Time-to-event modeling of subreddits transitions to r/SuicideWatch

Xueying Liu
Computational Biology
St. Jude Children's Research Hospital
Email: Xueying.Liu@stjude.org

Shiaofen Fang
Computer and Information Science
Indiana University - Purdue University
Indianapolis
Email: sfang@iu.edu

George Mohler
Computer Science
Boston College
Boston
Email: mohlerg@bc.edu

Joan Carlson
School of Social Work
Indiana University - Purdue University
Indianapolis
Email: joancarl@iu.edu

Yunyu Xiao
Population Health Sciences
Weill Cornell Medical College
Email: yux4008@med.cornell.edu

***Abstract*—Recent data mining research has focused on the analysis of social media text, content and networks to identify suicide ideation online. However, there has been limited research on the temporal dynamics of users and suicide ideation. In this work, we use time-to-event modeling to identify which subreddits have a higher association with users transitioning to posting on r/suicidewatch. For this purpose we use a Cox proportional hazards model that takes as input text and subreddit network features and outputs a probability distribution for the time until a Reddit user posts on r/suicidewatch. In our analysis we find a number of statistically significant features that predict earlier transitions to r/suicidewatch. While some patterns match existing intuition, for example r/depression is positively associated with posting sooner on r/suicidewatch, others were more surprising (for example, the average time between a high risk post on r/Wishlist and a post on r/suicidewatch is 10.2 days). We then discuss these results as well as directions for future research.**

I. Introduction

In 2019, approximately 47,500 deaths in the U.S. were attributed to suicide by the Center for Disease Control [1]. Given that suicide can be preventable by early intervention, recent data mining research has focused on the analysis of social media text, content and networks to identify suicide ideation and to better understand social media user risk, trajectories, interactions, and potential interventions [2]–[4].

One line of recent research focuses on detecting suicide ideation in online user content on sites such as Twitter and Reddit [5]–[7]. Other research has focused on modeling data from text messages [8] and surveys [9], [10]. While some studies utilized text based features input into classical machine learning models, more recently deep learning has been used to detect suicide ideation in text data [11]–[13]. A comprehensive survey on machine learning for suicide detection can be found in [14], and [15] provides a survey on mining social networks to improve suicide prevention.

Reddit, in particular, has been the focus of recent data mining research on suicide, as several subreddits such as 'r/suicidewatch' provide forums for individuals thinking about suicide, drug addiction, and/or depression and who may be seeking help from others online. In [16], the authors analyzed discourse patterns of posts and comments on four Reddit online communities including r/depression, r/suicidewatch, r/anxiety and r/bipolar. In [17], detection methods were developed for suicide ideation in text on r/suicidewatch and related subreddits and in [18], the authors showed how to improve detection on r/suicidewatch by combining graph and language models. Other work has focused on determining the impact of the COVID-19 pandemic on suicide ideation on Reddit [19], creating an automated question answering system for suicide risk assesment using posts and comments extracted from r/suicidewatch [20], and predicting the degree of suicide risk on r/suicidewatch and related subreddits [21].

While a great deal of work has focused on detecting suicide ideation in online posts, there has been limited research on the temporal dynamics of users and suicide ideation. For example, a user who has suicidal thoughts may post on social media, at which point another may be able to intervene and provide mental health support. However, it is possible that earlier posts may have contained early indicators that could also have been points for interventions. In this work our goal is to better understand these earlier events through time-to-event survival analysis of transitions from other subreddit forums to r/suicidewatch.

In Figure 1, we show three example post sequences from Reddit that illustrate the type of dynamics we would like to model in the present paper. The first user posts on r/LongDistance several times, indicating that they feel sad and are having relationship problems due to long distance, the user then posts on r/teenagers a few times expressing their confusion and then later post on r/suicidewatch. Our goal is to identify which subreddits have a higher association with users transitioning to posting on r/suicidewatch, which text based features are associated with such transitions, and the time between posts from other forums and the first post on

This is the author's manuscript of the article published in final edited form as

X. Liu, S. Fang, G. Mohler, J. Carlson and Y. Xiao, "Time-to-event modeling of subreddits transitions to r/SuicideWatch," 2022 21st IEEE International Conference on Machine Learning and Applications (ICMLA), Nassau, Bahamas, 2022, pp. 974-979, doi: 10.1109/ICMLA55696.2022.00163.

