

# Machine Learning Analysis of Social Media Posts Written in Natural Language May Enhance Suicidal Ideation Detection in Romanian Adults with Major Depression

[Eduard P. Drima MD, PhD](#) \*

Posted Date: 18 July 2023

doi: 10.20944/preprints202307.1177.v1

Keywords: Suicidal Ideation, major depression in adults, natural language written texts, Romanian depression support forum, machine learning text mining,



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

*Article*

# Machine Learning Analysis of Social Media Posts Written in Natural Language May Enhance Suicidal Ideation Detection in Romanian Adults with Major Depression

Eduard P. Drima MD, PhD\*

Senior psychiatrist, "Elisabeta Doamna" Hospital, Galați; Associate Professor "Dunărea de Jos" Medical University, Galați, Romania; edi.drima@medacs.ca

\* Correspondence: edi.drima@medacs.ca

**Abstract:** Detecting suicidal ideation in adults with major depression is crucial for timely intervention and prevention of self-harm. As suicide is influenced by various biological, socio-cultural and psychological factors, traditional screening methods have accuracy and efficiency limitations. In certain cultures, societal stigma and marginalization can compel individuals with depression to conceal their suffering. Such individuals often turn to online social media platforms and share their experiences with peers under the protection of anonymity. Our research explored the potential of machine learning detection of suicidal ideation among Romanian adults with major depression that contributed to a web-based depression support forum. A trained algorithm (C4.5 decision tree) analyzed 125 posts fed to on a free access online support forum over 5 years (2014 – 2018) and classified them based on suicidal ideation content. 32 texts (25%) were identified as having a high probability of suicidal ideation content. 65% of the authors were male, with a mean age of  $36.7 \pm 10.3$  years and an average duration of illness of  $3.4 \pm 1.4$  years. Texts indicating positive suicidal ideation were generally shorter and elicited more general responses but fewer professional responses compared to those without suicidal ideation content. The study's main limitations include the relatively small number of classified texts, the absence of prospective information and the lack of qualitative evaluation of the excerpts' content. As socio-demographic and linguistic actuarial results were comparable to data reported by real life studies, we may consider basic text mining techniques as a screening tool that is able to detect suicidal ideation in texts written in unstructured Romanian language.

**Keywords:** suicidal ideation; major depression in adults; natural language written texts; Romanian depression support forum; machine learning text mining

## 1. Introduction

Depression, the most commonly diagnosed mental disease [1], is a significant public health concern [2] as it affects over 300 million individuals on a daily basis, worldwide [3]. The complex and varied symptomatology, along with the challenges associated with treatment and the high rate of recurrence result in a life-long health burden for patients [4,5] and a clinical challenge for healthcare providers [6]. Patients grappling with major depressive disorder (MDD) often highlight social uncertainty and feelings of helplessness as prominent factors influencing their struggle with their mental health problems [7]. The presence of social stigma and marginalization [8], negative, hate comments or franc emotional bullying [9] further exacerbates their self-esteem problems, potentially triggering or intensifying suicidal thoughts [10]. Due to all negative social repercussions, MDD subjects tend to conceal suicidal ideation (SI) when they interact with others [11].

Suicide is one of the major leading causes of death, worldwide [12]. More than 1% of all world reported deaths are a consequence of a suicidal act [13]. The suicidal act is a complex socio-medical

phenomenon where various biological, psychological, social and cultural factors may play different roles [14]. Suicide is described as a continuum that starts with suicide thoughts (active or passive) and ends with suicidal behaviors (from attempts to completed suicide). This concept is not unanimously accepted mainly because far less than one third of subjects with suicide ideation (SI) will actually attempt suicide [15] and only 3-4% of them will complete suicide [16]. The epidemiology of the suicidal act is multifaceted [17], the highest rate of completed suicide being observed in male subjects over 55 years old (the age and gender paradox – [18] and with a majority of nonfatal self-injuries seen in 15-20 years old females [19]. Early detection of suicidal behavior can lead to suicide prevention [20] but screening for SI is a serious challenge. Suicide is a subject that is strongly influenced by widespread socio-cultural and religious ideas [21] and subjects planning suicide will often adopt compensatory behaviors [22]. There is no suicide risk assessment tool or procedure that is able to correctly identify suicide behavior [23]. In 1983 Pokorny demonstrated in a large prospective study, that predicting a suicidal act is rather a matter of chance than a scientific deduction [24]. Hawgood and coworkers [25] considered that a change in the SI detection paradigm (universal screening) may lead to a better prevention of suicide but results are unreliable as both false positive and negative cases were high [26].

