Natural Language Processing
Assignment: 2

Karan Gupta (2021258) | Raghav Sakhuja (2021274) | Shivesh Gulati (2021286) | Rahul Oberoi (2021555)

Snippet for: Part-1 A


```
"b0311cba3aac4d909eec6e156c059617": {
    "text": "(See Principles of Statutory Interpretation by Justice G.P. Singh, 9th Edn., 2004 at p.  438.).",
    "labels": [
        "O",
        "O",
        "O",
        "O",
        "O",
        "O",
        "O",
        "B_JUDGE",
        "I_JUDGE",
        "O",
        "O",
        "O",
        "O",
        "O",
        "O",
        "O"
    ]
},
```


Snippet for: Part-1 B


```
{
    "0": {
        "text": "Boot time is super fast , around anywhere from 35 seconds to 1 minute .",
        "labels": [
            "B",
            "I",
            "O",
            "O",
            "O",
            "O",
            "O",
            "O",
            "O",
            "O",
            "O",
            "O",
            "O",
            "O",
            "O"
        ]
    },
```

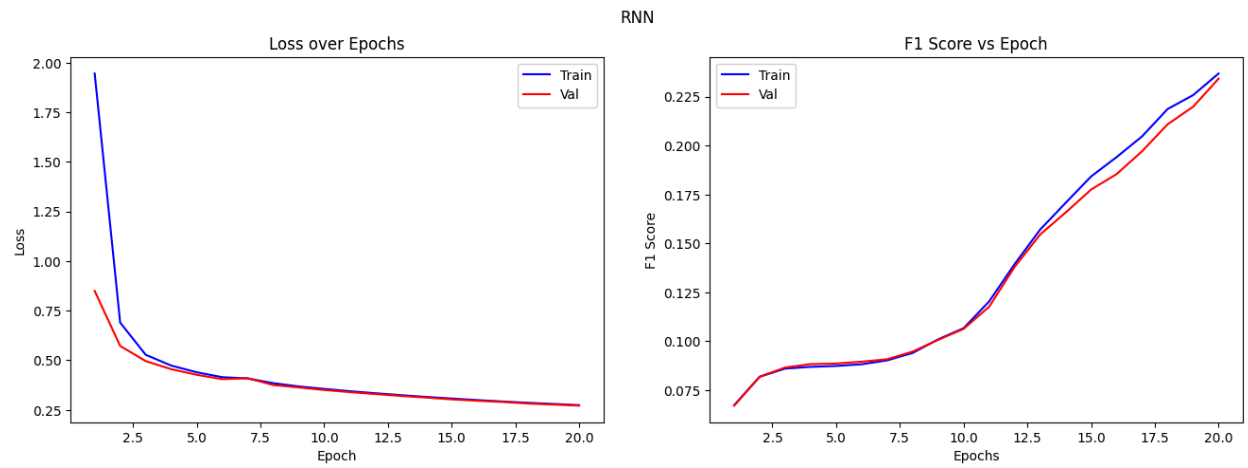

Preprocessing

1. Removed all escape sequences from the sequences such as “\n” “\t” etc.
2. Removed all multiple spaces between words and only allowed single space between each token
3. Considered full word for the BIO encoding even if only part of the word is given a label.

