a gap between the training and validation metrics, indicating some degree of overfitting, where the model performs better on the training data than on the unseen validation data.
6. The validation loss and F1-score appear to plateau towards the later epochs, suggesting that the model may be starting to overfit or reach its maximum performance on the given data.

3) Fast Text



The plot on the left displays the training loss (blue line) and validation loss (orange line) over 6 epochs. The training loss starts around 25 and decreases gradually in the first few epochs as the model learns from the training data. The validation loss starts very high around 170 and decreases rapidly, but remains higher than the training loss eventually. However, the fact that the validation loss starts significantly higher and remains consistently higher than the training loss implies that the model may be overfitting to some extent.

The plot on the right shows the training macro F1-score (blue line) and validation macro F1-score (orange line) over the same 6 epochs. Both the training and validation F1-scores start low but steadily increase as the model improves during training.

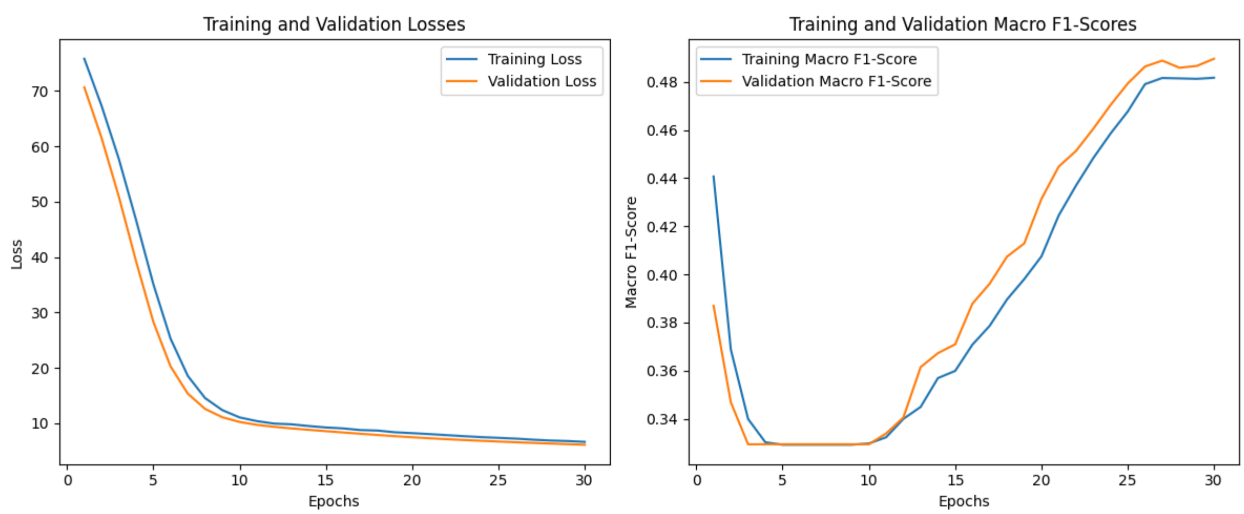
However, the validation F1-score lags behind the training score during the entire training process. Key observations from these plots are:

7. The model is learning effectively from the training data, as evidenced by the decreasing losses and increasing F1-scores.
8. There is a gap between the training and validation metrics, indicating some degree of overfitting, where the model performs better on the training data than on the unseen validation data.
9. The validation loss and F1-score appear to plateau towards the later epochs, suggesting that the model may be starting to overfit or reach its maximum performance on the given data.

Task 2 Plots

a) Vanilla RNN

1) Word2Vec



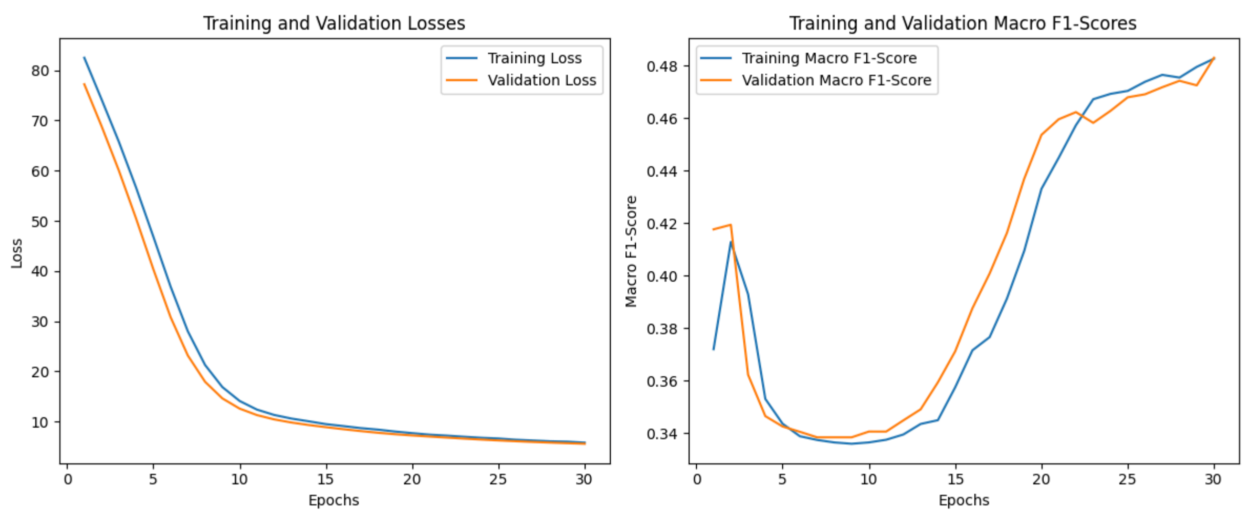
The plot on the left shows the training loss (blue curve) and validation loss (orange curve) over 30 epochs. As expected, the training loss starts high and decreases rapidly in the initial epochs as the model learns from the training data. However, the validation loss remains less than the training loss during the entire training process. This suggests that our model generalizes well on unseen data.