Psychiatric patients have an increased lifetime risk for committing suicide [27]. SI was often described in subjects with various forms of mental disorders [28] and a strong association between SI and MDD was constantly reported [29]. More than two-thirds of subjects that completed suicides were diagnosed with a form of mood disorder [30]. Detecting suicidal thoughts in psychiatric patients is a common but difficult task for the treating physician [31] and often physicians will not carefully test for SI as the process is time and resource consuming [32].

The abundance of health data available online provides great access to information pertinent to the management of many health conditions [33]. Online support groups and discussion forums provide a modern avenue for chronic patients to share their personal experience, to exchange treatment information and seek peer support [34]. For individuals with MDD, online support forums create a space with reduced social pressure, allowing them to connect with others who understand their struggles [35]. Although large written texts, such as forum posts, may not be as popular as they were a decade ago, they remain a significant means for MDD individuals to connect with their peers [36]. The anonymity provided in these forums [37] often encourages contributors to express their genuine thoughts and opinions, providing valuable insights into their ideation [38]. Conducting punctual analyses of these texts using computational linguistics techniques, such as text mining, can offer a deeper understanding of the emotional state of the individuals involved [39]. Anonymity also plays a crucial role in driving individuals with SI to seek peer interaction and support on social media platforms [40–42]. Multiple published reviews have explored the state of suicide prediction in social media, particularly in the context of various mental illnesses [43,44]. The consensus from these studies is that while predicting SI is possible, it is heavily influenced by factors such as social media access, language, culture, and education. Age often determines the choice of social media platforms [45], with newer platforms like Twitter, Reddit, and TikTok proving effective for predicting suicidal ideation in young, non-MDD individuals [46,47]. However, older individuals with chronic conditions often prefer specialized social channels like discussion forums that allow extensive text contributions and are not driven by commercial interests (controlled access). For non-native English speakers, posting short texts in English is non-natural [48]. Therefore, detecting SI in MDD individuals from diverse non-English speaking cultures [49] based on their social media contributions has become an increasingly important area of research [50,51].

The development of natural language processing algorithms had revolutionized computerized text analysis [52] providing valuable support for medical research. Text mining can adopt various tactics (terms tracking, automatic summarization, information retrieval or text categorization) in response of different needs. Algorithms used in text categorization (often named “classifiers”) can categorize large texts written in unstructured, natural languages based on distinct linguistic patterns which are associated with various sentiments or emotions. A classifier must be first trained using clear, predefined rules [53] which are considered the “ground truth”. After training, algorithms are

capable to detect similar patterns in unstructured texts, and predict association with a predefined class of sentiments. The whole process of text categorization is performed under “supervised learning” technique. In contrast, unsupervised learning methods allow algorithms to autonomously determine classification rules (using clustering algorithms) often leading to interesting results. However, the outcomes of unsupervised learning may vary when applied to texts written in non-English languages.

Supervised training text categorization were validated and have various applications (from email spam filtering to hate speech recognition). It allows a good SI prediction in social media posts [54]. There are several formal preliminary steps that have to be followed before text analysis (text preprocessing, feature extraction and feature selection). All these steps are grouped into the “training phase” of the algorithm [55]. Preprocessing consists in removing of words that are not important for information retrieval (like articles and prepositions – known as stop-words). Removing stop-words will reduce the volume of text from which the real information has to be extracted. Feature extraction and selection are important stages of training, and are associated with “the bag of words” method. After preprocessing, key words from text are organized as vectors (vector space model – Salton 1975) and finally the text will be classified. Several machine learning platforms (Knime, Orange, RapidMiner, Weka) are offering similar text mining algorithms. Decision to use one platform or another is more a personal preference. Weka (Waikato Environment for Knowledge Analysis [56]) is a collection of machine learning algorithms that has been designed to transform and preprocess raw text data and, based on attributes selection, to classify or cluster unstructured text. It also allows a substantial statistical evaluation of the results. It has a user-friendly graphic interface and may work well even with minimal programming.

In Romania, a South Eastern European country [57], 86% of population is practicing the Christian Orthodox religion under the lead of the authoritarian, autocephalous, Romanian Orthodox Christian Church. The Church is firmly condemning suicide considering it the most unpardonable capital sin. This idea is widely embraced by the older, religious population of the country. The strong socio-cultural opposition to suicide is often associated with a generalized stigmata concerning any forms of mental disorder. The overall percentage of completed suicide acts is lower than the European average and remained constant over the last decade [58]. Little is known about SI in Romanian MDD patients and the subject is difficult to investigate, as most of the subjects decide to avoid it in open medical discussions.