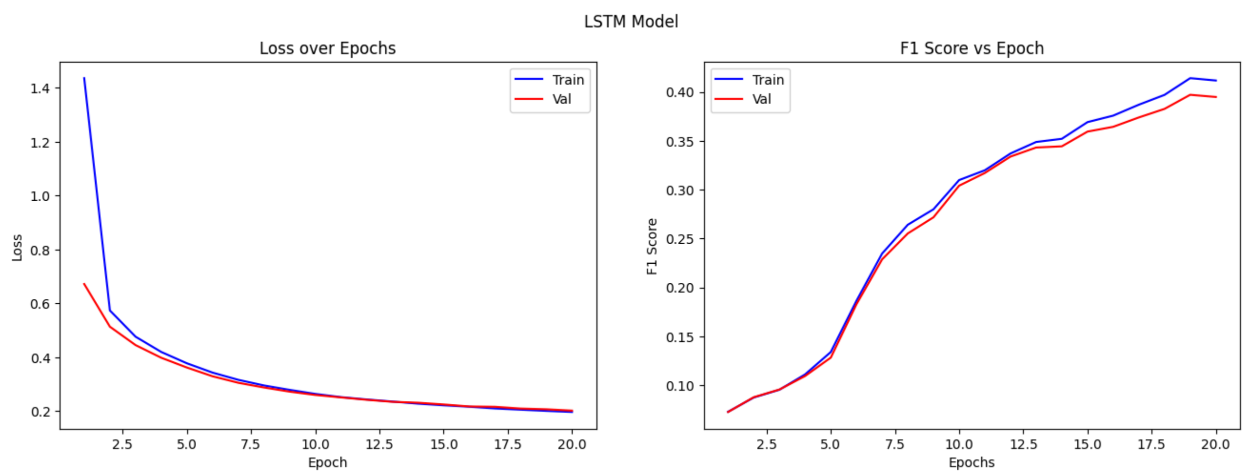
NER Dataset

Embedding: Fasttext

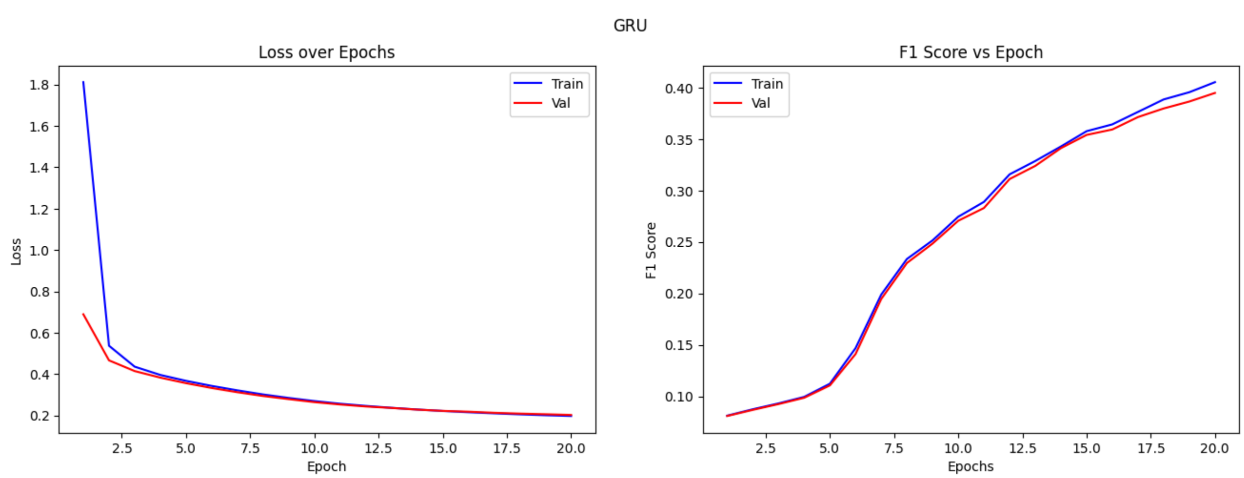
RNN



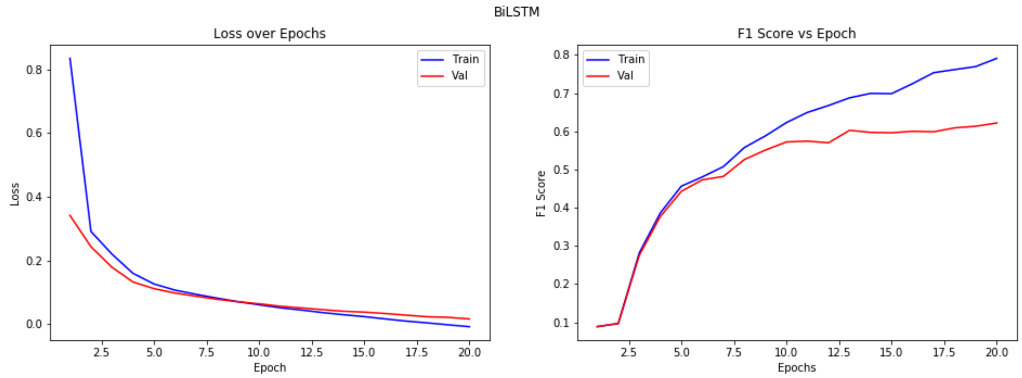
LSTM



GRU



BiLSTM

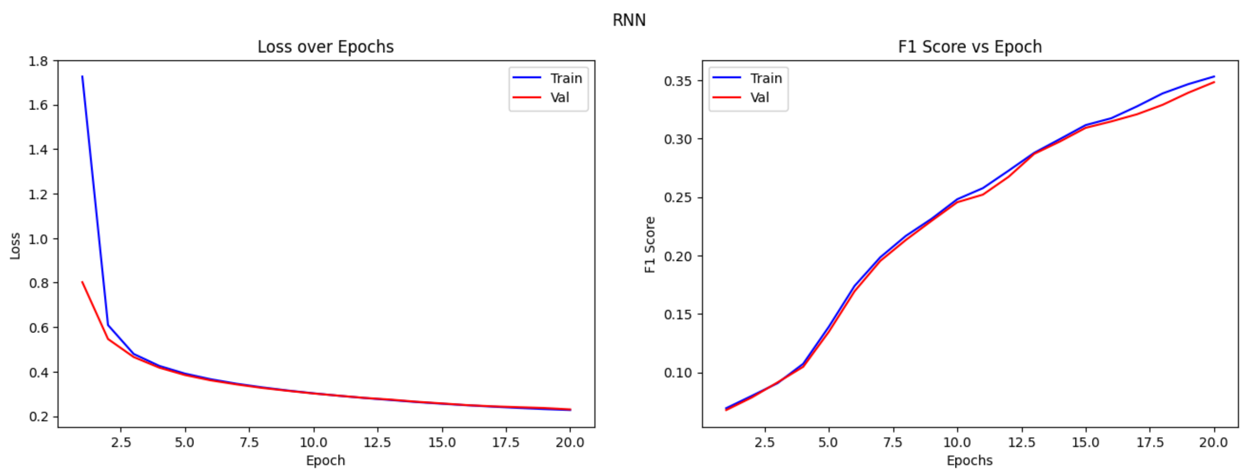


Explanation: Among the three types of word embeddings used, the FastText embedding exhibits notable performance. The RNN model could perform better at higher epochs, showcasing a steady improvement in F1 score and a reduction in loss with each successive epoch. Following closely, the LSTM model demonstrates the second-best performance, mirroring a similar trend. The GRU model secures the third position.

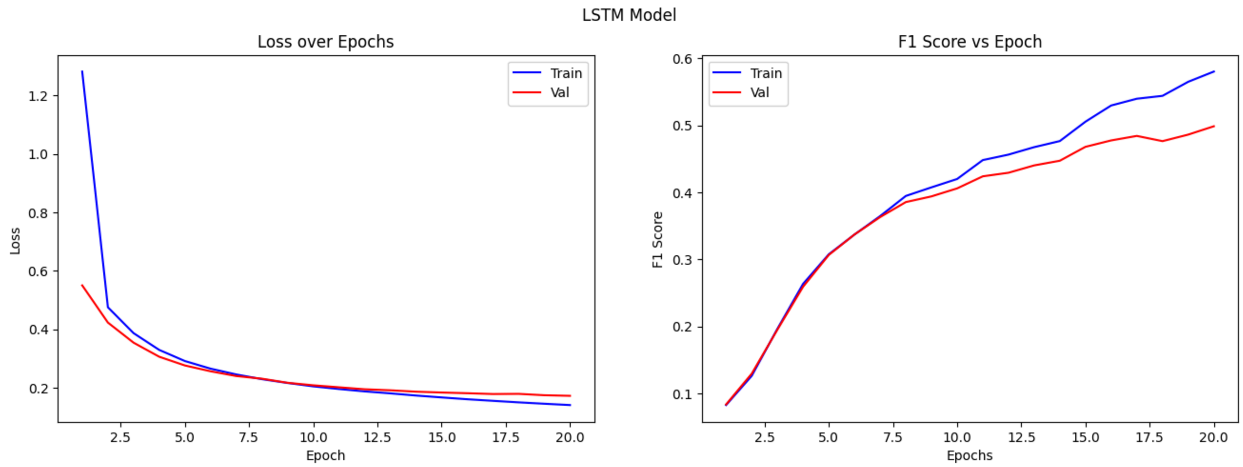
However, the Bidirectional LSTM (BiLSTM) + CRF model is the standout performer across all three word embeddings. This model outshines its counterparts, delivering the overall best performance in Named Entity Recognition (NER). The difference that we notice that there is a significant gap between the training F1 score and the validation F1 score in this case (rest of the models have slim to none difference between the training and validation F1 scores).

