SOCIAL MEDIA USER EMOTIONS DURING COVID19

Jan Strohschein, Ana María Lara-Palma, Heide Faeskorn-Woyke

TH Köln (Germany), Universidad de Burgos (Spain), TH Köln (Germany)

jan.strohschein@th-koeln.de; amlara@ubu.es; heide.faeskorn-woyke@th-koeln.de

ABSTRACT

Social networks are everywhere and a large part of users even frequents more than one platform (Pew Research Center, 2018). "Due to a constant presence in the lives of their users, social networks have a decidedly strong social impact" (Statista, 2019). However, several studies also suggest that social media usage is not beneficial for users health with symptoms ranging from sleep deprivation to anxiety and depression (Hogue & Mills, 2019; Hunt et al., 2018; Levenson et al., 2016).

This work uses a machine learning approach to study the emotions of a large group of social media users on Twitter during the Covid19 pandemic and compares the results to our previous research that evaluated 10 million tweets from 5000 users between 2015 - 2019.

It is possible to extract emotions of social media users from the text of their status updates as shown by Colneric and Demsar, and Tasoulis et al. (Colneric & Demsar, 2018; Tasoulis et al., 2018). This analysis is based on the work of Colneric and Demsar, who were kind enough to publish the resulting machine learning model. They utilized neural networks to generate a model that is able to detect emotions in English language. Neural networks are a supervised machine learning method and therefore the data needs annotations for the algorithm to learn from. As the authors learned on a massive dataset of 73 billion tweets it was infeasible to manually annotate the dataset. The authors exploited hashtags as annotations, an approach that was successfully used in several other natural language processing studies for sentiment classification (Go et al., 2009; Kouloumpis et al., 2011; Nodarakis et al., 2016), detecting sarcasm (Bamman & Smith, 2015; González-Ibáñez et al., 2011), studying personality traits (Plank & Hovy, 2015) and classifying emotions (Mohammad & Kiritchenko, 2015). As hashtags are selected by the author of a tweet they work well as indicators of their emotions.

Emotions can be modelled in a multitude of ways and popular emotion classification schemes were created by Paul Ekman, Robert Plutchik and Douglas McNair along with Maurice Lorr and Leo Droppleman (Ekman, 1999; McNair et al., 1971; Plutchik, 1982). The classification for this analysis is done with Ekmans scheme of basic emotions as it covers fear, disgust and anger, which have been previously identified as the most impactful emotions caused by the use of social media and should be investigated further.

In our previous work users have been grouped based on the number of status updates they publish and the amount of followers they have. Grouping users was more effective and showed more distinct results when based on the number of followers a user has. The prevalent expressed emotions on twitter from 2015-2019 were joy and surprise. Over the observed period of time, from 2015 to 2019, the values for joy remained consistent, while an increase in anger, disgust and fear could be verified for all user groups. It was noticeable that twitter users with the least amount of followers (<25%) expressed anger and fear most strongly. Even though it was expected that the positive emotions declined and the negative emotions increased during Covid19 the exact opposite happened. A detailed monthly analysis of the data suggests that the United States election had a big influence on the results.

INTRODUCTION

Most social media users frequent more than one platform (Pew Research Center, 2018). "Due to a constant presence in the lives of their users, social networks have a decidedly strong social impact" (Statista, 2019). "The blurring between offline and virtual life as well as the concept of digital identity and online social interactions are some of the aspects that have emerged in recent discussions. Approximately 2 billion, mostly young, internet users are using social networks and these figures are still expected to grow as mobile social network usage increasingly gains traction" (Statista, 2017, Statista, 2016, Pew Research Center, 2019). This paper builds upon earlier work that analyzed Twitter status updates from 2015-2019 (Strohschein et al., 2019). It investigates the development of emotions for different user groups on Twitter during the Covid19 pandemic utilizing machine learning and natural language processing techniques to create an automated approach.

