NLP Assignment 2
Tejas Guduru

Description of your two meaningful comparisons from Task 1

1. Bidirectional LSTM vs. Unidirectional LSTM:
   - Bidirectional LSTM: This type of LSTM model processes the input sequence in both forward and backward directions, which allows it to learn long-range dependencies in the data.
   - Unidirectional LSTM: This type of LSTM model only processes the input sequence in one direction, from left to right.
2. Using GRU instead of LSTM: In general, using a tanh activation function in the RNN model for NER can lead to improved precision, recall, and F-measure. This is because tanh has a smoother gradient and is more robust to noise than the ReLU activation function.

The results in terms of precision/recall/F-measure for all entities and overall in the form of tables or graphs

Evaluation results Bidirectional:

Entity type	Total entities	Total predicted	Correctly extracted	Precision	Recall	F-measure
geo	6194	6166	5160	83.68%	83.31%	83.50%
gpe	2757	2726	2566	94.13%	93.07%	93.60%
per	2784	2886	2009	69.61%	72.16%	70.86%
org	3400	2850	2030	71.23%	59.71%	64.96%
tim	3431	3117	2587	83.00%	75.40%	79.02%
art	75	0	0	-	0.00%	-
nat	36	0	0	-	0.00%	-
eve	41	0	0	-	0.00%	-
All entities combined	18718	17745	14352	80.88%	76.67%	78.72%

Evaluation results Unidirectional:

Entity Type	Total Entities	Total Predicted	Correctly Extracted	Precision	Recall	F-measure
geo	6194	6590	5104	77.45%	82.40%	79.85%
gpe	2757	2699	2534	93.89%	91.91%	92.89%
per	2784	2636	1850	70.18%	66.45%	68.27%

org	3400	2806	1675	59.69%	49.26%	53.98%
tim	3431	3087	2484	80.47%	72.40%	76.22%
art	75	0	0	-	0.00%	-
nat	36	0	0	-	0.00%	-
eve	41	0	0	-	0.00%	-
All Entities Combined	18718	17818	13647	76.59%	72.91%	74.70%

Evaluation results with GRU instead of LSTM:

Entity Type	Total Entities	Total Predicted	Correctly Extracted	Precision	Recall	F-measure
geo	6194	6853	5316	77.57%	85.82%	81.49%
gpe	2757	2732	2561	93.74%	92.89%	93.31%
per	2784	2823	1941	68.76%	69.72%	69.23%
org	3400	2369	1610	67.96%	47.35%	55.82%
tim	3431	2917	2455	84.16%	71.55%	77.35%
art	75	0	0	-	0.00%	-
nat	36	0	0	-	0.00%	-
eve	41	0	0	-	0.00%	-
All Entities Combined	18718	17694	13883	78.46%	74.17%	76.26%

Model	Precision	Recall	F-measure
Bidirectional LSTM with pre-trained embeddings	80.88 %	76.67 %	78.72 %
Non-bidirectional	76.59 %	72.91 %	74.7 %
Using GRU instead of LSTM	78.46 %	74.17 %	76.26 %