Our research interest is to improve SI detection in Romanian psychiatric patients. As classical medical screening is difficult, we evaluated how natural language posts sent to a Romanian social media forum can be used to early detect SI in MDD patients.

## 2. Materials and Methods

We used WEKA [59] text mining algorithms to detect SI in Romanian forum texts written by subjects having a clear MDD diagnostic. The “training dataset” included 15 posts that were extracted from a dedicated suicide prevention forum (<https://www.romedic.ro/forum/despre/sinuciderea>). Inclusion criteria were suicidal history and ideation (present ideation, previous suicidal tentative, emergency hospitalization after suicidal attempt, family history of suicide) and a mood disorder diagnostic (most of the times, a MDD suggested by diagnostic and or treatment). Authors were carefully balanced (age and sex) with 15 subjects that posted notes which did not include clear suicide ideation, sampled from a depression support forum (inclusion criteria – MDD in medical history and/or prescribed medication). The initial word corpus of the training set contained 4651 words (30 instances with 2 attributes – High Probability=HP and Low Probability=LP of SI). No socio-demographic information was used for training purposes.

In the “testing dataset” we included 125 unstructured, naturally written texts which were selected from a total of 657 posts sent over a 5 years period (2014-2018) to a dedicated Romanian depression support forum (<https://www.romedic.ro/forum/despre/depresia>). Inclusion criterium was a clear MDD that was established based on mentioned medical history and/or prescribed medication. Main text exclusion criteria were texts which suggest therapies or drugs positioning or

the absence of minimal socio-demographic information (sex and age). All posts were done by subjects with masked identity (pen name) but socio-demographic characteristics (age and sex) were available in all of the selected cases. All subjects included in this dataset were initially considered with a low probability of SI. The word corpus of the test dataset comprised 26872 words.

The text classification tactic was “supervised learning”. The training dataset was considered ground truth. The bag of words was trimmed using a Romanian stop-word dictionary (312 words – available @ <https://github.com/stopwords-iso/stopwords-ro>). Lemmatization and stemming were performed manually as existing dictionaries generated linguistic errors. After preprocessing, the training bag of words was reduced to 1073 words. Finally, we used a decision tree classifier (J48, the WEKA algorithm derived from C4.5) for the text classification. A “decision tree” is a non-parametric machine learning method able to predict the value of a target variable based on simple decision rules learned from the data features. It is easy to understand from a medical standpoint and can be graphically visualized [60]. Only cases with 100% prediction confidence were selected for the final analysis. Selected cases were analyzed from socio-demographic, peers’ interest (number of comments received), professional advice and support and texts statistics (number of words per text excerpt) perspectives.

**Statistics.** Results were analyzed using the PSPP software (GNU project, 2015). Only basic descriptive statistic (mean and standard deviation) is presented here. Scores were analyzed with Student’s T-test. P value was set at  $p=0.05$ .

**Ethics.** Both depression and suicide support forums are open to public. Both are totally anonymized and the de-identification process is performed by data curators. Our study used only non-identifiable data following the Canadian ethics research provisions for secondary data use studies (TCPs 2(2018)). The data investigation research was performed in accordance with declaration of Helsinki principles.

### 3. Results

The training dataset (30 subjects – 12 males and 18 females) had a mean age of  $34.56 \pm 7.8$  years. The mean age of the testing dataset (125 subjects – 51 males and 74 females) was  $30.8 \pm 10.6$  years. Detailed socio-demographic data are presented in Table 1. After preprocessing, feature selection and extraction from the training dataset, the algorithm generated a decision tree with 11 nodes and 6 leaves (Figure 1).

**Table 1.** Socio-demographic characteristics of training and testing datasets (m: males, f: females, yrs.: years; SI+ high probability of suicidal ideation and SI- low probability of suicidal ideation,  $*p=0.0026$ ).

