UNIWERSYTET HUMANISTYCZNO-PRZYRODNICZY IM. JANA DŁUGOSZA W CZĘSTOCHOWIE
Edukacyjna Analiza Transakcyjna 2020, nr 9

http://dx.doi.org/10.16926/eat.2020.09.09

Zbigniew WIECZOREK
https://orcid.org/0000-0001-5239-2171
Jan Dlugosz University in Czestochowa
e-mail:z.wieczorek@ujd.edu.pl

Symptoms of anxiety and depression in social media in connection with the threat of COVID-19

How to cite [jak cytować]: Wieczorek, Z. (2020). Symptoms of anxiety and depression in social media in connection with the threat of COVID-19. *Edukacyjna Analiza Transakcyjna*, *9*, 131–145.

Abstract

Professional literature usually perceives the Internet and social media from the perspective of threat. Many papers describe the risk of using the Internet, both practical one concerning threatened security or finances and psychological one pertaining to addiction or depression. However, more and more often the cyberspace is treated as the research subject in itself or an area where one can analyse behaviours of Internet users. This paper is an example of the latter approach. With the help of the Big Data analysis of social media, Kessler Psychological Distress Scale (K10) shall be used to compare how often suicidal behaviour symptoms occurred in Internet users' posts this year and the year preceding the COVID-19 threat.

Keywords: depression, social media, big data, COVID-19.

Internet use to diagnose depression risk

Every year the number of products of human activity uploaded in cyberspace grows. A rapid development of social media means that a large portion of these products includes content being the outcome of one's willingness to share their thoughts, emotions or experiences. The unique nature of Internet culture keeps generating new ways of expressing people's opinions. At the time of fast content consumption, the written word is replaced by an image or an image combined with content in the form of an Internet meme. The current political situation of a given country is reflected in shared memes commenting on everyday events. Knowing which memes someone sends, it is easy to define their political preferences or attitude to widely-known everyday issues. If we match it with opinions placed in the net, the picture becomes complete.

In the same way, one may try to diagnose more complex personality features or, within the framework of preventive measures, to look for symptoms of mental disorders. The analysis can focus on the occurrence of words and phrases which are predicates defining particular personal features, often linked with a particular psychological theory. A good example is the use of Transactional Analysis terminology in search of depression symptoms (Wieczorek, 2018, 2019). A more practical action consists in looking for depression symptoms and targeting appropriate advertisements to a diagnosed person:

> Social media (SM) offers a promising avenue for targeting information about third sector mental health services to people who need them. SM sites such as Facebook already use algorithms to target advertisement to the most appropriate users, for example, by using search keywords from the history of search engines and links that users have previously clicked on. As machine learning and other computer science techniques have become more advanced, it is increasingly possible to identify or predict specific characteristics, such as mood or depression, of SM users, from the content they post on sites such as Facebook or Twitter. This may involve sentiment analysis (the valence of the emotion or mood of their words), analyzing posted images, or recognizing changes in the quantity and frequency of a user's content . Previous research has shown that users disclose depressive symptoms on SM sites such as Facebook and Twitter; in some cases, users disclose enough information for researchers to make a diagnosis of a major depressive episode (Ford, Curlewis, Wongkoblap, Curcin, 2019).

Artificial intelligence can also contribute to the analysis of content of posted photographs.

> Photographs posted to Instagram offer a vast array of features that might be analyzed for psychological insight. The content of photographs can be coded for any number of characteristics: Are there people present? Is the setting in nature or indoors? Is it night or day? Image statistical properties can also be evaluated at a per-pixel level, including values for average color and brightness. Instagram metadata offers additional information: Did the photo receive any comments? How many 'Likes' did it get? Finally, platform activity measures, such as usage and posting frequency, may also yield clues as to an Instagram user's mental state. We incorporated only a narrow subset of possible features into our predictive models, motivated in part by prior research into the relationship between mood and visual preferences (Reece, Danforth, 2017).

The process of the very analysis in the aforesaid research is quite complicated, requires ample means and using advanced technology.

> Data collection was crowdsourced using Amazon's Mechanical Turk (MTurk) crowdwork platform. Separate surveys were created for depressed and healthy individuals. In the depressed survey, participants were invited to complete a survey that involved passing a series of inclusion criteria, responding to a standardized clinical depression survey, answering questions related to demographics and history of depression, and sharing

> social media history […]Qualified participants were asked to share their Instagram usernames and history. An app embedded in the survey allowed participants to securely log into their Instagram accounts and agree to share their data (Reece, Danforth, 2017).

This kind of research requires cooperation of people submitted to it and a significant level of trust. The next example of machine learning is research on early detection of depression by means of analysing Reddit posts. The process of data collection is similar to the previous one.

> Data were extracted from Reddit, Inc using the Reddit, Inc's application program interface (API), and the resulting dataset consists of a collection of tuples of the form (id, writing), such that id is a unique identifier for each social network user and writing represents a writing instance in the social network. At the same time, each writing was described as a tuple of the form (title, date, info, and text), whereby title indicates the title of the post or comment, date denotes the date and time when the writing was performed, info identifies the social network (in this case, only Reddit, Inc is considered), and text comprises the actual post or comment provided by the user (Cacheda, Fernandez, Novoa, Carneiro, 2019).