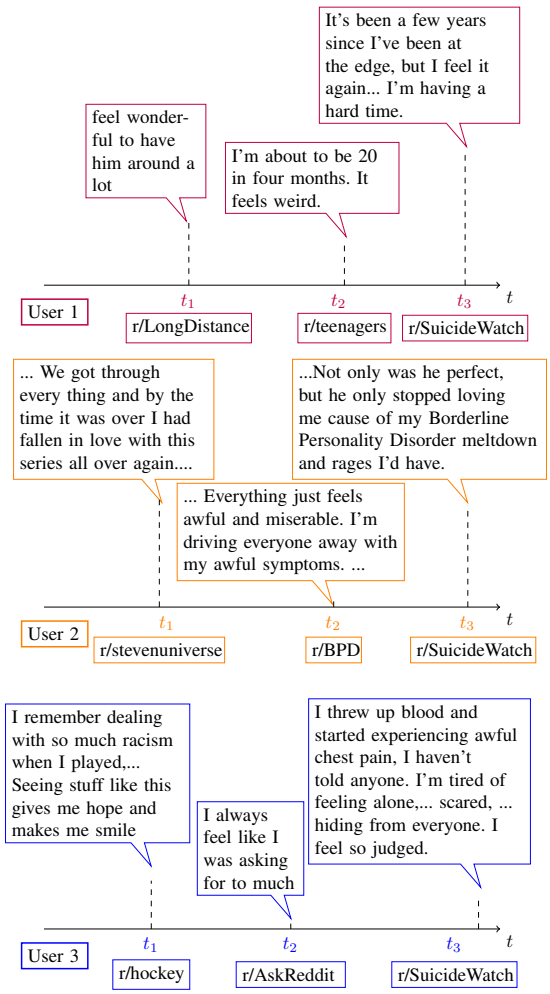

Fig. 1: Posting sequences of 3 Reddit users.

r/suicidewatch. We note that temporal dynamics of suicide ideation on Reddit were considered in [22], however the authors analyzed day of week and hour of day trends in the times of posts, rather than analyzing the inter-event time dynamics of transitions to r/suicidewatch.

The outline of the paper is as follows. In Section II, we describe our survival analysis approach, using Cox proportional hazards modeling. In Section III, we provide details on the data we collected from Reddit (r/suicidewatch and connected subreddits). In Section IV, we present our results of time-to-event modeling of transitions to r/suicidewatch, including the important features that indicate transitions. In Section V, we discuss our results and directions for future work.

II. Model

Survival analysis [23] is a statistical method for analyzing the expected duration until an event occurs. The survival function $S(t)$, defined as $S(t) = P(T \geq t)$, gives the probability that the time to the event occurs later than an observed time t. The cumulative distribution function (CDF) of the time to event gives the cumulative probability for a given t:

$$F(t) = P(T < t) = 1 - S(t)$$

The hazard function $h(t)$ is defined as the probability that an event will occur in the time interval $[t, t + \Delta t)$ given that the event has not occurred before:

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t} = \frac{f(t)}{S(t)},$$

where $f(t)$ is the probability density function (PDF) of the time to event.

One feature of survival analysis is censoring of event times, as some users observation windows may not be large enough to have fully observed an event outcome. If for a given user an event of interest has occurred, then the survival time is known (fully observed), whereas for those that the events has not (yet) occurred, we only know that the waiting time exceeds the observation time [24]. These events with unknown survival time are referred to as censored data. In this study we restrict our analysis to users who post or comment on r/suicidewatch at least once.

The Cox proportional hazards model [25] is a standard Survival model that allows for the incorporation of covariates. The idea behind the Cox model is that the log-hazard of an individual is a linear function of a covariate vector x and parameter vector β and a population-level baseline hazard $h_0(t)$ that changes over time. The Cox's proportional hazard model has the form,

$$h(t|\mathbf{x}) = h_0(t)\exp[g(\mathbf{x})], \quad g(\mathbf{x}) = \beta^T \mathbf{x}, \tag{1}$$

and has been used previously to model transitions to drug addiction and recovery on Reddit [26].