	Training Dataset 30 cases		Testing Dataset 125 cases	
	Selected SI+	Selected SI-	All presumed SI-	
<b>NR</b>	15	15	125	
<b>Sex</b>	6m/9f	6m/9f	51 m	74 f
<b>Age (yrs.)</b>	$33.4 \pm 6$	$35.6 \pm 8$	$29.25 \pm 10.1$	$31.01 \pm 10.6$
<b>Depression duration (yrs.)</b>	$2.15 \pm 1.4$	$2.41 \pm 1.7$	$3.04 \pm 3.21$	$4.37 \pm 6.1$
<b>Words/excerpt</b>	$161.1 \pm 93.8(*)$	$149.9 \pm 106.2$	$301.69 \pm 180(*)$	

32 subjects (25%) from the testing dataset were classified as having a high probability of SI. 91 subjects remained with a low probability of SI. 2 cases were rejected from LP cohort analysis (both females) as the text and socio-demographic info were identical but texts were posted under different pen names. Details concerning both cohorts after text classification are presented in Table 2.



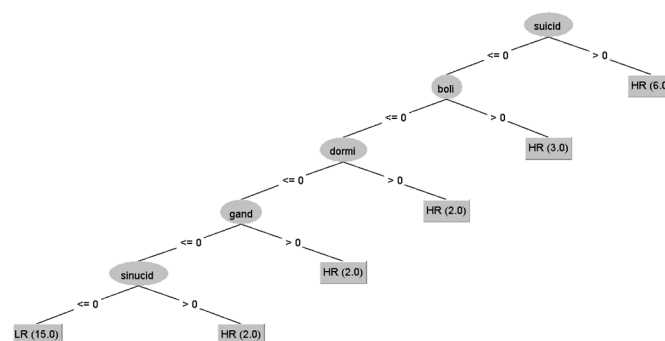
**Table 2.** Socio-demographic characteristics of test dataset subjects after text classification (\*p=0.055, \*\*p=0.0003, 2 subjects were rejected from the predicted SI- cohort analysis) .

Testing dataset (125 subjects)	Predicted SI+	Predicted SI-
NR	32	91
Sex	21m/11f	30m/61f
Age (yrs.)	34.1±10.2(*)	30.2±10.2(*)
Depression duration (yrs.)	3.54±3.4	4.0±5.52
Words/excerpt	221.4±157.25(**)	329.9±179.7(**)

Table 3 shows comparative data extracted from the two predicted cohorts (SI+ and SI-), after text classification. Beside socio-demographics (sex and age), medical (disease duration) and linguistics (number of words from texts excerpt) statistics are showed. The triggered general responses and the professional responses are compared as well.

**Table 3.** Comparing Predicted High Probability (SI+) vs Low Probability (SI-) cohorts: socio-demographic, medical and computational linguistic analysis results (\*p=0.05, \*\*p=0.0011, \$p<0.0001, #p=0.0003, &p=0.03).

	Predicted SI+			Predicted SI-		
	All	F	M	All	F	M
Number	32	11	21	91	63	28
Age	34.1±10.2(*)	29.2±8.4	36.7±10.3(**)	30.0±10.3(*)	31.4±10.89	27.5±8.23(**)
Disease duration	3.5±3.4	3.1±2.4	3.8±3.84	4.0±5.52	4.5±6.42	2.6±2.6
General Responses	3.75±1.79(&)	4.8±2.89	3.4±1.4	3.1±1.24(&)	3.2±1.4	3.0±1.0
Medical Responses	1±0.8(\$)	1.1±1.04	1±0.9	2.1±1.1(\$)	2.1±1.2	2.1±1.0
Words/excerpt	221.5± 157.25(#)	226.4± 122.6	218.9± 175.5	329.9± 179.8(#)	337.5± 189.1	313.2± 158.9



**Figure 1.** C4.5 Decision tree structure.

#### 4. Discussion

Depression is the most commonly diagnosed mental disease, worldwide. Depressed patients may face serious social stigmata and marginalization, circumstances that may originate suicidal thoughts. In cultures where there is a franc opposition against suicide and apprehension towards

mental illness, MDD patients with SI will conceal their mental health issues, making SI screening a notoriously difficult task. Depressed subjects having SI will rather accept to discuss their mental health problems in texts sent to social media platforms, where their true identity can be masked. In Romania, the older population of Christian Orthodox faith may show apprehension against both suicide and mental health issues. Mid-aged subjects with major depression and SI will often avoid any open discussion about their problems and will prefer the use of social media platforms for discussions and advice under the protection of anonymity.