Machine learning linked with a prior psychological diagnosis made it possible to diagnose a series of behaviours related to depression, such as focusing too much on the description of one's own posts or using a bigger number of more general expressions. Other examples of using social media in diagnosing depression include research on the analysis of Facebook posts (Eichstaedt et al., 2018), social media (Chancellor, Choudhury 2020; Kim, Lee, Park, 2020; Guntuku, et al., 2017).

The behaviour analysis of people diagnosed with depression combined with artificial intelligence makes it possible to find analogous behaviour patterns in other people, allowing for an early diagnosis of depression.

Methodology of own research

This research shall use the method requiring more modest financial means. Artificial intelligence will be replaced with Big Data technology, which allows for collecting and processing a big amount of information posted in social media. The Big Data analysis allows for processing huge amounts of data collected among others in social media (Żulicki, 2017; Janczyk, 2016, Wieczorek, 2018, 2019). Analyses were conducted with the help of Unamo service (https://unamo.com/) and Kessler psychological distress scale K10 (Kessler et al., 2002, 2003, 2010; Furukawa et al., 2003).

Kessler psychological distress scale allows for diagnosing depression symptoms. It consists of ten questions and answers on a Lickert scale.

Table 1
Kessler psychological distress scale and question counterparts for Internet research

Kessler psychological distress scale, In the past 4 weeks...	Big Data test – substitute
About how often did you feel tired out for no good reason?	I feel tired
About how often did you feel nervous?	I feel nervous
About how often did you feel so sad that nothing could cheer you up?	I feel sad
About how often did you feel hopeless?	I feel hopeless
About how often did you feel restless or fidgety?	I feel restless I feel fidgety
About how often did you feel so restless you could not sit still? About how often did you feel so nervous that nothing could calm you down?	I can't calm down
About how often did you feel depressed?	I feel depressed
About how often did you feel that everything was an effort?	I feel that everything was an effort
About how often did you feel worthless?	I feel worthless

Source: own research.

The answers to these questions can be found on a Lickert scale, which includes the following options: None of the time, A little of the time, Some of the time, Most of the time and All of the time. During a traditional test or while using an Internet form, a given person chooses appropriate answers. In case of the Big Data analysis, it is impossible to conduct such a test as it deals with archival data. To make use of the K10 scale, the phrases in question were shortened so that they could be found in Internet posts; their substitutes can be found in Table 1.

Two subsequent years were analysed and the factor that was considered significant was the COVID-19 pandemic likely to influence Internet users' mental state. A hypothesis was adopted that if the number of occurring phrases is statistically different for the year without the pandemic and for the year when the COVID-19 pandemic resurfaced, it can be assumed that it is possible to use scales for social media analysis. The Unamo platform allows for collecting information concerning the number of phrases occurring every day, the number of comments, likes and shares. For the needs of further analysis only the numbers of occurring comments will be used. The analysis concerns the periods of time from 23rd November 2018 to 22nd November 2019 and from 23rd November 2019 to 21st November 2020. The analysed period begins in November and for the reasons of clarity the first period shall be referred to as the year 2019 and the second one as the year 2020. As the year 2020 is a leap year, the second period of time is moved by one day. It makes no difference in case of statistical calculations as two cycles, 365 days each, are compared. In terms of graphs this difference is almost invisible. Table 2 presents the comparison of the number of searched expressions in social media.

Table 2
The number of searched expressions in subsequent years

An expression in a given year	A number of expressions found in social media
I feel tired 2019	163860
I feel tired 2020	211096
I feel that everything was an effort 2019	8476
feel that everything was an effort 2020	7381
I feel nervous 2019	11787
I feel nervous 2020	9889
I feel hopeless 2019	42597
I feel hopeless 2020	41036
I feel restless 2019	13079
I feel restless 2020	18200
I feel fidgety 2019	1098
I feel fidgety 2020	936
I feel depressed 2019	3046
I feel depressed 2019	2906
I feel worthless 2019	3606
I feel worthless 2020	1704
I feel sad 2019	55325
I feel sad 2020	75057
I can't calm down 2019	4196
I can't calm down 2020	6918
Total:	**682193**

Source: own research.

For the needs of further analysis, a series of statistical calculations was performed. The calculations focus on the same variables in two time cycles, the variables are arranged on the ratio scale. After the analysis of the variables' distribution and the conclusion that they have a normal distribution – Table 3, there was a T-Student test for dependent samples – Table 4. The regularity of the distribution was verified with the help of the Kolmogorov–Smirnov test for one sample and the Lilliefors test.