The plot on the right displays the training macro F1-score (blue curve) and validation macro F1-score (orange curve) over the same 30 epochs. The training and validation F1-score initially decreases, reaching a minimum around epoch 10 before eventually increasing. This behavior is contrary to the expected pattern, where the F1 score typically starts low and then gradually improves.

There could be several potential reasons for this behavior:

> Overfitting in Early Epochs: The initial high F1 score might be a result of overfitting to the training data. In the subsequent epochs, as the model encounters validation data, its inability to generalize leads to a dip in performance. However, as training progresses, the model learns to generalize better, leading to a gradual increase in the F1 score.
>
> Learning rate issues: Inappropriate learning rate could cause the model to get stuck in a suboptimal state initially.

2) Glove



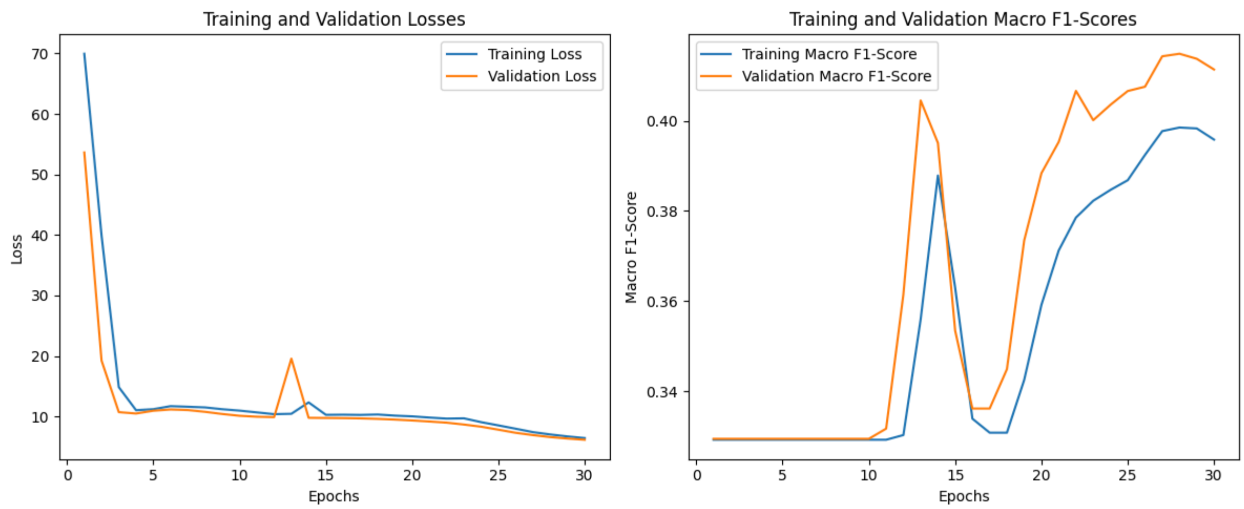
The plot on the left shows the training loss (blue curve) and validation loss (orange curve) over 30 epochs. As expected, the training loss starts high and decreases rapidly in the initial epochs as the model learns from the training data. However, the validation loss remains less than the training loss during the entire training process. This suggests that our model generalizes well on unseen data.

The plot on the right displays the training macro F1-score (blue curve) and validation macro F1-score (orange curve) over the same 30 epochs. The training and validation F1-score initially decreases, reaching a minimum around epoch 10 before eventually increasing. This behavior is contrary to the expected pattern, where the F1 score typically starts low and then gradually improves.

There could be several potential reasons for this behavior:

> Overfitting in Early Epochs: The initial high F1 score might be a result of overfitting to the training data. In the subsequent epochs, as the model encounters validation data, its inability to generalize leads to a dip in performance. However, as training progresses, the model learns to generalize better, leading to a gradual increase in the F1 score.
>
> Learning rate issues: Inappropriate learning rate could cause the model to get stuck in a suboptimal state initially.

