

Bankrupt in covid times.

Gerjon Schuurman, S4836022

Abstract

During the covid pandemic lock-downs, many companies got decreased income, and we think that might be visible in the amount of tweets with the word "failliet" in it. We take two months, April 2019 and April 2020, and counted the amount of tweets containing "failliet", and calculated the percentile usage compared to the full amount of tweets. We concluded a remarkable increase in the percentages, which is very likely to be the cause of the covid pandemic.

1 Introduction

The covid pandemic has put the whole world upside down. Governments decided their whole countries should go into lockdown. Likewise the Netherlands, where in March, the streets turned empty. Shops and schools closed, jobs were done from home, and events were canceled. Due to this lockdown, many companies lost their customers, and thus their income. We are curious to see whether this is visible in social media, by counting the occurrences of the word "failliet" in twitter data, during and before the pandemic. We are expecting to see a increase of usage of the word "failliet" in twitter data during the covid pandemic, compared to before the pandemic.

2 Related Work

The usage of twitter has become different in the Netherlands, as well as in other countries (Dubey, 2020). Users show not only their positive views, but also their negative thoughts. That alone could be a reason for twitter users to tweet more about "failliet" than times before the lockdown.

3 Data

For our research, we have made use of the Karora Twitter2 corpus. We have decided to take a month, April 2020, of twitter data, right after the lock-down kicked in, and compare that to the same month in 2019. This way we make sure not to get invalid results caused for any reason by the time of the year. We downloaded the data using the scp commands given in the file `download_data.sh`, which is included in the GitHub repository¹. In order to replicate the research, man might have to request permission to enter the Karora secure shell. The data gets saved into two different folders, therefor it is easy to keep April 2019 and April 2020 separated. The data consists of gzipped files, each containing an hour of tweets, each tweet on it's own line in .json format. With the data collected, we then iterated through the tweets, both for the 2019 and the 2020 folders, and we counted the tweets with the occurrences of "failliet" using the `zgrep` and `wc` commands per month, and we counted the total amount of tweets per month using the `zcat` and `wc` commands. That way, we can later combine the results into percentages, and get clearer results to analyze.

Table 1 provides a summary of the directories the data used in this study is stored in once downloaded.

resmet_countfailliet
2019/04/*.gz
2020/02/*.gz

Table 1: Tree structure of the stored dataset once downloaded with the `download_data.sh` file.

Pre-processing The data is hardly pre-processed. We only put the twitter data in

¹<https://github.com/GSrman/failliet>

strategically named directories, but for the research, we use the raw copied .gz files. The actual unzipping happens in the same command as the counting of "failliet".

4 Analysis and Results

Once we have downloaded and prepared the data, we have run the in the GitHub repository included count.sh file, which counts the amount of tweets and the amount of tweets with "failliet" in them. Table 2 shows the results given in the file's output, and the calculated percentages, rounded to three decimals.

Results Table 2 summarizes the results, and shows an increase of frequency in the percentile usage of the word "failliet" in twitter data in the month April of the year 2020 compared to the same month in the year 2019.

month	"failliet"	total	%
04'19	14793	14184602	0.104%
04'20	34906	18264471	0.191%

Table 2: The count of tweets with "failliet", the total tweet count, and the percentage of tweets with "failliet" per month.

5 Conclusion

This study aimed at showing if there was a visible increase in the usage of the word "failliet" in twitter data during the month April in the year 2020, compared to the same month in 2019. The results show an increase from 0.104% to 0.191%, which is quite remarkable. We think it is save to include this is for the majority the result of the covid pandemic lockdown, as also significant more companies went bankrupt in April 2020 (Centraal Bureau voor de Statistiek, 2020). It has to be said that possibly not all occurrences of the word "failliet" are related to the bankruptcy of companies. We hope this research support future research on the correlation of twitter usage on current events.

In the future, it might be important to see a full view on the usage of the word "failliet" on twitter. Rather than only two months, a graph that shows the usage on every month for multiple years, including the covid pandemic years.

References

- Centraal Bureau voor de Statistiek, C. (2020, May). Aantal faillissementen gestegen in april.
- Dubey, A. D. (2020). Twitter sentiment analysis during covid-19 outbreak. *Available at SSRN 3572023*.