Table 3
Testing the regularity of variable distribution

The Kolmogorov–Smirnov test for one sample								
	N	Average	Standard deviation	Absolute value	Positive	Nega-tive	Test sta-tistics	Asymptotic significance (2 sided)
I feel tired 2019	365	448.93	174.603	0.076	0.076	−0.062	0.076	.000c
I feel tired 2020	365	578.35	293.666	0.153	0.153	−0.108	0.153	.000c
I feel that every-thing was an effort 2019	365	23.22	49.018	0.325	0.297	−0.325	0.325	.000c
feel that everything was an effort 2020	365	20.22	21.427	0.214	0.214	−0.2	0.214	.000c
I feel nervous 2019	365	32.29	90.884	0.369	0.341	−0.369	0.369	.000c
I feel nervous 2020	365	27.09	60.764	0.34	0.33	−0.34	0.34	.000c
I feel hopeless 2019	365	116.7	126.053	0.245	0.245	−0.245	0.245	.000c
I feel hopeless 2020	365	112.43	58.272	0.154	0.154	−0.149	0.154	.000c
I feel restless 2019	365	35.83	56.715	0.279	0.259	−0.279	0.279	.000c
I feel restless 2020	365	49.86	51.251	0.195	0.195	−0.193	0.195	.000c
I feel fidgety 2019	365	3.01	15.766	0.424	0.393	−0.424	0.424	.000c
I feel fidgety 2020	365	2.56	6.688	0.351	0.315	−0.351	0.351	.000c
I feel depressed 2019	365	8.35	28.849	0.386	0.332	−0.386	0.386	.000c
I feel depressed 2019	365	7.96	13.63	0.28	0.268	−0.28	0.28	.000c
I feel worthless 2019	365	9.88	72.889	0.446	0.393	−0.446	0.446	.000c
I feel worthless 2020	365	4.67	9.541	0.312	0.267	−0.312	0.312	.000c
I feel sad 2019	365	151.58	131.715	0.235	0.235	−0.216	0.235	.000c
I feel sad 2020	365	205.64	528.809	0.378	0.347	−0.378	0.378	.000c
I can't calm down 2019	365	11.5	19.799	0.289	0.289	−0.287	0.289	.000c
I can't calm down 2020	365	18.95	37.109	0.323	0.322	−0.323	0.323	.000c
Lilliefors significance correction								

Source: own research.

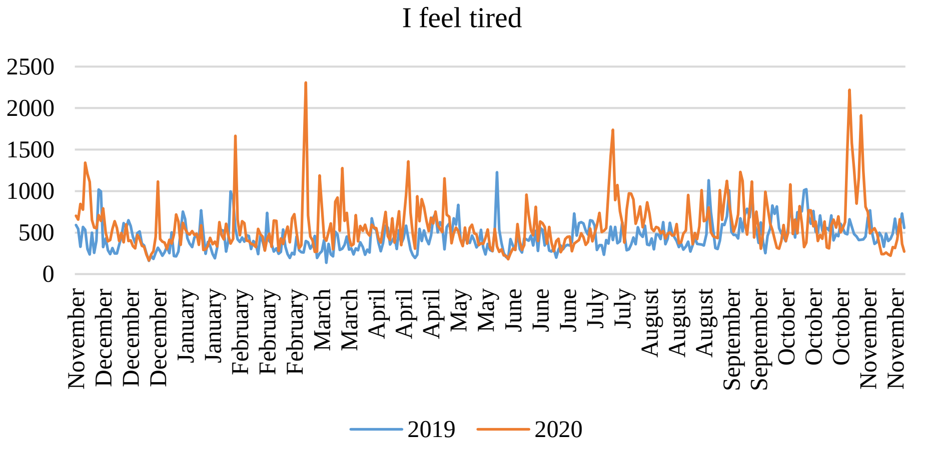
The Student t test was conducted for all the pairs of the variables, assuming that the first measurement is the analysis of 2019, and the second one is the analysis of 2020, the analysis was regarded as dependent samples. The analysis results are presented in Table 4. The pairs of variables "I feel tired," "I feel restless," and "I can't calm down" turned out statistically significant. Taking into account the nature of the COVID-19 pandemic, the differences in these variables reflect what could be expected after the change in reaction. A type of threat does not translate into depression indicators such as "I feel hopeless" or „I fell worthless" and others. It is directly related to the change in one's functioning, work or limitations in travelling. There were no detailed analyses of people's comments as in case of such a huge number of observations it is practically impossible without the use of artificial intelligence. However, it can be supposed that a big number of responses such as "I feel tired" is related to the epidemiological situation and e.g. the need to work from home.

Table 4
Statistical analysis, Student t test for dependent samples

	Average	Standard deviation	Average standard error	The lower limit	The upper limit	t	df	Significance (2 sided)
I feel tired 2019 – I feel tired 2020	−129.414	323.473	16.931	−162.709	−96.118	−7.643	364	0
I feel that everything was an effort 2019 – feel that everything was an effort 2020	3	49.084	2.569	−2.052	8.052	1.168	364	0.244
I feel nervous 2019 – I feel nervous 2020	5.2	107.353	5.619	−5.85	16.25	0.925	364	0.355
I feel hopeless 2019 – I feel hopeless 2020	4.277	138.258	7.237	−9,954	18.508	0.591	364	0.555
I feel restless 2019 – I feel restless 2020	−14.03	74.357	3.892	−21.684	−6.376	−3.605	364	0
I feel fidgety 2019 – I feel fidgety 2020	0.444	17.353	0.908	−1.342	2.23	0.489	364	0.625
I feel depressed 2019 – I feel depressed 2019	0.384	31.744	1.662	−2.884	3.651	0.231	364	0.818
I feel worthless 2019 – I feel worthless 2020	5.211	73.8	3.863	−2.385	12.807	1.349	364	0.178
I feel sad 2019 – I feel sad 2020	−54.06	543.859	28.467	−110.04	1.92	−1.899	364	0.058
I can't calm down 2019 – I can't calm down 2020	−7.458	41.185	2.156	−11.697	−3.218	−3.459	364	0.001
95% confidence interval for average differences								

Source: own research.