3) Fasttext

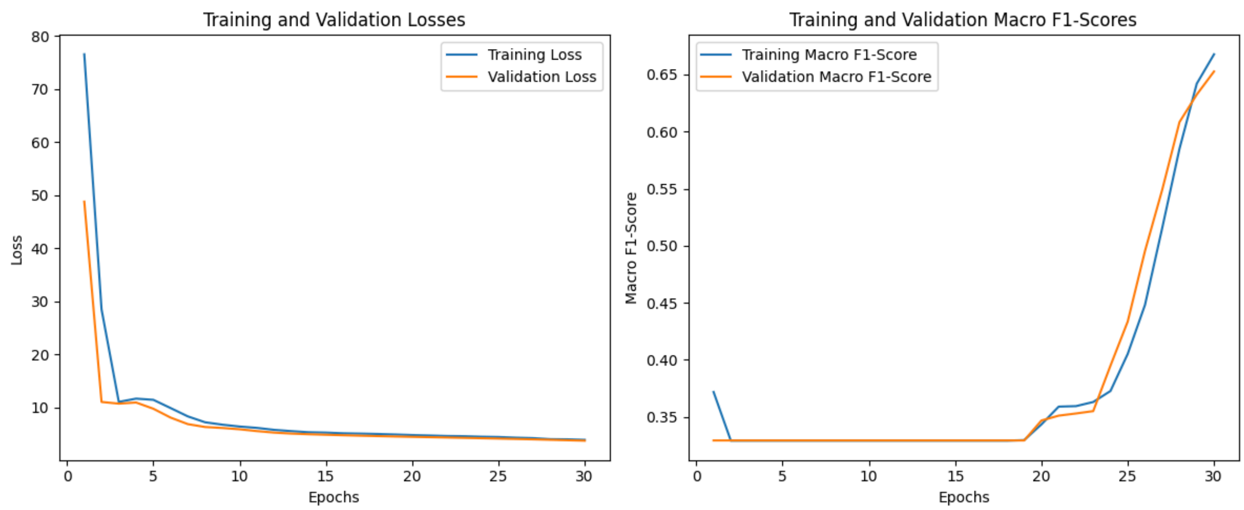


The left subplot shows the training loss (blue line) and validation loss (orange line) over the course of 30 epochs. We can observe that both losses start relatively high but decrease rapidly in the initial epochs as the model learns from the training data. The validation loss remains slightly lower than the training loss, which indicates that the model generalizes well to unseen data.

The right subplot shows the training macro F1-score (blue line) and validation macro F1-score (orange line) over the same 30 epochs. The F1 score remains constant initially as the model begins learning patterns from the data. Then there is a sharp increase in the F1 score indicating significant learning progress, where the model effectively captures relevant patterns. Then there is a rapid decrease in the F1 score following the peak, suggesting issues such as overfitting or encountering difficult samples. Towards the end of training, the F1 score gradually increases again, indicating the model's adjustments to generalize better.

b) Model- LSTM

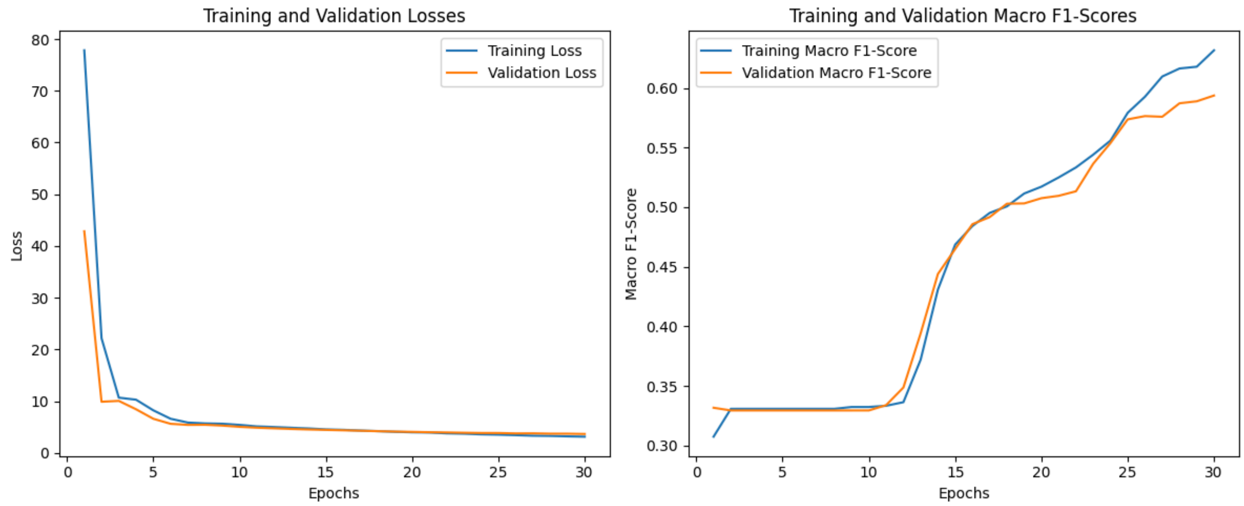
1) Word2Vec



The left subplot shows the training loss (blue line) and validation loss (orange line) over the course of 30 epochs. We can observe that both losses start relatively high but decrease rapidly in the initial epochs as the model learns from the training data. The validation loss remains slightly lower than the training loss, which indicates that the model generalizes well to unseen data.