Our research interest is to improve SI screening in Romanian MDD patients. As classical clinical screening often fails, we postulated that social media posts that are written in natural language can help the early detection of SI in MDD patients. Forum posts excerpts are usually long but for a machine learning algorithm perspective, text classification is fast once the algorithm is trained. 125 posts from a public depression support forum were included in a “testing dataset”. The text mining algorithm (decision tree C4.5) was trained using a “training dataset” (30 posts). In 15 of them, MDD patients shared their suicidal thoughts. These cases were carefully balanced (sex and age) with 15 posts provided by MDD subjects without any suicidal ideation. After the full text preprocessing, the trained decision tree algorithm classified the posts from the “testing dataset” based on six specific words (suicide wish, sleep, disease, thoughts, death, self-annihilation – translated from Romanian). These words were similar with those mentioned by Tadesse and coworkers when they used different deep learning text classification methods to detect SI in English speaking population (2020). After classification, 32 posts (25%) were classified as having a high probability of suicidal ideation content (SI+) (21 m/11f having a mean age of  $34.1 \pm 10.2$  years and a disease duration of  $3.75 \pm 1.79$  years). 91 posts were considered with a low probability of suicidal ideation content (SI-) (30m/61f, mean age  $30.2 \pm 10.2$  years, disease duration  $3.05 \pm 1.24$  years). Between the two cohorts, only age was significantly different ( $p=0.05$ ).

We evaluated the clinical precision of the algorithm classification based on socio-demographic (sex, age and disease duration) and linguistic criteria (excerpts length). Finally, we assessed how posts from both cohorts triggered general (peers) and professionals' responses.

Both training and testing datasets initially included 40% males. After text classification, the SI+ cohort contained 65% males (mean age of  $36.7 \pm 10.3$  years). This group was significantly older than the equivalent group included in the predicted SI- cohort (28% males, mean age  $27.5 \pm 8.23$  years). Our confidence in the text classification precision increased as this observation is in line with “the gender and age paradox” observed in the real-life studies (data sampled from depressed subjects that actually committed suicide). The predicted SI+ males had also a longer major depression history (mean  $3.8 \pm 3.84$  years), significantly different than what was disclosed by males without SI (mean  $2.6 \pm 2.6$  years). As expected, the group of depressed females without SI showed the longest history of disease (average  $4.5 \pm 6.42$  years).

From a computational linguistic perspective, the average excerpt length observed in the predicted SI+ cohort ( $221.5 \pm 157.25$  words) was significantly shorter than what was seen in the SI- cohort ( $329.9 \pm 179.8$  words). A similar finding was reported by Lehrman and coworkers based on English texts analysis (2012). The use of the word “disease” between the six classification items may suggest an association with other various clinical diagnostics but there was no qualitative evaluation of the written texts so this observation has to be deepened in a further study.

We evaluated the number of triggered answers coming from peers and from medical professionals. The number of general answers was significantly higher in the SI+ cohort ( $3.75 \pm 1.79$  comments). The number of comments coming from medical professionals was significantly lower ( $1 \pm 0.8$  advices) in the SI+ cohort. This observation is important as several researchers already stated that clinicians are reluctant to treat individuals with SI because of a suboptimal response to therapy [61]. Loss of one patient due to suicide is seen as one of the most stressful events in the life of a clinician/therapist [62,63]. Our observation concerning the low professional involvement in possible SI cases reflected by social media has to be elucidated in the future.

### *Limitations*

A serious limitation of any text mining study in psychiatry or psychology is coming from the known reality that no known actuarial algorithm can capture the full complexity of a person's mental state. The low number of texts used in the classification process is another important limitation of our study but it comes from the low number of Romanian depression support forums in use. The absence of any prospective long-term information and the strict quantitative analysis of the excerpts are also limitations of this study.

## 5. Conclusions

Large text posts written in natural language can be used for suicidal ideation detection in Romanian subjects with major depression. Early detection of suicidal thoughts can prevent a possible suicidal behavior. Trained algorithms can classify social media forums posts (using a text mining approach) and express a probability of suicidal ideation content. Actuarial methods are reliable and fast screening tools, allowing healthcare professionals to early identify individuals at risk and to provide appropriate support and intervention. Future studies are needed to better understand how suicidal thoughts will evolve in time in identified depressed subjects that used web support forums and how social pressure that sometimes exists in the on-line environment can further influence these patients.

**Acknowledgment:** The study had no funding support to report, and there is no conflict of interests to report, as well. The author thanks Dr. Marius D Gangal (@medacs.ca) for his constant involvement, scientific support and advice. The support provided by OpenAI's GPT-3.5 model in the revision of this article must also be acknowledged.