The Cox model in Equation 1 is fit to data in two steps [27]. First, the exponential part is fitted by maximizing the Cox partial likelihood (Equation 2), which does not depend on the baseline hazard, then the baseline hazard $h_0(t)$ is estimated using Breslow's method.

For observation i, let T_i denote the censored event time and R_i denote the set of all observations at risk at time T_i. The Cox partial likelihood is defined as

$$L_{cox} = \prod_i \left(\frac{\exp[g(\mathbf{x_i})]}{\sum_{j \in R_i} \exp[g(\mathbf{x_j})]} \right)^{D_i}, \tag{2}$$

and the negative partial log-likelihood, which can be used as a loss function, is

$$\text{ll}_{cox} = \sum_i D_i \log \left(\sum_{j \in R_i} \exp[g(\mathbf{x_j}) - g(\mathbf{x_i})] \right). \tag{3}$$

Let

$$S_x(t) = S(t|X) = P(T > t|X) \tag{4}$$

be the survival probability at time t, then the baseline probability is defined as follows:

$$S_0(t) = e^{-\int_0^t h_0(t')dt'} = e^{-H_0(t)}. \tag{5}$$

For an individual with features X,

$$S_x(t) = e^{-\int_0^t h(t'|x)dt'} = [S_0(t)]^{exp(\beta^T X)}. \quad (6)$$

Let $\hat{\beta}$ be the value of β that optimizes (2) and (3). Then the cumulative baseline hazard function can be estimated by the Breslow estimator [28]:

$$\widehat{H_0(t)} = \sum_{i=1}^{n} \frac{D_i}{\sum_{j \in R_i} \exp[g(\mathbf{x_j})]}. \quad (7)$$

Note here D_i is an indicator variable that event i is uncensored, and equals 1 in our model for all i.

We use the lifelines CoxPHFitter[1] in Python to fit the Cox model and estimate coefficients β and baseline hazard.

III. Data

The data is collected from Reddit[2], using PushShift[3] and PRAW[4] APIs. We first obtain a list of users who posted on r/suicidewatch between 1/1/2019 and 12/31/2021. We then randomly sample 2000 users and download their posts over the 3-year period, along with comments from these users that posted on r/suicidewatch and their comments and posts from other subreddits.

After dealing with exceptions on PRAW API and removing deleted posts, we collected more than 163k posts from over 1k users. We retained information including user name, post time, post content, post title, and on which subreddit the post was made. We then filtered out users with only one post, and posts that occurred after the user already posted on r/suicidewatch. We further cleaned the data by removing special tokens, detecting and translating posts in foreign languages into English, and performing spell check and making corrections. The data we use for analysis throughout this paper contains over 61k posts from 751 users.

We cut the data at 2020/12/31 23:59:59, and assign posts and comments prior to this time as training data, the rest being the test one. For the users who have posted by the cutoff time but post on r/SuicideWatch afterwards, their corresponding posts in the training data are labeled as `censored`. There are 129 censored users and 4398 posts.

A. Suicide ideation detection model

For each post, we estimated a probability score as to whether a post contained language associated with suicide ideation using a pre-trained model [29]. The model is trained on text data collected from r/suicidewatch and r/depression and utilizes a LSTM neural network based on text embeddings with 'suicidal' vs. 'non-suicidal' binary labels. We use this model to estimate a score between 0 and 1 that represents the probability that a post in our dataset is associated with suicide. We then define posts to be 'high risk' if the score is higher than 0.95, and low otherwise.

[1]https://lifelines.readthedocs.io/en/latest/fitters/regression/CoxPHFitter.html
[2]https://www.reddit.com
[3]https://github.com/pushshift/api
[4]https://praw.readthedocs.io/en/stable/

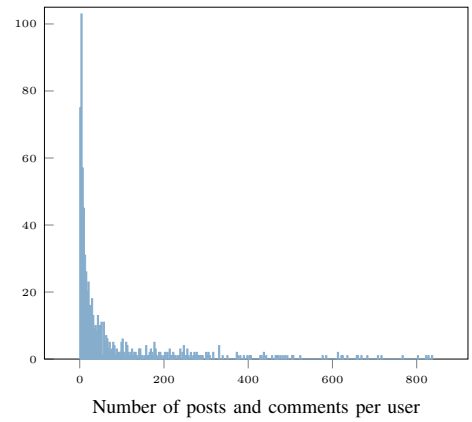

Fig. 2: Histogram of users' posts and comments.

