Natural Language Processing 2024
Assignment-2 Report

Group No. 41

TASK 1: DATA SAMPLES AND PRE-PROCESSING

Part 1A

Here two data samples from each of the datasets: NER_train and NER_test.

1 NER_train sample

Sample ID: 635588b86e5a4c7a89a2832f40450a80

Text: "The Moopil Nayar or senior member of this family for the time being was the ruler of the Kavalappara State."

Labels:

```
O
B_OTHER_PERSON
I_OTHER_PERSON
O
O
O
O
O
O
O
O
O
O
O
O
O
O
O
B_GPE
I_GPE
```

2 NER_test sample

Sample ID: 03f3901e95ed493b866bd7807f623bc0

Text: "True, our Constitution has no 'due process' clause or the VIII Amendment; but, in this branch of law, after R.C. Cooper v. Union of India, (1970) 1 SCC 248 and Maneka Gandhi v. Union of India, (1978) 1 SCC 248, the consequence is the same."

Labels:

```
O
O
B_STATUTE
O
O
O
O
O
O
O
O
O
O
O
O
O
O
O
O
B_PRECEDENT
I_PRECEDENT
I_PRECEDENT
I_PRECEDENT
I_PRECEDENT
I_PRECEDENT
I_PRECEDENT
I_PRECEDENT
I_PRECEDENT
I_PRECEDENT
O
B_PRECEDENT
I_PRECEDENT
I_PRECEDENT
I_PRECEDENT
I_PRECEDENT
I_PRECEDENT
I_PRECEDENT
I_PRECEDENT
I_PRECEDENT
I_PRECEDENT
O
O
O
O
O
```

Part 1B

Here two data samples from each of the datasets: ATE_train and ATE_test.