The right subplot shows the training macro F1-score (blue line) and validation macro F1-score (orange line) over the same 30 epochs. We can see that both the training and validation F1-scores remain constant initially, indicating that the model learns slowly initially. But these scores gradually increase as the model training progresses, with the validation F1-score lagging slightly behind the training F1-score after 30 epochs.

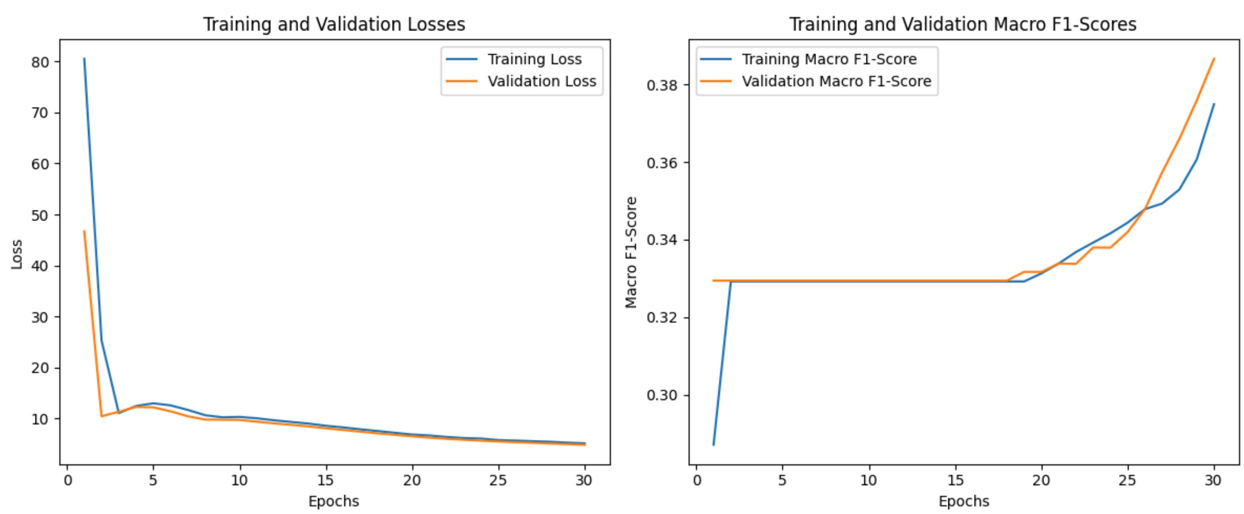
2) Glove



The left subplot shows the training loss (blue line) and validation loss (orange line) over the course of 30 epochs. We can observe that both losses start relatively high but decrease rapidly in the initial epochs as the model learns from the training data. The validation loss remains slightly lower than the training loss, which indicates that the model generalizes well to unseen data.

The right subplot shows the training macro F1-score (blue line) and validation macro F1-score (orange line) over the same 30 epochs. We can see that both the training and validation F1-scores remain constant initially, indicating that the model learns slowly initially. But these scores gradually increase as the model training progresses, with the validation F1-score lagging slightly behind the training F1-score after 30 epochs.

