Shootings in Sweden

From the eye of the Global Database of Events, Language and Tone (GDELT)

UPPSALA UNIVERSITET

Members:
Yaser Kaddoura
Mohammad Mustakim Ur Rahman

Supervisor:
Raazesh Sainudiin

Introduction

The media generates a massive amount of data every second. The GDELT project pushes us a step forward towards obtaining the holy grail of consuming the knowledge of the media in its entirety. We present the results of one of our experiments in this poster that shows the media exposure for shootings in Sweden.

GDELT

The GDELT project processes the media across the world extracting events, entities, locations, themes, and more from text-based and TV sources. It covers news in 100 languages while indicating the tone of the text using sentiment analysis. Many academic researchers in social sciences find it an attractive source to conduct all kinds of researches.

Method

We use a naive pipeline to indicate if the event is related to shooting or not. Apache Spark is used to handle the large amount of data presented in GDELT. It comprises the following steps:
0- Connect to the analytics/AI-ready GDELT delta lake house.
1- Fetch relevant articles presented in GDELT.
2- Translate the article to English.
3- Perform NLP techniques to extract terms.
4- Check if terms related to the shooting is present (e.g. shot, gunshot)

Reliability

The approach is naive and prone to errors since it's not sufficient to indicate if the text is related to shooting from the terms it has. For instance, an article about COVID shots or sports could be identified as shooting incidents. To remedy this problem, we use the Goldstein scale supplied by GDELT which gives a score from -10 to 10 identifying the theoretical potential impact of the event on the stability of the country. Only events that have negative scores are put into consideration to reduce the error significantly.
Another problem comes from GDELT itself since it is liable to make errors such as generating multiple entries for the same event or misidentifying an event. For example, an article that talks about COVID shots could be identified as an event related to shooting from GDELT's perspective.

Fig. 1 Shooting events and mentions in the media for 2021

Results

We processed Swedish articles that talk about shootings in Sweden fo 2021. We showcase the number of shooting events across the year accompanied by the number of times they are mentioned in the media [Figure 1]. The geo bubble map indicates the number of shooting events in Swedish cities where the cities with the highest shooting counts are Stockholm, Harjedale, and Goteborg with 568, 425, and 194, respectively [Figure 2].

Other Work

In addition to shooting events in Sweden, we use GDELT to visualize the average Goldstein scale for events across the cities in Sweden while comparing the results with other countries such as Norway. We also identify popular events, themes, and cities mentioned in the media.

For future work, we can increase the accuracy of the pipeline that identifies shooting events by using sophisticated Natural Language Processing techniques, including latest models such as GPT3, on the actual text of the articles scraped after identifying their URLs to be of interest on the basis of our naive preliminary analysis on the GDELT meta-data in the delta lake house.

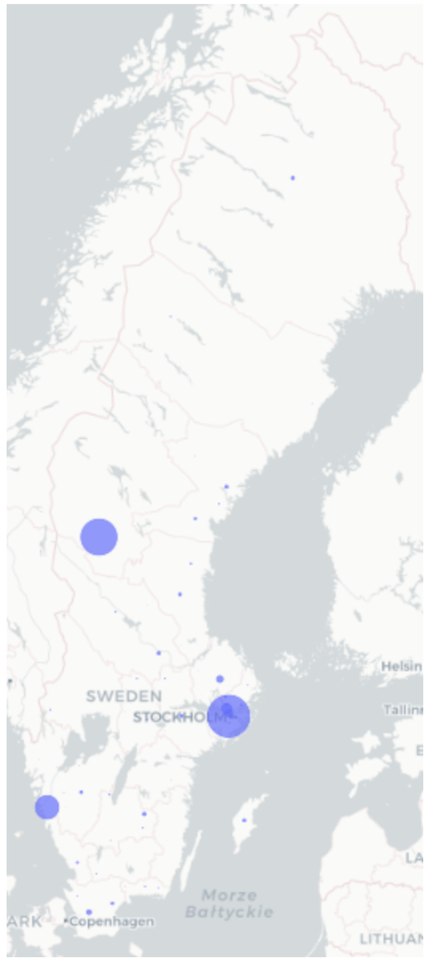

Fig. 2 Shootings in Sweden for 2021

Acknowledgment

This project was supported by *Combient Mix AB* through *Data Science Project Fellowships* between 2021-10 and 2021-12 to Mohammad Mustakim Ur Rahman and Yaser Kaddoura at *Combient Competence Centre for Data Engineering Sciences*, Department of Mathematics, Uppsala University, Uppsala, Sweden, and *databricks University Alliance* with infrastructure credits from *AWS*.