

Câu 1

test acc: 0.9608 | test f1_macro: 0.5288 | test f1_micro: 0.9608

classification report:

	precision	recall	f1-score	support
B-art	1.00	0.00	0.00	56
B-eve	0.78	0.16	0.26	45
B-geo	0.82	0.87	0.85	5663
B-gpe	0.93	0.94	0.94	2372
B-nat	0.00	0.00	0.00	35
B-org	0.73	0.57	0.64	2893
B-per	0.86	0.77	0.81	2642
B-tim	0.93	0.78	0.85	3086
I-art	1.00	0.00	0.00	34
I-eve	0.33	0.03	0.05	39
I-geo	0.77	0.80	0.78	1059
I-gpe	1.00	0.32	0.48	25
I-nat	1.00	0.00	0.00	11
I-org	0.82	0.69	0.75	2478
I-per	0.88	0.83	0.85	2643
I-tim	0.82	0.66	0.73	1027
0	0.98	0.99	0.99	132778
accuracy			0.96	156886
macro avg	0.80	0.49	0.53	156886
weighted avg	0.96	0.96	0.96	156886

Câu 2

test acc: 0.8396 | test f1_macro: 0.4628 | test f1_micro: 0.8396
 classification report:

	precision	recall	f1-score	support
B-AGE	0.93	0.77	0.85	547
B-DATE	0.76	0.85	0.80	1613
B-GENDER	0.87	0.57	0.69	436
B-JOB	1.00	0.00	0.00	172
B-LOCATION	0.66	0.24	0.35	4305
B-NAME	0.88	0.45	0.59	302
B-ORGANIZATION	0.75	0.13	0.22	759
B-PATIENT_ID	0.87	0.86	0.87	1904
B-SYMPTOM_AND_DISEASE	0.77	0.43	0.56	1110
B-TRANSPORTATION	0.91	0.78	0.84	190
I-AGE	1.00	0.00	0.00	6
I-DATE	0.95	0.99	0.97	1674
I-JOB	1.00	0.00	0.00	114
I-LOCATION	0.65	0.28	0.39	4735
I-NAME	1.00	0.00	0.00	10
I-ORGANIZATION	0.69	0.38	0.49	1978
I-PATIENT_ID	1.00	0.00	0.00	27
I-SYMPTOM_AND_DISEASE	0.79	0.37	0.51	1455
I-TRANSPORTATION	0.88	0.58	0.70	66
0	0.85	0.97	0.91	61094
[SEP]	0.00	1.00	0.00	0
accuracy			0.84	82497
macro avg	0.82	0.46	0.46	82497
weighted avg	0.82	0.84	0.81	82497

INFO:root:task done!!!

Câu 3.1

test acc: 0.9649 | test f1_macro: 0.5166 | test f1_micro: 0.9649

classification report:

	precision	recall	f1-score	support
B-art	0.00	0.00	0.00	56
B-eve	0.57	0.09	0.15	45
B-geo	0.84	0.88	0.86	5663
B-gpe	0.94	0.92	0.93	2372
B-nat	1.00	0.06	0.11	35
B-org	0.79	0.70	0.74	2893
B-per	0.82	0.84	0.83	2642
B-tim	0.88	0.85	0.86	3086
I-art	1.00	0.00	0.00	34
I-eve	0.17	0.03	0.04	39
I-geo	0.80	0.79	0.80	1059
I-gpe	0.85	0.44	0.58	25
I-nat	1.00	0.00	0.00	11
I-org	0.81	0.77	0.79	2478
I-per	0.82	0.91	0.86	2643
I-tim	0.76	0.75	0.75	1027
0	0.99	0.99	0.99	132778
[SEP]	0.00	1.00	0.00	0
accuracy			0.96	156886
macro avg	0.72	0.56	0.52	156886
weighted avg	0.96	0.96	0.96	156886

Câu 3.2

test acc: 0.7776 | test f1_macro: 0.2611 | test f1_micro: 0.7776
classification report:

	precision	recall	f1-score	support
B-AGE	0.73	0.43	0.54	510
B-DATE	0.88	0.48	0.62	1499
B-GENDER	0.75	0.55	0.64	403
B-JOB	1.00	0.00	0.00	154
B-LOCATION	0.41	0.13	0.20	3718
B-NAME	1.00	0.00	0.00	261
B-ORGANIZATION	0.67	0.35	0.46	716
B-PATIENT_ID	0.79	0.37	0.50	1698
B-SYMPTOM_AND_DISEASE	0.34	0.03	0.06	968
B-TRANSPORTATION	1.00	0.00	0.00	170
I-AGE	1.00	0.00	0.00	6
I-DATE	0.87	0.54	0.67	1497
I-JOB	1.00	0.00	0.00	102
I-LOCATION	0.41	0.17	0.24	4013
I-NAME	1.00	0.00	0.00	8
I-ORGANIZATION	0.64	0.45	0.52	1871
I-PATIENT_ID	1.00	0.00	0.00	23
I-SYMPTOM_AND_DISEASE	0.31	0.11	0.16	1191
I-TRANSPORTATION	1.00	0.00	0.00	59
0	0.80	0.96	0.87	53149
[SEP]	0.00	1.00	0.00	0
accuracy			0.78	72016
macro avg	0.74	0.27	0.26	72016
weighted avg	0.74	0.78	0.73	72016

Nhận xét:

- Bộ dữ liệu tiếng anh: cả LSTM và BERT đều đạt đến mức độ chính xác (Accuracy) cao là 0.96. Điều này có thể chỉ ra rằng cả hai mô hình đều tốt trên tập dữ liệu tiếng Anh và không có sự chênh lệch lớn về hiệu suất giữa chúng
- Bộ dữ liệu tiếng việt: LSTM có vẻ vượt trội hơn BERT với mức độ chính xác (Accuracy) là 0.84 so với 0.78. Điều này có thể do nhiều yếu tố, bao gồm đặc tính của dữ liệu, kích thước dữ liệu, kiến trúc mô hình, và cách tiếp cận huấn luyện.
- Trong quyết định lựa chọn giữa LSTM và BERT, nên xem xét các yếu tố như kích thước dữ liệu, đặc tính của ngôn ngữ, và tài nguyên tính toán. BERT thường được coi là một trong những mô hình mạnh mẽ nhất cho NLP tasks nhưng đòi hỏi nhiều tài nguyên tính toán hơn so với LSTM. LSTM, mặc dù đôi khi không thể đạt đến

độ phức tạp như BERT, nhưng nó có thể hoạt động tốt trên những tập dữ liệu nhỏ và tài nguyên tính toán hạn chế.