Embedding: Glove

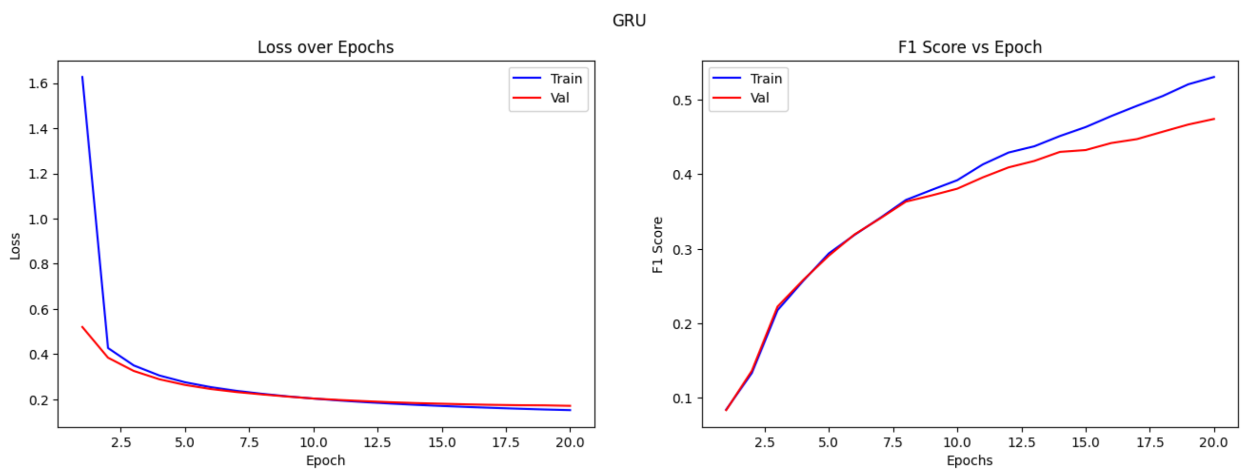
RNN



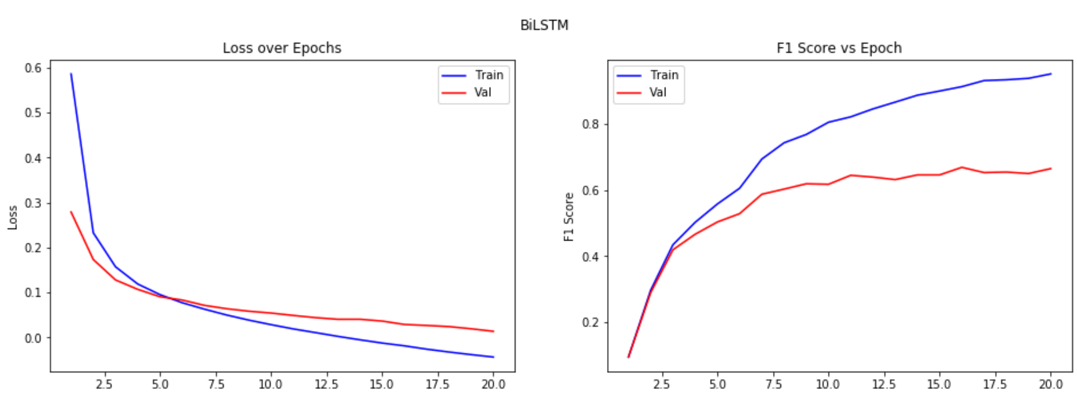
LSTM



GRU



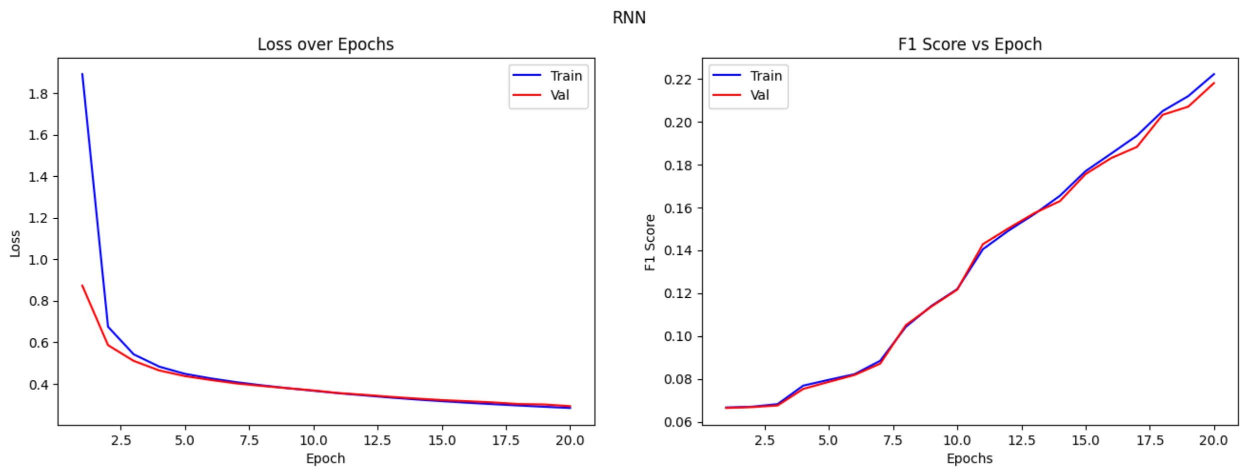
BiLSTM



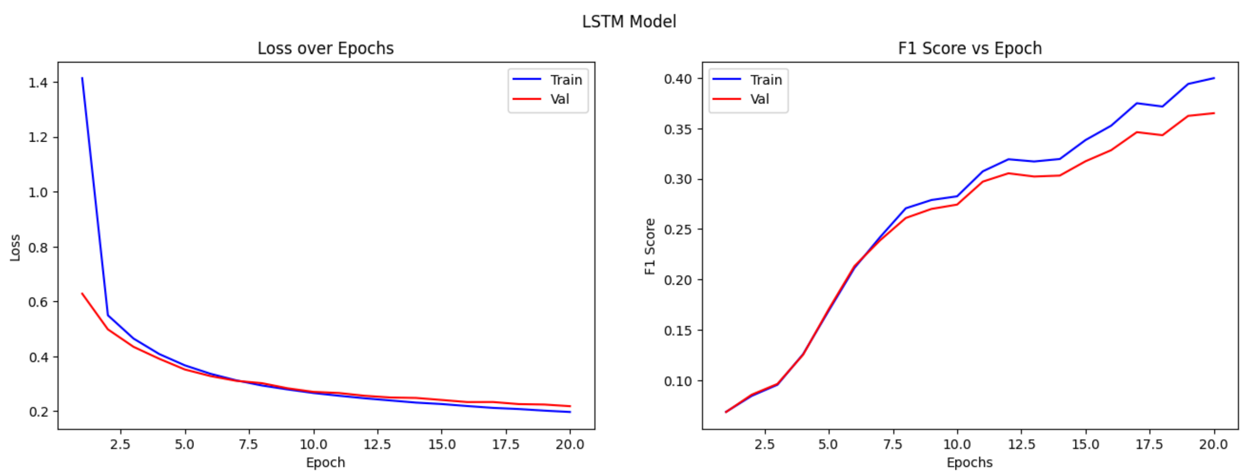
Explanation: Similar to the observed pattern with FastText embedding, the GloVe embedding also exhibits a consistent trend. The RNN model consistently ranks as the least effective, demonstrating the poorest performance. In contrast, both the LSTM and GRU models follow similar patterns however the validation and training F1 scores have a wider gap and the validation loss is more at the end than the training loss. We notice a similar pattern in the BiLSTM model and again it performs better than rest of the 3 models.