B. Summary statistics and figures

In Table I, we provide average suicidal scores of some most frequently posted subreddits, where on r/suicidewatch, the score is noticeably higher. In Table IIa we provide a summary of the number of posts of each user, where the average number of posts is 53.4 per user. In Table IIb we provide a summary of the length of posts with high and low suicidal scores, where we find that high risk scores are associated with longer posts. We also find that posts with high suicidal scores are more likely to occur on weekends (Friday and Saturday). Figure 2 shows a histogram of users' posts, where 10% of users have only 2 posts and 76% of users have less than 50 posts each. Figure 3 shows the distribution of inter-event times between posts on other sub-reddits and r/suicidewatch. More than 3000 of the posts were made within 3.5 days of posting on r/suicidewatch. The longest waiting time before posint on r/suicidewatch is 698 days.

Subreddit	Average score	Subreddit	Average score
AskReddit	0.1730	dankmemes	0.1631
teenagers	0.1455	AmItheAsshole	0.2027
SuicideWatch	**0.8036**	trees	0.0958
memes	0.1485	FortNiteBR	0.0963
depression	0.6200	selfharm	0.4478
AskOuija	0.1416	awakened	0.4240
relationship_advice	0.3657	Advice	0.4374
unpopularopinion	0.1425	NoFap	0.2781

TABLE I: Average suicidal score of posts on popular subreddits.

Max	838
Min	2
Mean	81.34

(a) Number of posts and comments per user.

	High	Low
Mean length	539.18	164.38
% on weekend	27.69	27.94

(b) Posts and comments grouped by suicidal scores.

TABLE II: Summary of data 1

C. Topic models

We analyze the topic models of text data by first utilizing SentenceTransformers [30] - a BERT-based pretrained model

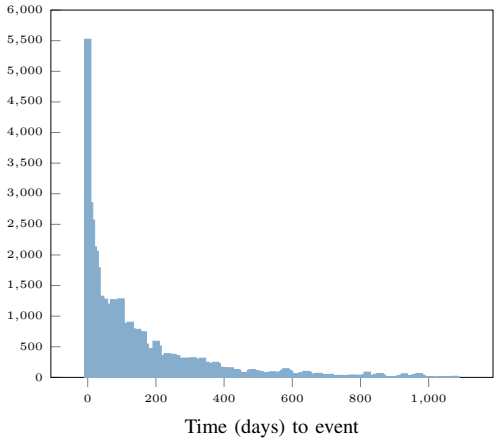

Fig. 3: Histogram of time to event.

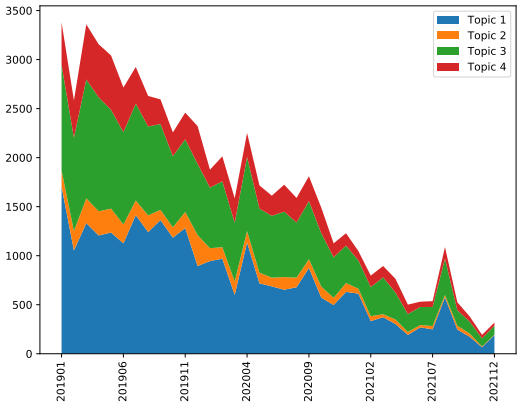

Fig. 4: Popularity of each topic over time.

that derives semantically meaningful sentence embeddings. We then perform KMeans with the embeddings that associate with high suicidal score (i.e. > 0.95). We search for the optimal K using the "elbow" method, which suggests $K = 4$. Moreover, we extract keywords of each topic with the help of spaCy library in Python. The keywords are displayed in Table III.