1 ATE_train sample

Sample ID: 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 ATE_test Sample

Sample ID: 1

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
```

TASK 2: RESULTS AND ANALYSIS

In this section, we present the results and analysis of our experiments. We trained a total of 4 models using 3 different embeddings, resulting in a total of 12 model variations. Each model variation produced 2 graphs, leading to a total of 48 graphs.

Part 1A

The following subsections present the graphs generated for each model variation:

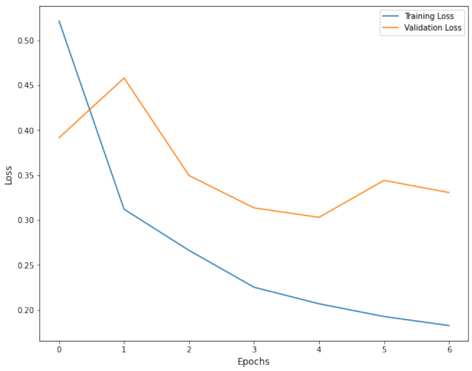

Figure 1: Loss [VanillaRNN [w2v]]

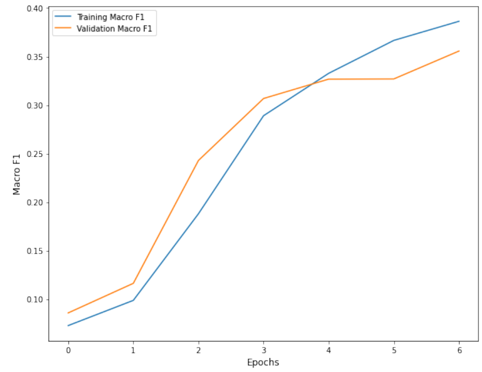

Figure 2: Macro F1 [VanillaRNN [w2v]]

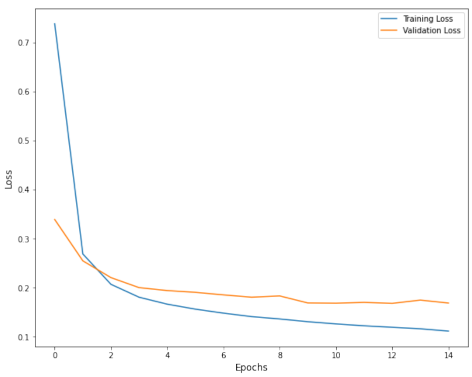

Figure 3: Loss [GRU [w2v]]

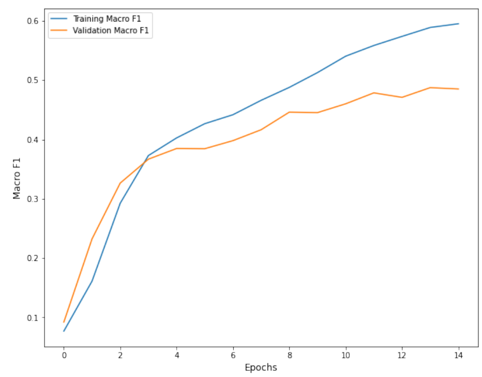

Figure 4: Macro F1 [GRU [v]]

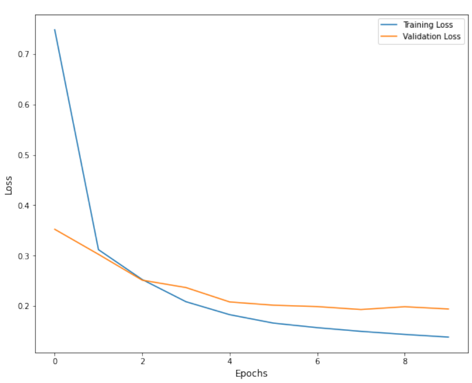

Figure 5: Loss [LSTM [w2v]]

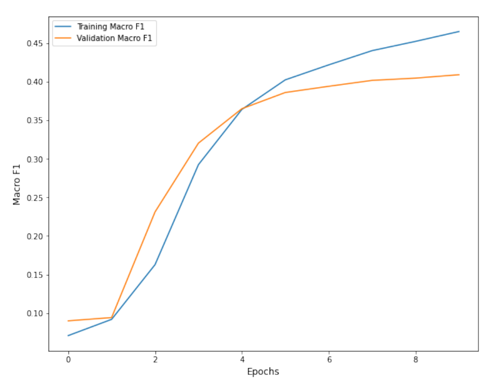

Figure 6: Macro F1 [LSTM [w2v]]

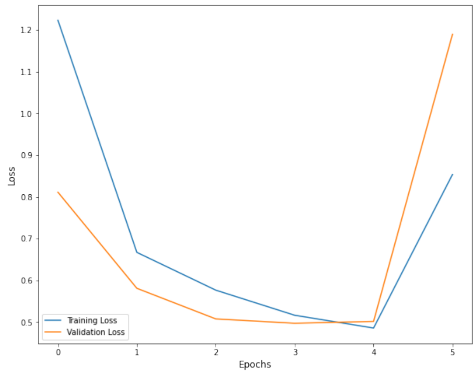

Figure 7: Loss [Bi-LSTM CRF[w2v]]

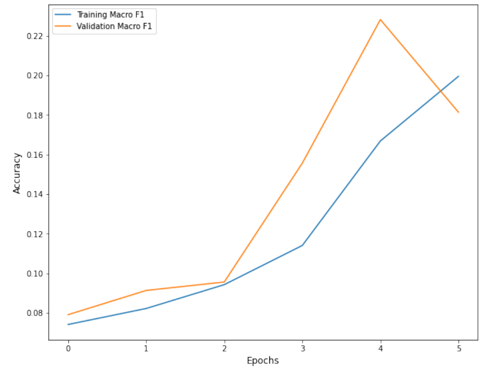

Figure 8: Macro F1 [Bi-LSTM CRF[w2v]]

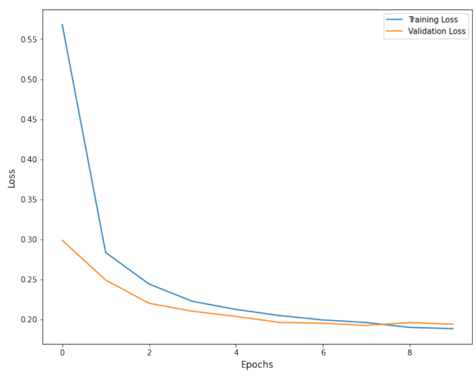

