

### Question 3.2

- 1) Comparing the difference in the output of the evaluation between logistic regression and conditional random field is shown below using just the base features.

**Logistic Regression (without features):** To remove redundancy, I have just added the result from the conllevall evaluation.

#### CONLL evaluation for twitter\_dev.ner.pred(without features)

```
guest-wireless-207-151-035-009:Homework-6 RishabhTyagi$ data/conllevall.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
```

#### CRF: Conditional Random Field (without features):

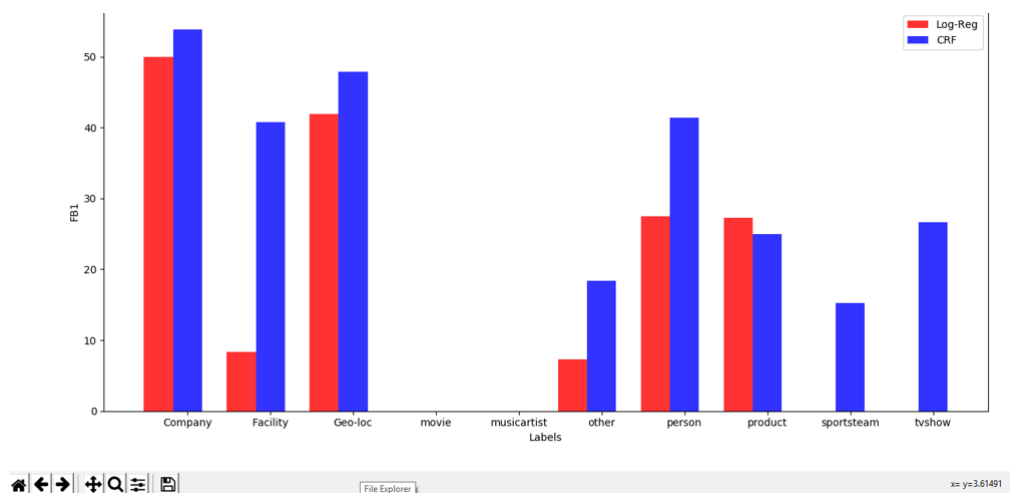
#### CONLL evaluation for twitter\_dev.ner.pred(without features)

```
guest-wireless-207-151-035-009:Homework-6 RishabhTyagi$ data/conllevall.pl -d \\t < twitter_dev.ner.pred
processed 11537 tokens with 373 phrases; found: 165 phrases; correct: 100.
accuracy: 95.77%; precision: 60.61%; recall: 26.81%; FB1: 37.17
company: precision: 87.50%; recall: 38.89%; FB1: 53.85 16
facility: precision: 47.62%; recall: 35.71%; FB1: 40.82 21
geo-loc: precision: 70.00%; recall: 36.36%; FB1: 47.86 40
movie: precision: 0.00%; recall: 0.00%; FB1: 0.00 1
musicartist: precision: 0.00%; recall: 0.00%; FB1: 0.00 2
other: precision: 53.85%; recall: 11.11%; FB1: 18.42 13
person: precision: 57.38%; recall: 32.41%; FB1: 41.42 61
product: precision: 60.00%; recall: 15.79%; FB1: 25.00 5
sportsteam: precision: 50.00%; recall: 9.09%; FB1: 15.38 2
tvshow: precision: 50.00%; recall: 18.18%; FB1: 26.67 4
```

Accuracy of Logistic Regression: 95.54%; FB1: 25.20

Accuracy of Conditional Random Field: 95.77%; FB1: 37.17

#### Graph for CRF vs LogReg for twitter\_dev.ner (without features):



## CONLL evaluation for twitter\_dev\_test.ner.pred(without features)

```

guest-wireless-207-151-035-009:Homework-6 RishabhTyagi$ 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