Embedding: Word2Vec

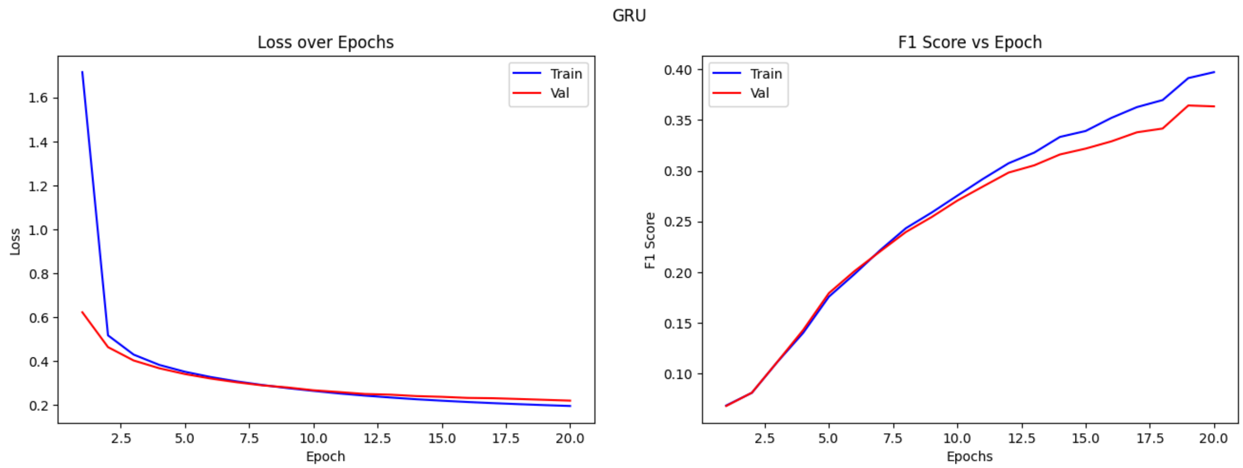
RNN



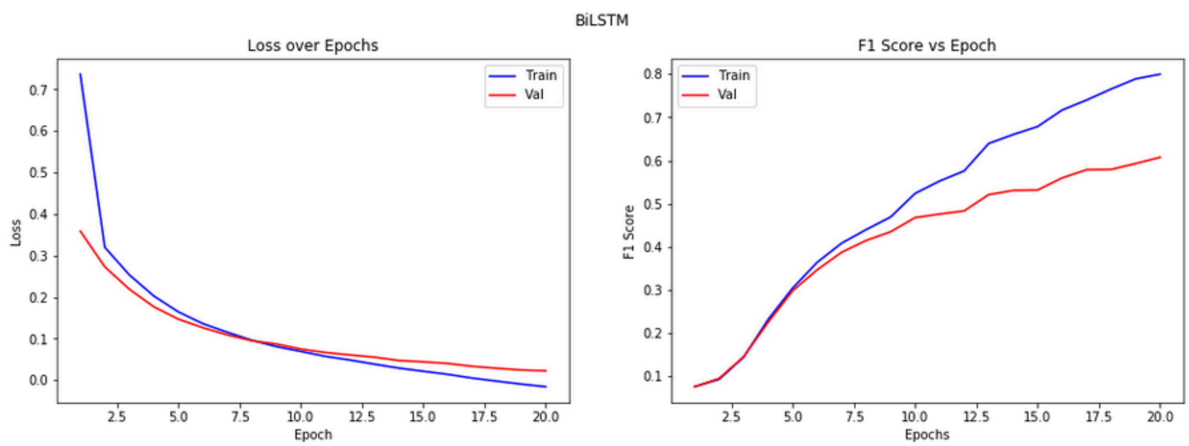
LSTM



GRU



BiLSTM

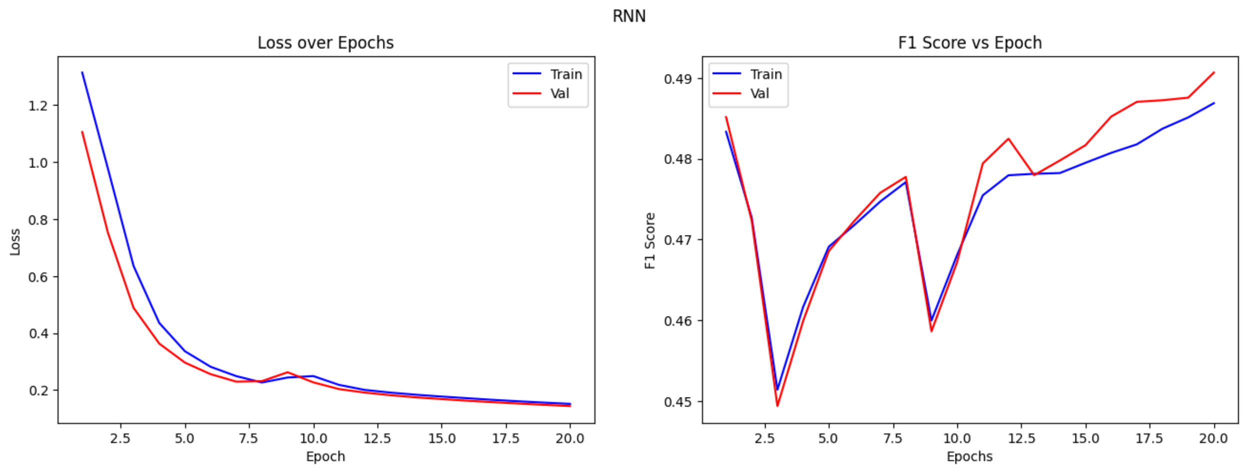


Explanation: In a parallel fashion to the trends observed with GloVe embedding, Word2Vec embedding also follows a comparable pattern across all the models but performs worse in every model.