3) Fasttext

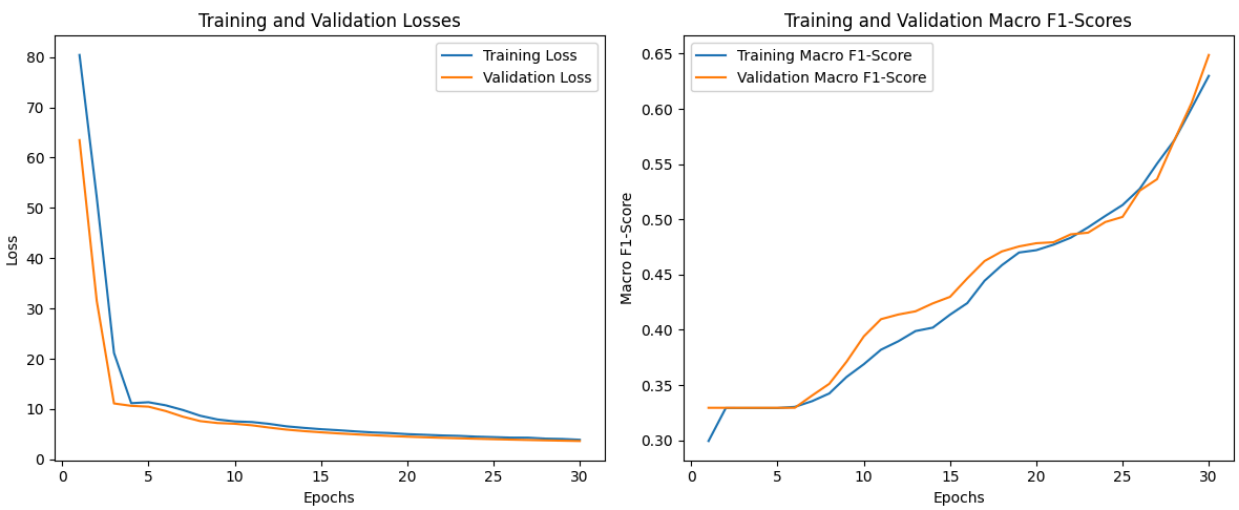


The left subplot shows the training loss (blue line) and validation loss (orange line) over the course of 30 epochs. We can observe that both losses start relatively high but decrease rapidly in the initial epochs as the model learns from the training data. The validation loss remains slightly lower than the training loss, which indicates that the model generalizes well to unseen data.

The right subplot shows the training macro F1-score (blue line) and validation macro F1-score (orange line) over the same 30 epochs. We can see that both the training and validation F1-scores remain constant initially, indicating that the model learns slowly initially. But these scores gradually increase as the model training progresses, with the training F1-score lagging slightly behind the validation F1-score after 30 epochs. This means that the model generalizes well on unseen data.

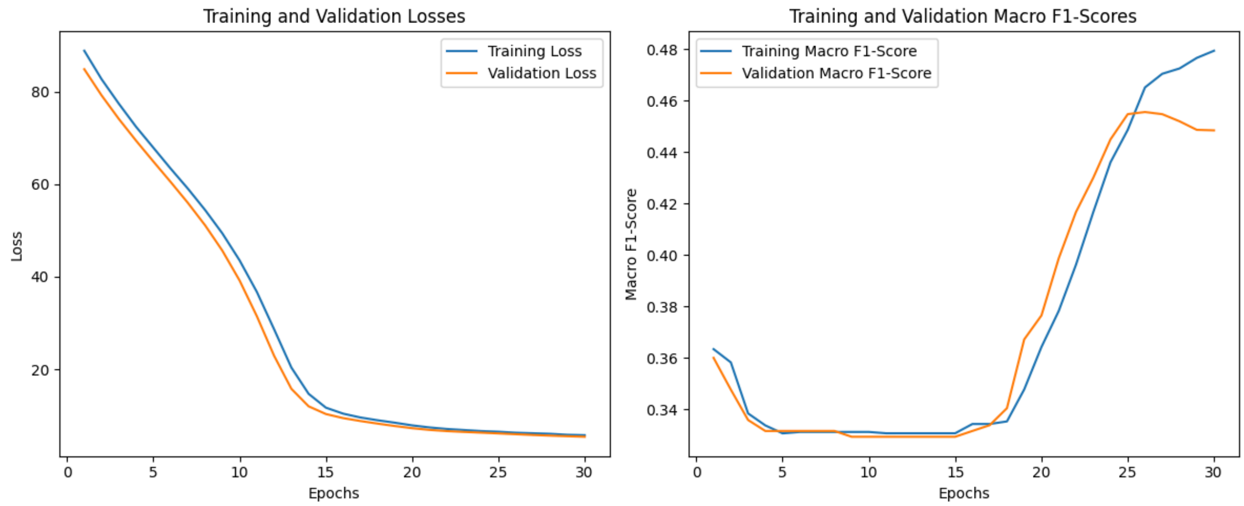
c) Model - GRU

1) Word2Vec



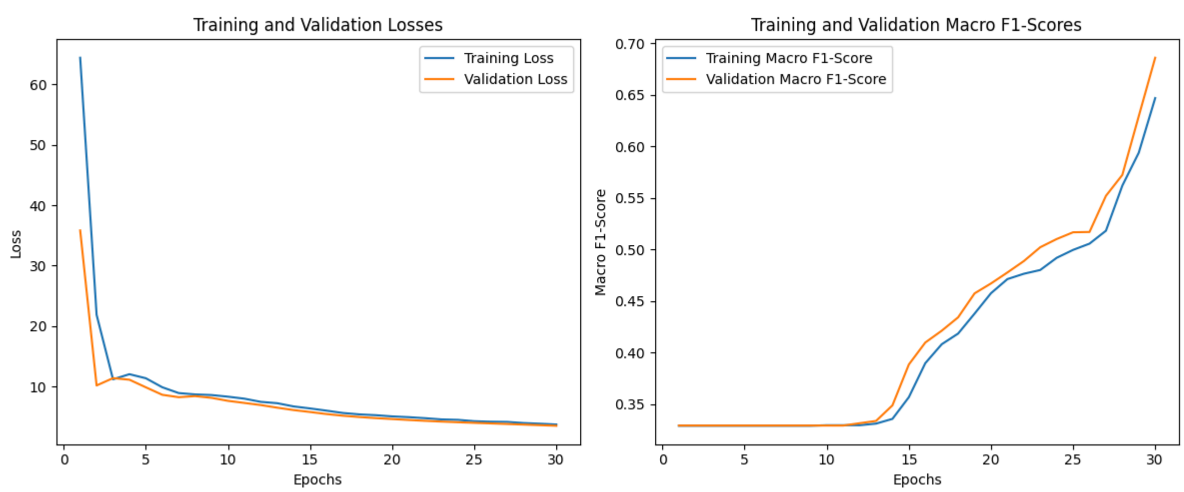
The left plot displays training and validation losses over 30 epochs. The training loss starts high but rapidly decreases as the model learns, while the validation loss is almost equal to the training loss indicating the model's generalization performance on unseen data. The right plot shows training and validation macro F1-scores, both initially low but steadily increasing as the model improves. The validation score is similar to the training score throughout Overall, the plots suggest the model effectively learns from the training data, demonstrated by decreasing losses and increasing F1-scores.