Figure 9: Loss [VanillaRNN [glove]]

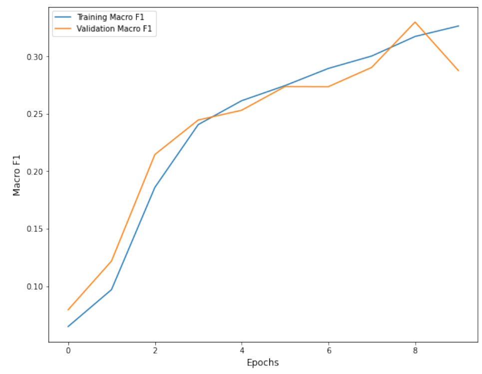

Figure 10: Macro F1 [VanillaRNN [glove]]

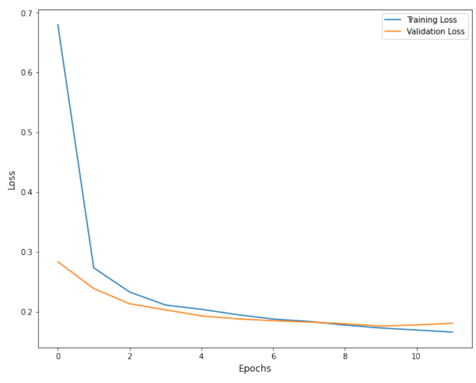

Figure 11: Loss [GRU [glove]]

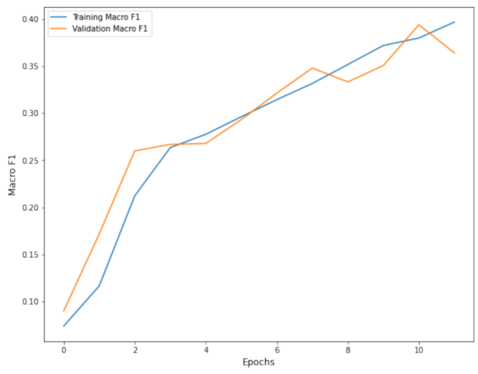

Figure 12: Macro F1 [GRU [glove]]

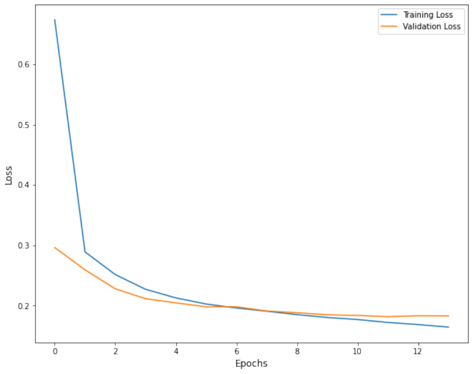

Figure 13: Loss [LSTM [glove]]

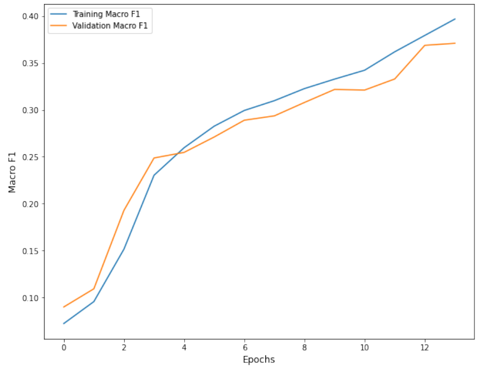

Figure 14: Macro F1 [LSTM [glove]]

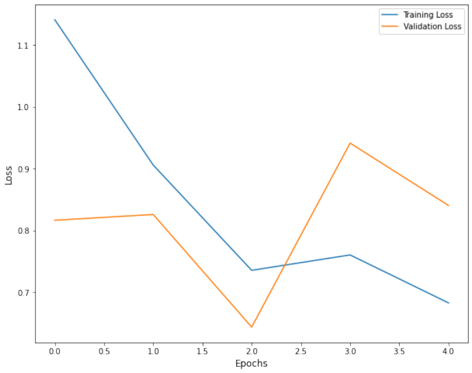

Figure 15: Loss [Bi-LSTM CRF[glove]]

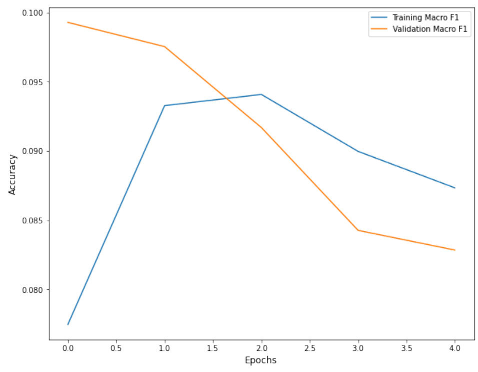

Figure 16: Macro F1 [Bi-LSTM CRF[glove]]

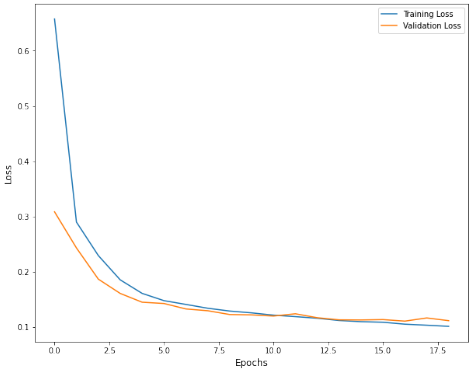

Figure 17: Loss [Vanilla RNN [fasttext]]

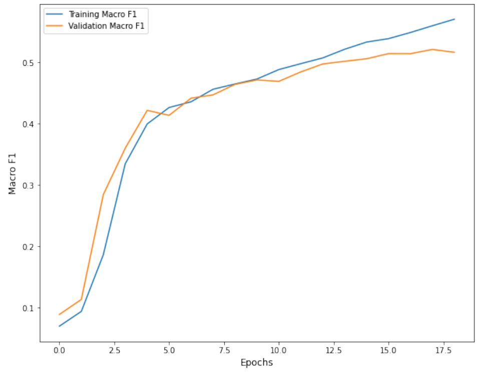

Figure 18: Macro F1 [Vanilla RNN [fasttext]]

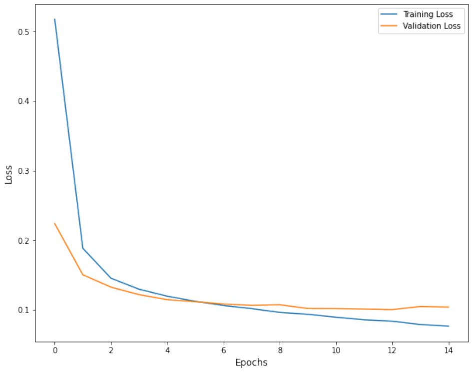

Figure 19: Loss [GRU[fasttext]]

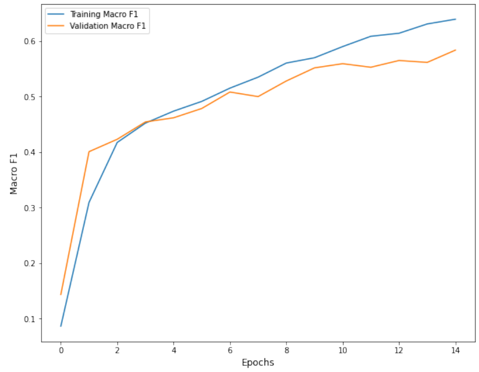

Figure 20: Macro F1 [GRU[fasttext]]

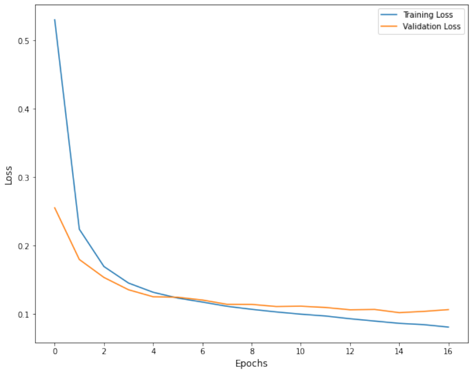

Figure 21: Loss [LSTM[fasttext]]

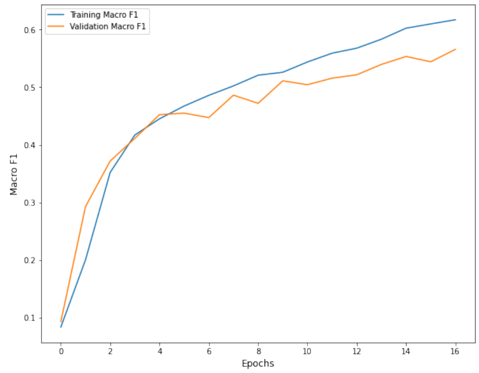

Figure 22: Macro F1 [LSTM[fasttext]]

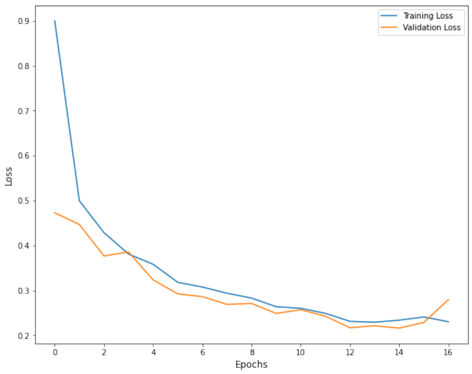

Figure 23: Loss [Bi-LSTM CRF[fasttext]]

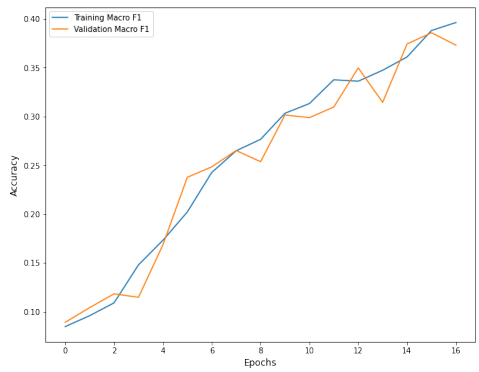