To make the analysis more thorough, one can use the graphs comparing the frequency with which searched expressions occur in social media. The pairs of the variables for which differences turned out statistically significant shall be presented first.



Graph 1
I feel tired

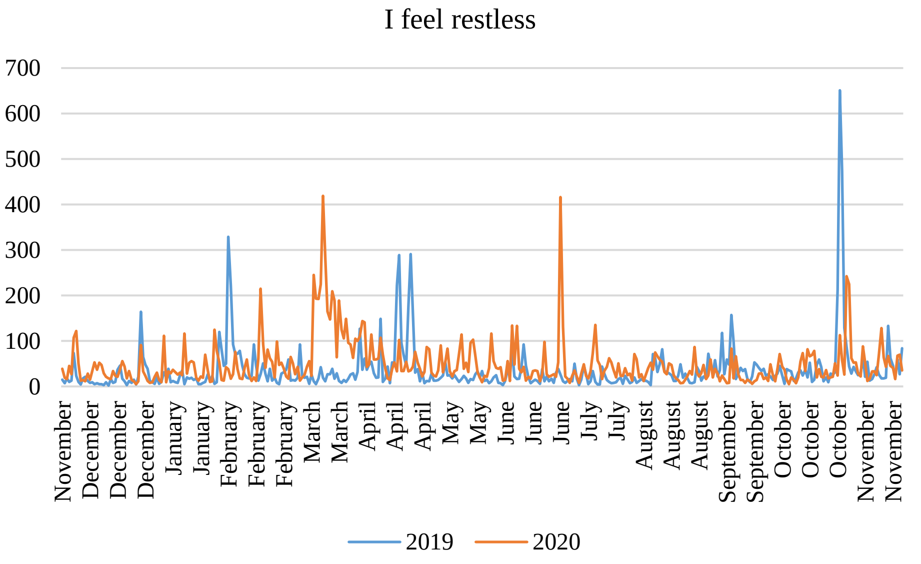
Source: own research.

Graph 1 with the variable „I feel tired" illustrates the biggest difference between the year 2019 and 2020. One can see clearly an increase in posts occurrence for the lockdown months, i.e. March 2020 and October/November 2020.

Graph 2 illustrating the variable "I feel restless" presents a similar increase for March 2020. Contrary to Graph 1, one can discern a clear increase in posts occurrence also for the year 2019, but most probably it was caused by a natural cycle of the year and periods of time that are more difficult from the point of view of depression indicators, such as Easter – time in April, Christmas – December, the end of the winter semester – February. Such an increase is characteristic for the annual frequency analysis of some utterances (Wieczorek, 2028, 2029).

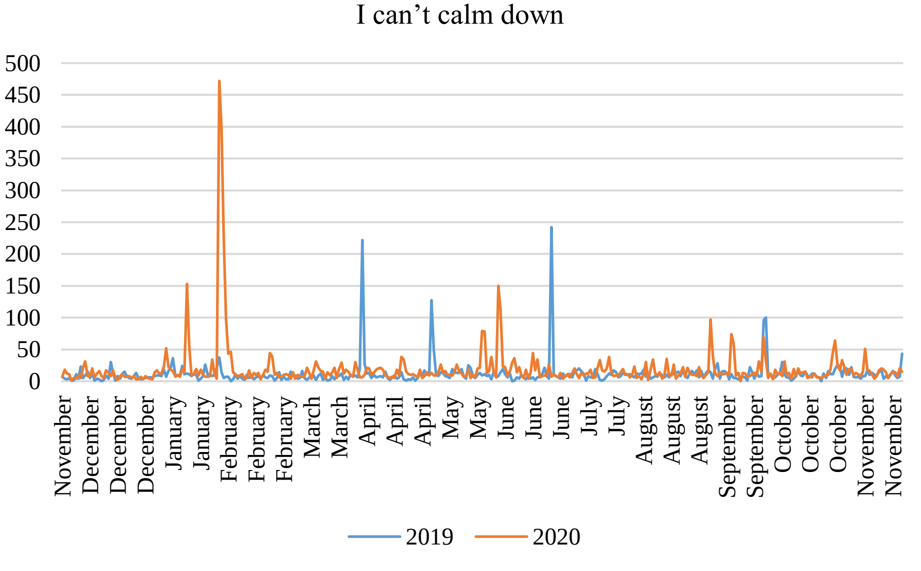
In case of the variable "I can't calm down" presented in Graph 3, one can see similar dependencies. An increase in post occurrence takes place near February – it was then that Internet users faced the sanitary regime, which naturally increased their anxiety.