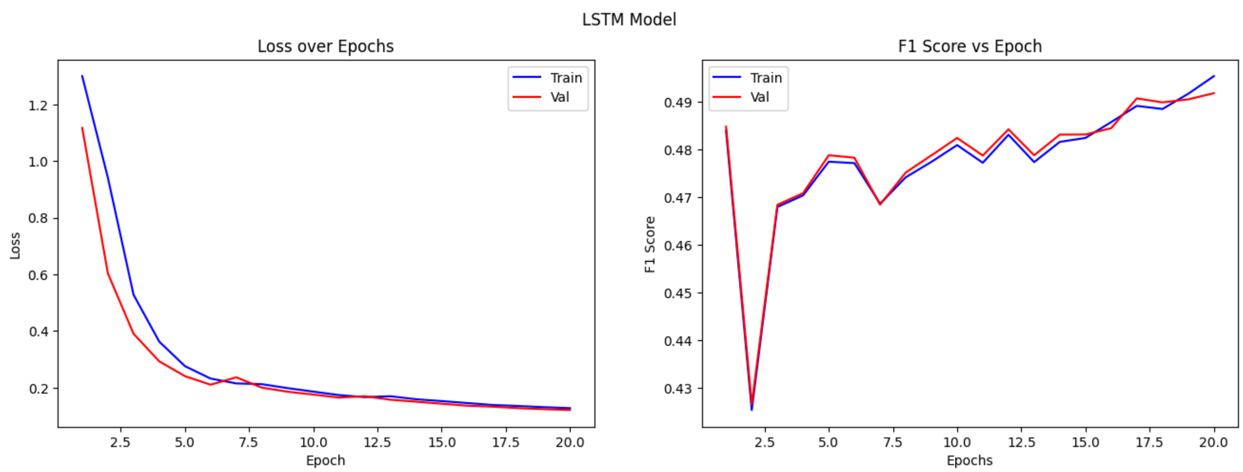
ATE Dataset

Embedding: Fasttext

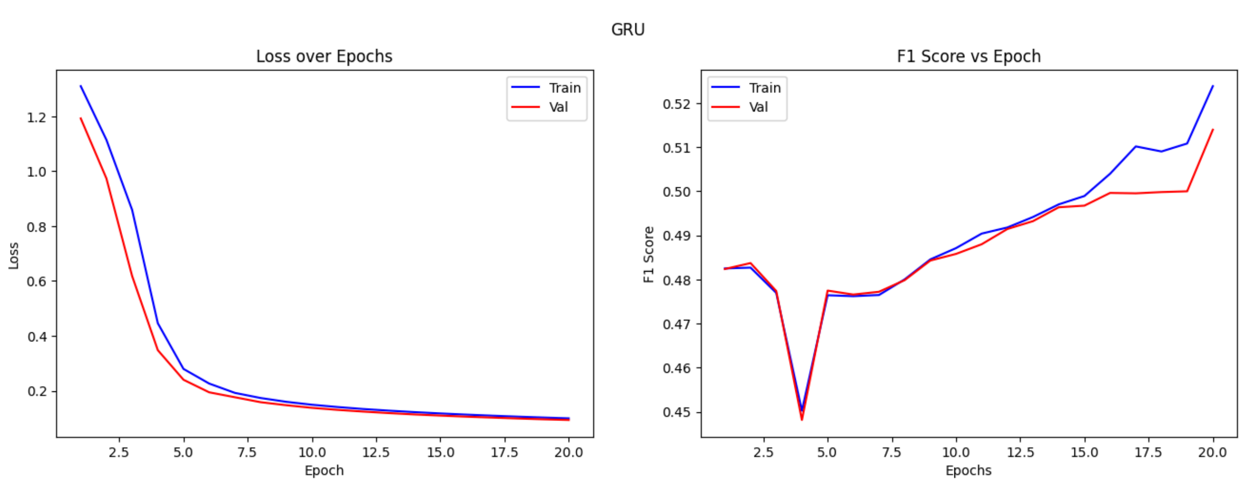
RNN



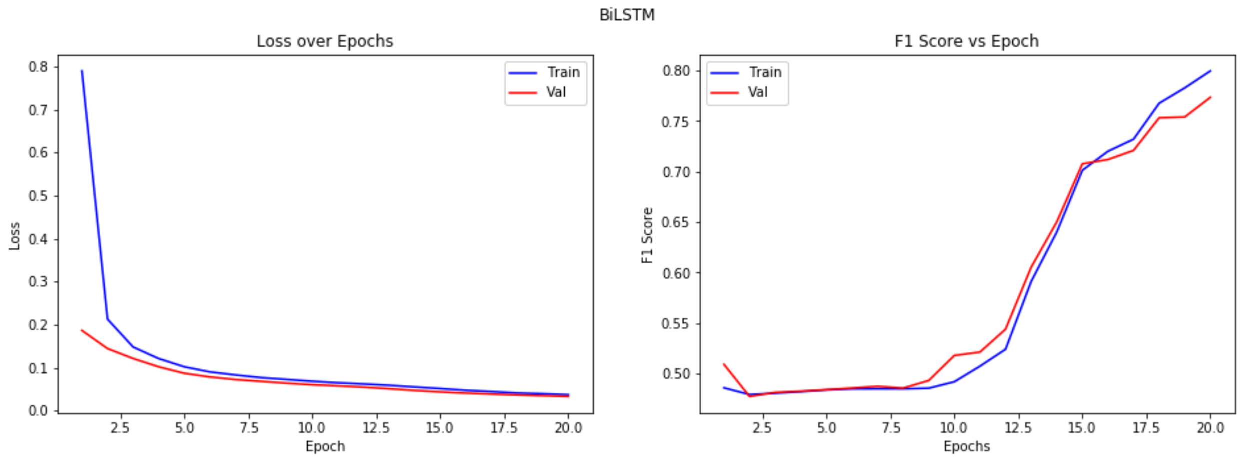
LSTM



GRU



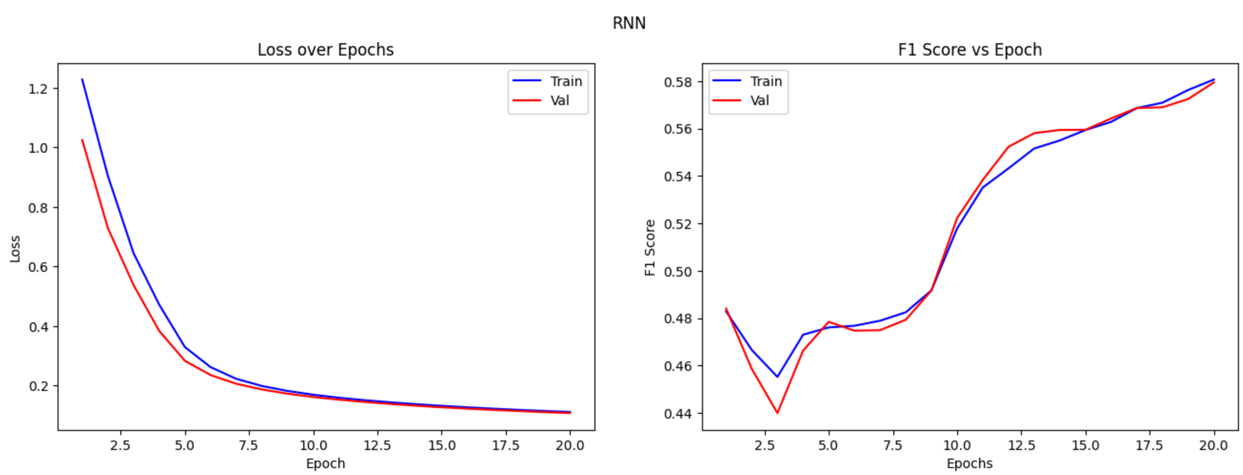
BiLSTM



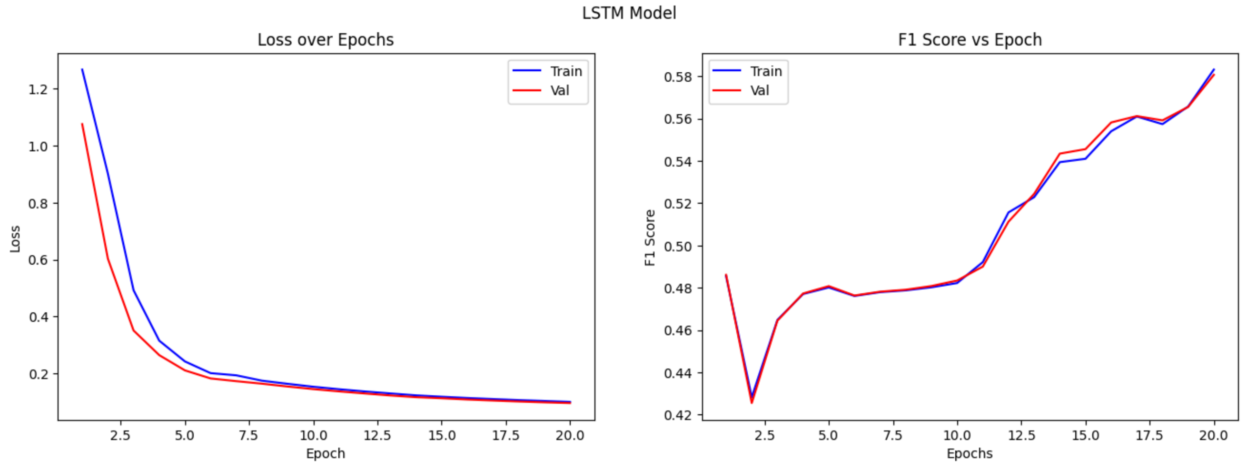
Explanation: Unlike the NER dataset, in the ATE dataset we notice sudden increases/drops in the F1 scores, which then seem to stabilize. RNN generates similar results like the LSTM model here (1% F1 score difference), and GRU performs the second best. BiLSTM is again the best-performing model among the three.

