

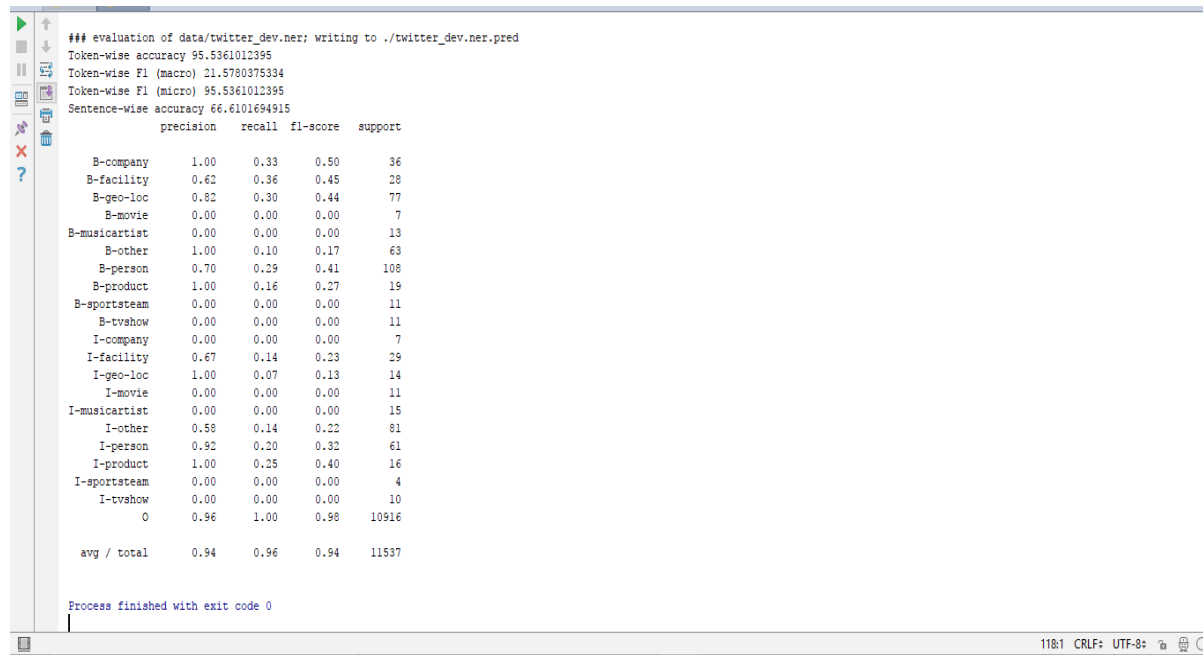
QUESTION:

Implement the features you described and evaluate how much they helped on the dev and dev test set for the logistic regression tagger (you should just need to execute main.py after making your changes in feat_gen.py). Briefly write down your improvements.

ANSWER:

1. Twitter_dev.ner

a. Without Features



```
### evaluation of data/twitter_dev.ner; writing to ./twitter_dev.ner.pred
Token-wise accuracy 95.5361012395
Token-wise F1 (macro) 21.5780375334
Token-wise F1 (micro) 95.5361012395
Sentence-wise accuracy 66.6101694915
```

	precision	recall	f1-score	support
B-company	1.00	0.33	0.50	36
B-facility	0.62	0.36	0.45	28
B-geo-loc	0.82	0.30	0.44	77
B-movie	0.00	0.00	0.00	7
B-musicartist	0.00	0.00	0.00	13
B-other	1.00	0.10	0.17	63
B-person	0.70	0.29	0.41	108
B-product	1.00	0.16	0.27	19
B-sportsteam	0.00	0.00	0.00	11
B-tvshow	0.00	0.00	0.00	11
I-company	0.00	0.00	0.00	7
I-facility	0.67	0.14	0.23	29
I-geo-loc	1.00	0.07	0.13	14
I-movie	0.00	0.00	0.00	11
I-musicartist	0.00	0.00	0.00	15
I-other	0.58	0.14	0.22	81
I-person	0.92	0.20	0.32	61
I-product	1.00	0.25	0.40	16
I-sportsteam	0.00	0.00	0.00	4
I-tvshow	0.00	0.00	0.00	10
O	0.96	1.00	0.98	10916
avg / total	0.94	0.96	0.94	11537

