**Project Proposal**

COMP 7118

Data Mining

Spring 2020

Suraj Sharma

[sshrmspk@memphis.edu](mailto:sshrmspk@memphis.edu)

**What is the problem/application?**

* Word Suggestion

**What methods will be tested or implemented?**

* Dynamically suggest the next word to the user while typing

**What data sets will be used?**

* English Corpora (<https://www.corpusdata.org/>) <https://www.corpusdata.org/formats.asp>

**What are potential challenges for implementation?**

* Building an application to learn the English language from the large collection of data
* Personalized writing preferences
* Real time response (next word) analysis (engineering)
* Resources (memory/space) to process the large data.

**What are the expected deliverables?**

* Datasets used
* References on the similar work/research done on the area.
* Algorithm to learn and predict to next word.
* Gadget that will predict/recommend the next word based on the previous n-words while typing.
* Analyze if any potential future works could be there.