Embedding: Word2Vec

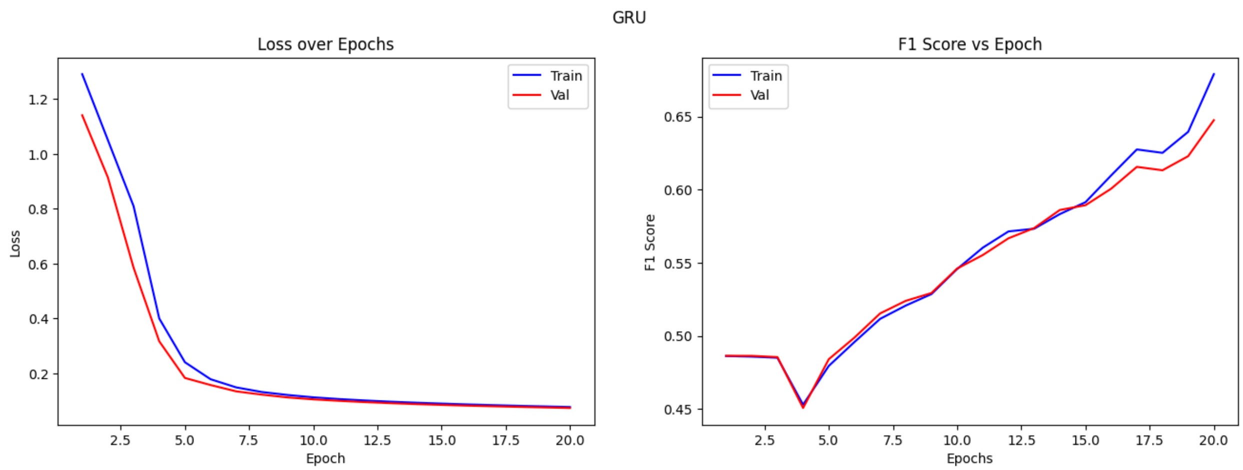
RNN



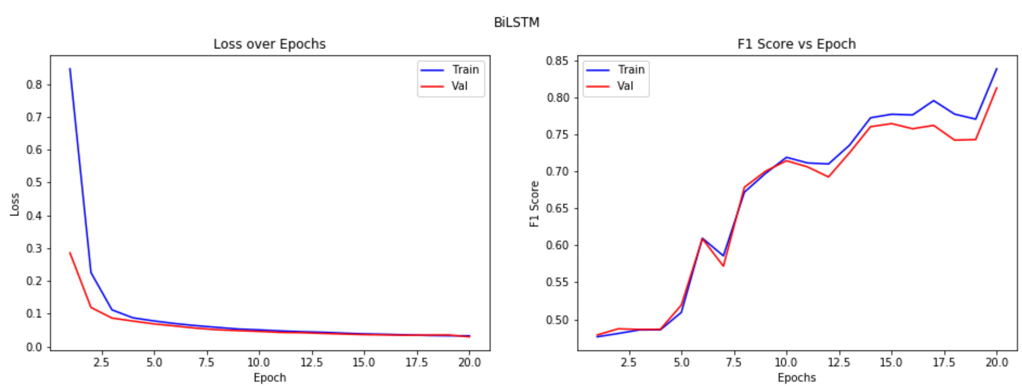
LSTM



GRU



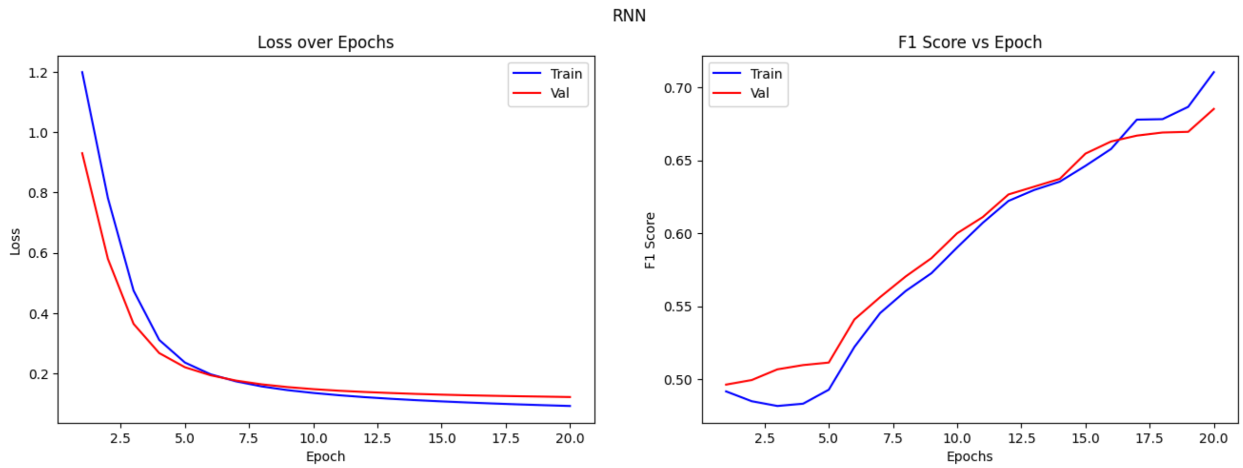
BiLSTM



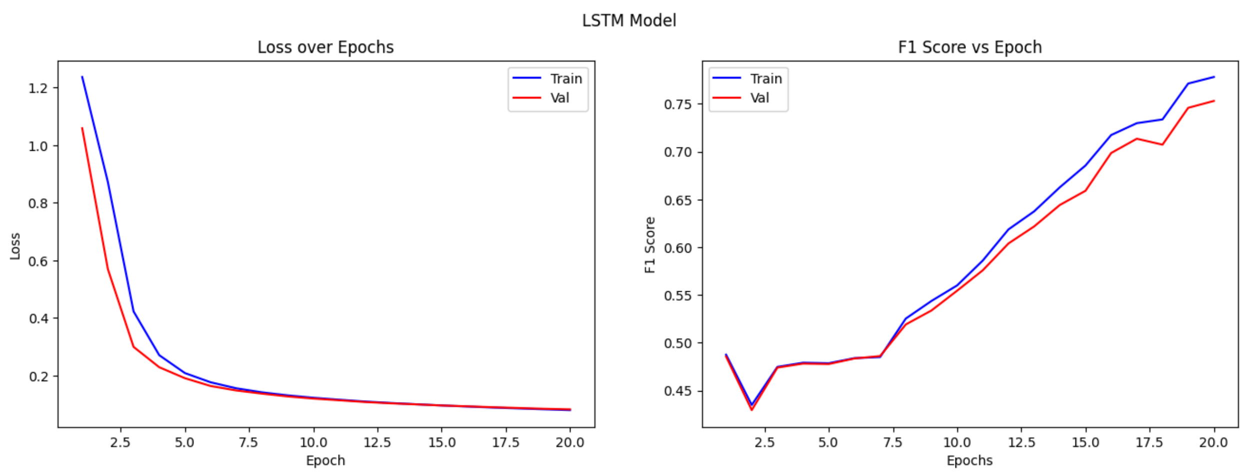
Explanation: In the context of the ATE dataset, the Word2Vec embedding paired with BiLSTM emerges as the most effective combination, delivering the best overall performance. LSTM model secures the third position in terms of performance, while the Gated Recurrent Unit (GRU) model takes the second spot.