The other variables presented in graphs 4–10 are not statistically significant, yet they illustrate the dynamics of posts occurrence in social media. There are characteristic peaks near the time of Christmas, the beginning of a new calendar year, semesters related to education and many others. A more thorough analysis might determine which situations provoke Internet users' posts, but that might require the use of artificial intelligence.



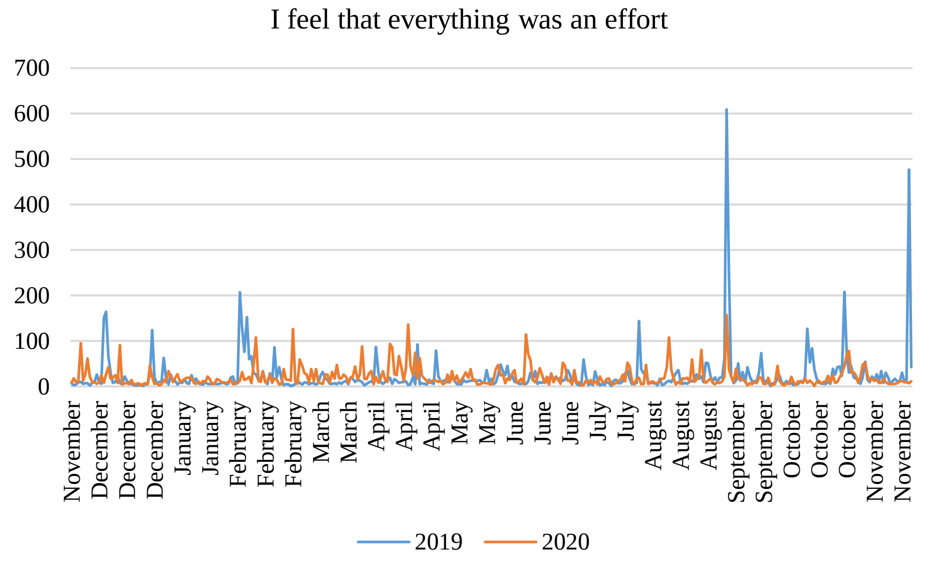
Graph 2
I feel restless

Source: own research.



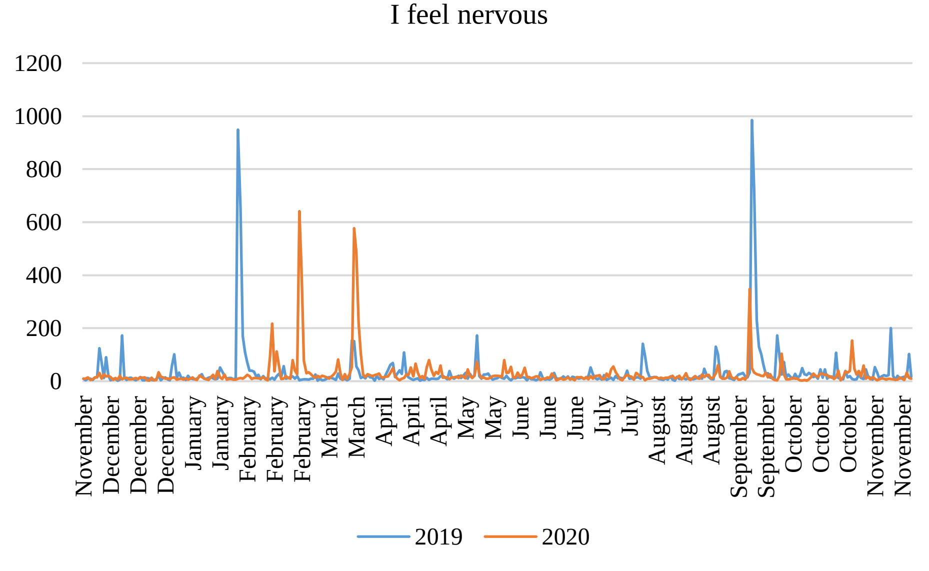
Graph 3
I can’t calm down

Source: own research.



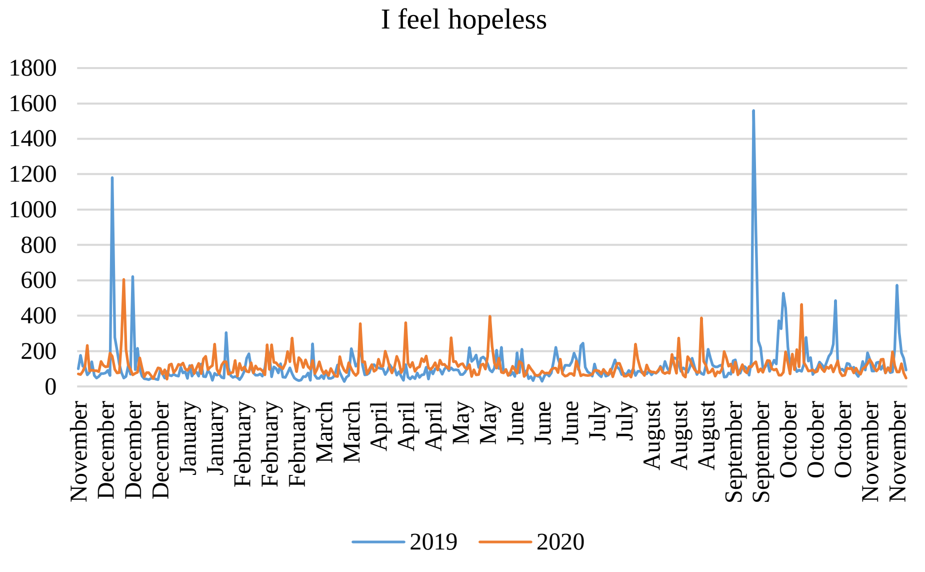
Graph 4
I feel that everything was an effort