Social Media Platforms

The usability of social media platforms has been tested in multiple fields being beneficial in a large number of indicators, such as learning (Hortigüela-Alcalá et al., 2019), inclusion, or socialization. However, unfortunately, for a percentage of the world's population, the technological mirror in which people look returns the image of sadness, fear, worry, and hopelessness. Many studies suggest that social media usage is not beneficial for user´s health with symptoms ranging from sleep deprivation to anxiety and depression (Levenson et al., 2016; Hunt et al., 2018; Hogue & Mills, 2019). Regarding these emotional consequences, Yoon, Kleinman, Mertz, and Brannick (Yoon et al., 2019) in their meta-analysis study on the correlation between social networks and symptoms of depression, highlight “Our results are consistent with the notion of ‘Facebook depression phenomenon’ and with the theoretical importance of social comparisons as an explanation”. But the effects of sleep deprivation and depression also persist during the workday and companies fear for their organizational productivity. Several groups of researchers studied the effects of social media on the productivity of students and workers alike. Brooks conducted research on students and supposes that being in the classroom can be analogous to being in a work environment as the students have to efficiently perform different tasks. He found that inefficiencies result from time spent on the interruption but also the time necessary to fully concentrate on the task again (Brooks, 2015). Lau as well as Flanigan and Babchuk also concluded that social media usage decreases motivation and hinders academic performance (Lau, 2017; Flanigan & Babchuk, 2015). Ali-Hassan, Nevo, and Wade studied the effects of social media in the workplace and found that social use of technology can have an indirect positive effect on job performance by building networks in the workplace and sharing knowledge but also discovered a direct negative impact on task routine performance when workers spent their time in social networks instead (Ali-Hassan et al., 2015). While research shows no clear results whether or not social media platforms hinder work performance, some companies don’t want to risk a loss in productivity and try to ban social media from the workplace (Gaudin, 2009).

Users and User groups

Social media users and digital natives have been subject to a lot of studies. Oblinger & Oblinger (Oblinger & Oblinger, 2005) characterized them as active experiential learners, proficient in multitasking, dependent on communication technology to access information and interacting with others. Kennedy, Judd, Delgarno and Waycott define digital natives based on several parameters, e.g., the number of devices they regularly use, their formal education, gender and age, to create the following user groups: power user, ordinary user, irregular user and basic user (Kennedy et al., 2010).

Unfortunately, those parameters are not easy to obtain in an automated approach. Twitter profiles do not contain any information about the formal education of a user. Specifying a location or a birth date is completely optional and even if a location is specified, it does not have to be in a standardized format so “Berlin” could mean the German capital or one of several cities with this name in the USA or Australia. Therefore, for this analysis users are grouped solely based on their number of followers, who want to read new status updates because this approach has been effective in the previous work.

Emotion Modeling

Emotions can be modeled in a multitude of ways and popular emotion classification schemes were created by Paul Ekman, Robert Plutchik and Douglas McNair along with Maurice Lorr and Leo Droppleman (Ekman, 1999; Plutchik, 1980; McNair et al., 1971). The classification for this analysis is done with Ekmans scheme of basic emotions as it covers fear, disgust and anger, currently the most researched emotions in association with social media networks according to the comprehensive literature review of our previous work. The six basic emotions are explained by Ekman and Cordaro (Ekman, Cordaro, 2011) in a later article as follows:

Anger: the response to interference with our pursuit of a goal we care about. Anger can also be triggered by someone attempting to harm us (physically or psychologically) or someone we care about. In addition to removing the obstacle or stopping the harm, anger often involves the wish to hurt the target.

Fear: the response to the threat of harm, physical or psychological. Fear activates impulses to freeze or flee. Often fear triggers anger.

Surprise: the response to a sudden unexpected event. It is the briefest emotion.

Sadness: the response to the loss of an object or person to which you are very attached. The prototypical experience is the death of a loved child, parent, or spouse. In sadness there is resignation, but in can turn into anguish in which there is agitation and protest over the loss and then return to sadness again.

Disgust: repulsion by the sight, smell, or taste of something; disgust may also be provoked by people whose actions are revolting or by ideas that are offensive.

