A2 Group 20 Report

Task 1

- Padded all input to a fixed length.

Dataset after BIO Encoding


```
{
    "dcf83b5e00704559a3a63a85d1e3e18a": {
        "text": "The exercise of revisional jurisdiction in such a case is taken away by the proviso inserted under sub-section (1) of Section 115 CPC.",
        "labels": [
            "O",
            "O",
            "I_PROVISION",
            "O",
            "O",
            "O",
            "O",
            "O",
            "O",
            "O",
            "O",
            "O",
            "O",
            "O",
            "O",
            "O",
            "O",
            "B_PROVISION",
            "I_PROVISION",
            "I_PROVISION",
            "I_PROVISION",
            "I_PROVISION",
```

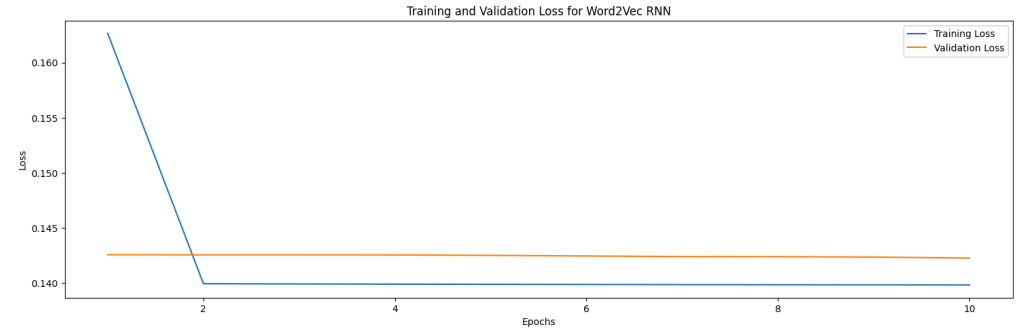
```
            "O"
        ]
    },
    "6fa04d6a7dec438fa124a4f851165abc": {
        "text": "His Lordship has observed as follows:\n\"The amount
received by the claimant on the Life Insurance of the deceased is not
deductible from the compensation computed under the Motor Vehicles
Act.\"",
        "labels": [
            "O",
            "O",
            "O",
            "O",
            "O",
            "O",
            "O",
            "O",
            "O",
            "O",
            "O",
            "O",
            "O",
            "O",
            "O",
            "O",
            "O",
            "O",
            "O",
            "O",
            "O",
            "O",
            "O",
            "O",
            "O",
            "O",
            "O",
            "O",
            "O",
            "B_STATUTE",
            "I_STATUTE",
            "O"
```