Figure 24: Macro F1 [Bi-LSTM CRF[fasttext]]

Part 1B

The following subsections present the graphs generated for each model variation:

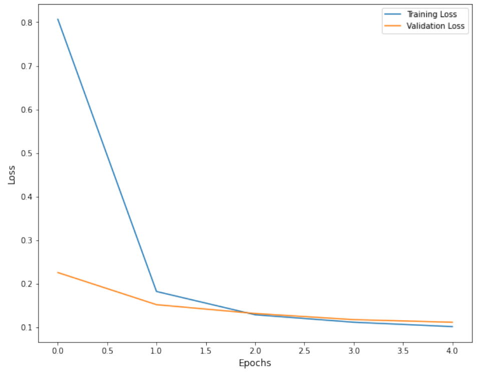

Figure 25: Loss [VanillaRNN [w2v]]

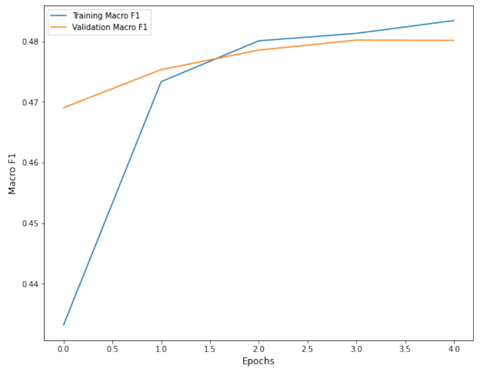

Figure 26: Macro F1 [VanillaRNN [w2v]]

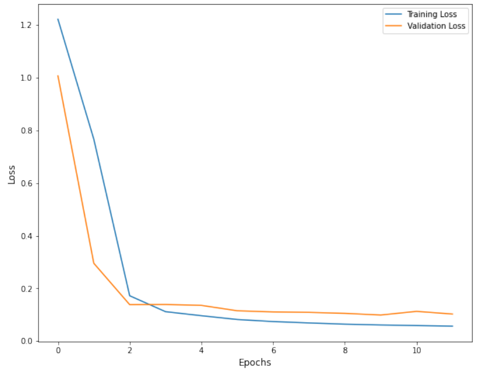

Figure 27: Loss [GRU [w2v]]

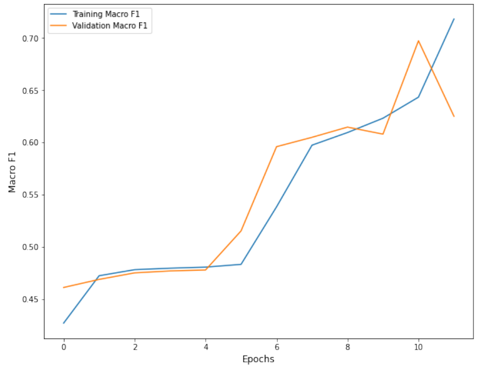

Figure 28: Macro F1 [GRU [w2v]]

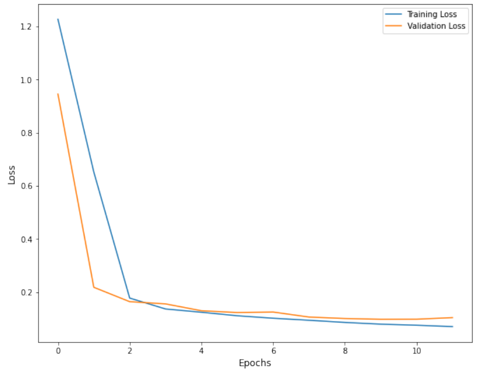

Figure 29: Loss [LSTM [w2v]]

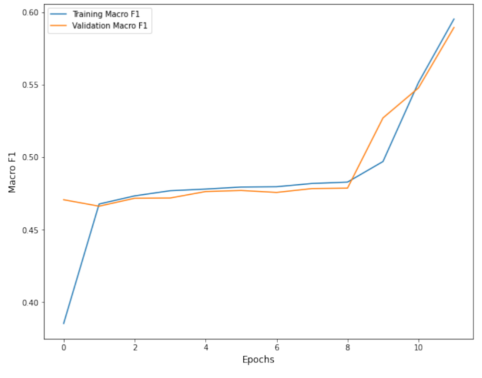

Figure 30: Macro F1 [LSTM [w2v]]

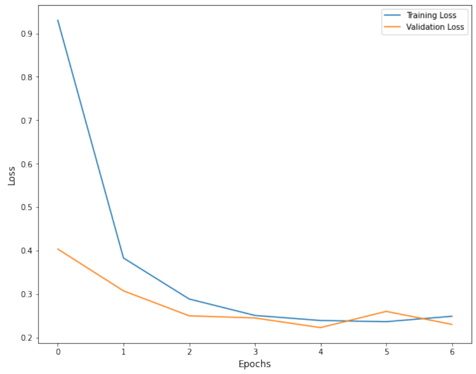

Figure 31: Loss [Bi-LSTM CRF[w2v]]

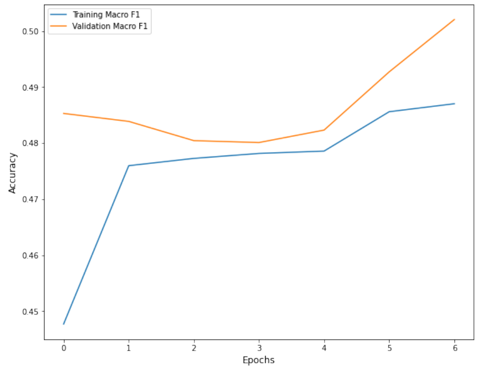

Figure 32: Macro F1 [Bi-LSTM CRF[w2v]]

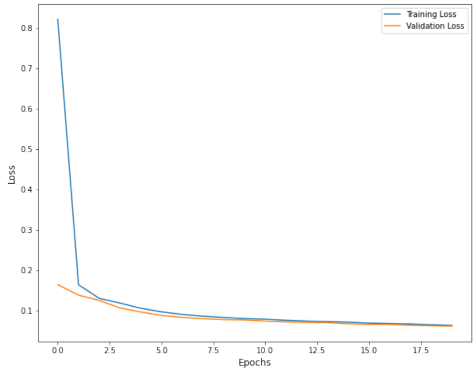

Figure 33: Loss [VanillaRNN [glove]]

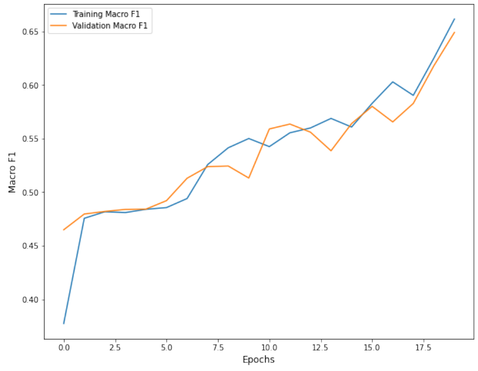

Figure 34: Macro F1 [VanillaRNN [glove]]

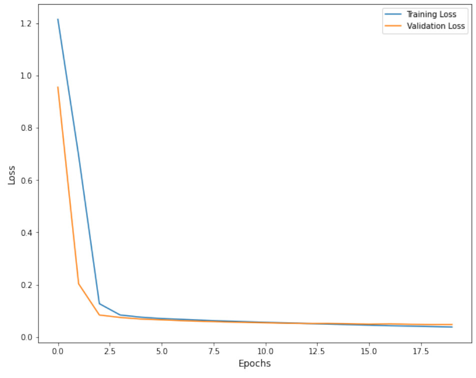

Figure 35: Loss [GRU [glove]]

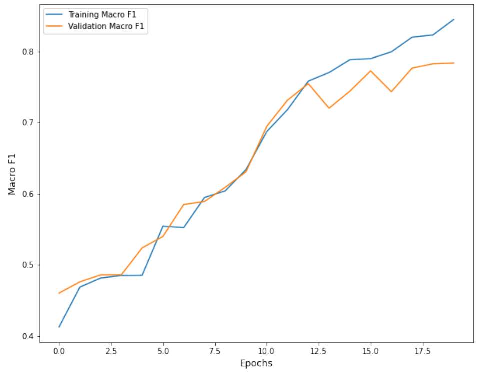

Figure 36: Macro F1 [GRU [glove]]

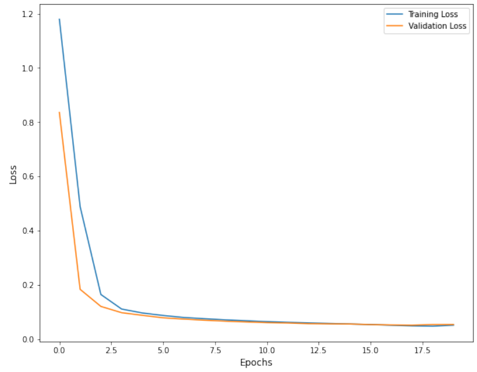

Figure 37: Loss [LSTM [glove]]

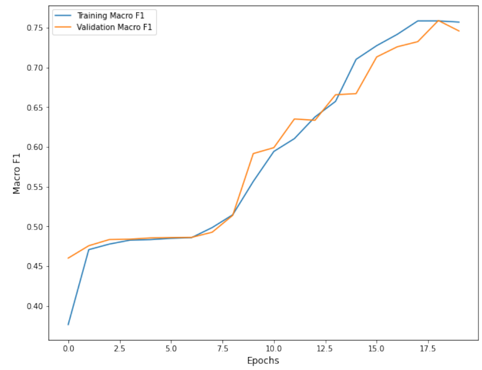

Figure 38: Macro F1 [LSTM [glove]]