2) Glove



The left plot displays training and validation losses over 30 epochs. The training loss starts high but rapidly decreases as the model learns, while the validation loss is almost equal to the training loss indicating the model's generalization performance on unseen data. The right plot shows training and validation macro F1-scores, both initially low but steadily increasing as the model improves. The validation score is similar to the training score throughout except at the end where it dips due to over-fitting.

3) Fasttext

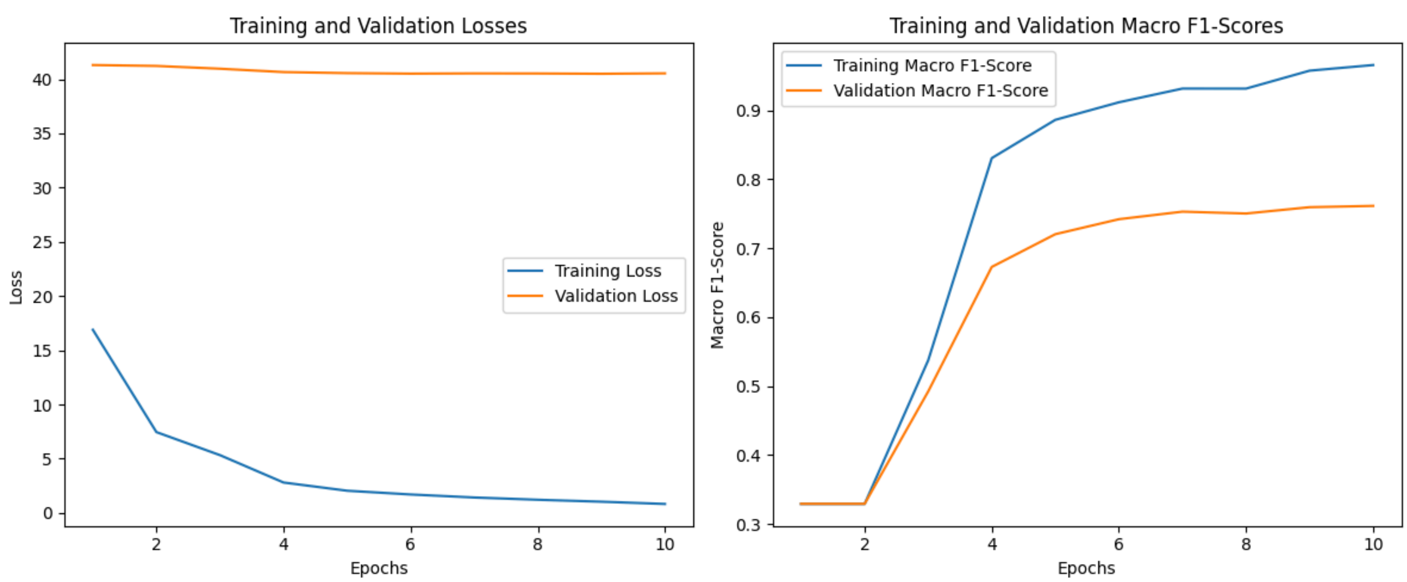


The left plot displays training and validation losses over 30 epochs. The training loss starts high but rapidly decreases as the model learns, while the validation loss is almost equal to the training loss indicating the model's generalization performance on unseen data. The right plot shows training and validation macro F1-scores, both initially low but steadily increasing as the model improves. The validation score is similar to the training score throughout except at the end where it beats the training F1. Overall, the plots suggest the model effectively learns from the training data, demonstrated by decreasing losses and increasing F1-scores.