**Data Availability Statement:** All data used in this research is available upon request.

## References

1. World Health Organization. 2018. [2018-04-23]. Depression: Key Facts <https://www.who.int/news-room/fact-sheets/detail/depression>
2. Kessler RC, Chiu WT, et al. Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. *Arch Gen Psychiatry*. 2005; 62:617–627
3. World Health Organization. Geneva, Switzerland: World Health Organization; 2017. [2020-10-10]. Depression and other common mental disorders: global health estimates. <https://apps.who.int/iris/handle/10665/254610>
4. Rehm J, Shield KD. Global Burden of Disease and the Impact of Mental and Addictive Disorders. *Curr Psychiatry Rep*. 2019 Feb 7;21(2):10. doi: 10.1007/s11920-019-0997-0.
5. Schoevers RA, Smit F, et al. Prevention of late-life depression in primary care: do we know where to begin? *The American journal of psychiatry* 2006, 163(9):1611-1621
6. Touloumis C. The burden and the challenge of treatment-resistant depression. *Psychiatriki*. 2021 Dec;32(Supplement I):11-14. English, Greek, Modern. doi: 10.22365/jpsych.2021.046.
7. Sik, D. (2018). From mental disorders to social suffering: Making sense of depression for critical theories. *European Journal of Social Theory*, 22(4), 477–496
8. Barney LJ, Griffiths KM, Jorm AF, Christensen H. Stigma about depression and its impact on help-seeking intentions. *Aust N Z J Psychiatry*. 2006 Jan; 40 (1):51–4. doi: 10.1080/j.1440-1614.2006.01741.x.
9. Giumetti GW, Kowalski RM. Cyberbullying via social media and well-being. *Curr Opin Psychol*. 2022 Jun;45:101314. doi: 10.1016/j.copsyc.2022.101314.
10. Woods HC, Scott H. #Sleepyteens: social media use in adolescence is associated with poor sleep quality, anxiety, depression and low self-esteem. *J Adolesc*. 2016 Dec; 51:41–9. doi: 10.1016/j.adolescence.2016.05.008
11. Oexle N, Ajdacic-Gross V, et al, Mental illness stigma, secrecy and suicidal ideation. *Epidemiol Psychiatr Sci*. 2017 Feb;26(1):53-60. doi: 10.1017/S2045796015001018.
12. WHO suicide worldwide in 2019 report available at <https://www.who.int/publications/i/item/9789241506021>
13. Värnik P (March 2012). "Suicide in the world". *International Journal of Environmental Research and Public Health*. 9 (3): 760–71. doi:10.3390/ijerph9030760
14. Raue P, Ghesquire A., et al Suicide Risk in Primary Care: Identification and Management in Older Adults, *Curr Psychiatry Rep*. 2014 September ; 16(9): 466. doi:10.1007/s11920-014-0466-8.
15. Klonsky, E.D.; May, A.M. Differentiating suicide attempters from suicide ideators: A critical frontier for suicidology research. *Suicide Life-Threat. Behav*. 2014, 44, 1–5.



