

**NAME :** Hardik Mittal

**ROLL NUMBER :** 2021114016

# ASSIGNMENT 1 REPORT

- **STATISTICAL METHODS**

- OBSERVATIONS**

- It was very evident that if both the training and testing set for both the models were same, then the perplexity was always in a single digit, which should be the case as the model has been trained on the same data and thus should perform well on the same set.
    - If we were training on one corpus but were testing on another corpus the perplexity was increasing drastically, which should be the case as there would be a whole new vocabulary while testing which will mess up a lot.

# PERPLEXITY SCORES

```
2021114016_LM1_test-perplexity.txt X
home > hardk > AMNESIA > College > Sem 4 > NLP > 202111401_Assignment1 > 2021114016_LM1_test-perplexity.txt
1 Average perplexity for Kneser-Ney language model is: 8.808333
2 Sentence 1: 7.335435
3 Sentence 2: 5.500927
4 Sentence 3: 6.891333
5 Sentence 4: 10.453390
6 Sentence 5: 8.095579
7 Sentence 6: 5.134219
8 Sentence 7: 7.102490
9 Sentence 8: 5.148747
10 Sentence 9: 9.712614
11 Sentence 10: 7.031899
12 Sentence 11: 6.524343
13 Sentence 12: 6.903291
14 Sentence 13: 11.210895
15 Sentence 14: 34.557594
16 Sentence 15: 7.121790
17 Sentence 16: 11.464308
18 Sentence 17: 7.889917
19 Sentence 18: 6.572348
20 Sentence 19: 7.896244
21 Sentence 20: 5.790771
22 Sentence 21: 44.601771
23 Sentence 22: 7.495570
24 Sentence 23: 6.635355
25 Sentence 24: 9.000732
26 Sentence 25: 6.265804
27 Sentence 26: 5.380864
28 Sentence 27: 7.085475
29 Sentence 28: 9.123525
30 Sentence 29: 6.319186
31 Sentence 30: 10.785097
32 Sentence 31: 6.042107
```

```
2021114016_LM2_test-perplexity.txt X
2021114016_LM2_test-perplexity.txt
1 Average perplexity for Witten-Bell language model is: 183.018632
2 Sentence 1: 23.240052
3 Sentence 2: 49.528458
4 Sentence 3: 55.578883
5 Sentence 4: 95.414616
6 Sentence 5: 89.043758
7 Sentence 6: 8172.272313
8 Sentence 7: 49.582145
9 Sentence 8: 49.810511
10 Sentence 9: 126.561745
11 Sentence 10: 50.499819
12 Sentence 11: 44.045800
13 Sentence 12: 63.219328
14 Sentence 13: 121.839481
15 Sentence 14: 106.419850
16 Sentence 15: 108.155085
17 Sentence 16: 72.331784
18 Sentence 17: 49.443068
19 Sentence 18: 184.977747
20 Sentence 19: 69.630095
21 Sentence 20: 30.619133
22 Sentence 21: 61.733599
23 Sentence 22: 17.885279
24 Sentence 23: 284.224005
25 Sentence 24: 39.123079
26 Sentence 25: 39.965228
27 Sentence 26: 13.220400
28 Sentence 27: 46.817156
29 Sentence 28: 114.123555
30 Sentence 29: 37.493029
31 Sentence 30: 33.993744
32 Sentence 31: 64.223224
```

```
2021114016_LM3_test-perplexity.txt X
2021114016_LM3_test-perplexity.txt
1 Average perplexity for Kneser-Ney language model is: 18.355698
2 Sentence 1: 7.366471
3 Sentence 2: 9.091004
4 Sentence 3: 10.649801
5 Sentence 4: 9.221365
6 Sentence 5: 9.414490
7 Sentence 6: 13.858502
8 Sentence 7: 8.772674
9 Sentence 8: 99.533918
10 Sentence 9: 20.148646
11 Sentence 10: 15.687151
12 Sentence 11: 7.480639
13 Sentence 12: 13.066367
14 Sentence 13: 6.412753
15 Sentence 14: 9.125080
16 Sentence 15: 21.114366
17 Sentence 16: 9.543268
18 Sentence 17: 104.233672
19 Sentence 18: 25.425673
20 Sentence 19: 11.748339
21 Sentence 20: 35.769136
22 Sentence 21: 96.818652
23 Sentence 22: 7.591867
24 Sentence 23: 9.118021
25 Sentence 24: 8.702248
26 Sentence 25: 6.235133
27 Sentence 26: 6.255954
28 Sentence 27: 12.892243
29 Sentence 28: 7.436754
30 Sentence 29: 5.832132
31 Sentence 30: 5.921672
32 Sentence 31: 6.215220
```

```
2021114016_LM4_test-perplexity.txt X
2021114016_LM4_test-perplexity.txt
1 Average perplexity for Witten-Bell language model is: 1600.939132
2 Sentence 1: 355.921999
3 Sentence 2: 1038.028137
4 Sentence 3: 116.457133
5 Sentence 4: 294.696386
6 Sentence 5: 2582.409577
7 Sentence 6: 146.503875
8 Sentence 7: 138.821826
9 Sentence 8: 392.878927
10 Sentence 9: 80.411423
11 Sentence 10: 1067.923506
12 Sentence 11: 54.868953
13 Sentence 12: 261.920739
14 Sentence 13: 62.453909
15 Sentence 14: 485.572459
16 Sentence 15: 12.879737
17 Sentence 16: 473.005242
18 Sentence 17: 180.309038
19 Sentence 18: 42.060277
20 Sentence 19: 4319.045353
21 Sentence 20: 1168.903491
22 Sentence 21: 177333.667259
23 Sentence 22: 2025.754246
24 Sentence 23: 77.552224
25 Sentence 24: 716.626409
26 Sentence 25: 209.114898
27 Sentence 26: 278.888368
28 Sentence 27: 197.966417
29 Sentence 28: 45.225626
30 Sentence 29: 139.527536
31 Sentence 30: 893.623137
32 Sentence 31: 10.008082
```

- **NEURAL METHOD**

- On github : [https://github.com/mhardik003/NLP\\_Assignment1\\_NeuralNetwork.git](https://github.com/mhardik003/NLP_Assignment1_NeuralNetwork.git)