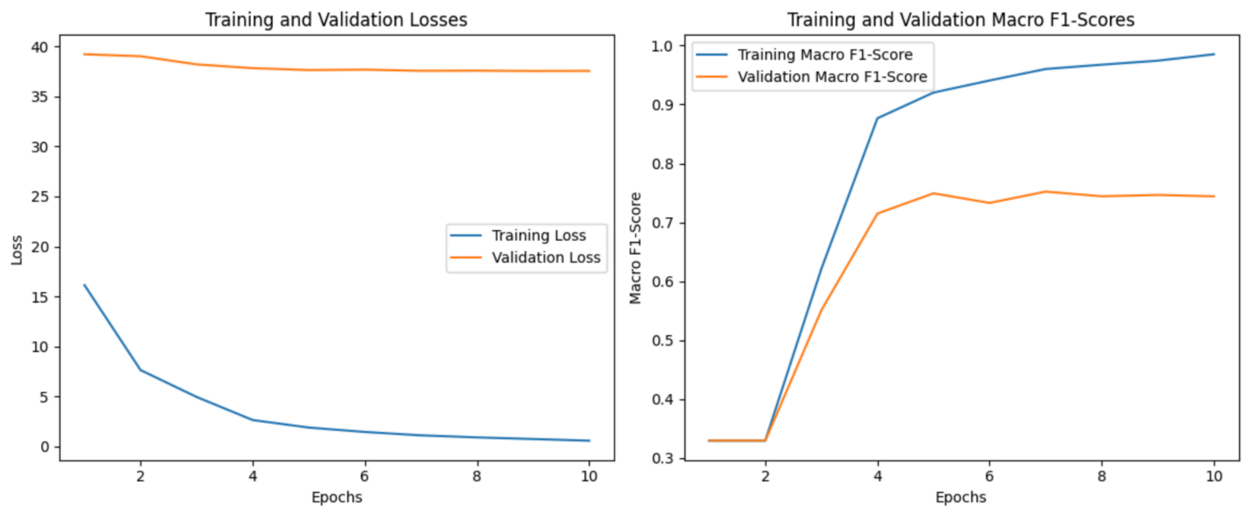
d) Model- BiLSTM CRF

1) Word2Vec

The plot on the left displays the training loss (blue line) and validation loss (orange line) over 10 epochs. The training loss starts around 15 and decreases gradually in the first few epochs as the model learns from the training data
The plot on the right shows the training macro F1-score (blue line) and validation macro F1-score (orange line) over the same 10 epochs. Both the training and validation F1-scores start low but steadily increase as the model improves during training. However, the validation F1-score lags behind the training score during the entire training process.