Joy: feelings that are enjoyed, that are sought by the person. There are a number of quite different enjoyable emotions, each triggered by a different event, involving a different signal and likely behavior. The evidence is not as strong for all of these as it is for the emotions listed above.

Emotion Classification

It is possible to extract emotions of social media users from the text of their status updates as shown by Tasoulis et al. and Colneric and Demsar (Tasoulis et al., 2018; Colneric & Demsar, 2018). This analysis is based on the work of Colneric and Demsar, who utilized deep learning of neural networks to generate a model that is able to detect emotions in english language. Neural networks are a supervised machine learning method and therefore the data needs annotations for the algorithm to learn from. As the authors learned on a massive dataset of 73 billion tweets it was infeasible to manually annotate the dataset. The authors exploited hashtags as annotations, an approach that was successfully used in several other natural language processing studies for sentiment classification (Go, et al., 2009; Nodarakis, et al., 2016; Kouloumpis, et al., 2011), detecting sarcasm (Gonzalez-Ibanez, et al., 2011; Bamman & Smith, 2015), studying personality traits (Plank & Hovy, 2015) and classifying emotions

(Mohammad & Kiritchenko, 2015). As hashtags are selected by the author of a tweet they work well as indicators of their emotions. The machine learning algorithm analyses the given text and tries to derive the hashtag as target variable.

Hypothesis

The following hypotheses are used to investigate the development of user emotions on the Twitter social media platform during the COVID19 pandemic in comparison to the previous years:

H1: Joy will decline for users.

H2: Anger will increase for users.

H3: Disgust will increase for users.

H4: Fear will increase for users.

H5: Sadness will increase for users.

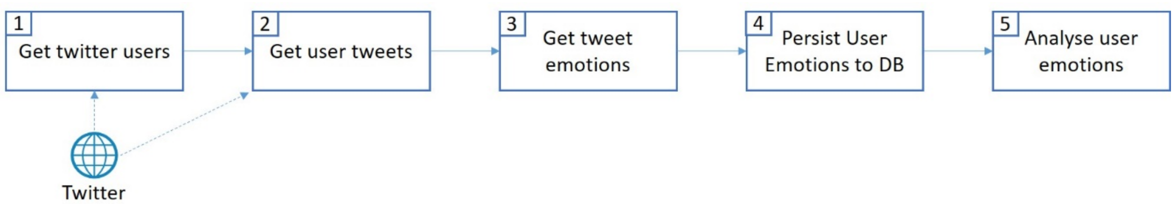
H6: Surprise will increase for users.

H7: It is possible to differentiate emotions between user groups, based on their number of followers.

METHODOLOGY

This paper uses an automated approach to study the emotions of a larger group of social media users over time. Colneric and Demsar published their resulting machine learning model on GitHub (Colneric & Demsar, 2018). This machine learning model is implemented in a real-time architecture to collect and analyze tweets and create a classification for the emotions expressed in the text. The CAAI architecture, developed by our workgroup, consists of virtualized processing, storage or analysis building blocks that communicate via messaging to create a highly modular data analytics pipeline, depicted in Figure 1.

Figure 1. Data collection and analysis pipeline.



In the first step trending Twitter topics are observed and active users, who write status updates on these topics, extracted. For every user their following users were extracted to increase the user sample-size. In a second step for every user the tweets were collected through the Twitter API. So called “retweets”, where a user quotes the status update of someone else, were not collected to analyse only tweets that the user wrote himself. The API imposes limits on the amount of requests per time interval and the number of status updates that can be downloaded for any given user. Therefore, only the last ~3.200 status updates can be retrieved for each user. Surprisingly this turned out to be an important constraint, because for very active users this was just a fraction of their status update history.

The pre-trained model from Colneric and Demsar was used in the third step to analyze the status updates regarding expressed emotions. This step was run in parallel as the process was very time-consuming and the results were reconciled and persisted in a database (step 4) for further analysis (step 5). For the analysis the emotions of all users were evaluated over time.