```

## CONLL evaluation for twitter\_dev\_test.ner.pred(without features)

```

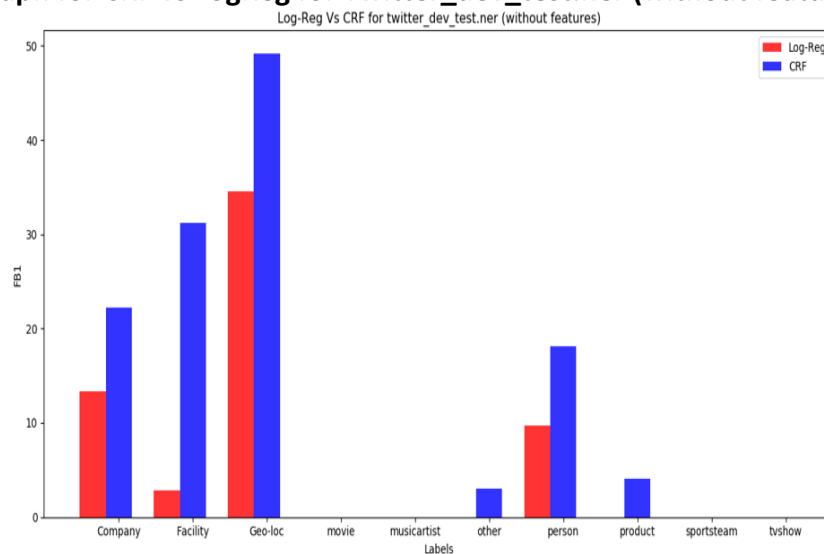
tvshow: precision: 50.00%; recall: 18.18%; FB1: 26.67 4
guest-wireless-207-151-035-009:Homework-6 RishabhTyagi$ data/conlleval.pl -d \\t < twitter_dev_test.ner.pred
processed 11308 tokens with 644 phrases; found: 220 phrases; correct: 103.
accuracy: 91.31%; precision: 46.82%; recall: 15.99%; FB1: 23.84
company: precision: 82.35%; recall: 12.84%; FB1: 22.22 17
facility: precision: 38.71%; recall: 26.09%; FB1: 31.17 31
geo-loc: precision: 72.84%; recall: 37.11%; FB1: 49.17 81
movie: precision: 0.00%; recall: 0.00%; FB1: 0.00 0
musicartist: precision: 0.00%; recall: 0.00%; FB1: 0.00 1
other: precision: 14.29%; recall: 1.69%; FB1: 3.03 14
person: precision: 21.43%; recall: 15.62%; FB1: 18.07 70
product: precision: 20.00%; recall: 2.27%; FB1: 4.08 5
sportsteam: precision: 0.00%; recall: 0.00%; FB1: 0.00 0
tvshow: precision: 0.00%; recall: 0.00%; FB1: 0.00 1
guest-wireless-207-151-035-009:Homework-6 RishabhTyagi$

```

Accuracy of Logistic Regression: 91.02%; FB1: 13.51

Accuracy of Conditional Random Field: 91.31%; FB1: 23.84

## Graph for CRF vs LogReg for Twitter\_dev\_test.ner (without features):



- 2) Comparing the difference in the output of the evaluation between logistic regression and conditional random field is shown below using baseline along with appended features.

### CONLL evaluation for twitter\_dev.ner.pred(with feature) : Using Logistic Regression Model

```
guest-wireless-207-151-035-009:Homework-6 RishabhTyagi$ data/conlleval.pl -d \\t < twitter_dev.ner.pred
processed 11537 tokens with 373 phrases; found: 249 phrases; correct: 130.
accuracy: 96.34%; precision: 52.21%; recall: 34.85%; FB1: 41.80
  company: precision: 93.33%; recall: 38.89%; FB1: 54.90 15
  facility: precision: 16.13%; recall: 17.86%; FB1: 16.95 31
  geo-loc: precision: 68.42%; recall: 50.65%; FB1: 58.21 57
  movie: precision: 0.00%; recall: 0.00%; FB1: 0.00 1
  musician: precision: 0.00%; recall: 0.00%; FB1: 0.00 2
  other: precision: 25.71%; recall: 14.29%; FB1: 18.37 35
  person: precision: 59.00%; recall: 54.63%; FB1: 56.73 100
  product: precision: 60.00%; recall: 15.79%; FB1: 25.00 5
  sportsteam: precision: 0.00%; recall: 0.00%; FB1: 0.00 1
  tvshow: precision: 50.00%; recall: 9.09%; FB1: 15.38 2
guest-wireless-207-151-035-009:Homework-6 RishabhTyagi$
```