Topic	Keywords
Topic 1	right, tear, know
Topic 2	think, sending, affected, died, unjustified, death, help, threaten, pills, life, traumatising, emotionally,...
Topic 3	tried, turning, fix, way, find
Topic 4	longer, like, want, future, realized, world, care, depression, time,method, meant, planned, fought, struggled, torture, hope,...

TABLE III: Keywords extracted from posts with high suicidal score.

Figure 4 demonstrates the popularity of each topic over the 3-year observed timeframe.

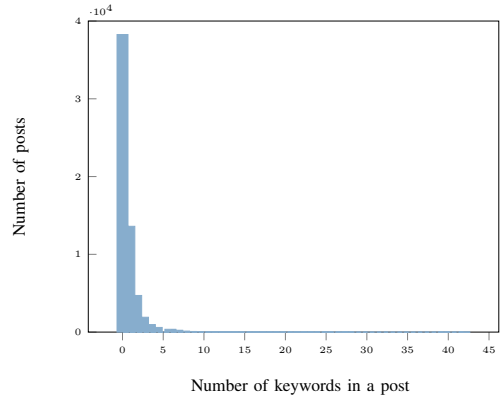

Fig. 5: Histogram of number of keywords.

D. Feature Selection

1) Keyword expansion: We use keyword expansion to determine a list of keywords related to suicide. Starting with a manually selected list of 39 keywords, we then use cosine similarity of word vectors to find the 10 most similar words to each. We use the word2vec implementation in gensim [31] to create the 100-dimensional word vector representations. This process results in a keyword list of length 331. We next create dummy variables that indicate if a post contains each of these keywords. Figure 5 shows the histogram of number of keywords present in a post or comment. 40 most frequently occurred keywords are displayed in Table IV.

word	count	word	count	word	count
like	8823	soon	565	abuse	329
good	3462	pain	550	depressed	326
way	2708	relationship	534	therapy	306
life	2220	damn	527	red	251
thanks	1899	check	457	cry	238
work	1657	anxiety	455	therapist	228
friends	1254	health	448	account	224
friend	985	kid	423	upset	198
idea	799	important	381	weed	186
place	778	death	372	ending	182
god	696	eat	356	toxic	174
mental	690	type	354	emotional	169
women	589	bring	336	party	164
using	569				

TABLE IV: Top 40 most frequent keywords.

2) Sources connecting to subreddit r/suicidewatch: We select the top 50 frequent subreddits (excluding r/suicidewatch) and create dummy variables that indicate if a post is from one of these subreddits. In Figure 6, we show the most frequently posted 15 subreddits. On average, the data contains 13.75 posts and comments from each subreddit.

3) Index of topics: The topic indices obtained from part III-C are transferred into dummy variables.

IV. Results

We fit a Cox proportional hazards model to our data, with a penalizer term of 5, and within the summary table of results, we focus on the variables with a p-value less than

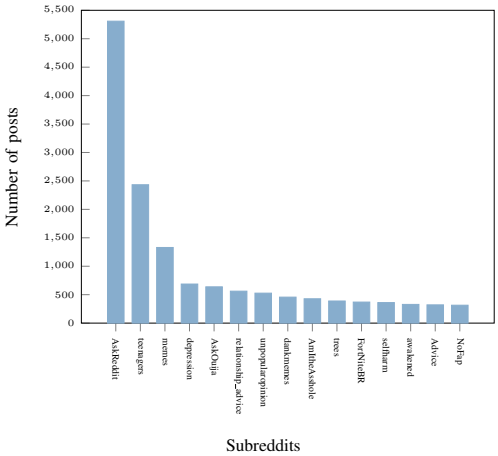

Fig. 6: 15 mostly posted subreddits.

Subreddit indicators			
depression	0.06	LivestramFail	-0.28
teenagers	0.05	fireemblem	-0.08
relationship_advice	0.05	MortalKombat	-0.12
awakened	0.15	AvPD	-0.37
selfharm	0.09	Sweden	-0.23
MadeOfStyrofoam	0.11	Traaa...nnnns	-0.1
Wishlist	0.17	ftm	-0.19
BPD	0.08	PurplePillDebate	-0.29
Eminem	0.14		
cats	0.1		
unpopularopinion	0.06		
BreakUps	0.12		
Keyword indicators		**Suicidal score**	
"pain"	0.04	score	0.04
"women"	-0.07		
"life"	0.02		

TABLE V: Coefficients of significant variables.