RESULTS

For each user the last ~3200 tweets have been analysed, if the user wrote that many status updates. Over the course of several weeks roughly 11 million tweets from ~6000 users have been collected and analysed for the whole of 2020 and January to March of 2021.

Data Overview

The analysis is based on two related datasets, the users and their tweets with the associated emotional classification, that have been collected in separate steps but are joined to increase the available information. The two tables below describe the datasets and the available fields.

Table 1. User Features and descriptions.

User Feature	Description
user_id (integer)	Automatically generated identification number for every user.
user_name (string)	The user can choose the nickname to display.
user_location (string)	The user can specify his/her location.
account_created_at (timestamp)	The system records the day and time of account creation.
statuses_count (integer)	The amount of status updates a user has written, including retweets.
favorites_count (integer)	The number of Tweets this user has liked.
followers_count (integer)	The number of followers this account currently has.
friends_count (integer)	The number of users this account is following.
verified (boolean)	If the user´s identity has been verified by twitter the value is “True”, otherwise “False”.

Table 2. Tweet Analysis Features and Descriptions.

Tweet Analysis Feature	Description
status_id (integer)	Automatically generated identification number for every tweet.
user_id (integer)	Automatically generated identification number for every user.
status_created_at (timestamp)	The system records the day and time of tweet creation.
text (string)	The tweet text written by the user and analyzed for emotions.
retweet_count (integer)	Number of times this status update has been “retweeted”.
anger, disgust, fear, joy, sadness, surprise (float)	The calculated percentage value for a particular emotion in a users tweet.

An example for such a classification is shown below in Table 3, only the columns relevant for the classification are shown. The status update regards a sport event and the detected prevalent emotion is joy, followed by surprise.

Table 3. Tweet Analysis multi-class Classification Example.

Text	Anger	Disgust	Fear	Joy	Sadness	Surprise
'Anthony Davis makes his debut with the Hornets dropping 21 points and grabbing 7 rebounds.'	0.0109	0.0028	0.0425	0.7640	0.0551	0.1244

Distribution of tweets in the dataset per year

The previous analysis consisted mostly of status updates for 2018 and 2019 with few status updates for the earlier years. A possible reason is, that for very active users analysing 3200 tweets is just not enough. Another possibility may be that just active users are analysed and users from the earlier years stopped using the platform. The current evaluation contains several million status updates for the years 2020 and 2021 and should give a credible insight into user emotions during Covid19.

Table 4. Distribution of tweets in the dataset by year.

First Evaluation					Current Evaluation	
2015	2016	2017	2018	2019	2020	2021
138.984	134.807	273.948	1.512.112	8.128.685	6.482.840	4.627.022

Analysing the average emotions of all users over time

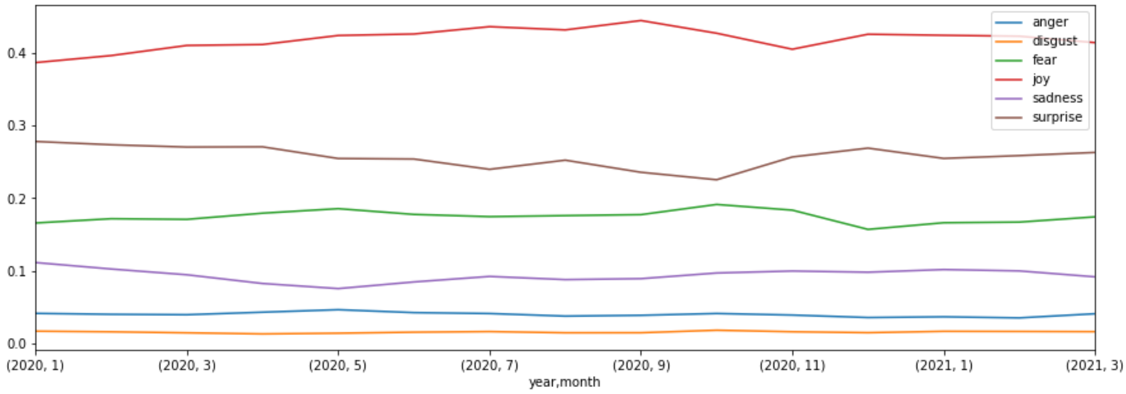
The first aggregation contains all user tweets and the associated emotions for each year. Table 5 shows the mean values for each emotion with joy and surprise as the dominant emotions, joy even more so for the years 2020 and 2021. It is noticeable that sadness is declining while anger and fear are rising until 2019 and then also declining.