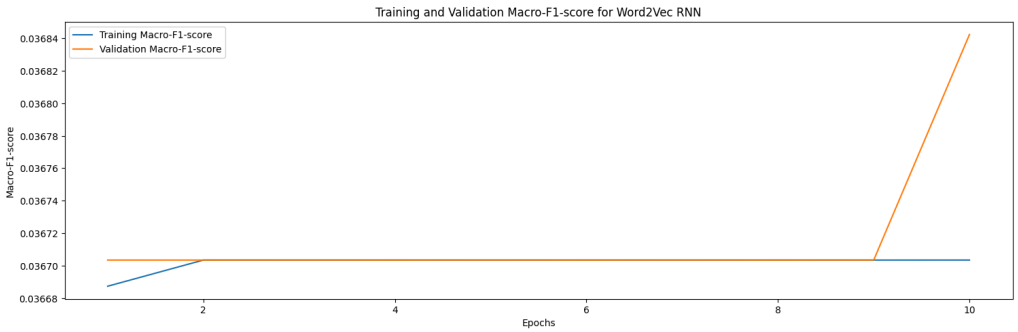
```
            ]
        }
}
```

Plots

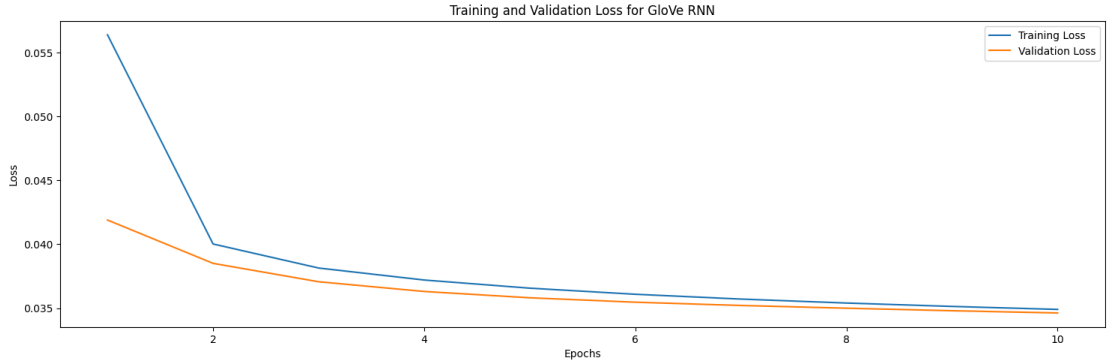
RNN

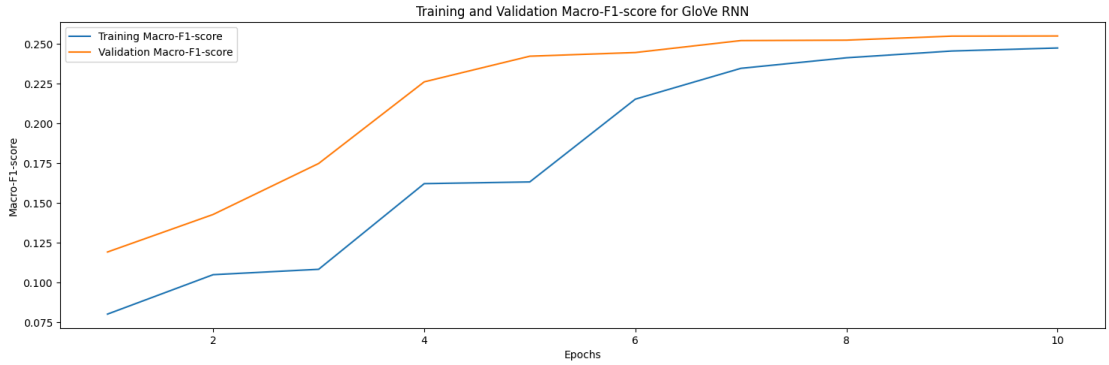
- **Word2Vec :**



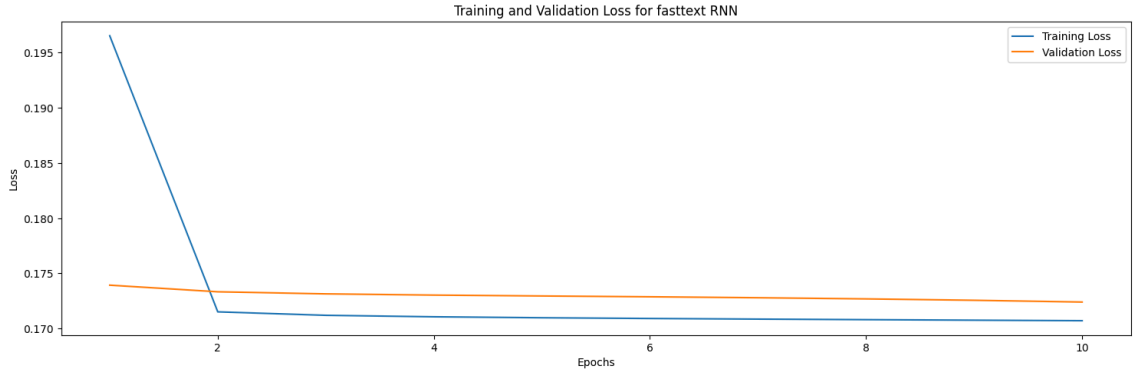


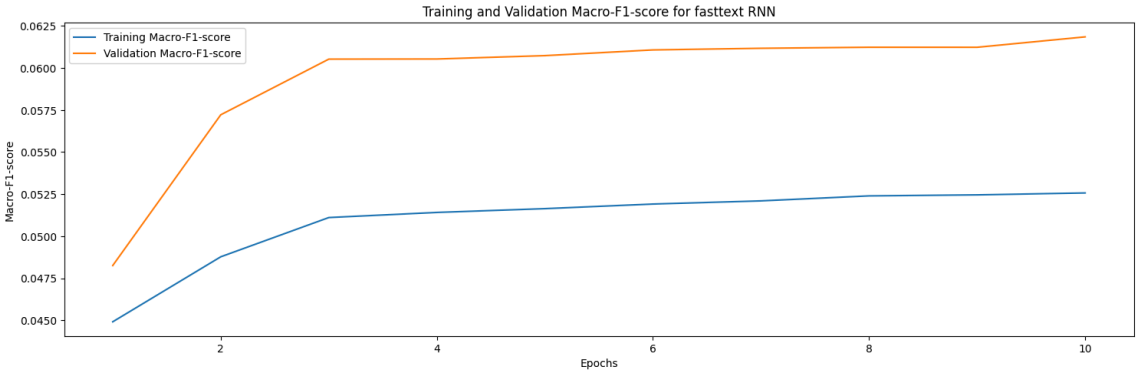
- **GloVe :**





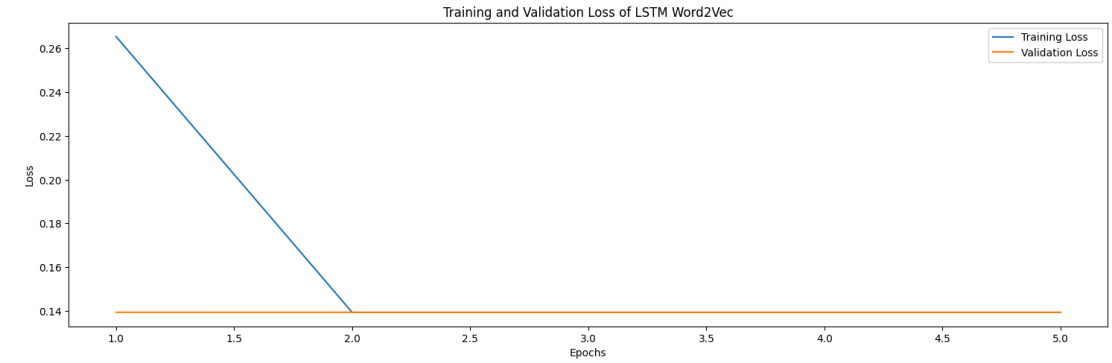
- **Fasttext :**

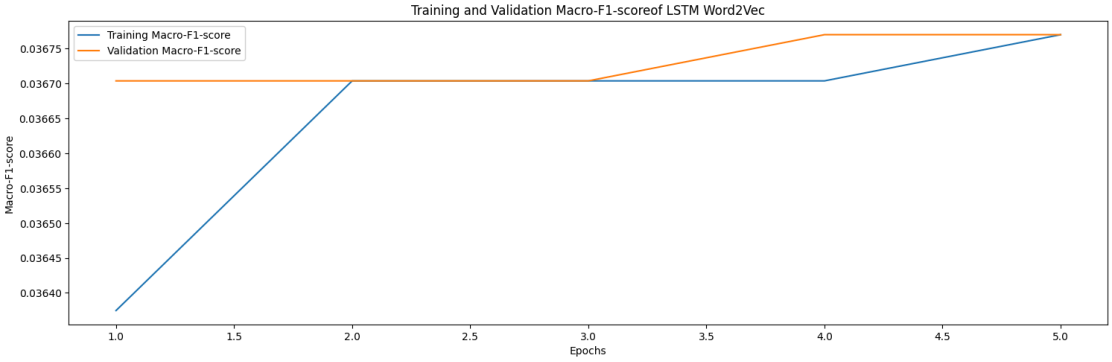




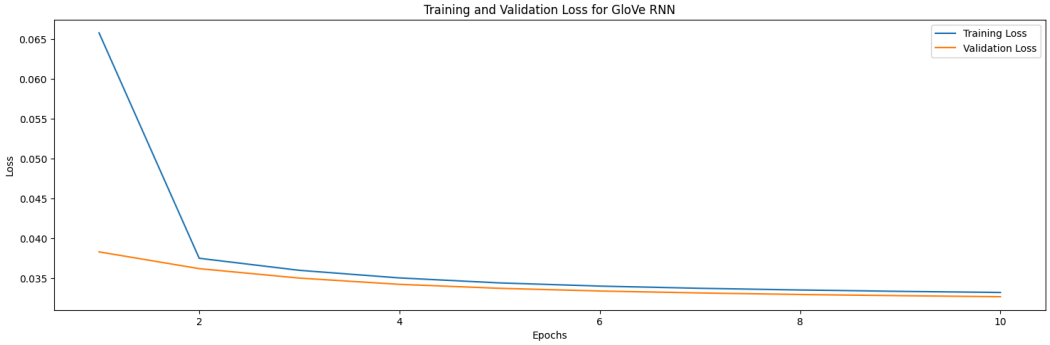
LSTM

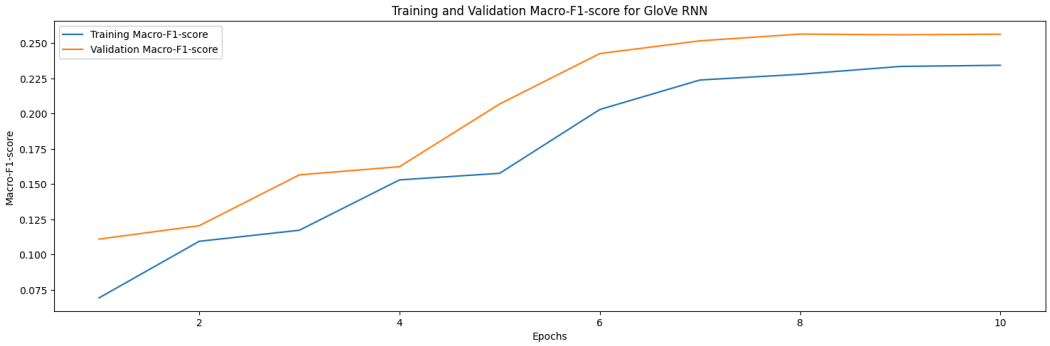
- **Word2Vec :**