### CRF: Conditional Random Field:

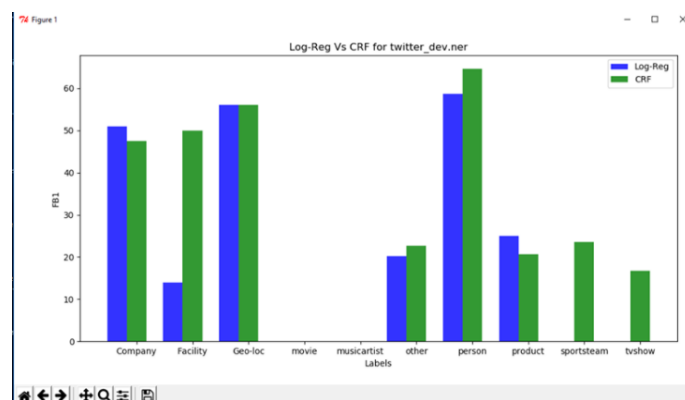
### CONLL evaluation for twitter\_dev.ner.pred(with features)

```
guest-wireless-207-151-035-009:Homework-6 RishabhTyagi$ data/conlleval.pl -d \\t < twitter_dev.ner.pred
processed 11537 tokens with 373 phrases; found: 277 phrases; correct: 160.
accuracy: 96.32%; precision: 57.76%; recall: 42.90%; FB1: 49.23
  company: precision: 63.64%; recall: 38.89%; FB1: 48.28 22
  facility: precision: 50.00%; recall: 50.00%; FB1: 50.00 28
  geo-loc: precision: 56.16%; recall: 53.25%; FB1: 54.67 73
  movie: precision: 0.00%; recall: 0.00%; FB1: 0.00 4
  musician: precision: 0.00%; recall: 0.00%; FB1: 0.00 1
  other: precision: 54.17%; recall: 20.63%; FB1: 29.89 24
  person: precision: 66.06%; recall: 66.67%; FB1: 66.36 109
  product: precision: 37.50%; recall: 15.79%; FB1: 22.22 8
  sportsteam: precision: 33.33%; recall: 18.18%; FB1: 23.53 6
  tvshow: precision: 50.00%; recall: 9.09%; FB1: 15.38 2
guest-wireless-207-151-035-009:Homework-6 RishabhTyagi$
```