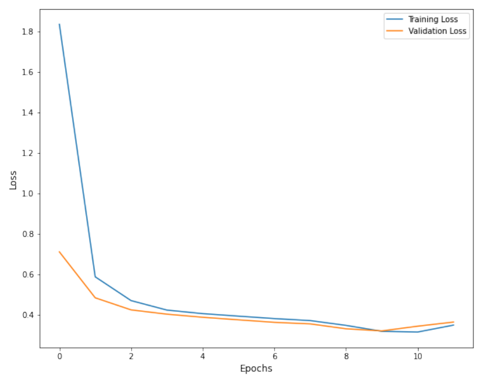

Figure 39: Loss [Bi-LSTM CRF[glove]]

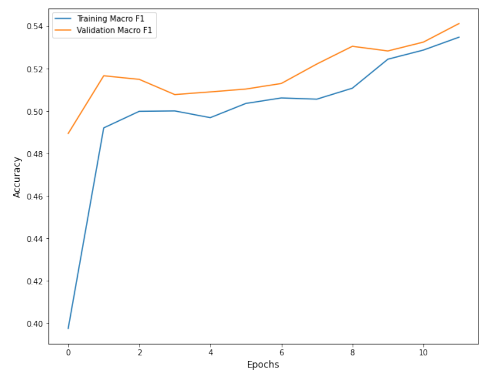

Figure 40: Macro F1 [Bi-LSTM CRF[glove]]

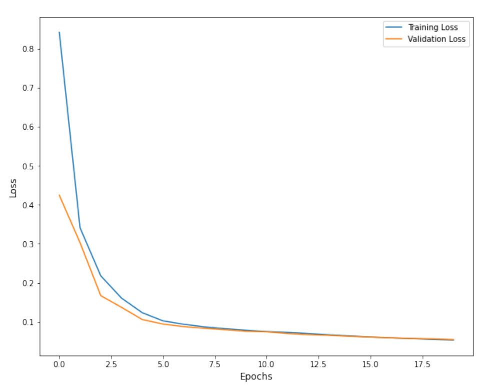

Figure 41: Loss [Vanilla RNN [fasttext]]

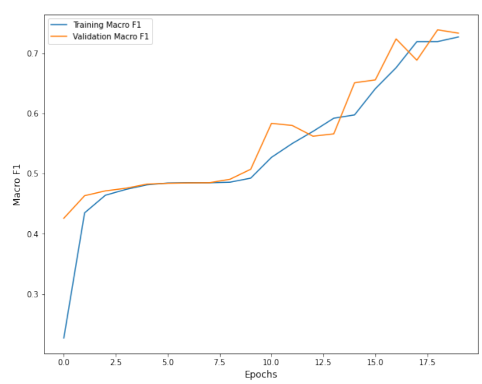

Figure 42: Macro F1 [Vanilla RNN [fasttext]]

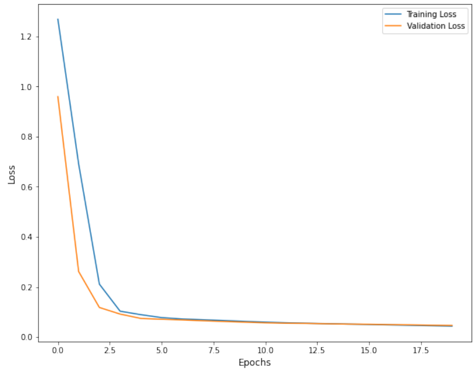

Figure 43: Loss [GRU[fasttext]]

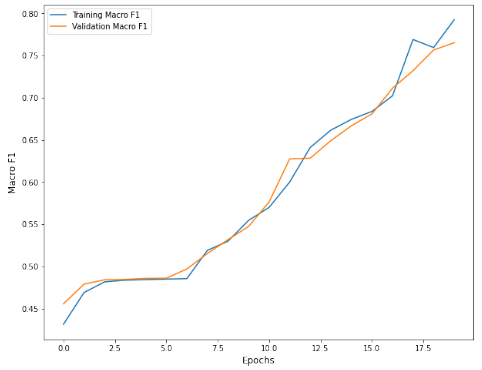

Figure 44: Macro F1 [GRU[fasttext]]

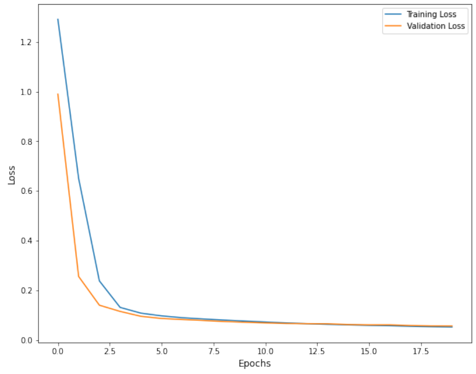

Figure 45: Loss [LSTM[fasttext]]

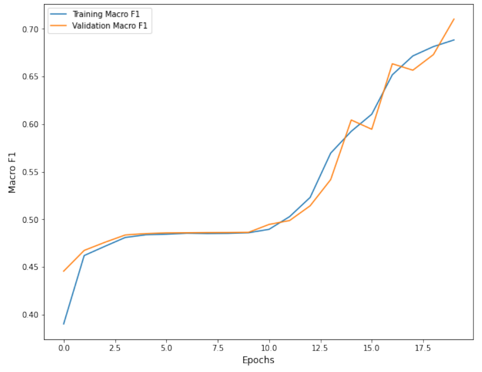

Figure 46: Macro F1 [LSTM[fasttext]]

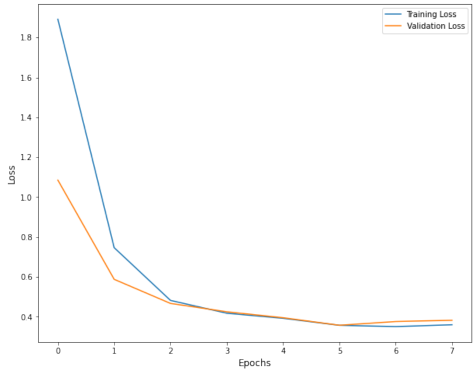

Figure 47: Loss [Bi-LSTM CRF[fasttext]]

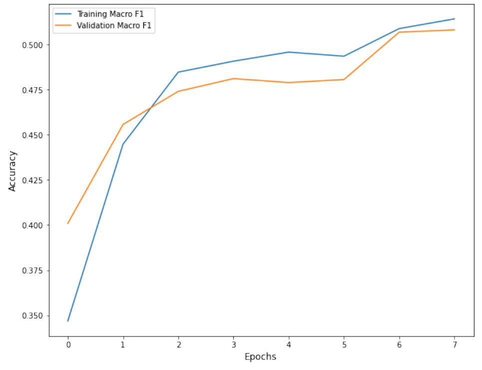

Figure 48: Macro F1 [Bi-LSTM CRF[fasttext]]

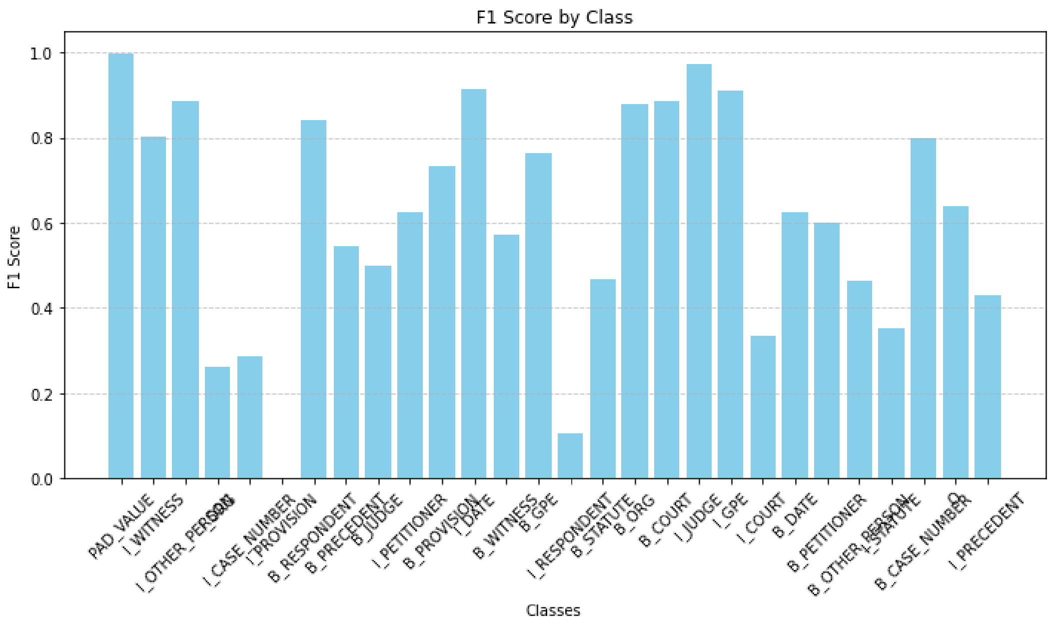

Figure 49

PERFORMANCE OF MODELS ON TEST DATA

Model No.	Model	Embedding Used	Macro F1	Accuracy
1	Vanilla RNN	Word2Vec	0.355	0.943
2	LSTM	Word2Vec	0.409	0.951
3	GRU	Word2Vec	0.485	0.956
4	Bi-LSTM CRF	Word2Vec	0.181	0.823
5	Vanilla RNN	GloVe	0.287	0.947
6	LSTM	GloVe	0.371	0.951
7	GRU	GloVe	0.363	0.951
8	Bi-LSTM CRF	GloVe	0.082	0.85
9	Vanilla RNN	FastText	0.516	0.966