16. Shea, S. C. (2011). *The practical art of suicide assessment: A guide for mental health professionals and substance abuse counselors*. Hoboken, NJ: Mental Health Presses.
17. Nock M., Borges G., et al *Suicide and Suicidal Behavior*, 2008, *Epidemiologic reviews*, 30, DOI: 10.1093/epirev/mxn002
18. Schrijvers DL, Bollen J, Sabbe BG. The gender paradox in suicidal behavior and its impact on the suicidal process. *J Affect Disord*. 2012;138(1-2):19-26
19. Piscopo K., Lipari R, *Suicidal Thoughts and Behavior among Adults: Results from the 2015 National Survey on Drug Use and Health*, NSDUH Data Review from <http://www.samhsa.gov/data/>
20. Mann JJ, Michel CA, Auerbach RP. Improving Suicide Prevention Through Evidence-Based Strategies: A Systematic Review. *Am J Psychiatry*. 2021 Jul;178(7):611-624. doi: 10.1176/appi.ajp.2020.20060864.
21. Chappell PB, Stewart M, et al, Assessment of Suicidal Ideation and Behavior: Report of the International Society for CNS Clinical Trials and Methodology Consensus Meeting. *J Clin Psychiatry*. 2017 Jun;78(6):e638-e647. doi: 10.4088/JCP.16cs11417
22. Simon, R. I. (2011). Improving suicide risk assessment. *Psychiatric Times*, 28(11), 16–21.
23. Chan, M. K. Y., Bhatti, H., et al, (2016). Predicting suicide following self-harm: A systematic review of risk factors and risk scales. *British Journal of Psychiatry*. E-Publication. 10.1192/bjp.bp.115.170050
24. Pokorny, A. D. (1983). Prediction of suicide in psychiatric patients. Report of a prospective study. *Archives of General Psychiatry*, 40(3), 249–257.
25. Hawgood J and De Leo D., *Suicide Prediction – A Shift in Paradigm Is Needed* *Crisis* 2016 37:4, 251-255
26. Nestadt PS, Triplett P, et al., Universal screening may not prevent suicide. *Gen Hosp Psychiatry*. 2020 Mar-Apr;63:14-15. doi: 10.1016/j.genhosppsych.2018.06.006. Epub 2018 Jun 25. PMID: 30072036; PMCID: PMC7112159.
27. Arsenault-Lapierre G, Kim C, Turecki G. Psychiatric diagnoses in 3275 suicides: a meta-analysis. *BMC Psychiatry*. 2004 Nov 4;4:37. doi: 10.1186/1471-244X-4-37.
28. de Beurs D, Ten Have M, et al The longitudinal association between lifetime mental disorders and first onset or recurrent suicide ideation. *BMC Psychiatry*. 2019 Nov 6;19(1):345. doi: 10.1186/s12888-019-2328-8.
29. Powell J., Geddes J, et al, *Suicide in psychiatric hospital in-patients: Risk factors and their predictive power*, *The British Journal of Psychiatry* 2000 176: 266-272, doi:10.1192/bjp.176.3.266
30. Isometsä E. Suicidal behaviour in mood disorders--who, when, and why? *Can J Psychiatry*. 2014 Mar;59(3):120-30. doi: 10.1177/070674371405900303
31. Ratkowska, K. A., Grad, O., et al, (2013). Traumatic bereavement for the therapist: The aftermath of a patient suicide. In D. De LeoA. CimitanK. DyregrovO. GradK. AndriessenEds., *Bereavement after traumatic death: Helping the survivors* (pp.105–114). Göttingen, Germany: Hogrefe.
32. Draper, B. (2012). Isn't it a bit risky to dismiss suicide risk assessment? *Australian and New Zealand Journal of Psychiatry*, 46(4), 385–386. 10.1177/0004867411433222
33. Lee K, Hoti K, et al, Dr Google and the consumer: a qualitative study exploring the navigational needs and online health information-seeking behaviors of consumers with chronic health conditions. *J Med Internet Res*. 2014 Dec 2;16(12):e262. doi: 10.2196/jmir.3706
34. Baumann E, Czerwinski F, Reifegerste D. Gender-specific determinants and patterns of online health information seeking: results from a representative German health survey. *J Med Internet Res* 2017 Dec 04;19(4):e92 doi: 10.2196/jmir.6668
35. Slavich GM, O'Donovan A, et al, Black sheep get the blues: a psychobiological model of social rejection and depression. *Neurosci Biobehav Rev*. 2010 Sep; 35 (1):39–45. doi: 10.1016/j.neubiorev.2010.01.003.
36. Nemeth R., Sik D., et al *Machine Learning of Concepts Hard Even for Humans: The Case of Online Depression Forums* *International Journal of Qualitative Methods* Volume 19: 1–8 DOI: 10.1177/1609406920949338
37. Yan L, Tan Y. Feeling blue? go online: an empirical study of social support among patients. *Inf Syst Res*. 2014Dec; 25 (4):690–709. doi: 10.1287/isre.2014.0538
38. Laukka E, Rantakokko P, Suhonen M. Consumer-led health-related online sources and their impact on consumers: an integrative review of the literature. *Health Informatics J* 2017 Apr 01 doi: 10.1177/1460458217704254
39. Kiritchenko, S., Zhu, X., Mohammad, S. M. 2014, Sentiment analysis of short informal texts. *Journal of Artificial Intelligence Research*, 50, 723–762.
40. Tadesse M., Lin H., et al, *Detection of Suicide Ideation in Social Media Forums Using Deep Learning*, *Algorithms* 2020, 13, 7; doi:10.3390/a13010007
41. Yeskuatov E, Chua SL, Foo LK, *Leveraging Reddit for Suicidal Ideation Detection: A Review of Machine Learning and Natural Language Processing Techniques*, *Int. J. Environ. Res. Public Health* 2022, 19, 10347. <https://doi.org/10.3390/ijerph191610347>
42. Birjali M., Beni-Hassane A., Erritali M., *Machine Learning and Semantic Sentiment Analysis based Algorithms for Suicide Sentiment Prediction in Social Networks*, *Procedia Computer Science* 113 2017, 65-72 doi 10.1016/j.procs.2017.08.290

