

Data Intake Report

Name: [NLP: Resume Extraction](#)

Report date: 03/21/2023

Internship Batch: LISUM30

Version: [1.0](#)

Data intake by: Sanjana Naidu Gedela

Data intake reviewer:

Data storage location: <https://gist.github.com/Rahulrky/b57ad459545c896231c4b770bd8d22ef>

Tabular data details:

| | |
|-------------------------------------|--------|
| Total number of observations | 200 |
| Total number of files | 1 |
| Total number of features | 2 |
| Base format of the file | .txt |
| Size of the data | 1.1 MB |
| | |

Problem Description: Resumes often contain excessive information, making it challenging for HR professionals to manually shortlist candidates efficiently.

Business Understanding: Implementing a NER system will streamline the process by automatically identifying and classifying relevant entities in resumes.

Dataset Attributes: The dataset includes 'content' as resume text and 'label' as annotated tagged entities, facilitating the training of NER models.