Table 5. Average Emotions of all users per year.

Year	Anger (%)	Disgust (%)	Fear (%)	Joy (%)	Sadness (%)	Surprise (%)
2015	3,7	1,9	14,7	32,1	14,6	33,0
2016	4,5	1,9	16,4	32,8	12,6	31,8
2017	5,1	1,6	16,5	34,9	11,7	30,2
2018	5,4	1,8	17,6	35,1	11,4	28,7
2019	5,8	2,0	18,0	32,9	11,8	29,6
2020	4,0	1,6	17,6	42,2	9,3	25,3
2021	3,7	1,7	16,9	42,0	9,8	25,8

Figure 2 shows the user emotions for each analyzed month. It can be seen that fear rises when the first cases of Covid19 turn into a pandemic around March 2020 and before the United States presidential election in November 2020. Notably, expressed joy is rising throughout 2020 and just drops when the presidential election concludes, while surprise increases afterwards.

Figure 2. User emotions grouped by month.



Analyzing the emotions of user groups based on the user´s followers

The distribution of twitter users below is based on the number of followers a user has. The mean and median differ, which stems from outliers with an extreme amount of followers. The following analysis categorizes users based on the quartiles of this distribution. The amount of followers an average user has increased between the first analysis and the current analysis for all quartiles.

Table 6. Distribution based on the amount of followers.

Year	Mean	Std	Min	25%	50%	75%	Max
2015-2019	3.466	40.188	0	67	386	1.407	1.653.827
2020-2021	5.069	38.146	0	224	765	2.255	1.754.784

The distribution of emotions with user groups based on the amount of followers the users have, shows clear distinctions between the groups for all emotions in 2015-2019 and for all groups but anger for 2020-2021, e.g., expressed fear and joy continually decreases from the users with the least amount of followers to the users with the most followers. The emotions disgust, sadness and surprise are expressed more strongly the more followers a user has, with the highest values for the user group with >75% of followers.

Table 7. Users with less than 25% of followers.

2015-2019 / 2020-2021	Anger (%)		Disgust (%)		Fear (%)		Joy (%)		Sadness (%)		Surprise (%)	
Mean	7,5	3,8	1,0	1,0	20,5	18,0	36,2	48,0	7,7	4,3	27,0	24,9
Std	13,4	6,5	2,4	3,6	17,9	19,6	24,4	31,1	11,6	9,3	21,7	27,3
Min	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
25%	1,3	0,2	0,2	0,0	7,0	4,0	16,4	19,9	1,6	0,1	11,3	1,9
50%	2,9	1,7	0,5	0,1	15,4	11,2	32,7	45,4	4,1	0,9	21,0	15,5
75%	7,1	4,9	1,1	0,5	29,1	24,8	51,6	75,0	9,1	3,7	37,6	38,9
Max	100	100	96,9	98,5	100	100	100	100	100	100	100	100

Table 8. Users with followers >25% and <50%

2015-2019 / 2020-2021	Anger (%)		Disgust (%)		Fear (%)		Joy (%)		Sadness (%)		Surprise (%)	
Mean	5,8	3,9	1,9	1,3	17,7	17,8	33,9	45,4	11,5	6,5	29,0	25,1
Std	11,0	7,0	4,2	4,2	18,2	19,9	27,8	31,8	16,3	12,6	24,2	26,8
Min	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
25%	0,8	0,3	0,2	0,0	4,6	3,6	11,1	16,1	1,8	0,2	9,2	2,4
50%	2,4	1,7	0,7	0,1	11,8	10,8	26,1	41,2	6,0	1,3	23,3	16,3
75%	5,5	4,6	1,9	0,9	24,4	24,5	50,6	72,7	14,1	7,1	44,4	39,4
Max	100	100	99,5	98,4	100	100	100	100	100	100	100	100

Table 9. Users with followers >50% and <75%.

