Bankruptcy 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 country 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. Unknown is the impact of this on social media and therefor we are investigating how on social media platform Twitter the occurrence count of the word "failliet" (bankrupt), which is the dependent variable, has changed on a month time span during the covid pandemic compared to the same month prior to the pandemic, which is the independent variable. We compare data produced during covid times with data produced before covid times. The outcome we expect to see, is an increase of usage of the word "failliet" in twitter data during the covid pandemic, compared to the data prior to the covid pandemic.

2 Related Work

According to study by Chakraborty et al. (2020), the sentiment on USA twitter has changed during covid times. People showed confused behavior and portrayed emotions likewise. Dubey (2020) shows besides trust (19.42%), a large amount of USA tweets shows fear (17.23%)(Dubey, 2020). For Dutch tweets, 23.04% of tweets build on fear(Dubey, 2020). Both these studies show remarkable sentiment in social media usage during and caused by the covid pandemic. However, both do not investigate what subjects those differences consist of. We do not have time nor ability to construct a list of all those subjects, therefor we have decided to investigate one subject we find likely to have had noteworthy influence: bankruptcy.

3 Data

For our research, we used of the Karora Twitter2 corpus[1]. We decided to take a month of twitter data: April 2020[2], right after the lockdown kicked in, and compare that to the same month in 2019[3]. 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 repository[4]. In order to replicate the research, man might have to request permission to enter the Karora secure shell. Using the scripts provided, 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.

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

[1] ssh s1234567@karora.let.rug.nl/net/corpora/twitter2/Tweets
[2] s1234567@karora.let.rug.nl:/net/corpora/twitter2/Tweets/2020/04/
[3] s1234567@karora.let.rug.nl:/net/corpora/twitter2/Tweets/2019/04/
[4] https://github.com/GSrman/failliet

April 2019	April 2020
2019/04/*.gz	2020/04/*.gz

Table 1: Structure of the stored data set once downloaded with the download_data.sh file. Each month contains 720 .gz files, one for every hour in said month.

Pre-processing The data undergoes barely any pre-possessing. We put the twitter data in .gz files in strategically named directories, but for the research, we use those raw copied .gz files.

4 Analysis and Results

For counting the occurrences of "failliet" per month, we iterate through the different .gz files. For each file, we count the total amount of tweets of said month using bash: the zcat and the wc commands. For counting the tweets that contain the word "failliet", we used the zgrep and the wc commands. With both the total count and the "failliet" count. This can be done using the count.sh file, which is included in the GitHub repository mentioned previously. This file gives 4 outputs: the total count of tweets per month and the occurrence count of "failliet", for April 2019 and April 2020. We can calculate the percentage per month of tweets that contain the word "failliet", for analyzing purposes.

Results The output of count.sh as well as the calculated percentages are included in Table 2, rounded to three decimals. 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 calculated percentage of tweets with "failliet", all per month.

5 Conclusion

The research aimed at investigating the impact of the covid lockdown on social media usage. We have counted the occurrences of the word "failliet" in twitter data and compared it to data prior to the covid lockdown. The results show an increase, as anticipated. The occurrence count grows from 0.104% to 0.191%, for April 2019 and April 2020, respectively. We think this increase is quite remarkable. According to the Centraal Bureau voor de Statestiek (2020), a significant larger amount of companies went bankrupt during April 2020, and its reflection is shown in social media usage. However, it has to be said, not all occurrences of the word "failliet" may correspond to the bankruptcy of companies.

Future research might be focused on a full view of the occurrence count of the word "failliet" during the entire covid lockdown, to find out how it correlates with the amount of bankrupt companies and possibly with the covid support money for companies from the dutch government. We foresee similar research on correlation between twitter usage and other events currently happening, for which we hope our research may be a pedestal.

References

Centraal Bureau voor de Statestiek (2020, May). Aantal faillissementen gestegen in april.

Chakraborty, K., S. Bhatia, S. Bhattacharyya, J. Platos, R. Bag, and A. E. Hassanien (2020). Sentiment analysis of covid-19 tweets by deep learning classifiers—a study to show how popularity is affecting accuracy in social media. *Applied Soft Computing 97*, 106754.

Dubey, A. D. (2020). Twitter sentiment analysis during covid-19 outbreak. *Available at SSRN 3572023*.