Source: own research.



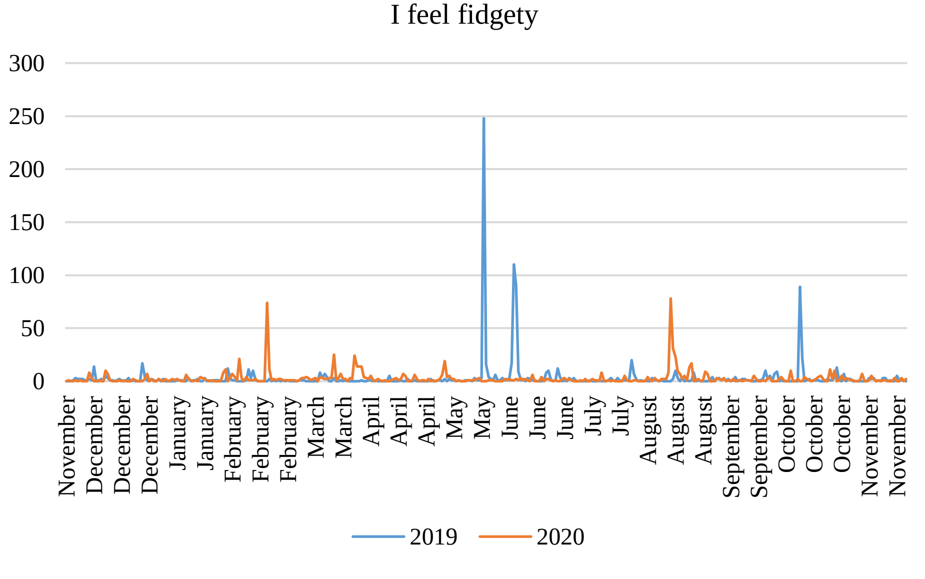
Graph 5
I feel nervous

Source: own research.



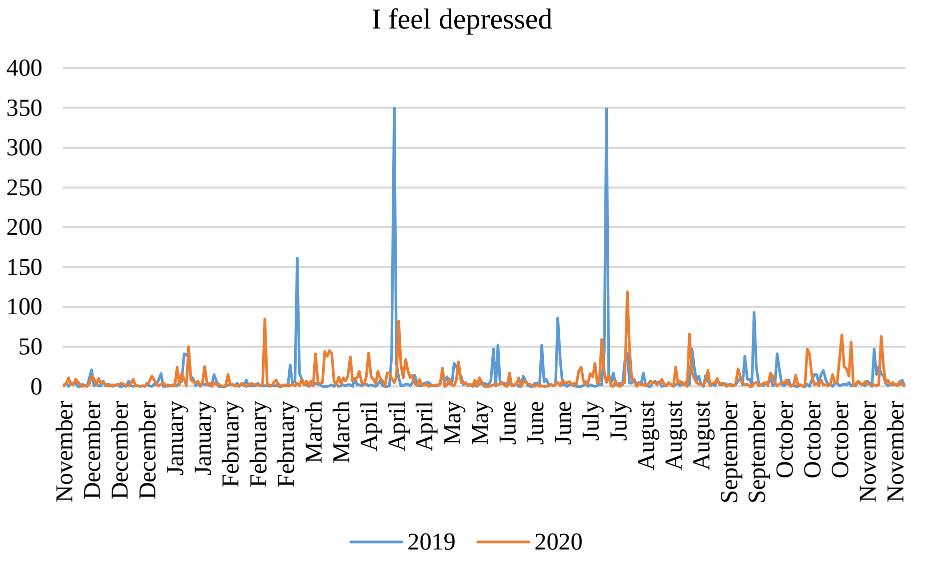
Graph 6
I feel hopeless

Source: own research.