2015-2019 / 2020-2021	Anger (%)		Disgust (%)		Fear (%)		Joy (%)		Sadness (%)		Surprise (%)	
Mean	5,4	4,0	2,1	1,7	17,2	17,3	33,0	40,6	12,4	10,5	29,8	25,9
Std	10,3	8,0	4,6	4,8	18,3	20,2	28,2	32,2	17,1	16,0	24,8	25,7
Min	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
25%	0,7	0,3	0,2	0,0	4,4	3,2	10,1	11,6	1,9	0,5	9,1	3,7
50%	2,2	1,6	0,8	0,4	11,0	10,0	24,2	32,5	6,6	4,1	24,6	18,5
75%	5,3	4,2	2,1	1,5	23,4	23,6	49,2	66,9	15,1	13,8	46,3	40,6
Max	100	100	99,5	98,7	100	100	100	100	100	100	100	100

Table 10. Users with followers >75%.

2015-2019 / 2020-2021	Anger (%)		Disgust (%)		Fear (%)		Joy (%)		Sadness (%)		Surprise (%)	
Mean	5,0	3,9	2,3	1,9	16,9	16,9	31,1	40,2	13,7	11,5	31,1	25,7
Std	9,2	8,2	4,6	5,0	17,7	20,2	26,6	32,6	17,3	16,5	24,2	25,1
Min	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
25%	0,8	0,3	0,3	0,1	4,6	2,9	9,8	11,2	2,8	0,9	11,5	3,9
50%	2,3	1,4	0,9	0,5	11,0	9,4	22,8	30,9	8,0	5,3	26,4	18,8
75%	5,0	3,8	2,4	1,7	22,8	22,8	46,0	67,2	16,9	14,9	47,2	40,5
Max	100,0	100	99,3	99,1	100	100	100	100	100	100	100	100

Average emotions of user groups per year

The following tables show the analysis of all tweets of a user group and the resulting mean values per emotion for a certain year. Joy and surprise are the emotions with the highest mean across all user groups and years. The users with the least amount of followers (<25%) expressed anger, fear and joy most strongly in the years 2015-2019. Joy was communicated even more strongly across all user groups in 2020 and 2021. Contrary to this, the expressed sadness and disgust increases the more followers a user has. It is also noticeable that across all user groups the values for surprise and sadness decrease over the years.

Table 11. Less than 25% of followers.

Year	Anger (%)	Disgust (%)	Fear (%)	Joy (%)	Sadness (%)	Surprise (%)
2015	6,6	0,9	19,6	36,7	8,6	27,6
2016	6,7	0,9	18,6	34,3	7,5	32,0
2017	6,2	1,0	18,6	37,4	8,5	28,3
2018	6,4	1,0	19,4	37,5	7,9	27,9
2019	7,9	1,0	20,9	35,9	7,6	26,7
2020	4,0	0,9	18,4	47,9	4,3	24,4
2021	3,4	1,0	16,8	48,2	4,3	26,4

Table 12. Followers >25% and <50%.

Year	Anger (%)	Disgust (%)	Fear (%)	Joy (%)	Sadness (%)	Surprise (%)
2015	3,6	1,7	14,2	32,6	14,2	33,8
2016	4,9	1,8	17,5	33,4	11,6	30,7
2017	5,1	1,6	16,4	35,2	11,6	30,1
2018	5,3	1,9	17,6	35,2	11,5	28,5
2019	5,9	1,9	17,8	33,7	11,5	29,1
2020	4,0	1,3	18,3	45,1	6,7	24,6
2021	3,6	1,1	16,8	46,1	6,1	26,3

Table 13. Followers >50% and <75%.