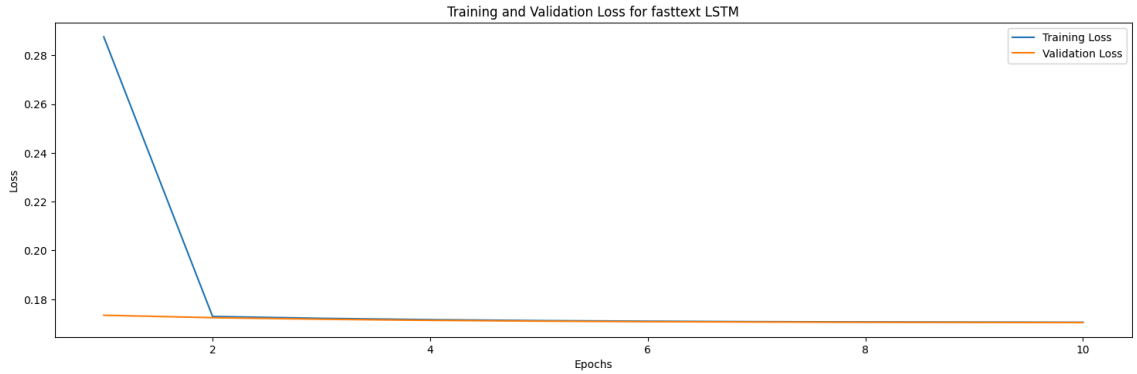


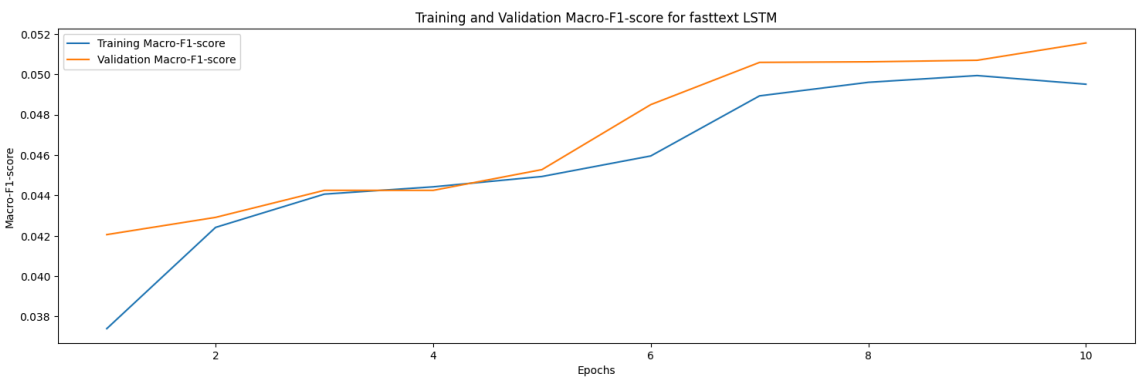
- **GloVe :**





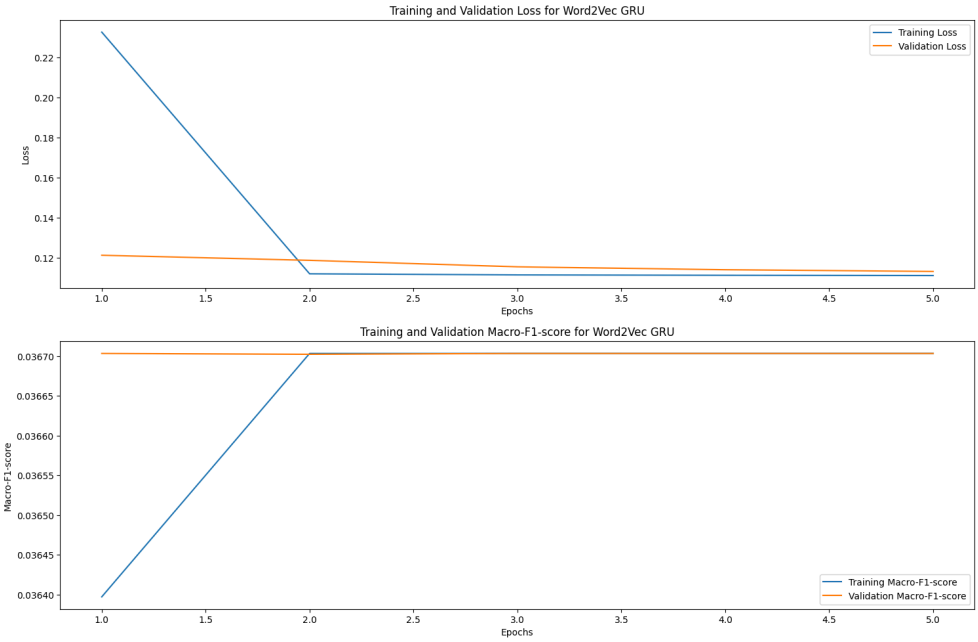
- **Fasttext :**



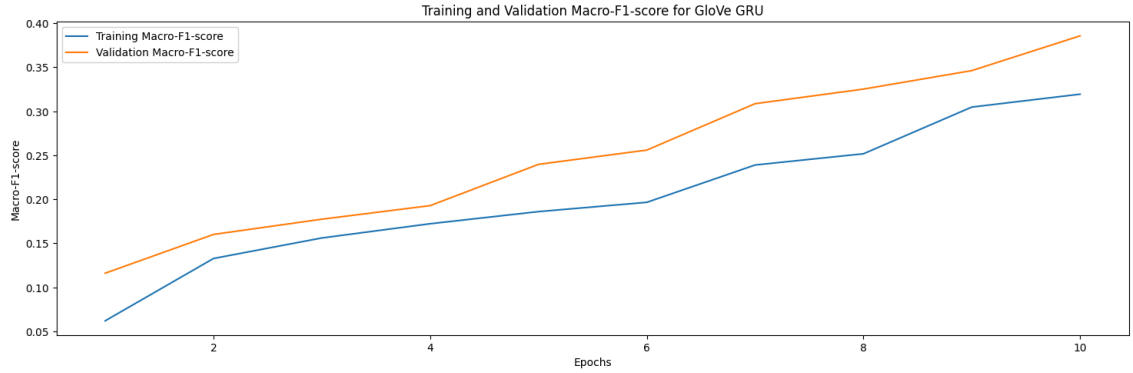


GRU

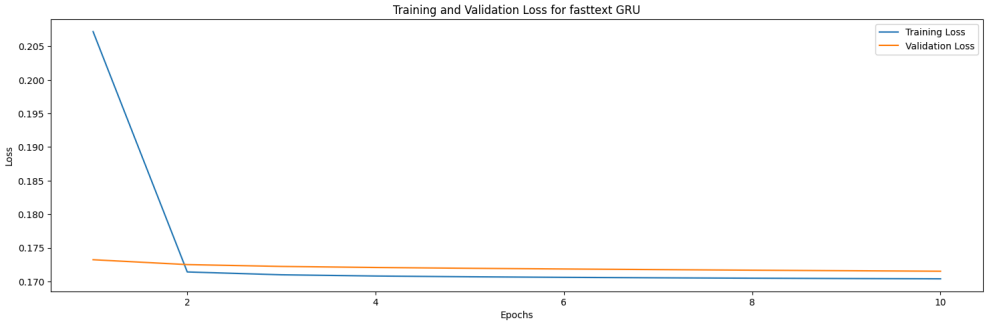
- **Word2Vec :**

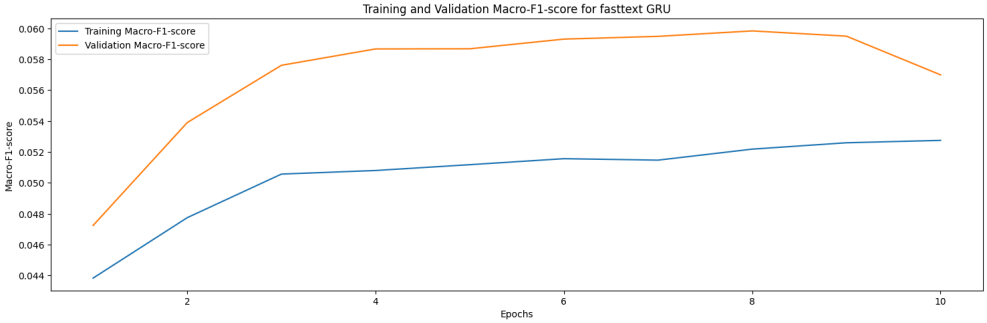


- **GloVe :**



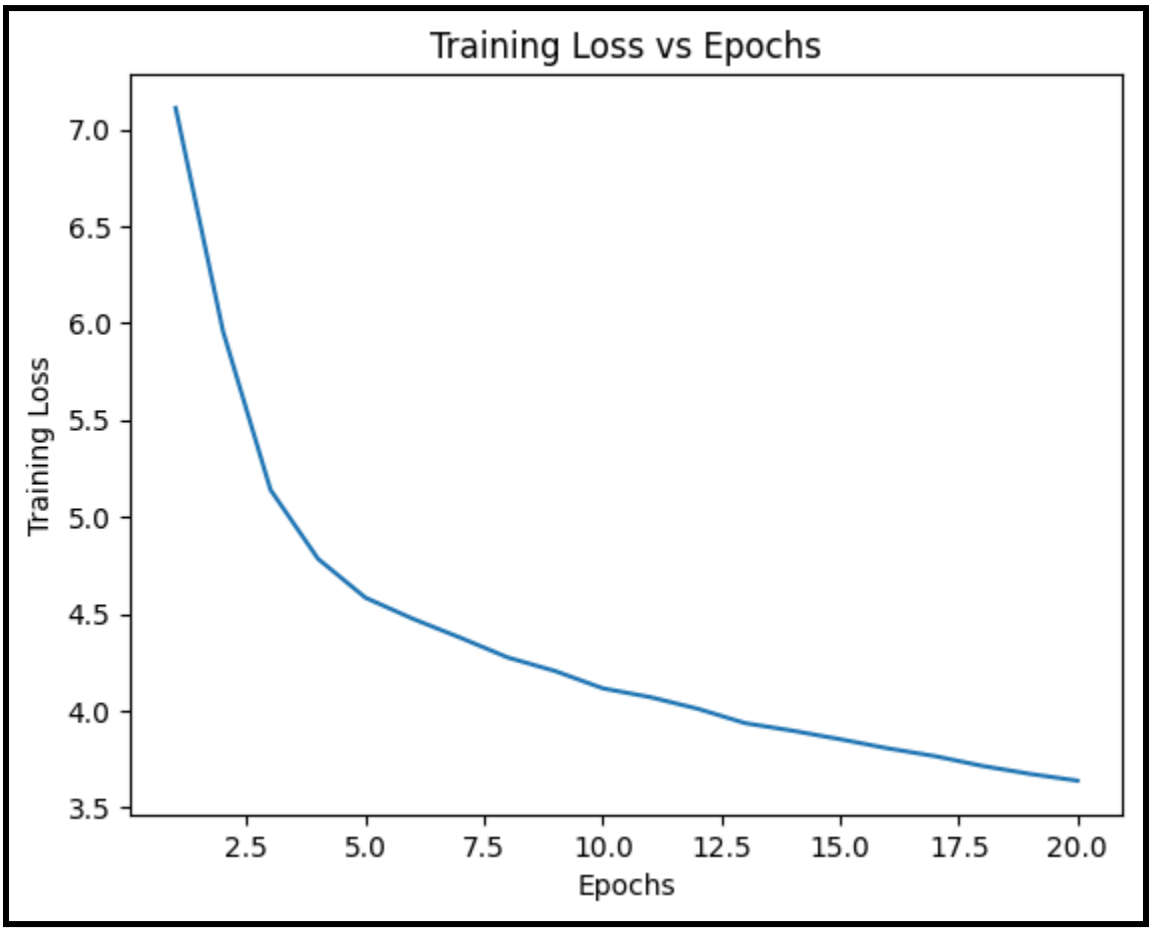
- **Fasttext :**



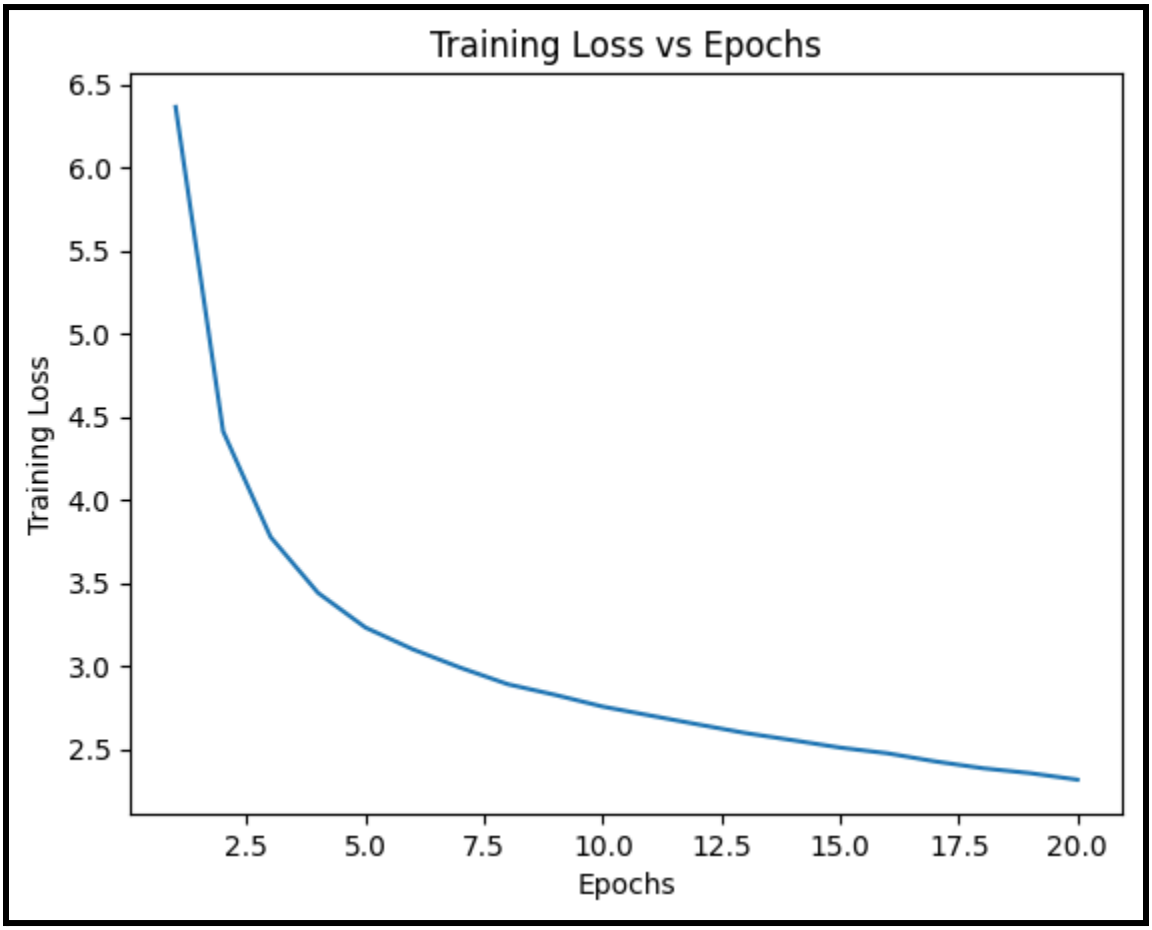


BiLSTM CRF

- **Word2Vec :**



- **GloVe :**


Training Loss vs Epochs