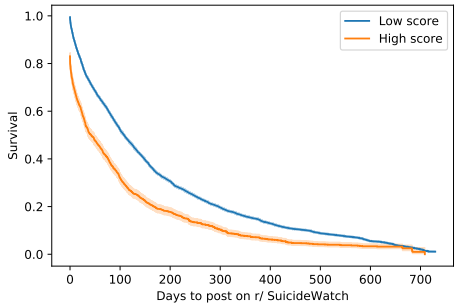

Fig. 7: Kalpan Meier estimates by suicidal score

0.05. In Table V, we show the coefficients of these statistically significant variables. The subreddits with the highest coefficients (indicating sooner transition to r/suicidewatch) include r/selfharm, r/Wishlist, r/awakened, r/BreakUps and r/MadeOfStyrofoam (which is a forum for selfharm discussion). Subreddits associated with longer time-to-event intervals between an initial post and a subsequent post on r/suicidewatch include r/LivestreamFail, r/AvPD, r/ftm and r/PurplePillDebate (a forum to discuss sex and gender issues). While a majority of the keywords were not statistically significant, both 'pain' and 'life' were associated with shorter time-to-event periods, as was the high risk category based on the suicide detection model described above. The keyword 'women' appeared to be associated with a longer time-to-event interval. Topic feature is not statistically significant.

In Figure 7, we display the estimated Kaplan Meier curve for the distribution of time-to-event disaggregated by high vs. low suicidal scores. Here we find that the time-to-event distribution has a shorter tail for higher risk scores, indicating that posts with high risk scores are associated with subsequent r/suicidewatch posts occuring sooner. In Figure 8, we display the transition network from other subreddits (yellow indicating a positive association, blue indicating a negative association) to r/suicidewatch along with the average transition time between the final post preceding a post on r/suicidewatch and their first post on r/suicidewatch. On each edge, the number represents the average number of days between subreddit and r/suicidewatch posts when the suicidal score is high (low). No post is found from r/cats, r/MortalKomat, r/sweden and r/Eminem with a high suicidal score.

We predict expected remaining lifetime of each censored user and compute the concordance between prediction and ground truth, the model obtains a concordance index of 0.5123. We also compute AUC by dividing the entire future into 30-day interval. We label an interval 1 if transition to r/SuicideWatch occurred in the interval. This gives us an AUC of 0.8214.

V. Discussion

In this research, we collected a large corpus of suicide related posts from r/suicidewatch, along with earlier posts made by users on other subreddits. We then fit a Cox proportional hazards model to predict the time-to-event between earlier posts and later posts on r/suicidewatch. We found statistically significant features using indicators for subreddit, keyword, or suicide risk. While some patterns match existing intuition, for example r/teenagers and r/relationship_advice are positively associated with posting sooner on r/suicidewatch, others were more surprising. For example, the average time between a high risk post on r/Wishlist and a consecutively following post on r/suicidewatch is 10.2 days (less than the 97.2 day average time between events on r/depression and r/suicidewatch). Our results indicate potential points of earlier intervention and analysis of associated subreddits to suicide may yield new hypotheses for suicide researchers to investigate.

Future research may improve upon our work in several ways. While we used a Cox proportional hazards model, a deep learning based survival model [32] may yield improvements to accuracy. Also, we did not explore the social networks of individuals and how interactions on the network may be predictive of transitions to suicide ideation. In future work we hope to analyze posting behavior of network connections and people whom a given person interacts with, and determine how those interactions may act to protect against or lead to higher risk of suicide ideation observed online.