Year	Anger (%)	Disgust (%)	Fear (%)	Joy (%)	Sadness (%)	Surprise (%)
2015	3,2	1,8	13,7	32,8	15,4	33,1
2016	5,0	1,8	18,1	34,0	11,5	29,5
2017	4,9	1,7	15,5	35,1	12,2	30,6
2018	5,2	2,0	17,2	34,5	12,1	28,9
2019	5,5	2,1	17,3	32,7	12,4	30,0
2020	4,1	1,7	17,5	39,9	10,8	26,0
2021	3,8	1,7	17,1	41,9	10,0	25,5

Table 14. Followers >75%.

Year	Anger(%)	Disgust (%)	Fear (%)	Joy (%)	Sadness (%)	Surprise (%)
2015	3,5	2,1	14,6	31,5	15,3	32,9
2016	3,9	2,1	15,5	32,2	13,9	32,4
2017	4,4	2,1	15,3	32,7	13,9	31,6
2018	5,1	2,1	16,4	33,6	13,3	29,5
2019	5,5	2,3	17,1	30,6	13,7	31,2
2020	4,0	1,8	17,0	40,2	11,5	25,5
2021	3,8	1,9	16,8	40,2	11,5	25,8

DISCUSSION

The implemented real-time data analysis pipeline collected 11 million tweets from roughly 6000 users over the time span of several weeks. The analysis of this dataset highlights joy and surprise as the most expressed emotions among all users and all years, with joy even increasing in 2020 and 2021. The evaluation of the collected data without user groups shows that expressed sadness and surprise decline each year among all users. It is reasonable to think that the election has a major impact on the outcome, as suggested in the aggregated monthly data.

The users have been categorized for a follow-up analysis based on their interaction with the social media platform and each other to get more detailed insights. The amount of followers a given user has was used as metric, that describes a user's reach on the platform and the interest other users have in her or him. Classifying the users based on the amount of followers showed differences between expressed emotions of user groups and suggests that this criterion is characteristic. The results allow the evaluation of the constructed hypothesis as follows:

- The expressed joy (H1) is strongly increased for 2020 and 2021 in comparison to the earlier timespan from 2015-2019. This is true for the aggregated mean of all users but also for each of the individual user groups. Thus, H1 can be rejected.

- The negative emotions, i.e., anger, disgust, fear and sadness, but also surprise, decline in 2020/2021 despite Covid19. Thus, H2-H6 can be refuted.
- The analysis of user groups based on the amount of followers clearly differentiates the groups, just as in previous work. Despite Covid19 the values for expressed emotions clearly increase / decrease from each group to the next and it is possible to verify H7.

CONCLUSION

The analysis of ~11 million tweets by 6000 users over a time period of a few weeks allows several conclusions. The KOARCH architecture enabled us to easily setup a big data pipeline, consisting of virtualized building blocks, with continuous operation under high load.

For the analysis users have been grouped based on the number of followers they have.

The prevalent expressed emotions on twitter were joy and surprise. Over the observed period of time, from 2015 to 2019, the values for joy remained consistent, while an increase in anger, disgust and fear could be verified for all user groups. Sadness on the other hand declined, maybe it was transformed into anger or fear. It is noticeable that twitter users with the least amount of followers (<25%) expressed anger and fear most strongly. However, for 2020 and 2021 during the Covid19 pandemic and the United States election the joy was massively increased while the negative emotions declined. Aggregating the data for the individual months helped to show differences that could be explained with the election.

There are several limitations to this study. Access through the official twitter interface was limited to the last 3.200 tweets of any given user. While this is enough for the majority of users, for some of the power users this was just a fraction of their status update history. Collecting data several times, as done for this study, and combining the resulting datasets avoids this limitation. Predictions derived from a machine learning model are always just as good as the underlying model. Even though the results of Colneric and Demsar are really impressive, there may be a bias towards a certain emotion in their model. It is also notable, that this model was trained on a massive dataset of English text, but therefore it can be used to classify text written in English only.