10	LSTM	FastText	0.565	0.969
11	GRU	FastText	0.584	0.969
12	Bi-LSTM CRF	FastText	0.373	0.952

Table 1: Performance of Models on Dataset_1

Model No.	Model	Embedding Used	Macro F1	Accuracy
1	Vanilla RNN	Word2Vec	0.480	0.966
2	LSTM	Word2Vec	0.584	0.969
3	GRU	Word2Vec	0.625	0.971
4	Bi-LSTM CRF	Word2Vec	0.502	0.959
5	Vanilla RNN	GloVe	0.649	0.978
6	LSTM	GloVe	0.746	0.981
7	GRU	GloVe	0.784	0.983
8	Bi-LSTM CRF	GloVe	0.541	0.953
9	Vanilla RNN	FastText	0.733	0.980
10	LSTM	FastText	0.710	0.981
11	GRU	FastText	0.765	0.983
12	Bi-LSTM CRF	FastText	0.508	0.946

Table 2: Performance of Models on Dataset_2

NLP ANALYSIS OF MODEL PERFORMANCE

Dataset 1 Analysis:

- **Embedding Impact:**

- Fasttext embeddings generally outperform Word2Vec and GloVe embeddings, with higher macro F1 scores and accuracies.

- **Model Performance:**
  - Vanilla RNN, LSTM, and GRU models using FastText embeddings demonstrate the best overall performance, with consistently high macro F1 scores and accuracies.
  - Bi-LSTM CRF models with Word2Vec and Glove embedding show comparatively lower performance, indicating that the CRF layer may not be effectively capturing the sequence information in this dataset.
