

NLP case for recruitment

Task: Exploring Covid-19 and non-Covid vaccine reports VAERS 2020

For the second interview, we would like to ask you to analyse a dataset containing reports of suspected side-effects with free-text information.

During investigations of safety concerns with medicinal products, pharmacovigilance experts will evaluate the causality between the medicinal product and the observed side effect. For this, they need as much information as possible to understand each report and what happened to that patient in as much detail as possible. They do not usually have the possibility to reach the patient and ask follow up questions. They therefore rely not only on structured information but also on the narrative descriptions of the side effect if available. These can include information about the course of events, the symptoms, the severity and quality of life of the patient but also information about concomitant factors such as other drugs or other differential diagnosis. Reports of suspected side effects can be sent by different reporters, for example the patient themselves, health care professionals or pharmaceutical companies.

For this assignment, we would like you to look at reports from **VAERS in 2020**, the US database for adverse events reported for vaccines. You can download it here: <https://vaers.hhs.gov/data/datasets.html>. Use the CSV files VAERS DATA and VAERS VACCINE (both files contain a unique ID VAERS_ID for linking the tables). The narrative field is in the VAERS DATA file in column SYMPTOM_TEXT but there exist other free-text fields on the reports. The vaccine information can be found in the VAERS VACCINE/VAX file the column VAX_NAME contains the vaccine name for the report and you can identify the COVID19 reports using this column. Since the data is from the US in 2020, it includes reports for Covid-19 vaccines and other vaccines. The question that we would like you to answer is: Are COVID-19 vaccine narratives different from narratives for other vaccines?

Some suggestions for questions that you may want to answer during your exploration are:

- Are more or less narratives written in the first person?
- Is there more redundancy in covid-19 vaccine narratives?