Accuracy of Logistic Regression: 96.34%; FB1: 41.80

Accuracy of Conditional Random Field: 96.32%; FB1: 49.23

### Graph for CRF vs LogReg for twitter\_dev.ner (with features):



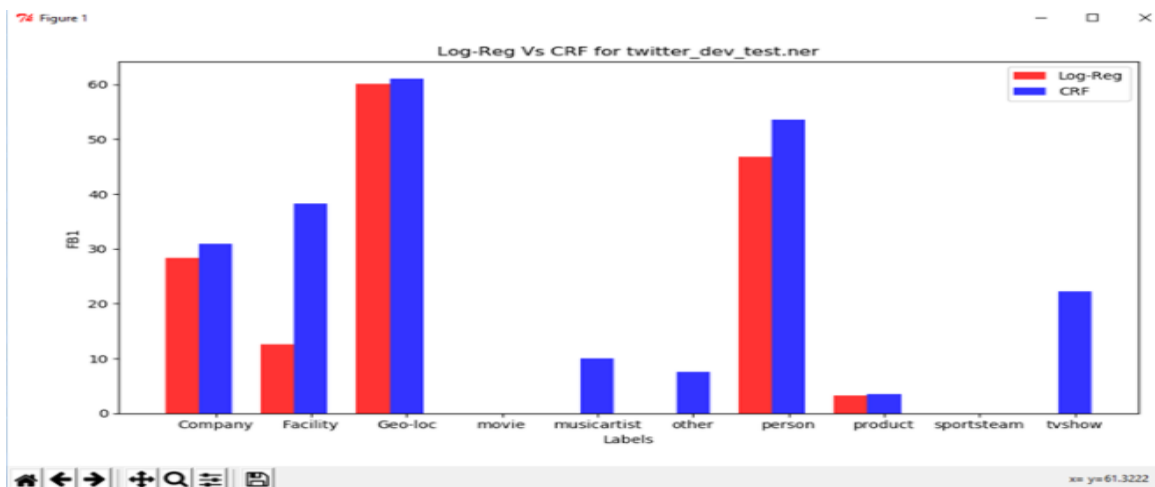
## CONLL evaluation for twitter\_dev\_test.ner.pred(with feature) : Logistic Regression Model

```
guest-wireless-207-151-035-009:Homework-6 RishabhTyagi$ data/conlleval.pl -d \\t < twitter_dev_test.ner.pred
processed 11308 tokens with 644 phrases; found: 430 phrases; correct: 173.
accuracy: 92.42%; precision: 40.23%; recall: 26.86%; FB1: 32.22
company: precision: 68.97%; recall: 18.35%; FB1: 28.99 29
facility: precision: 10.20%; recall: 10.87%; FB1: 10.53 49
geo-loc: precision: 64.19%; recall: 59.75%; FB1: 61.89 148
movie: precision: 0.00%; recall: 0.00%; FB1: 0.00 1
musicartist: precision: 0.00%; recall: 0.00%; FB1: 0.00 4
other: precision: 0.00%; recall: 0.00%; FB1: 0.00 56
person: precision: 41.27%; recall: 54.17%; FB1: 46.85 126
product: precision: 9.09%; recall: 2.27%; FB1: 3.64 11
sportsteam: precision: 0.00%; recall: 0.00%; FB1: 0.00 6
tvshow: precision: 0.00%; recall: 0.00%; FB1: 0.00 0
guest-wireless-207-151-035-009:Homework-6 RishabhTyagi$
```

## CONLL evaluation for twitter\_dev\_test.ner.pred(with features)

```
guest-wireless-207-151-035-009:Homework-6 RishabhTyagi$ data/conlleval.pl -d \\t < twitter_dev_test.ner.pred
processed 11308 tokens with 644 phrases; found: 442 phrases; correct: 207.
accuracy: 92.41%; precision: 46.83%; recall: 32.14%; FB1: 38.12
company: precision: 58.97%; recall: 21.10%; FB1: 31.08 39
facility: precision: 40.62%; recall: 28.26%; FB1: 33.33 32
geo-loc: precision: 59.76%; recall: 63.52%; FB1: 61.59 169
movie: precision: 0.00%; recall: 0.00%; FB1: 0.00 0
musicartist: precision: 0.00%; recall: 0.00%; FB1: 0.00 5
other: precision: 17.65%; recall: 5.08%; FB1: 7.89 34
person: precision: 46.27%; recall: 64.58%; FB1: 53.91 134
product: precision: 7.14%; recall: 2.27%; FB1: 3.45 14
sportsteam: precision: 7.14%; recall: 3.23%; FB1: 4.44 14
tvshow: precision: 0.00%; recall: 0.00%; FB1: 0.00 1
guest-wireless-207-151-035-009:Homework-6 RishabhTyagi$
```

## Graph for CRF vs LogReg for twitter\_dev\_test.ner (with features):



The accuracy of both the model is almost the same but the F1 score is significantly larger in the CRF model, which proves that CRF is a better model as the number of false positives and false negatives are significantly less in the CRF model. This is the reason why CRF works better than logistic regression. As per me, the CRF model works better in sequence tagging problems, mainly because this model will account for the labels of the previously observed inputs as well as the features of the current input.

using Viterbi algorithm. On the other hand, the logistic regression, will only take features of each token into account at a time, making it more susceptible to missing out on context of the word.