Data Intake Report

Name: NLP: Document Classification
Report date: 15.11.2022.
Internship Batch: LISUM14
Version: 1.0
Data intake by: Marija Babić
Data intake reviewer:
Data storage location: http://qwone.com/~jason/20Newsgroups/

Tabular data details:

Total number of observations	18828
Total number of files	18828
Total number of features	1
Base format of the file	txt
Size of the data	32.3 MB

Proposed Approach:
- There are no duplicates or missing values in the data