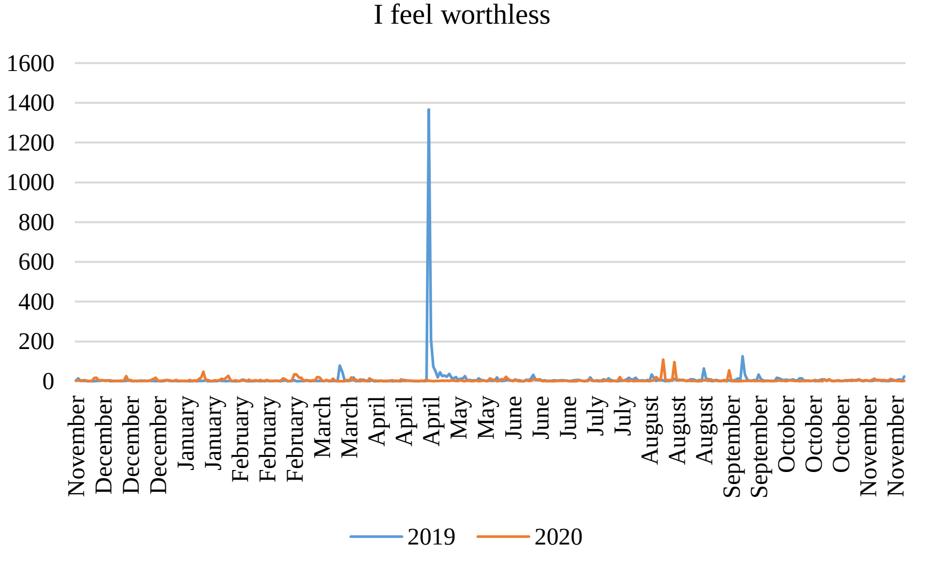
Graph 7
I feel fidgety

Source: own research.



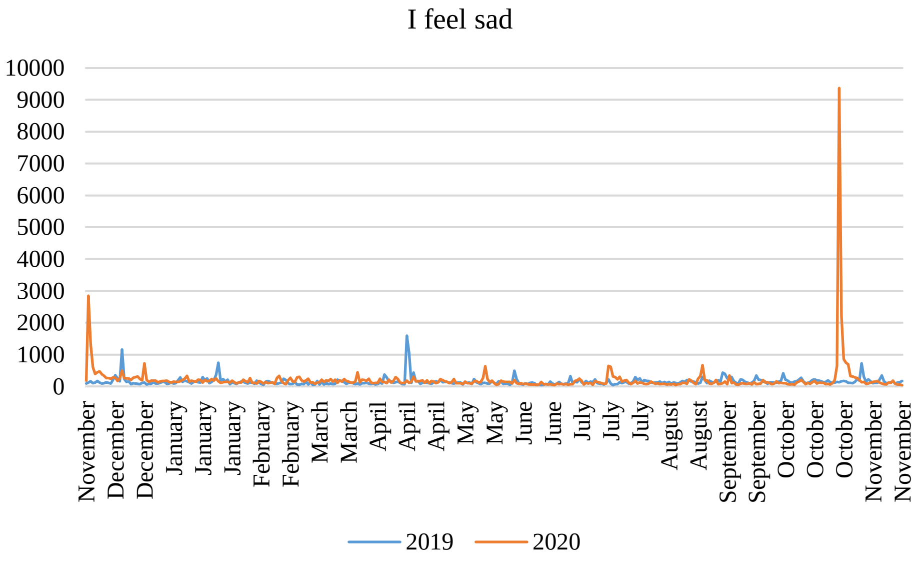
Graph 8
I feel depressed

Source: own research.



Graph 9
I feel worthless

Source: own research.



Graph 10
I feel sad

Source: own research.

Conclusions

The research in question was an attempt to use Kessler psychological distress scale K10 to analyse Internet users' behaviours. The analysis concerned two areas. The first one regarded the use of the modified scale to diagnose depression. The number of expressions found in social media suggests that this direction of the research has development potential. We deal here with a reversed research procedure. In an ordinary test, we can ask an examined person how often they feel tired. In the Big Data analysis I check how often Internet users write "I feel tired" in their Internet posts. In the research of that type it is not possible to find people at risk of depression, but it is possible to diagnose moods in the whole population.

The second area regards a diagnosis of how the COVID-19 situation is reflected in Internet users' behaviour. Abundant press news or everyday observation let us suppose that the epidemiological threat, limited interpersonal contact or working from home experienced by a vast group of people may be reflected in a worsened mood and an increase in depression symptoms. The research is not able to determine how depressive observed Internet users are as the applied methodology lacks a reference system. The difference between the year without the pandemic and with the pandemic was sought. Significant statistical differences occurred with three variables: "I feel tired," "I feel restless" and "I can't calm down." It shows that the year 2020 differs significantly from the previous one and the COVID-19 pandemic is one of possible factors. The difference in these areas reflects the situation of threat and changed lifestyle. At the same time, statistical significance in three variables out of ten leads to the conclusion that Internet users are tired of the situation but not as much as to increase the possibility of permanent mood worsening and depression.

References

Cacheda F., Fernandez D., Novoa F.J., Carneiro V. (2019). Early Detection of Depression: Social Network Analysis and Random Forest Techniques. *J Med Internet Res*, *21*(*6*): e12554. https://doi.org/10.2196/12554.

Chancellor, S., De Choudhury, M. (2020). Methods in predictive techniques for mental health status on social media: a critical review. *npj Digital Medicine, 3*, *43*. https://doi.org/10.1038/s41746-020-0233-7.