Process finished with exit code 0

```
bash: data/twitter_dev.ner.pred: No such file or directory
root@Roshani:/mnt/d/Homework-6# data/conlleval.pl -d \t < twitter_dev.ner.pred
processed 11537 tokens with 373 phrases; found: 127 phrases; correct: 63.
accuracy: 95.54%; precision: 49.61%; recall: 16.89%; FB1: 25.20
company: precision: 100.00%; recall: 33.33%; FB1: 50.00 12
facility: precision: 10.00%; recall: 7.14%; FB1: 8.33 20
geo-loc: precision: 78.57%; recall: 28.57%; FB1: 41.90 28
movie: precision: 0.00%; recall: 0.00%; FB1: 0.00 0
musicartist: precision: 0.00%; recall: 0.00%; FB1: 0.00 0
other: precision: 15.79%; recall: 4.76%; FB1: 7.32 19
person: precision: 46.67%; recall: 19.44%; FB1: 27.45 45
product: precision: 100.00%; recall: 15.79%; FB1: 27.27 3
sportsteam: precision: 0.00%; recall: 0.00%; FB1: 0.00 0
tvshow: precision: 0.00%; recall: 0.00%; FB1: 0.00 0
root@Roshani:/mnt/d/Homework-6#
```

b) With Additional features:

```
### evaluation of data/twitter_dev.ner; writing to ./twitter_dev.ner.pred
Token-wise accuracy 96.316200052
Token-wise F1 (macro) 29.3101228151
Token-wise F1 (micro) 96.316200052
Sentence-wise accuracy 68.9830508475
```

	precision	recall	f1-score	support
B-company	0.82	0.39	0.53	36
B-facility	0.67	0.43	0.52	28
B-geo-loc	0.77	0.53	0.63	77
B-movie	0.00	0.00	0.00	7
B-musicartist	0.00	0.00	0.00	13
B-other	0.88	0.22	0.35	63
B-person	0.74	0.66	0.70	108
B-product	0.75	0.16	0.26	19
B-sportsteam	0.00	0.00	0.00	11
B-tvshow	0.33	0.09	0.14	11
I-company	0.00	0.00	0.00	7
I-facility	0.47	0.31	0.38	29
I-geo-loc	0.25	0.07	0.11	14
I-movie	0.00	0.00	0.00	11
I-musicartist	0.50	0.07	0.12	15
I-other	0.59	0.21	0.31	81
I-person	0.89	0.64	0.74	61
I-product	0.80	0.25	0.38	16
I-sportsteam	0.00	0.00	0.00	4
I-tvshow	0.00	0.00	0.00	10
O	0.97	1.00	0.98	10916
avg / total	0.95	0.96	0.95	11537

Process finished with exit code 0

```
Selectroot@Roshani:/mnt/d/Homework-6
sportsteam: precision: 0.00%; recall: 0.00%; FB1: 0.00 0
tvshow: precision: 0.00%; recall: 0.00%; FB1: 0.00 0
root@Roshani:/mnt/d/Homework-6# data/ conlleval.pl -d \t < twitter_dev.ner.pred
processed 11537 tokens with 373 phrases; found: 252 phrases; correct: 128.
accuracy: 96.32%; precision: 50.79%; recall: 34.32%; FB1: 40.96
company: precision: 73.68%; recall: 38.89%; FB1: 50.91 19
facility: precision: 13.79%; recall: 14.29%; FB1: 14.04 29
geo-loc: precision: 66.67%; recall: 46.75%; FB1: 54.96 54
movie: precision: 0.00%; recall: 0.00%; FB1: 0.00 2
musicartist: precision: 0.00%; recall: 0.00%; FB1: 0.00 3
other: precision: 29.41%; recall: 15.87%; FB1: 20.62 34
person: precision: 59.80%; recall: 56.48%; FB1: 58.10 102
product: precision: 60.00%; recall: 15.79%; FB1: 25.00 5
sportsteam: precision: 0.00%; recall: 0.00%; FB1: 0.00 1
tvshow: precision: 0.00%; recall: 0.00%; FB1: 0.00 3
root@Roshani:/mnt/d/Homework-6#
```