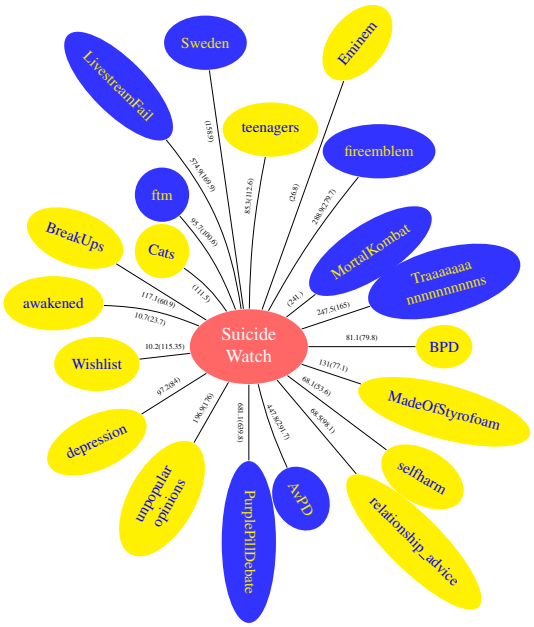

Fig. 8: Average days from the most recent subreddit transitions to r/ SuicideWatch with a high (low) suicidal score

ACKNOWLEDGMENTS

This research was supported in part by AFOSR MURI grant FA9550-22-1-0380 and a seed grant from the IUPUI Institute of Integrative AI.

REFERENCES

[1] "https://www.cdc.gov/mmwr/volumes/70/wr/mm7008a1.htm."

[2] M. A. Lindsey, A. H. Sheftall, Y. Xiao, and S. Joe, "Trends of Suicidal Behaviors Among High School Students in the United States: 1991–2017," Pediatrics, vol. 144, 11 2019. e20191187.

[3] H. Li, Y. Han, Y. Xiao, X. Liu, A. Li, and T. Zhu, "Suicidal ideation risk and socio-cultural factors in china: A longitudinal study on social media from 2010 to 2018," International Journal of Environmental Research and Public Health, vol. 18, no. 3, 2021.

[4] Y. Xiao, J. Cerel, and J. J. Mann, "Temporal Trends in Suicidal Ideation and Attempts Among US Adolescents by Sex and Race/Ethnicity, 1991-2019," JAMA Network Open, vol. 4, pp. e2113513–e2113513, 06 2021.

[5] A. Mbarek, S. Jamoussi, A. Charfi, and A. B. Hamadou, "Suicidal profiles detection in twitter," in 15th International Conference on Web Information Systems and Technologies, January 2019.

[6] B. O'Dea, S. Wan, P. J. Batterham, A. L. Calear, C. Paris, and H. Christensen, "Detecting suicidality on twitter," Internet Interventions, vol. 2, no. 2, pp. 183–188, 2015.

[7] S. Ji, C. P. Yu, S. fu Fung, S. Pan, and G. Long, "Supervised learning for suicidal ideation detection in online user content," Complexity, vol. 2018, pp. 1–10, 2018.

[8] A. Nobles, J. J. Glenn, K. Kowsari, B. Teachman, and L. E. Barnes, "Identification of imminent suicide risk among young adults using text messages," Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, 2018.

[9] A. M. May, E. K. Czyz, and B. T. West, "Differentiating adolescent suicide attempters and ideators: A classification tree analysis of risk behaviors," Journal of Adolescent Health, vol. 67, no. 6, pp. 837–850, 2020.

[10] Y. Xiao, M. Romanelli, and M. Lindsey, "A latent class analysis of health lifestyles and suicidal behaviors among us adolescents," Journal of Affective Disorders, vol. 255, pp. 116–126, Aug. 2019. Publisher Copyright: © 2019 Elsevier B.V.

[11] Y. Ophir, R. Tikochinski, C. S. C. Asterhan, I. Sisso, and R. Reichart, "Deep neural networks detect suicide risk from textual facebook posts," Scientific Reports, no. 1, p. 16685, 2020.

[12] M. M. Tadesse, H. Lin, B. Xu, and L. Yang, "Detection of suicide ideation in social media forums using deep learning," Algorithms, vol. 13, no. 1, 2020.

[13] M. Matero, A. Idnani, Y. Son, S. Giorgi, H. Vu, M. Zamani, P. Limbachiya, S. C. Guntuku, and H. A. Schwartz, "Suicide risk assessment with multi-level dual-context language and BERT," in Proceedings of the Sixth Workshop on Computational Linguistics and Clinical Psychology, (Minneapolis, Minnesota), pp. 39–44, Association for Computational Linguistics, June 2019.