Eichstaedt, J.C., Smith, R.J., Merchant, R.M., Ungar, L.H., Crutchley, P., Preoţiuc-Pietro, D., Asch, D.A., Schwartz, H.A. (2018). Facebook language predicts depression in medical records, *Proceedings of the National Academy of Sciences Oct 2018*, *115* (*44*), 11203–11208. https://doi.org/10.1073/pnas.1802331115.

Ford, E., Curlewis, K., Wongkoblap, A., Curcin, V. (2019). Public Opinions on Using Social Media Content to Identify Users With Depression and Target Mental Health Care Advertising: Mixed Methods Survey. *Journal of Medical Internet Research*, 6 (*11*): e12942. https://doi.org/10.2196/12942.

Furukawa, T.A., Kessler, R.C., Slade, T., Andrews, G. (2003). The performance of the K6 and K10 screening scales for psychological distress in the Australian National Survey of Mental Health and Well-Being. *Psychological Medicine*, 33, 357–362.

Guntuku, S., Yaden, D., Kern, M., Ungar, L., Eichstaedt, J. (2017). Detecting depression and mental illness on social media: an integrative review. *Current Opinion in Behavioral Sciences*, *18*, 43–49, ISSN 2352-1546. https://doi.org/10.1016/j.cobeha.2017.07.005.

Janczyk, J. (2016). Big Data w relacji do procesów zmian w edukacji. *Dydaktyka Informatyki*, *11*, 100–108. http://dx.doi.org/10.15584/di.2016.11.13.

Kessler, R.C., Andrews, G., Colpe, L.J., Hiripi, E., Mroczek, D.K., Normand, S.-L.T., Walters, E.E., & Zaslavsky, A. (2002). Short screening scales to monitor population prevalances and trends in nonspecific psychological distress. *Psychological Medicine*, *32*(*6*), 959–976.

Kessler, R.C., Barker, P.R., Colpe, L.J., Epstein, J.F., Gfroerer, J.C., Hiripi, E., Howes, M.J, Normand, S-L.T., Manderscheid, R.W., Walters, E.E., Zaslavsky, A.M. (2003). Screening for serious mental illness in the general population. *Archives of General Psychiatry*, *60*(*2*), 184–189.

Kessler, R.C., Green, J.G., Gruber, M.J., Sampson, N.A., Bromet, E., Cuitan, M., Furukawa, T.A., Gureje, O., Hinkov, H., Hu, C.Y., Lara, C., Lee, S., Mneimneh, Z., Myer, L., Oakley-Browne, M., Posada-Villa, J., Sagar, R., Viana, M.C., Zaslavsky, A.M. (2010). Screening for serious mental illness in the general population with the K6 screening scale: results from the WHO World Mental Health (WMH) survey initiative. *International Journal of Methods in Psychiatric Research*, *19*(*S1*), 4–22. Erratum in International Journal of Methods in Psychiatric Research 2011 Mar; 20(1): 62.

Kim, J., Lee, J., Park, E. (2020). A deep learning model for detecting mental illness from user content on social media. *Scientific Reports*, *10*, article: 11846. https://doi.org/10.1038/s41598-020-68764-y.

Reece, A.G., Danforth, C.M. (2017). Instagram photos reveal predictive markers of depression. *EPJ Data Sci*, *6*, *15*. https://doi.org/10.1140/epjds/s13688-017-0110-z.

Wieczorek, Z. (2018). Wykorzystanie analizy transakcyjnej w badaniach mediów społecznościowych. *Edukacyjna Analiza Transakcyjna*, *7*, 133–159. https://doi.org/10.16926/eat.2018.07.09.

Wieczorek, Z. (2019). Depresja w świetle teorii analizy transakcyjnej – analiza mediów społecznościowych przy pomocy Big Data. *Edukacyjna Analiza Transakcyjna*, *8*, 91–115. https://doi.org/10.16926/eat.2019.08.07.

Żulicki, R. (2017). Potencjał big data w badaniach społecznych. *Studia Socjologiczne*, *3*(*226*), 175–207.

Objawy lęku i depresji w mediach społecznościowych w związku z zagrożeniem COVID-19

Streszczenie

W piśmiennictwie naukowym Internet i media społecznościowe postrzegane są zazwyczaj z perspektywy zagrożenia. Wiele opracowań opisuje ryzyko korzystania z Internetu, zarówno praktyczne, które wiąże się z możliwością zagrożenia bezpieczeństwa czy finansów, jak i psychologiczne związane uzależnieniem czy depresją. Coraz częściej jednak cyberprzestrzeń jest traktowana, jako obiekt badań sam w sobie lub przestrzeń, w której można badać zachowania internautów. Niniejsze opracowanie jest przykładem tego drugiego podejścia. Za pomocą analizy Big Data mediów społecznościowych dokonana będzie próba wykorzystania Kessler Psychological Distress Scale (K10) do porównania, jak często w wypowiedziach internautów pojawiały się sygnały zachowań depresyjnych w roku bieżącym i poprzedzającym zagrożenie COVID-19.

Słowa kluczowe: depresja, media społecznościowe, big data, COVID-19.