Datasheets for Dataset 6: Web scraping from Indeed.com

Motivation for Dataset Creation

Why was the dataset created?
Dataset 6 is scrapped from Indeed.com in order to get the job opening number for engineering related job in states cross USA.

Has the dataset been used for any tasks already?
Yes, for Duke University AIPI 510 Sourcing Data for Analytics class project.

What (other) tasks could the dataset be used for?
Get job opening number for each state for job titles related with "Engineer".

Dataset Composition

What are the instances?
Job opening for each state for each job title.

Are relationships between instances made explicit in the data?
No.

How many instances of each type are there?
55 instances in total.

What data does each instance consist of?
Job opening for each state for each job title.

Is everything included or does the data rely on external resources?
Yes, rely on Indeed.com.

Are there recommended data splits or evaluation measures?
No.

Data Collection Process

How was the data collected?
Scrap the total job opening number with the search keyword "Engineer" and filtered by state name on Indeed.com. For example, to get the current job opening number for North Carolina, "Engineer" and "NC" are entered. The count of results is scrapped as job opening.

Who was involved in the data collection process?
Students. Group 8 of AIPI 510 course.

Over what time-frame was the data collected?
Current data scrapped at November 13, 2021, 9:26:09 AM (EST).

Dataset Preprocessing

What preprocessing/cleaning was done?
No preprocessing/cleaning was done.

Does this dataset collection/processing procedure achieve the motivation for creating the dataset stated in the first section of this datasheet?
Yes.

Dataset Distribution

How is the dataset distributed?
Not distributed. Only for internal analysis.

When will the dataset be released/first distributed?
Will not be distributed.

Are there any copyrights on the data? Are there any fees or access/export restrictions?
No.

Dataset Maintenance

Who is supporting/hosting/maintaining the dataset?
Students. Group 8 of AIPI 510 course.

Will the dataset be updated?
No.

Legal & Ethical Considerations

If the dataset relates to people or was generated by people, were they informed about the data collection?
It is not related with people.

Does the dataset comply with the EU General Data Protection Regulation (GDPR)?
Yes.

Does the dataset contain information that might be considered sensitive or confidential?
No.

Does the dataset contain information that might be considered inappropriate or offensive?
No.