CS7.401: Introduction to NLP | Assignment 1

Instructor: Manish Shrivastava

Submitted By : Somya Lalwani (2020201092)

REPORT

1. EuroParl Dataset:
☐ Train dataset avg perplexity for Kneyser-Ney smoothing : 7.856176819942633
☐ Test dataset avg perplexity for Kneyser-Ney smoothing : 9.316282743303592
☐ Train dataset avg perplexity for Witten-Bell smoothing : 11068.231272728524
☐ Test dataset avg perplexity for Witten-Bell smoothing : 15100.453487091025

2. On Medical Abstracts corpus:
☐ Train dataset avg perplexity for Kneyser-Ney smoothing : 6.493184779125741
☐ Test dataset avg perplexity for Kneyser-Ney smoothing : 6.419396576421301
☐ Train dataset avg perplexity for Witten-Bell smoothing : 13390.563651011613
☐ Test dataset avg perplexity for Witten-Bell smoothing : 12954.52914388947

CONCLUSION : The experiment showed that **Kneser-Ney** consistently outperforms Witten-Bell smoothing technique in terms of perplexity values.