[14] S. Ji, S. Pan, X. Li, E. Cambria, G. Long, and Z. Huang, "Suicidal ideation detection: A review of machine learning methods and applications," IEEE Transactions on Computational Social Systems, vol. 8, no. 1, pp. 214–226, 2021.

[15] J. Lopez-Castroman, B. Moulahi, J. Azé, S. Bringay, J. Deninotti, S. Guillaume, and E. Baca-Garcia, "Mining social networks to improve suicide prevention: A scoping review," Journal of neuroscience research, vol. 98, p. 616—625, April 2020.

[16] B. Silveira Fraga, A. P. Couto da Silva, and F. Murai, "Online social networks in health care: A study of mental disorders on reddit," in 2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI), pp. 568–573, 2018.

[17] A. E. Aladağ, S. Muderrisoglu, N. B. Akbas, O. Zahmacioglu, and H. O. Bingol, "Detecting suicidal ideation on forums: Proof-of-concept study," J Med Internet Res, vol. 20, p. e215, Jun 2018.

[18] A. Ruch, "Can x2vec save lives? integrating graph and language embeddings for automatic mental health classification," Journal of Physics: Complexity, no. 3, 2020.

[19] D. M. Low, L. Rumker, T. Talkar, J. Torous, G. Cecchi, and S. S. Ghosh, "Natural language processing reveals vulnerable mental health support groups and heightened health anxiety on reddit during covid-19: Observational study," J Med Internet Res, vol. 22, p. e22635, Oct 2020.

[20] A. Alambo, M. Gaur, U. Lokala, U. Kursuncu, K. Thirunarayan, A. Gyrard, A. Sheth, R. S. Welton, and J. Pathak, "Question answering for suicide risk assessment using reddit," in 2019 IEEE 13th International Conference on Semantic Computing (ICSC), pp. 468–473, 2019.

[21] A. Zirikly, P. Resnik, Ö. Uzuner, and K. Hollingshead, "CLPsych 2019 shared task: Predicting the degree of suicide risk in Reddit posts," in Proceedings of the Sixth Workshop on Computational Linguistics and Clinical Psychology, (Minneapolis, Minnesota), pp. 24–33, Association for Computational Linguistics, June 2019.

[22] R. Dutta, G. Gkotsis, S. Velupillai, I. Bakolis, and R. Stewart, "Temporal and diurnal variation in social media posts to a suicide support forum," BMC Psychiatry, no. 1, p. 259, 2021.

[23] J. Klein and M. Moeschberger, Survival analysis: techniques for censored and truncated data. Springer Science & Business Media, 2006.

[24] G. Rodríguez, "Survival models," in Lecture Notes on Generalized Linear Models, 2007.

[25] D. R. Cox, "Regression models and life-tables," Journal of the Royal Statistical Society. Series B (Methodological), vol. 34, no. 2, pp. 187–220, 1972.

[26] J. Lu, S. Sridhar, R. Pandey, M. A. Hasan, and G. Mohler, "Investigate transitions into drug addiction through text mining of reddit data," in Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 2367–2375, 2019.

[27] H. Kvamme, Ø. Borgan, and I. Scheel, "Time-to-event prediction with neural networks and cox regression," arXiv preprint arXiv:1907.00825, 2019.

[28] "Discussion on professor cox's paper," Journal of the Royal Statistical Society: Series B (Methodological), vol. 34, no. 2, pp. 202–220, 1972.

[29] A. Singh, "Suicidal thought detection." https://www.kaggle.com/abhijitsingh001/suicidal-thought-detection/notebook.

[30] N. Reimers and I. Gurevych, "Sentence-bert: Sentence embeddings using siamese bert-networks," in Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, 11 2019.

[31] R. Řehůřek and P. Sojka, "Software Framework for Topic Modelling with Large Corpora," in Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks, (Valletta, Malta), pp. 45–50, ELRA, May 2010.

[32] Y. Zhao, Q. Hong, X. Zhang, Y. Deng, Y. Wang, and L. Petzold, "Bertsurv: Bert-based survival models for predicting outcomes of trauma patients," arXiv preprint arXiv:2103.10928, 2021.