Comparison for Log-reg before and after adding features:

Twitter_dev.ner

Metrix for main.py with & without features

The Token wise accuracy for log-reg without adding features is 95.53% and the sentence wise accuracy as 66.61%

After adding features the changes reflected are quiet noticeable. Comparing the Token wise and Sentence Wise accuracy improves to 96 and 68% respectively.

Matrix for conllevl.pl with & without features

FB1 score after running conllevl is only about $\frac{1}{4}$ % before adding any additional features. FB1 of person is 27.45 and FB1 of Geo-location is 41.90 these two are the scores that get affected largely by adding feature to them.

We see that Categories "Company and Geo-Loc" have top FB1 score respectively.

After adding features the overall FB1 scores almost improves by 25%. And person and Geo location improves by 31% and 13% respectively.

The overall accuracy, precision and recall is improved to an extent by adding features.

2. Twitter_dev_test.ner

a) Without features

```
### evaluation of data/twitter_dev_test.ner; writing to ./twitter_dev_test.ner.pred
Token-wise accuracy 91.0152104705
Token-wise F1 (macro) 10.9195384447
Token-wise F1 (micro) 91.0152104705
Sentence-wise accuracy 48.6486486486
```

	precision	recall	f1-score	support
B-company	0.82	0.08	0.15	109
B-facility	0.56	0.30	0.39	46
B-geo-loc	0.72	0.30	0.42	159
B-movie	0.00	0.00	0.00	4
B-musicartist	0.00	0.00	0.00	33
B-other	0.00	0.00	0.00	118
B-person	0.22	0.10	0.14	96
B-product	0.00	0.00	0.00	44
B-sportsteam	0.00	0.00	0.00	31
B-tvshow	0.00	0.00	0.00	4
I-company	0.00	0.00	0.00	26
I-facility	0.00	0.00	0.00	60
I-geo-loc	1.00	0.03	0.05	37
I-movie	0.00	0.00	0.00	10
I-musicartist	0.00	0.00	0.00	15
I-other	0.62	0.08	0.14	123
I-person	0.14	0.02	0.03	58
I-product	0.00	0.00	0.00	88
I-sportsteam	0.00	0.00	0.00	7
I-tvshow	0.00	0.00	0.00	9
O	0.92	1.00	0.95	10231
avg / total	0.86	0.91	0.88	11308

Process finished with exit code 0

```
root@Roshani:/mnt/d/Homework-6
sportsteam: precision: 0.00%; recall: 0.00%; FB1: 0.00 0
tvshow: precision: 0.00%; recall: 0.00%; FB1: 0.00 0
root@Roshani:/mnt/d/Homework-6# data/ conlleval.pl -d \t < twitter_dev_test.ner.pred
processed 11308 tokens with 644 phrases; found: 170 phrases; correct: 55.
accuracy: 91.02%; precision: 32.35%; recall: 8.54%; FB1: 13.51
company: precision: 72.73%; recall: 7.34%; FB1: 13.33 11
facility: precision: 4.00%; recall: 2.17%; FB1: 2.82 25
geo-loc: precision: 58.21%; recall: 24.53%; FB1: 34.51 67
movie: precision: 0.00%; recall: 0.00%; FB1: 0.00 0
musicartist: precision: 0.00%; recall: 0.00%; FB1: 0.00 0
other: precision: 0.00%; recall: 0.00%; FB1: 0.00 17
person: precision: 14.29%; recall: 7.29%; FB1: 9.66 49
product: precision: 0.00%; recall: 0.00%; FB1: 0.00 1
sportsteam: precision: 0.00%; recall: 0.00%; FB1: 0.00 0
tvshow: precision: 0.00%; recall: 0.00%; FB1: 0.00 0
root@Roshani:/mnt/d/Homework-6#
```

