Data Intake Report

Name: Flask Deployment

Report date: Aug 28, 2023

Internship Batch: LISUM24

Version:1.0

Data intake by: Siobhan Hwang

Data intake reviewer:

Data storage location: [GitHub](https://github.com/shivurs/DataGlacierInternship/tree/main/Flask%20Deployment)

**Text data details:**

|  |  |
| --- | --- |
| **Total number of observations** | N/A |
| **Total number of files** | 6 |
| **Total number of features** | N/A |
| **Base format of the file** | .txt |
| **Size of the data** | 809KB |

**Shakespeare play scripts.**

**Proposed Approach:**

* Preprocess texts
* Concatenate processed texts
* Probabilistic language model predicts next words (saved to pickle file)
* Generate number of texts specified by the user (on web via flask)
  + Generated texts are of varying lengths