

DesinfoBotWatch

Spark didn't work with java 25, if you have issues with it, we recommend switching to java 17

This project is about analysing data from Twitter bots, We aimed at being able to uncover useful insights on this data using Spark.

Technologies used

- **Python:** It's easy to use, but most importantly, it's the standard in data analysis because it's overall the best language for this kind of task.
- **uv:** An extremely fast Python package and project manager, written in Rust, it's slowly becoming the new standard.
- **Spark:** For this kind of data analysis, we don't really need realtime streaming, which is why spark, a technology that's using batch processing, is perfectly fit for the job. We don't need either to go very low level, so there's really no need to go with MapReduce, it would be a lot more work for results that don't really differ.

Getting started

Download dataset

```
cd data
```

from there, uncomment the files you want to download in `dl_data.sh`, and execute the script.

Start analysis

We've decided to use `uv` for this project. It can be installed using `brew intall uv`.

In the root directory, run `uv run main.py` to automatically install dependencies to a new virtual environment and run the program.

After the analysis is done, a short summary is written to the `outputs` directory