  - GRU models generally perform slightly better than LSTM models across all embeddings.

- **Best Performing Model:**
  - GRU with FastText embeddings (Model 11) achieves the highest macro F1 score (0.584) and accuracy (0.969) in Dataset 1.

Dataset 2 Analysis:

- **Embedding Impact:**
  - GloVe embeddings consistently outperform Word2Vec embeddings across all models.
  - FastText embeddings still demonstrate superior performance compared to Word2Vec but show slightly lower performance compared to GloVe in this dataset.

- **Model Performance:**
  - Vanilla RNN, LSTM, and GRU models using GloVe embeddings exhibit the highest performance in terms of macro F1 scores and accuracies.
  - Bi-LSTM CRF models continue to show lower performance compared to other architectures, indicating potential issues with capturing complex sequence patterns in this dataset.

- **Best Performing Model:**
  - GRU with GloVe embeddings (Model 7) achieves the highest macro F1 score (0.784) and accuracy (0.983) in Dataset 2.

General Observations:

- **Embedding Quality:**
  - The choice of embedding greatly impacts model performance, with GloVe and FastText embeddings generally outperforming Word2Vec embeddings. This suggests that richer semantic information captured by GloVe and FastText embeddings contributes to better model understanding.

- **Model Complexity:**
  - Simple models like Vanilla RNNs perform reasonably well but are outperformed by more sophisticated architectures like LSTM, GRU, and Bi-LSTM CRF, indicating the importance of capturing long-range dependencies and contextual information in NLP tasks.

- **Trade-offs:**

- While Bi-LSTM CRF models offer potential for capturing complex sequential patterns, they come with higher computational costs and may require more tuning to achieve optimal performance compared to simpler models like GRU and LSTM.

- **Dataset Sensitivity:**
  - Performance variation across datasets suggests that models may generalize differently depending on the characteristics of the data. GRU models with GloVe embeddings show remarkable consistency across datasets, indicating robustness to variations in data distribution.

Credit Statement

- **Arjit Singh Arora -** Embedding Layer Definitions, CRF Model
- **Akshat Gupta -** Embedding Layer Definitions, Baseline Models, CRF Model
- **Kumar Aryan Singh -** CRF Model
- **Swati Sharma -** Baseline Models, CRF Model