Embedding: Glove

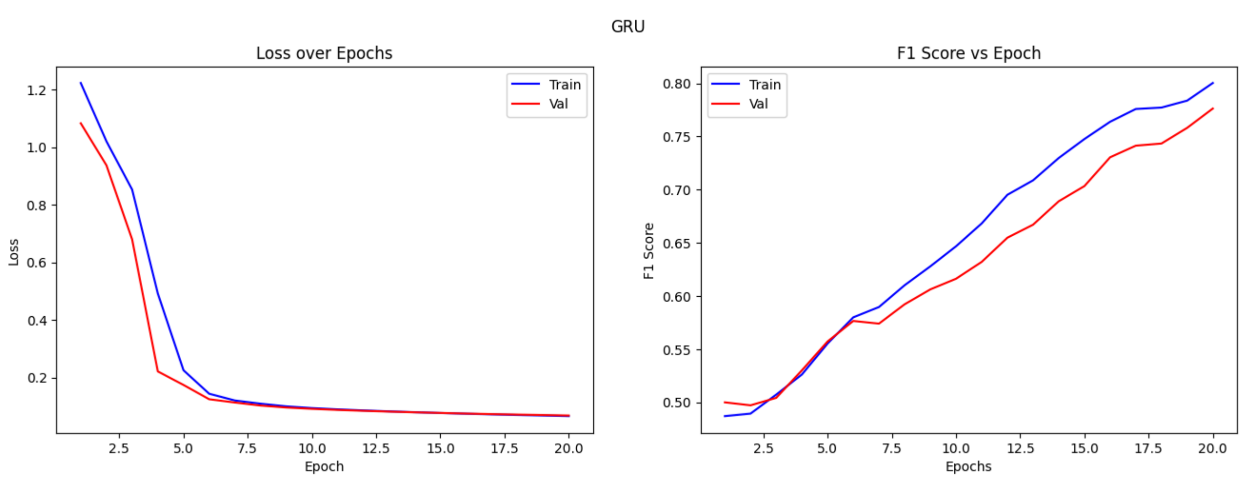
RNN



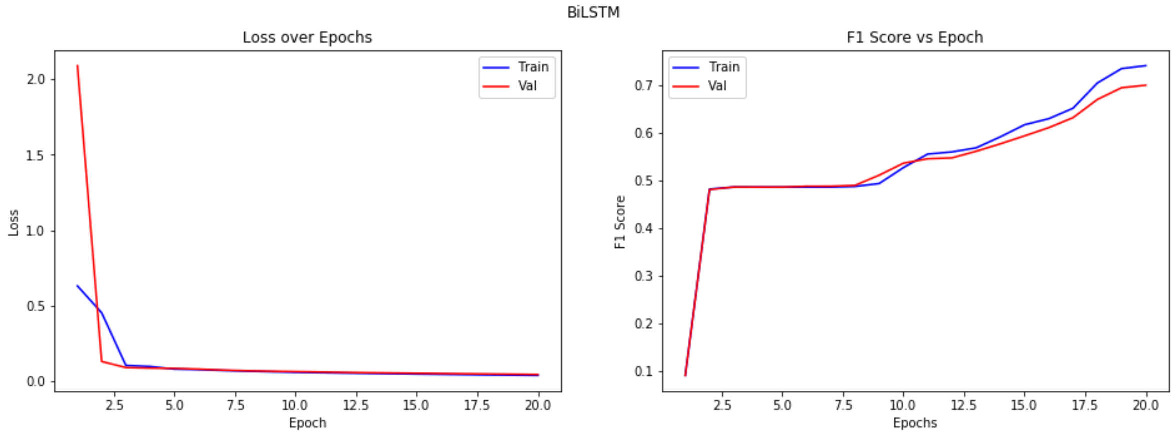
LSTM



GRU



BiLSTM



Explanation: GloVe embedding consistently outperforms both Word2Vec and FastText embeddings across various models, showcasing a significant superiority in performance. However, it's noteworthy that in the specific case of the Bidirectional LSTM (BiLSTM) model using GloVe isn't able to generate better results than Word2Vec with BiLSTM. Despite this, GloVe embedding still stands out as a reliable and practical choice, providing consistently good results across the board.

NER Dataset

Model Number	Embedding Used	Accuracy	Macro F1
1.1	Fasttext	0.92958003282547	0.23017192898728367
1.2	Fasttext	0.9454011917114258	0.38336041435071305
1.3	Fasttext	0.9455215930938721	0.38323361826543273
1.4	**Fasttext**	**0.9616793394088745**	**0.6432647566619955**
2.1	Glove	0.9415926337242126	0.34407264519555586
2.2	Glove	0.9506849050521851	0.48372834594521563
2.3	Glove	0.9515429735183716	0.45864538136044863
2.4	**Glove**	**0.9628621339797974**	**0.6347917032303547**
3.1	Word2Vec	0.9280596375465393	0.21580725181495963
3.2	Word2Vec	0.942059338092804	0.36683018906692705
3.3	Word2Vec	0.94183349609375	0.380250220419787
3.4	**Word2Vec**	**0.9584560394287109**	**0.5837450269977118**

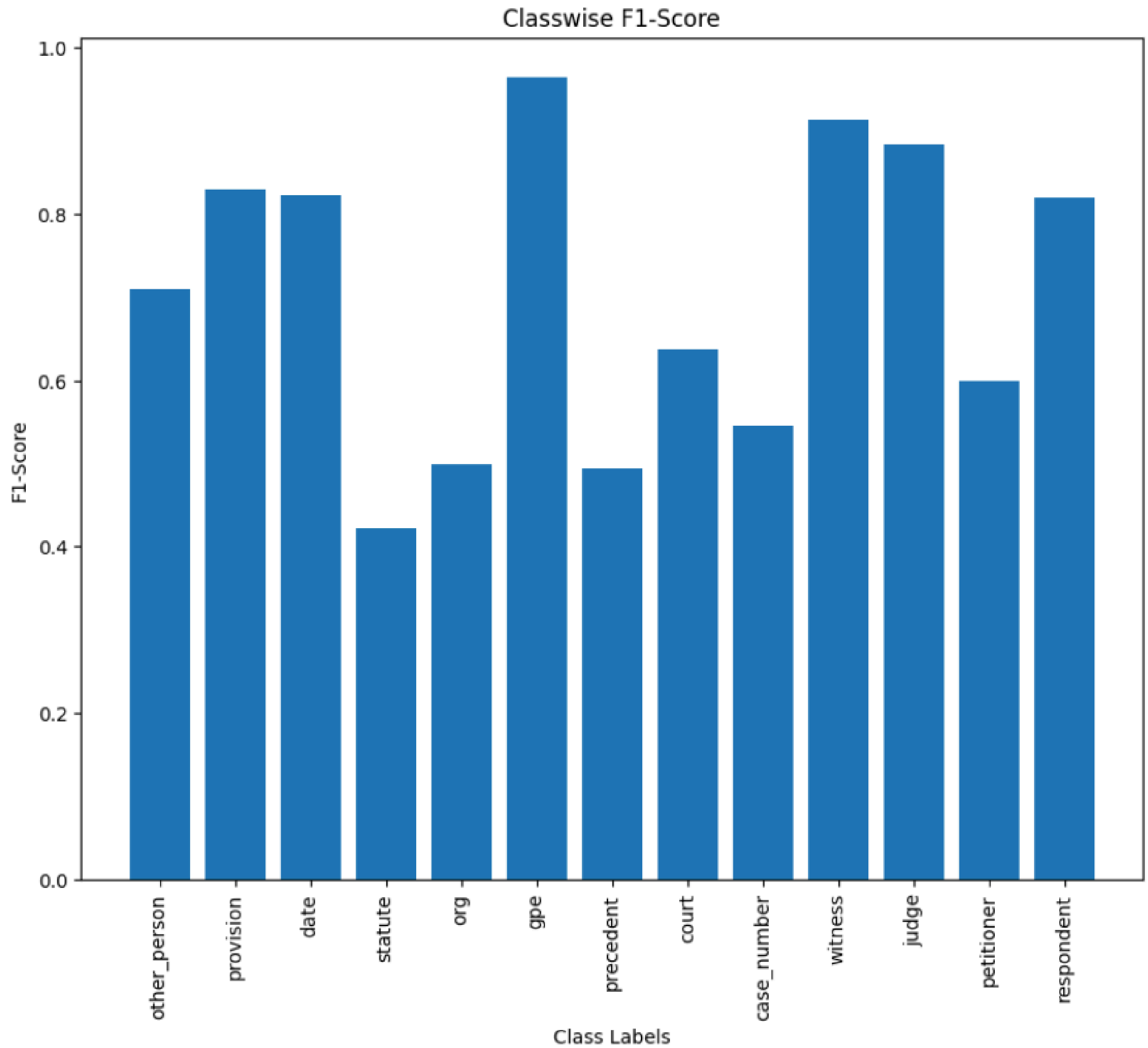
ATE Dataset

Model Number	Embedding Used	Accuracy	Macro F1
1.1	Fasttext	0.9618120789527893	0.48729728761153934
1.2	Fasttext	0.9673624038696289	0.4976203476901322
1.3	Fasttext	0.969473123550415	0.522530472724726
1.4	**Fasttext**	**0.9789844751358032**	**0.7565797324839114**
2.1	Glove	0.964274525642395	0.6561934162367418
2.2	Glove	**0.9797529578208923**	0.7249169421823067
2.3	Glove	0.979401171207428	**0.7337861599274077**
2.4	**Glove**	**0.9770178198814392**	**0.709294162181573**
3.1	Word2Vec	0.9718574285507202	0.5741517610582809
3.2	Word2Vec	0.9752970337867737	0.5819878469718709
3.3	Word2Vec	0.9778767824172974	0.6555275846443042
3.4	**Word2Vec**	**0.9816797375679016**	**0.7845944700367797**