Training Loss
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0
Epochs

- **Fasttext :**

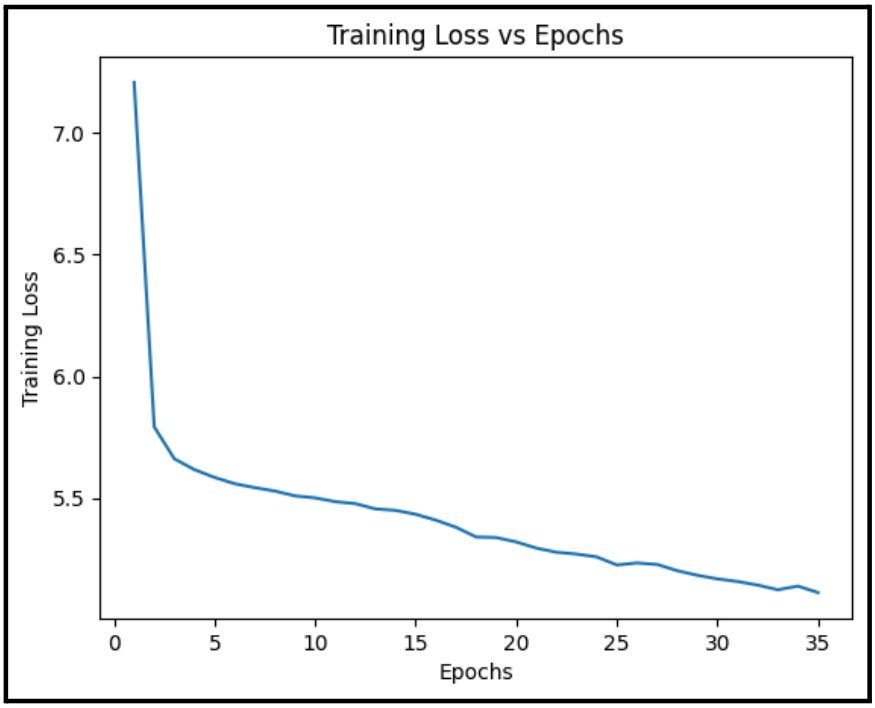
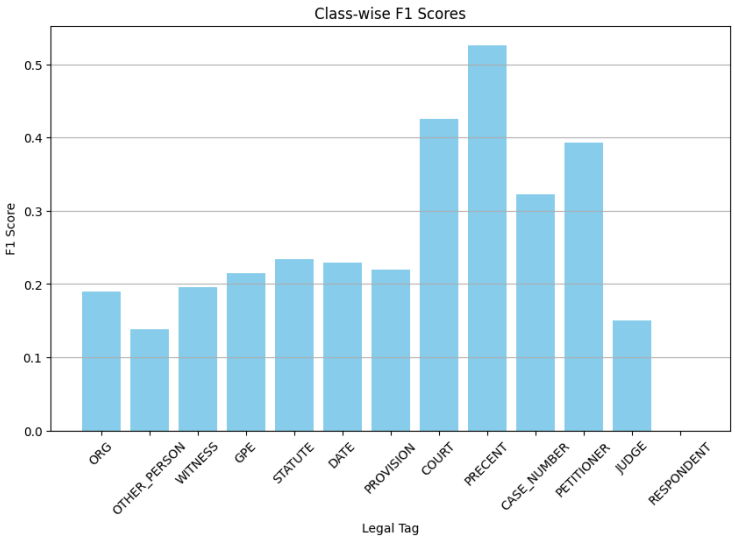


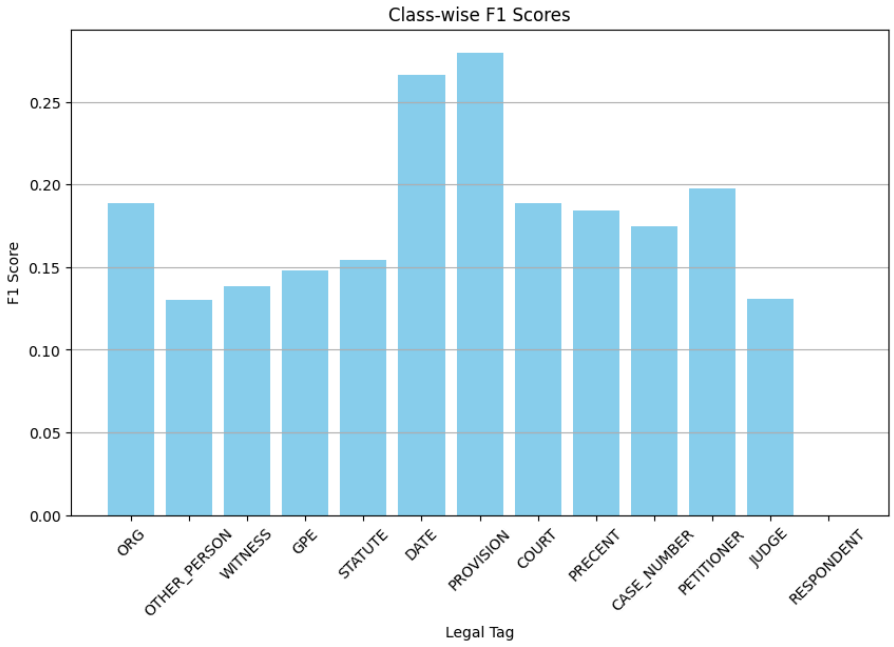
Table for Task-1

Model_Type	Embedding_used	Accuracy	Macro_F1
RNN	**Word2Vec**	0.979415718905563	0.03665188258262209
RNN	**GloVe**	0.9904315492411855	0.2549428886103641
RNN	**Fasttext**	0.971415767353415	0.05903285792539194
LSTM	**Word2Vec**	0.979415718905563	0.03665188258262209