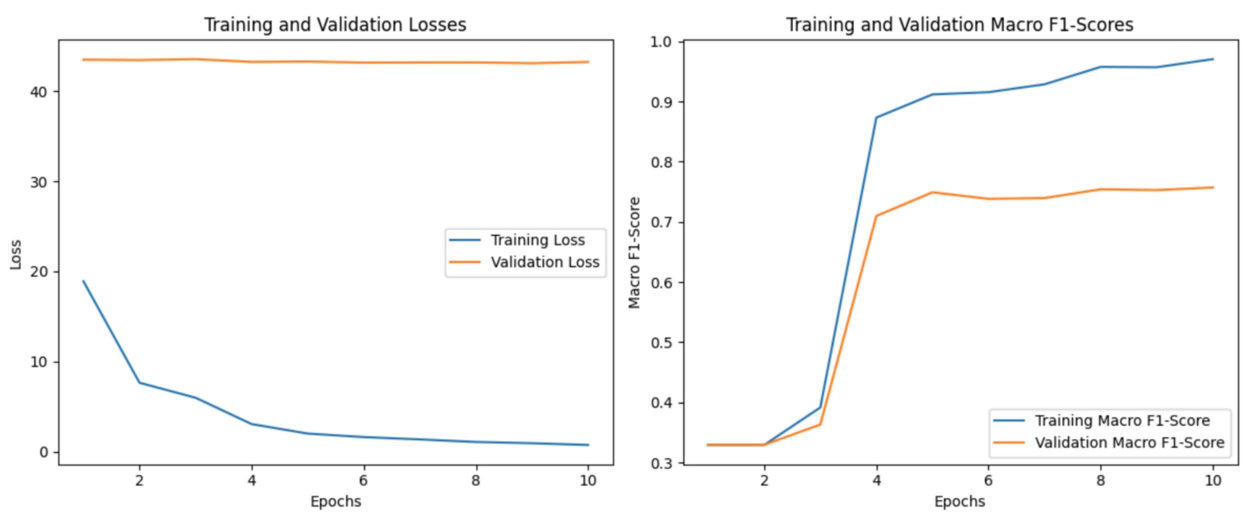


2)Glove



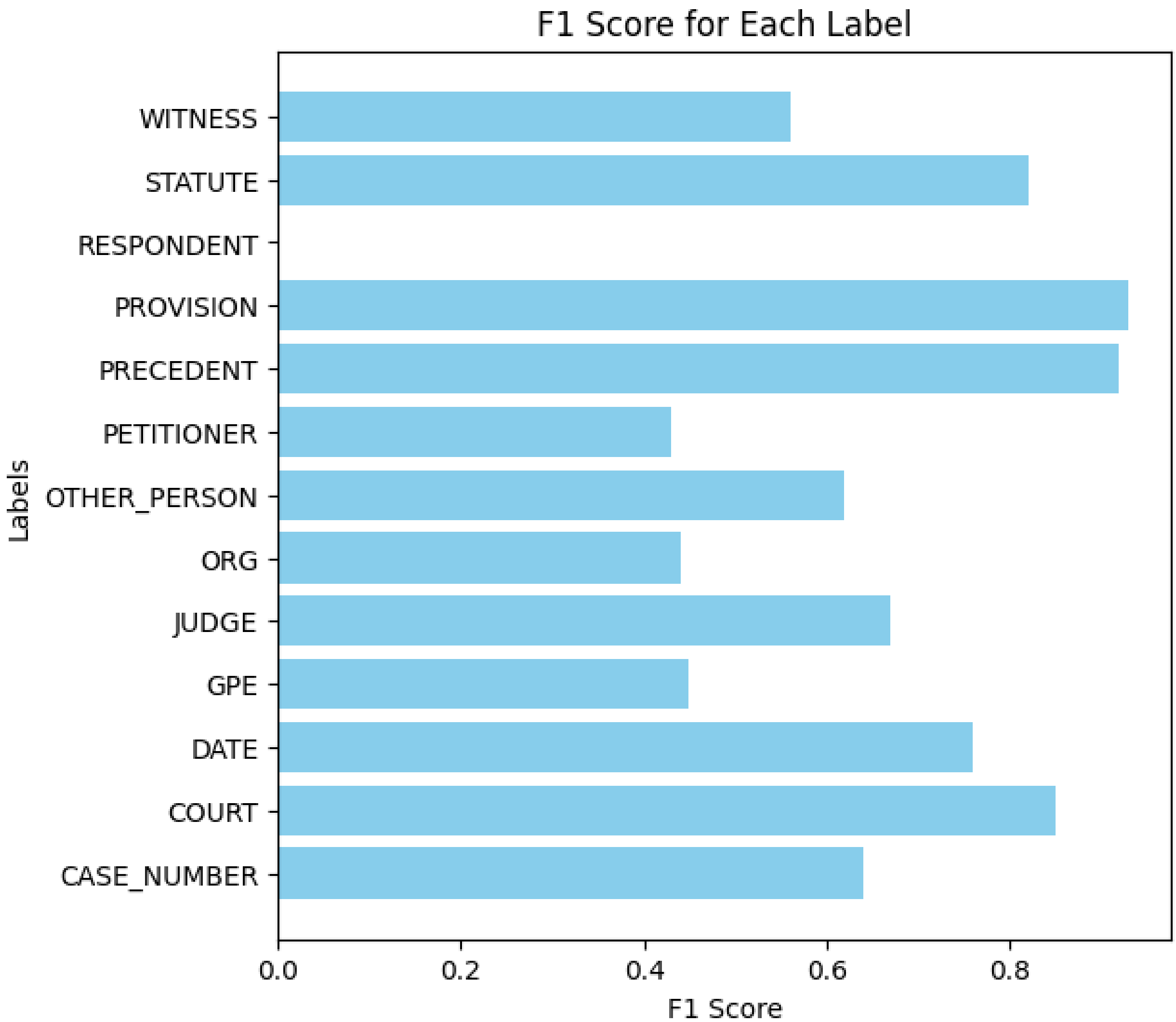
The plot on the left displays the training loss (blue line) and validation loss (orange line) over 10 epochs. The training loss starts around 15 and decreases gradually in the first few epochs as the model learns from the training data while the validation loss decreases slowly.
The plot on the right shows the training macro F1-score (blue line) and validation macro F1-score (orange line) over the same 10 epochs. Both the training and validation F1-scores start low but steadily increase as the model improves during training. However, the validation F1-score lags behind the training score during the entire training process

3) FastText



The plot on the left displays the training loss (blue line) and validation loss (orange line) over 10 epochs. The training loss starts around 20 and decreases gradually in the first few epochs as the model learns from the training data while the validation loss decreases slowly.
The plot on the right shows the training macro F1-score (blue line) and validation macro F1-score (orange line) over the same 10 epochs. Both the training and validation F1-scores start low but steadily increase as the model improves during training. However, the validation F1-score lags behind the training score during the entire training process

F1-Plot for Task-1



Contribution

1. Part-1 Task1 - Nalish Jain(2021543)
2. Part-1 Task2 - Shobhit Pandey(2021287), Aniket Malik(2021231)
3. Part-2 RNN - Aniket Malik(2021231)
4. Part-2 LSTM -Shobhit Pandey(2021287)
5. Part-2 GRU - Nalish Jain(2021543)
6. Part-3 - Sanmay Sood(2021095)
7. Inference - Nalish Jain(2021543) and Sanmay Sood(2021095)
8. Report - Aniket Malik(2021231), Shobhit Pandey(2021287)

CRF Layer Reference

1. https://github.com/ngoquanghuy99/POS-Tagging-BiLSTM-CRF/blob/main/preprocess.py
2. https://pytorch.org/tutorials/beginner/nlp/advanced_tutorial.html