It would be interesting to conduct the same analysis for the rest of 2021 and the following years to see if the recovery from Covid19 or the political scenario further influences the expressed emotions.

KEYWORDS: Social Media, Machine Learning, Natural Language Processing, Emotion Classification.

REFERENCES

Bamman, D., & Smith, N. A. (2015). Contextualized sarcasm detection on twitter. *Proceedings of the 9th International Conference on Web and Social Media, ICWSM 2015*, 574-577.

Colneric, N., & Demsar, J. (2018). Emotion Recognition on Twitter: Comparative Study and Training a Unison Model. *IEEE Transactions on Affective Computing*, *3045*(c). https://doi.org/10.1109/TAFFC.2018.2807817

Ekman, P. (1999). Basic emotions. In *Handbook of Cognition and Emotion*. John Wiley & Sons Ltd.

Go, A., Bhayani, R., & Huang, L. (2009). *Twitter Sentiment Classification using Distant Supervision*. 1-6. https://www-cs.stanford.edu/people/alecmgo/papers/TwitterDistantSupervision09.pdf

González-Ibáñez, R., Muresan, S., & Wacholder, N. (2011). *Identifying Sarcasm in Twitter: A Closer Look*. http://www.vidarholen.net/contents/interjections/

Hogue, J. V., & Mills, J. S. (2019). The effects of active social media engagement with peers on body image in young women. *Body Image*, *28*, 1-5. https://doi.org/10.1016/j.bodyim.2018.11.002

Hunt, M. G., Marx, R., Lipson, C., & Young, J. (2018). No more FOMO: Limiting social media decreases loneliness and depression. *Journal of Social and Clinical Psychology*, *37*(10), 751-768. https://doi.org/10.1521/jscp.2018.37.10.751

Kouloumpis, E., Wilson, T., & Moore, J. (2011). Twitter Sentiment Analysis: The Good the Bad and the OMG! *International AAAI Conference on Web and Social Media*.

Levenson, J. C., Shensa, A., Sidani, J. E., Colditz, J. B., & Primack, B. A. (2016). The association between social media use and sleep disturbance among young adults. *Preventive Medicine*, *85*, 36-41. https://doi.org/10.1016/j.ypmed.2016.01.001

McNair, D. M., Lorr, M., & Droppleman, L. F. (1971). EITS Manual for the Profile of Mood States. *Educational and Industrial Testing Service*, *3*(27), 1984.

Mohammad, S. M., & Kiritchenko, S. (2015). Using Hashtags to Capture Fine Emotion Categories from Tweets. *Computational Intelligence*, *31*(2), 301-326. https://doi.org/10.1111/coin.12024

Nodarakis, N., Sioutas, S., Tsakalidis, A., & Tzimas, G. (2016). *Using Hadoop for Large Scale Analysis on Twitter: A Technical Report*. http://arxiv.org/abs/1602.01248

Plank, B., & Hovy, D. (2015). *Personality Traits on Twitter—or—How to Get 1,500 Personality Tests in a Week*. 92-98. https://doi.org/10.18653/V1/W15-2913

Plutchik, R. (1982). A psychoevolutionary theory of emotions. *Social Science Information*, *21*(4), 529-553.

Strohschein, J., Lara Palma, A. M., & Faeskorn-Woyke, H. (2019). Detecting emotions in social media. a technological challenge to enhance youngest behavior. *28th AEDEM International Conference - Management in a Smart Society: Business and Technological Challenges*.

Tasoulis, S. K., Vrahatis, A. G., Georgakopoulos, S. V., & Plagianakos, V. P. (2018). Real Time Sentiment Change Detection of Twitter Data Streams. *2018 IEEE (SMC) International Conference on Innovations in Intelligent Systems and Applications, INISTA 2018*, *Us*. https://doi.org/10.1109/INISTA.2018.8466326