b) With Features

I-tvshow	1.00	0.62	0.76	21
O	0.99	1.00	0.99	33091
avg / total	0.99	0.99	0.99	34932
### evaluation of data/twitter_dev_test.ner; writing to ./twitter_dev_test.ner.pred				
Token-wise accuracy 92.3328616908				
Token-wise F1 (macro) 24.7557440656				
Token-wise F1 (micro) 92.3328616908				
Sentence-wise accuracy 51.920341394				
	precision	recall	f1-score	support
B-company	0.70	0.19	0.30	109
B-facility	0.59	0.37	0.45	46
B-geo-loc	0.74	0.60	0.67	159
B-movie	0.00	0.00	0.00	4
B-musicartist	0.00	0.00	0.00	33
B-other	0.13	0.03	0.04	118
B-person	0.47	0.59	0.53	96
B-product	0.25	0.05	0.08	44
B-sportsteam	0.00	0.00	0.00	31
B-tvshow	0.00	0.00	0.00	4
I-company	1.00	0.12	0.21	26
I-facility	0.56	0.30	0.39	60
I-geo-loc	0.69	0.54	0.61	37
I-movie	0.00	0.00	0.00	10
I-musicartist	0.00	0.00	0.00	15
I-other	0.37	0.19	0.25	123
I-person	0.66	0.69	0.67	58
I-product	0.25	0.02	0.04	88
I-sportsteam	0.00	0.00	0.00	7
I-tvshow	0.00	0.00	0.00	9
O	0.94	0.99	0.97	10231
avg / total	0.90	0.92	0.90	11308
Process finished with exit code 0				

```
root@Roshani: /mnt/d/Homework-6
tvshow: precision: 0.00%; recall: 0.00%; F1: 0.00 3
root@Roshani:/mnt/d/Homework-6# data/conlleval.pl -d \\t < twitter_dev_test.ner.pred
processed 11308 tokens with 644 phrases; found: 450 phrases; correct: 170.
accuracy: 92.33%; precision: 37.78%; recall: 26.40%; F1: 31.08
company: precision: 62.50%; recall: 18.35%; F1: 28.37 32
facility: precision: 12.00%; recall: 13.04%; F1: 12.50 50
geo-loc: precision: 65.22%; recall: 56.60%; F1: 60.61 138
movie: precision: 0.00%; recall: 0.00%; F1: 0.00 0
musicartist: precision: 0.00%; recall: 0.00%; F1: 0.00 4
other: precision: 0.00%; recall: 0.00%; F1: 0.00 66
person: precision: 39.85%; recall: 55.21%; F1: 46.29 133
product: precision: 6.25%; recall: 2.27%; F1: 3.33 16
sportsteam: precision: 0.00%; recall: 0.00%; F1: 0.00 8
tvshow: precision: 0.00%; recall: 0.00%; F1: 0.00 3
root@Roshani:/mnt/d/Homework-6#
```

Comparison for Log-reg before and after adding features:

Twitter_dev_test.ner

Metrix for main.py with & without features

The Token wise accuracy for log-reg without adding features is 91.0% and the sentence wise accuracy as 48%

After adding features the changes reflected are quiet noticeable. Comparing the Token wise and Sentence Wise accuracy improves to 92.33% and 51.92% respectively.

Matrix for conlleval.pl with & without features

FB1 score after running conlleval is only about 13.51% before adding any additional features. FB1 of person is 9.66 and FB1 of Geo-location is 34.51 these two are the scores that get affected largely by adding feature to them.

We see that Categories "Company and Geo-Loc" have top FB1 score respectively, before adding features but after adding features these top positions are taken by "Geo-Loc and Person"

After adding features the overall FB1 scores almost improves by 26%. And person Score improves by nearly 32%.

The overall accuracy, precision and especially recall is improved to an extent by adding features.