LSTM	**GloVe**	0.9906556205564235	0.2563795149159976
LSTM	**Fasttext**	0.9713612635199786	0.04736354529562763
GRU	**Word2Vec**	0.9826558711109284	0.03671303973957373
GRU	**GloVe**	0.9940816242653405	0.21563522579720295
GRU	**Fasttext**	0.971409711371922	**0**.052727638229631765
BiLISTM CRF	**Word2Vec**		
BiLISTM CRF	**GloVe**		
BiLISTM CRF	**Fasttext**		

Label-Wise F1 score for Dataset 1





Task 2

- Padded all input to a fixed length.

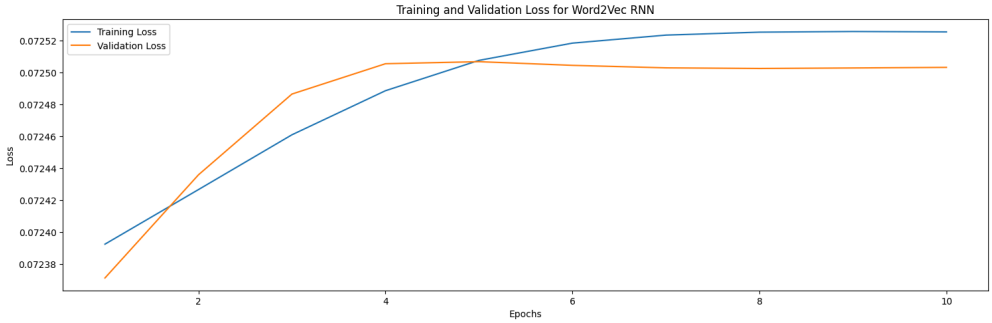
Dataset after BIO Encoding

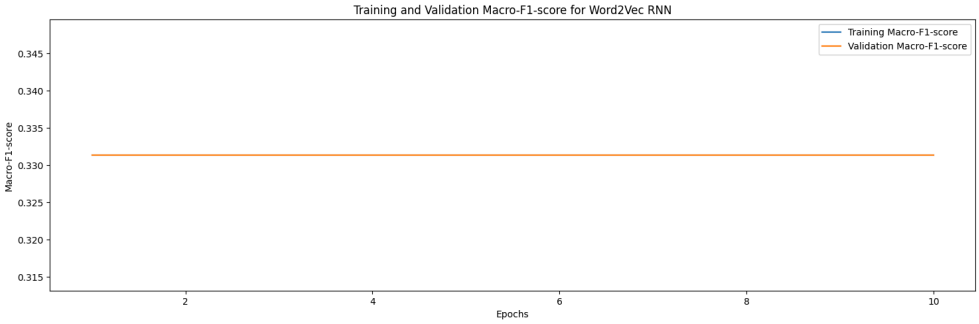
```
{1: {'text': 'I charge it at night and skip taking the cord with me
because of the good battery life .', 'labels': ['O', 'O', 'O', 'O', 'O',
'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'B', 'I', 'O']},
2: {'text': 'it is of high quality , has a killer GUI , is extremely
stable , is highly expandable , is bundled with lots of very good
applications , is easy to use , and is absolutely gorgeous .', 'labels':
['O', 'O', 'O', 'O', 'B', 'O', 'O', 'O', 'O', 'B', 'O', 'O', 'O', 'O',
'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'O', 'B', 'O', 'O',
'O', 'O', 'B', 'O', 'O', 'O', 'O', 'O', 'O']}...
```