Label Wise F1 Scores for Best Model for NER (BiLSTM with NER Fasttext)

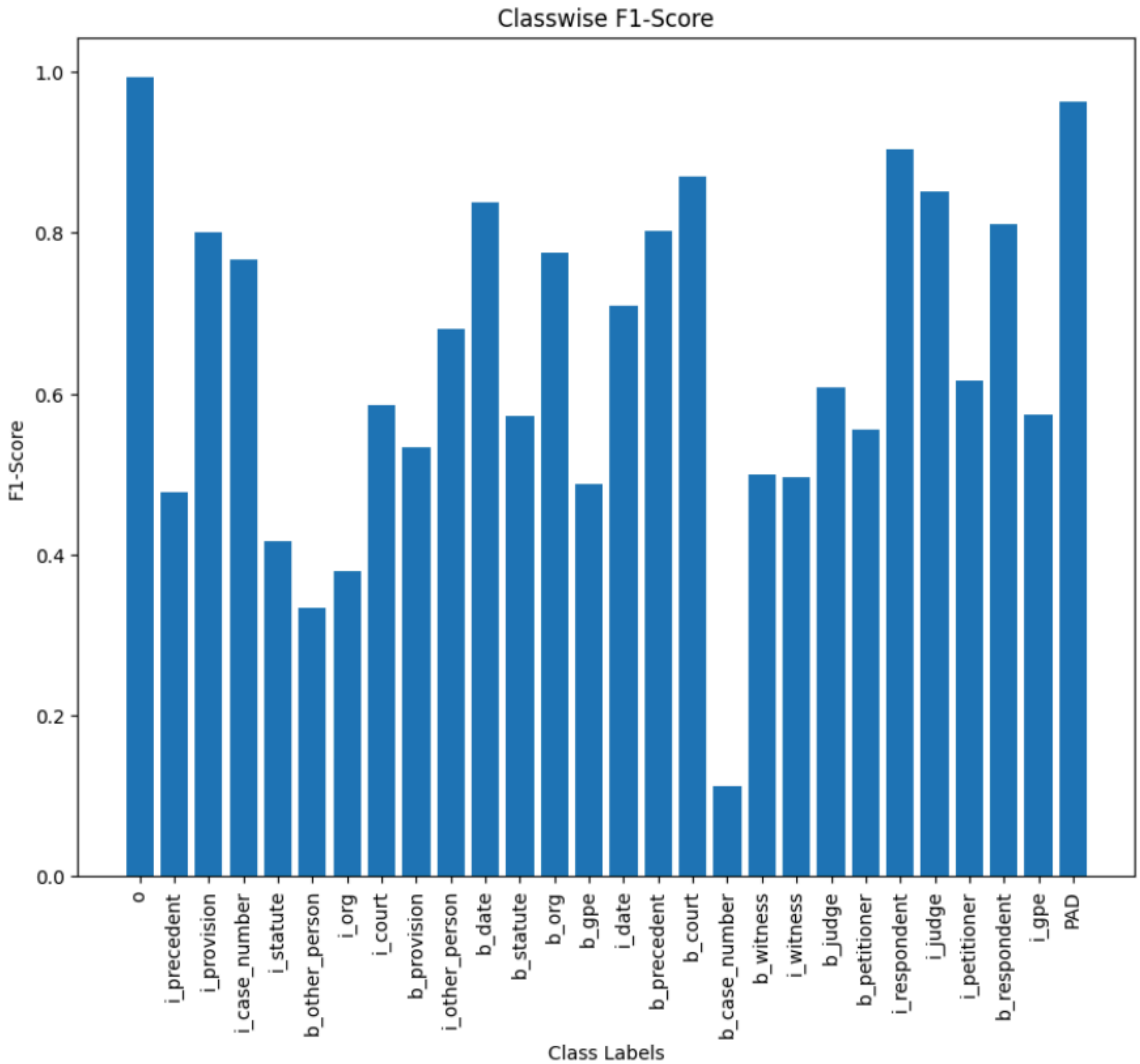
---

	precision	recall	f1-score	support
o	0.99	1.00	0.99	32840
other_person	0.71	0.71	0.71	465
provision	0.90	0.77	0.83	504
date	0.86	0.79	0.82	323
statute	0.50	0.37	0.42	230
org	0.67	0.40	0.50	15
gpe	0.96	0.97	0.96	27699
precedent	0.63	0.41	0.49	469
court	0.75	0.56	0.64	473
case_number	0.69	0.45	0.55	20
witness	0.91	0.91	0.91	1970
judge	0.92	0.85	0.88	696
petitioner	1.00	0.43	0.60	14
respondent	0.79	0.85	0.82	600
PAD	0.54	0.60	0.57	112
accuracy			0.96	66430
macro avg	0.79	0.67	0.71	66430
weighted avg	0.96	0.96	0.96	66430



Classwise F1-Score
F1-Score
1.0
0.8
0.6
0.4
0.2
0.0
other_person
provision
date
statute
org
gpe
precedent
court
case_number
witness
judge
petitioner
respondent
Class Labels

	precision	recall	f1-score	support
o	0.99	1.00	0.99	32840
i_precedent	0.54	0.43	0.48	121
i_provision	0.85	0.75	0.80	178
i_case_number	0.82	0.72	0.77	221
i_statute	0.48	0.37	0.42	183
b_other_person	0.50	0.25	0.33	8
i_org	0.58	0.28	0.38	159
i_court	0.72	0.49	0.59	275
b_provision	0.67	0.44	0.53	9
i_other_person	0.74	0.63	0.68	177
b_date	0.85	0.82	0.84	258
b_statute	1.00	0.40	0.57	5
b_org	0.82	0.74	0.78	220
b_gpe	0.49	0.48	0.49	58
i_date	0.68	0.74	0.71	344
b_precedent	0.88	0.74	0.80	326
b_court	0.86	0.88	0.87	102
b_case_number	0.16	0.09	0.11	47
b_witness	0.60	0.43	0.50	7
i_witness	0.59	0.43	0.50	310
b_judge	0.67	0.56	0.61	198
b_petitioner	0.71	0.45	0.56	11
i_respondent	0.90	0.91	0.90	1793
i_judge	0.89	0.82	0.85	438
i_petitioner	1.00	0.44	0.62	9
b_respondent	0.75	0.88	0.81	380
i_gpe	0.51	0.65	0.57	54
PAD	0.96	0.97	0.96	27699
accuracy			0.96	66430
macro avg	0.72	0.60	0.64	66430
weighted avg	0.96	0.96	0.96	66430



***How to run:**

1. Change embedding name and dataset. (example: `Embedding_name="fasttext"/"word2vec"/"glove" Dataset="ATE"/"NER"`)
2. Change path to the working directory. (example:`PATH='/kaggle/input/nernlp3/Assignment2/'` `OUTPATH='/kaggle/working/'`)

Contributions

---

1. **Karan Gupta (2021258):** Part 1 Data preparation, Saving and Loading models, Preprocessing data
2. **Raghav Sakhuja (2021274):** BiLSTM+CRF
3. **Shivesh Gulati (2021286):** RNN, GRU, LSTM and graph plotting
4. **Rahul Oberoi (2021555):** RNN, GRU and LSTM and graph plotting