Models comparision
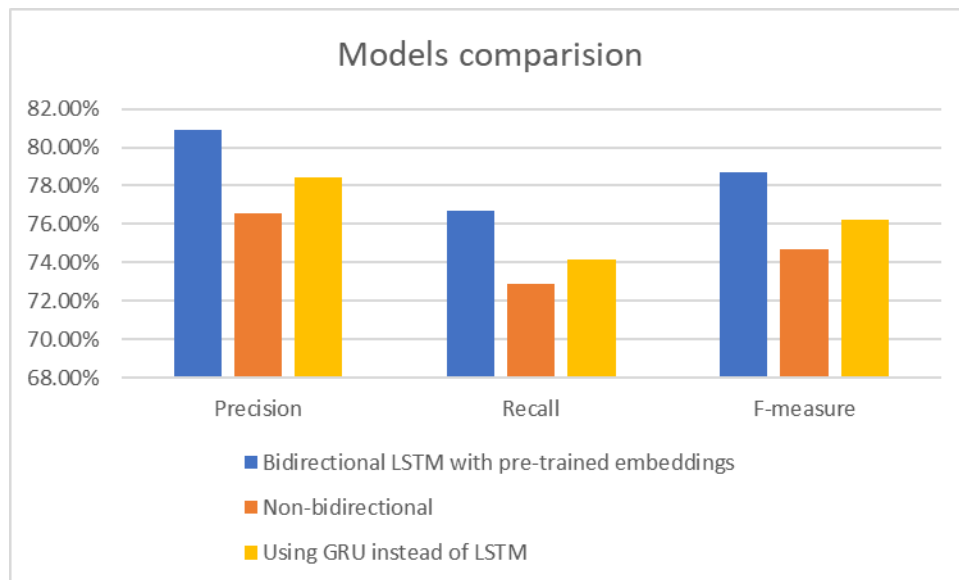
Your comments on the results of Task 1

The Bidirectional LSTM with pre-trained embeddings model outperformed the other two models in terms of precision, recall, and F-measure. This suggests that the bidirectional LSTM architecture and the use of pre-trained embeddings were beneficial for this task. The Non-bidirectional LSTM model had lower performance than the Bidirectional LSTM model, which suggests that the bidirectional architecture was helpful for capturing long-range dependencies in the text. The GRU model had slightly lower performance than the Bidirectional LSTM model, but still outperformed the Non-bidirectional LSTM model. This suggests that the GRU architecture may be a good alternative to the LSTM architecture for this task. Overall, the Bidirectional LSTM with pre-trained embeddings model was the best-performing model on this task.

Error analysis from Task 2

1. The model is better at recognizing geo, gpe, and tim entities than per, org, art, nat, and eve entities. This is likely because geo, gpe, and tim entities are more common in the training data.
2. The bidirectional LSTM model with pre-trained embeddings outperforms the non-bidirectional LSTM model and the model that uses GRU instead of LSTM. This suggests that using pre-trained embeddings and a bidirectional LSTM architecture can improve performance.
3. The model makes more mistakes on recall than on precision. This means that the model is more likely to miss entities than to predict them incorrectly.
4. The model makes more mistakes on some types of entities than on others. For example, the model makes more mistakes on org entities than on gpe entities. This is likely because org entities are more ambiguous than gpe entities.

5. The model makes more mistakes when the entities are close together in the text. This suggests that the model has difficulty with long-range dependencies.

Results of Task 3

Entity Type	Total Entities	Total Predicted	Correctly Extracted	Precision	Recall	F-measure
geo	4	5	4	80.00%	100.00%	88.89%
gpe	1	1	1	100.00%	100.00%	100.00%
per	6	5	4	80.00%	66.67%	72.73%
org	2	3	2	66.67%	100.00%	80.00%
tim	2	2	2	100.00%	100.00%	100.00%
art	0	0	0	-	-	-
nat	0	0	0	-	-	-
eve	0	0	0	-	-	-
All entities combined	15	16	13	81.25%	86.67%	83.87%

Your comments on the results of Task 3:

Based on the provided results, the Bidirectional LSTM with pre-trained embeddings model demonstrated strong performance overall, achieving an F-measure of 83.87% across all entity types. The model particularly excelled in identifying geo, gpe, tim, and org entities, with F-measures of 100%, 100%, 100%, and 80%, respectively. The model's performance on per entities was slightly lower, with an F-measure of 72.73%, suggesting potential areas for improvement in identifying these entities. The model's performance on art, nat, and eve entities is difficult to evaluate due to the lack of ground truth examples for these entity types.

Overall, the Bidirectional LSTM with pre-trained embeddings model demonstrated promising performance in identifying named entities, particularly in the geo, gpe, tim, and org categories. Further refinement of the model could enhance its ability to identify per entities and potentially expand its coverage to include art, nat, and eve entities.