Plots

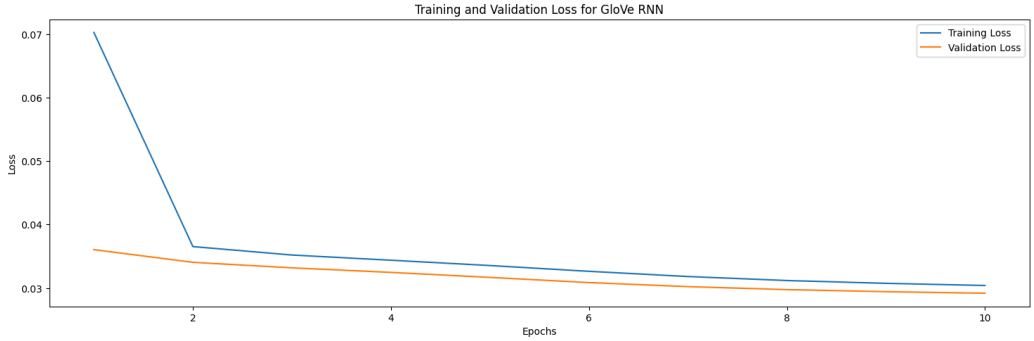
RNN

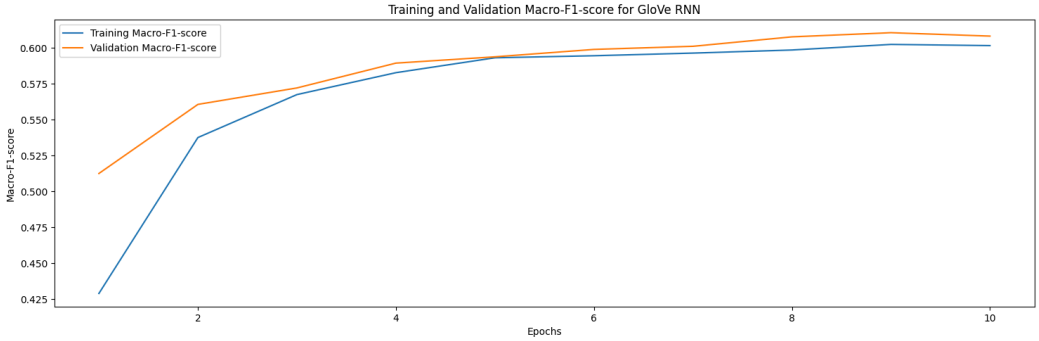
- **Word2Vec :**





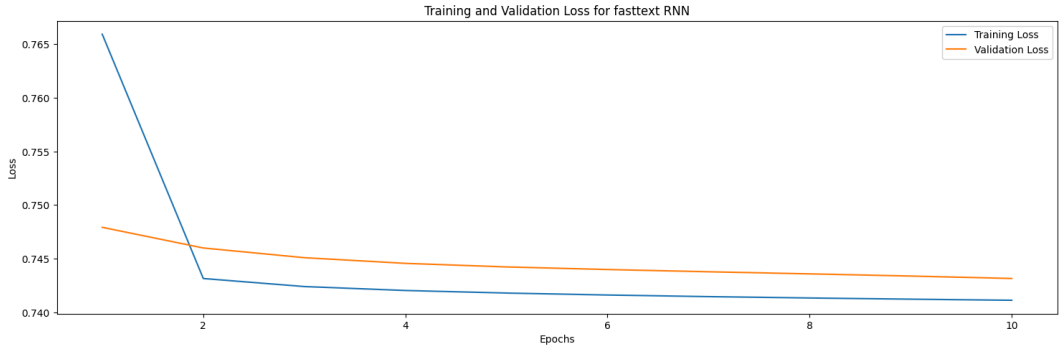
- **GloVe :**





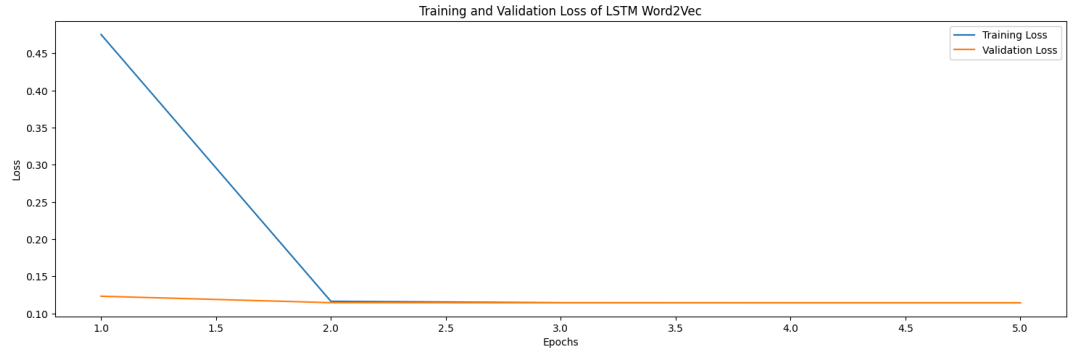
- **Fasttext :**

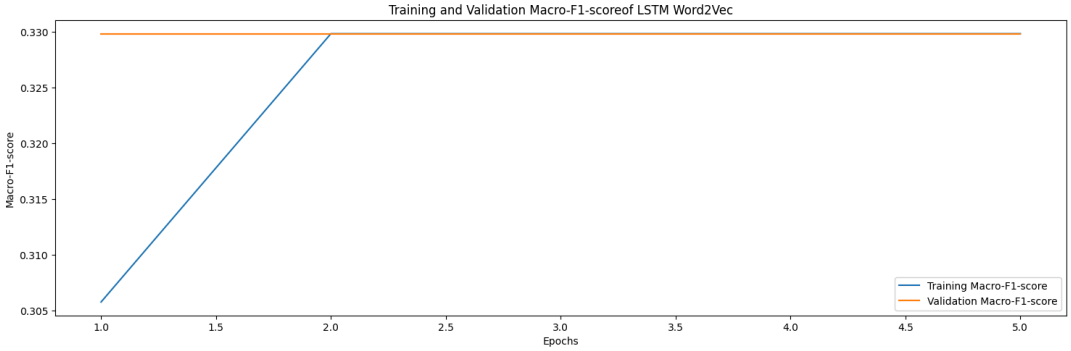
-



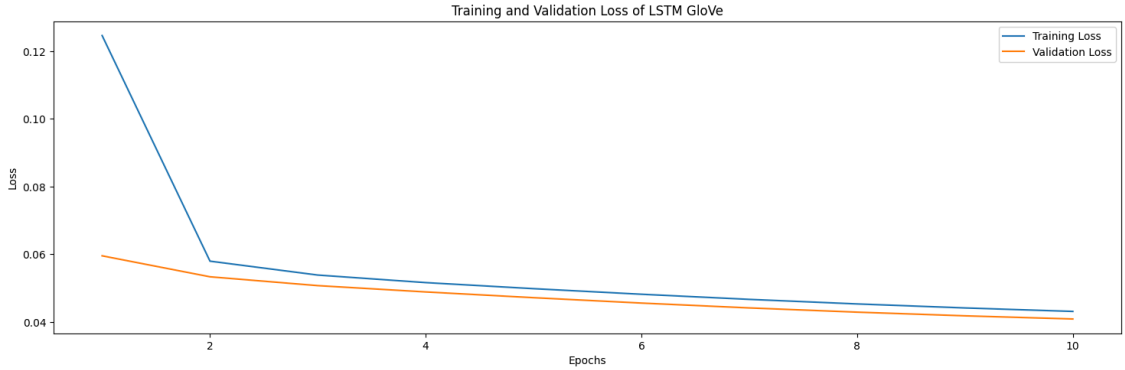
LSTM

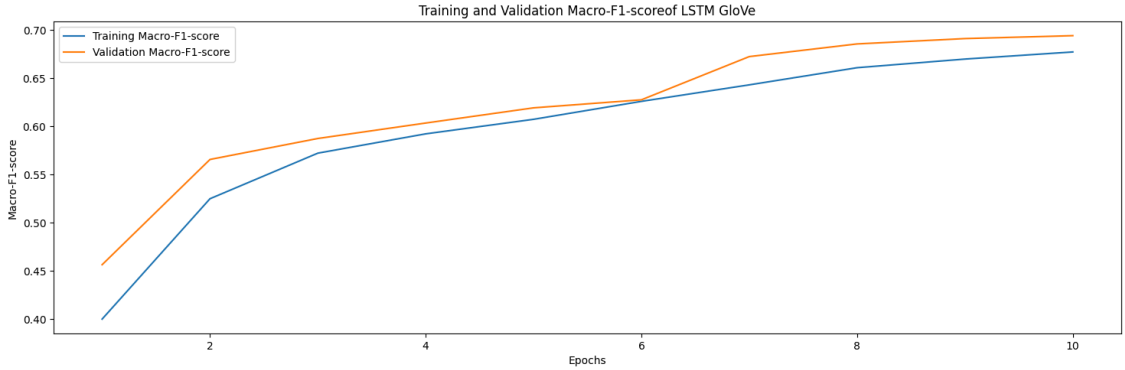
- **Word2Vec :**



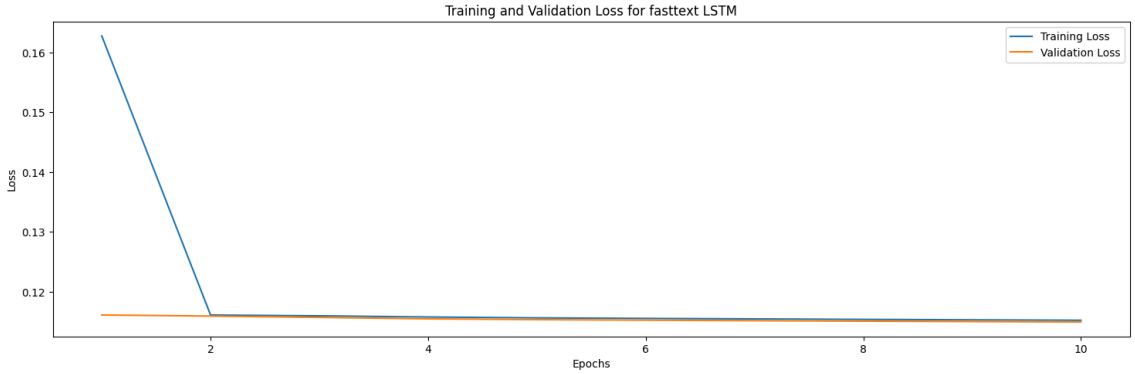


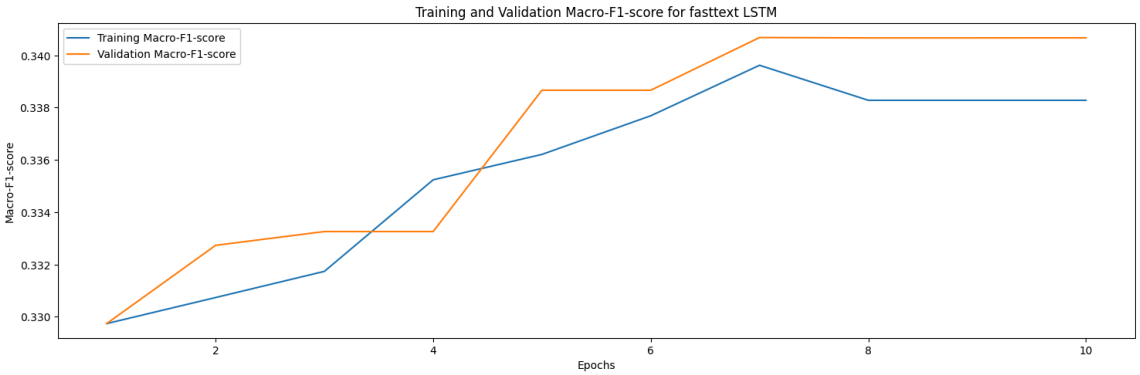
- **GloVe :**





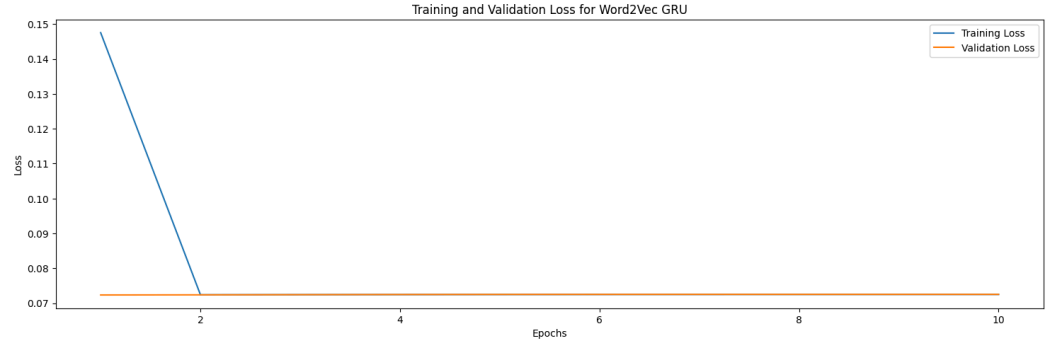
- **Fasttext :**

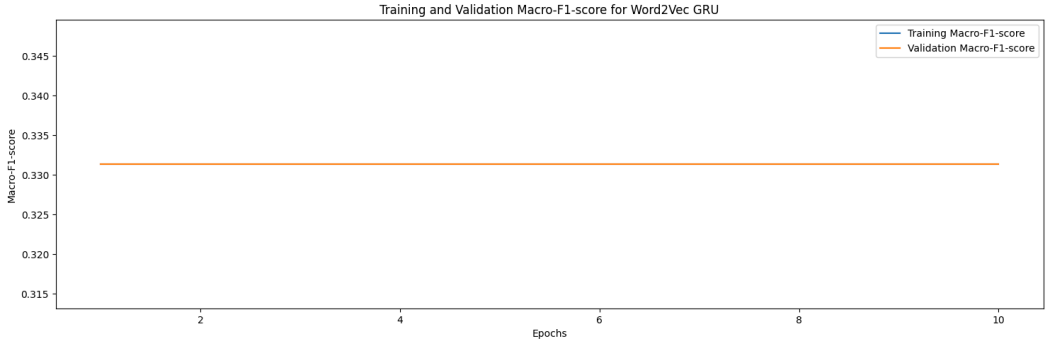




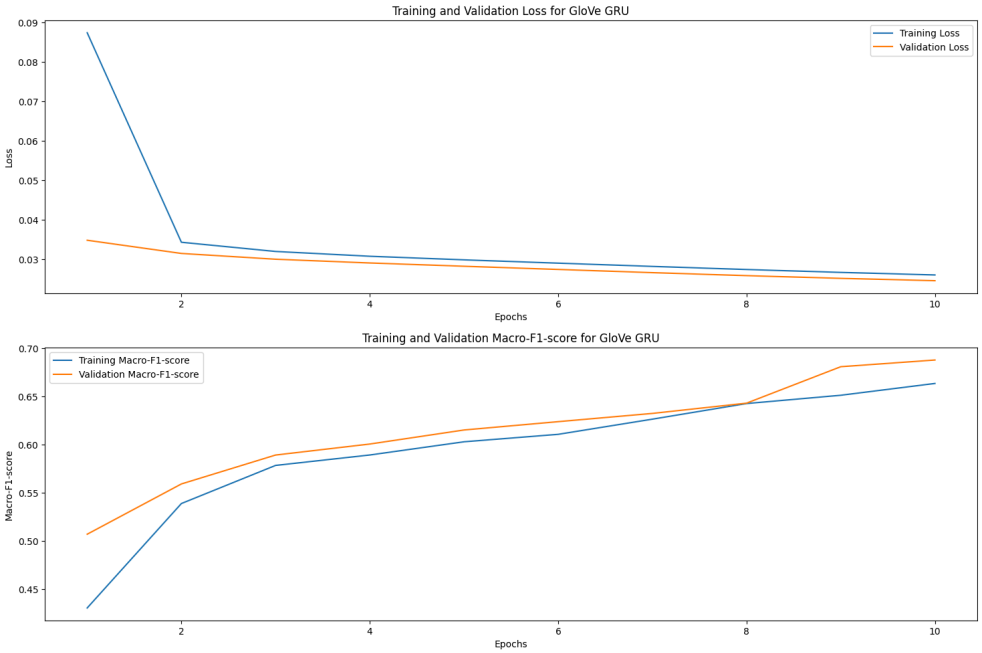
GRU

- **Word2Vec :**

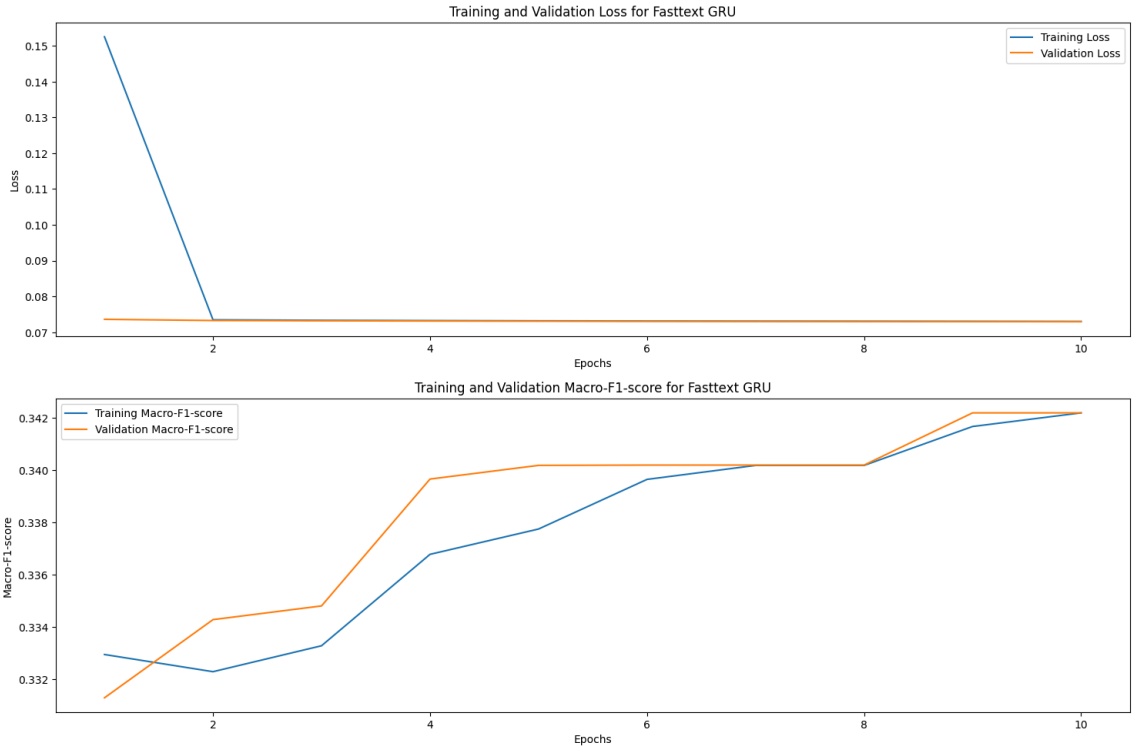




- **GloVe :**

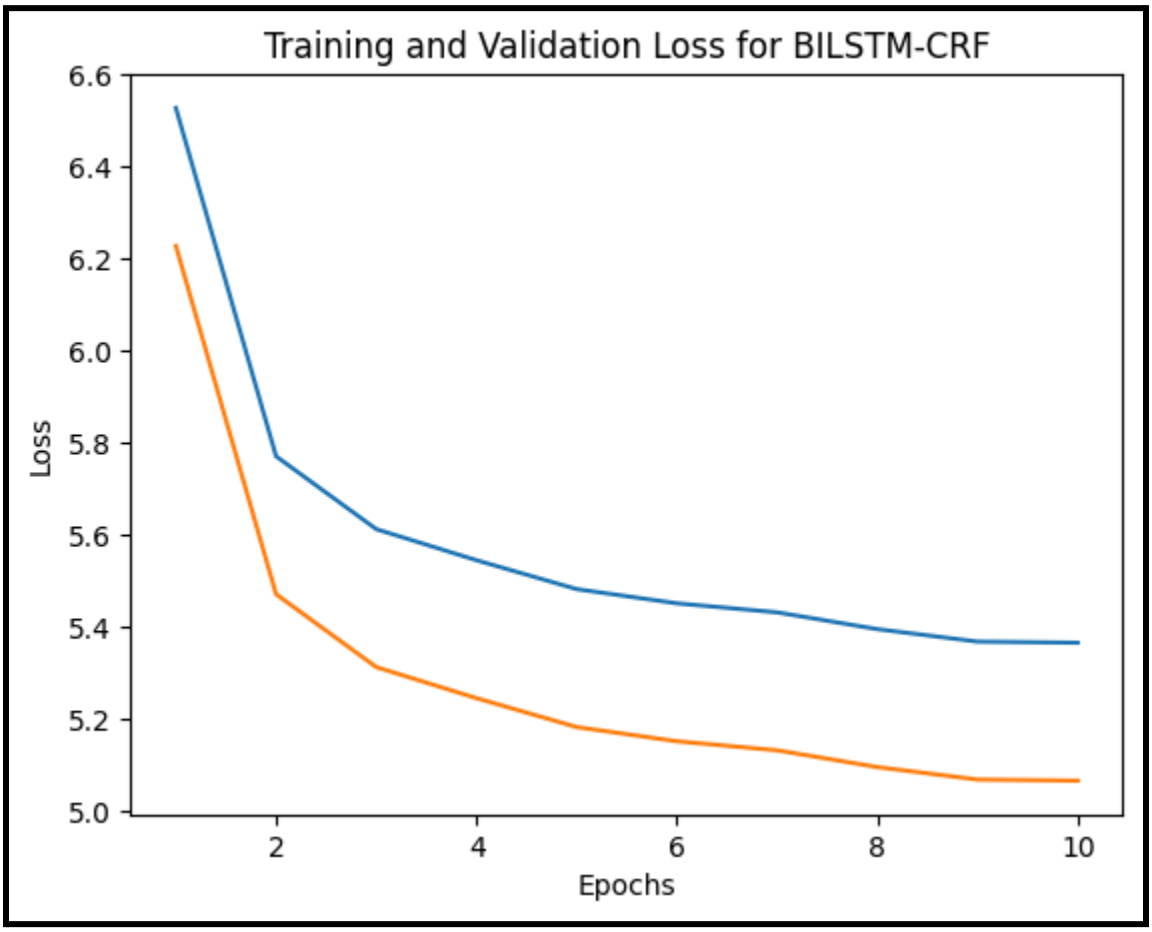


- **Fasttext :**

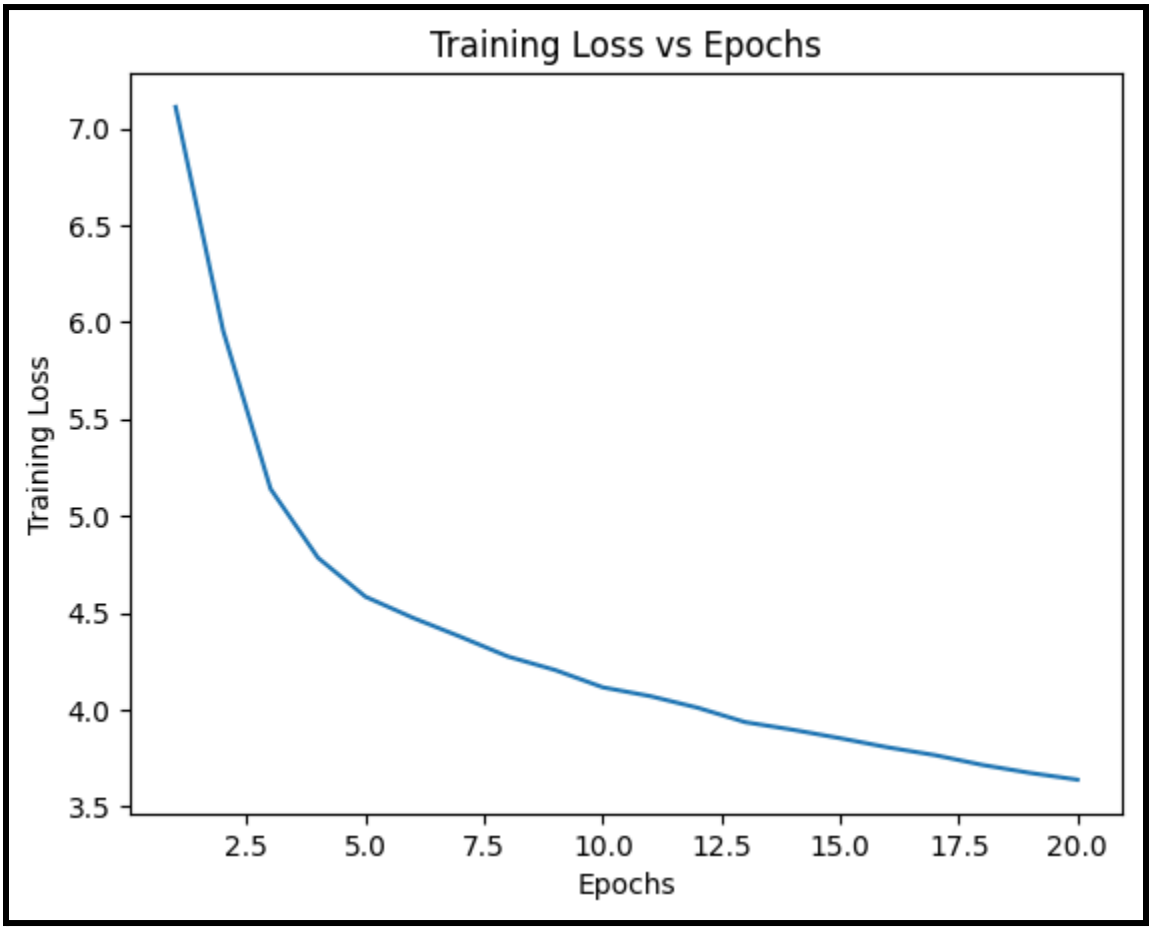


BiLSTM CRF

- **Word2Vec :**



- **GloVe :**


Training Loss vs Epochs
Training Loss
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0
Epochs

- **Fasttext :**

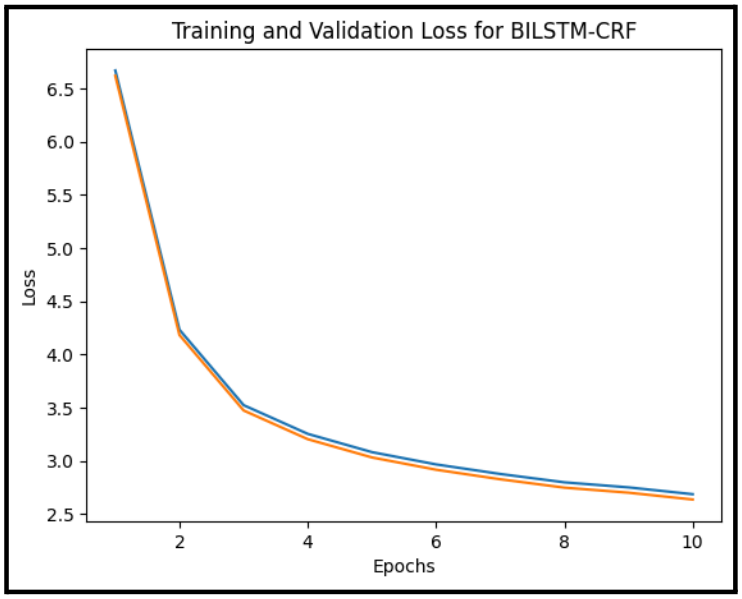


Table for Task-2

Model_No	Embedding_used	Accuracy	Macro_F1
RNN	**Word2Vec**	0.9790853658536586	0.329810723932177
RNN	**GloVe**	0.9916771797028315	0.5387312564922245
RNN	**Fasttext**	0.978750000000000	0.3510957579166409