43. Chancellor, S.; De Choudhury, M. Methods in Predictive Techniques for Mental Health Status on Social Media: A Critical Review. *NPJ Digit. Med.* 2020, 3, 43
44. Beriwal, M.; Agrawal, S. Techniques for Suicidal Ideation Prediction: A Qualitative Systematic Review. In *Proceedings of the 2021 International Conference on INnovations in Intelligent SysTems and Applications (INISTA)*, Kocaeli, Turkey, 25–27
45. Roy, A., Nikolitch, K., McGinn, R. et al. A machine learning approach predicts future risk to suicidal ideation from social media data. *npj Digit. Med.* 3, 78 (2020). <https://doi.org/10.1038/s41746-020-0287-6>
46. Braithwaite, S.R.; Giraud-Carrier, C.; West, J.; Barnes, M.D.; Hanson, C.L. Validating machine learning algorithms for Twitter data against established measures of suicidality. *JMIR Ment. Health* 2016, 3, e21
47. Lachmar M., Wittenborn A., et al #MyDepressionLooksLike: Examining Public Discourse About Depression on Twitter, *JMIR Ment Health.* 2017 Oct-Dec; 4(4): e43, doi:10.2196/mental.8141:10.2196/mental.8141
48. Naranowicz M, Jankowiak K, Kakuba P, Bromberek-Dyzman K, Thierry G. In a Bilingual Mood: Mood Affects Lexico-Semantic Processing Differently in Native and Non-Native Languages. *Brain Sci.* 2022 Feb 26;12(3):316. doi: 10.3390/brainsci12030316.
49. Richter T., Fishbain B., Using machine learning-based analysis for behavioral differentiation between anxiety and depression, 2021, *Scientific Reports*, 10:16381 <https://doi.org/10.1038/s41598-020-72289-9>
50. Kabir MK, Islam M, et al, Detection of Depression Severity Using Bengali Social Media Posts on Mental Health: Study Using Natural Language Processing Techniques. *JMIR Form Res.* 2022 Sep 28;6(9):e36118. doi: 10.2196/36118.
51. Aladağ AE, Muderrisoglu S, et al, Detecting Suicidal Ideation on Forums: Proof-of-Concept Study *J Med Internet Res* 2018;20(6):e215 doi: 10.2196/jmir.9840
52. Kao, A, Poteet RS (Editors), *Natural language processing and text mining*. London: Springer, [http://dx.doi.org/10.1007/978-1-84628-754-1\\_1](http://dx.doi.org/10.1007/978-1-84628-754-1_1).
53. Dhar, A., Mukherjee, H., Dash, N. S., & Roy, K. (2020). Text categorization: past and present. *Artificial Intelligence Review*, <http://dx.doi.org/10.1007/s10462-020-09919-1>.
54. Coppersmith, G.; Ngo, K.; Leary, R.; Wood, A. Exploratory analysis of social media prior to a suicide attempt. in *Proceedings of the Third Workshop on Computational Linguistics and Clinical Psychology*, San Diego, CA, USA, 16 June 2016; pp. 106–117
55. Salton, G., Wong, A., & Yang, C. S. (1975). A vector space model for automatic indexing. *Communications of the ACM*, 18(11), 613–620.
56. <https://www.cs.waikato.ac.nz/ml/weka/>
57. <https://en.wikipedia.org/wiki/Romania>
58. <https://legmed.ro/doc/dds2018.pdf>
59. Merlini D., Rossini M., 2021 Text categorization with WEKA, a survey., *Machine Learning with Applications*, 4, p.100033.
60. Loh WY. Fifty years of classification and regression trees. *Int Stat Rev.*2014; 82(3): 329-348 .doi: <http://dx.doi.org/10.1111/insr.12016>
61. Shea, S. C. (2011). *The practical art of suicide assessment: A guide for mental health professionals and substance abuse counselors*. Hoboken, NJ: Mental Health Presses.
62. Ratkowska, K. A., Grad, O., et al, (2013). Traumatic bereavement for the therapist: The aftermath of a patient suicide. In D. De LeoA. CimitanK. DyregrovO. GradK. AndriessenEds., *Bereavement after traumatic death: Helping the survivors* (pp.105–114). Göttingen, Germany: Hogrefe.
63. Seguin, M., Bordeleau, V, et al. (2014). Professionals' reactions following a patient's suicide: Review and future investigation. *Archives of Suicide Research*, 18, 340–362. Doi: 10.1080/13811118.2013.833151

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.