LSTM	**Word2Vec**	0.9790853658536586	0.329810723932177
LSTM	**GloVe**	0.9863109756097561	0.5908400099585648
LSTM	**Fasttext**	0.978750000000000	0.3510957579166409
GRU	**Word2Vec**	0.9879800953181945	0.3313179032476041
GRU	**GloVe**	0.9920626576955425	0.5873960406533144

GRU	Fasttext	0.9877698345948976	0.35252584739276643
BiLISTM CRF	Word2Vec	0.9217288725584993	0.39425416520368034
BiLISTM CRF	GloVe	0.9584293614779353	0.562529303598263
BiLISTM CRF	Fasttext	0.9736124540707793	0.7840855205598118

Analysis of Plots

1) **Loss Plot: Training Loss and Validation Loss V/s Epochs**
   In all cases training loss decreases with the number of epochs indicating that the model is learning the data on which it is being trained similarly in majority of the cases validation loss also decreases like those in bilstm and gru models for task-1 This indicates that the model performs well on unseen data, or effective learning. Conversely, in a few cases the validation loss increases but training loss continues to decrease, this may be due to a case of overfitting of the model.

2) **F1 Plot: Training Macro-F1-score and Validation Macro-F1-score V/s Epochs**
   The training and Validation Macro-F1-scores versus epochs provides insights into the model's performance throughout training. We observe in most cases like in task 2 that the macro f1 scores for training and validation increase with the number of epochs. Indicating a good performance. However, for a few cases of task 1 models the training/validation macro-f1 scores also decrease after a certain number of epochs. In some plots, we can observe that the validation F1 score first increases with each epochs, then starts decreasing at some point. This may be due to a case of overfitting of the model, where the model becomes overly sensitive to noise or specific characteristics of the training data that do not generalize well to unseen examples. As a result, while the F1 score continues to increase on the training data, it may start to decrease on the validation or test data, indicating that the model's performance is deteriorating.

   Furthermore, we observe that bilstm models significantly perform better than those used in part - 2 i.e. GRU, vanilla RNN & LSTM. This can be seen from cases of task 1 and task 2 where test macro - f1 scores of embeddings: word2vec, GloVe, and fasttext for bilstm increase as compared to part - 1 models.

Contributions

- Ayush Srivastava : BiLSTM
- Kartik Gupta : GRU + Fasttext + report
- Abhishek Sushil : LSTM + Word2vec + report
- Manas Narang : RNN + GloVe + report