Leveraging Reddit for Suicidal Ideation Detection

Suicide is a major public-health problem that exists in virtually every part of the world. Hundreds of thousands of people commit suicide every year. The early detection of suicidal ideation is critical for suicide prevention. However, there are challenges associated with conventional suicide-risk screening methods. At the same time, individuals contemplating suicide are increasingly turning to social media and online forums, such as Reddit, to express their feelings and share their struggles with suicidal thoughts.

1. Introduction

Suicide is a global public-health problem. According to the World Health Organization, approximately 703,000 people commit suicide every year [1]. It is the world's fourth leading cause of death among young people aged 15 to 29 years old. Moreover, it is estimated that there are more than 20 attempts for every completed suicide [2].

The causes of suicide are largely complicated and result from the interaction of multiple factors that can be grouped into three categories: health factors, environmental factors, and factors related to personal history, such as childhood abuse or previous suicide attempts [3][4]. Other examples of suicide risk factors include mental disorder, physical illness, substance abuse, domestic violence, bullying, relationship problems, and other stressful life events. Due to the complexity of the problem, no single risk factor can be reliably used to predict suicide [5]. For instance, despite the strong association between suicide and depression, a depression diagnosis alone has a limited ability to predict suicide. More recently, the issue of suicide has been further exacerbated by the impact of the COVID-19 pandemic [6]. In particular, social isolation—which resulted from measures imposed to curb the spread of the virus—was linked to increased suicide risk.

People with suicide risk fall into two classes: ideators and attempters [7]. Suicidal ideation is a broad term that describes thoughts and behaviors ranging from being preoccupied with death to planning a suicide attempt [8]. The suicidal ideations can be passive and active. Passive suicidal ideation involves thinking about suicide and wishing to be dead, whereas active suicidal ideation implies intending and planning an attempt to take one's own life [8]. While it is believed that passive suicidal ideation poses a lower risk, both types need to be carefully assessed by mental health professionals, since passive suicidal ideation can rapidly transform into the active form [9]. This can happen when a person's circumstances or health condition worsen.

The early detection of suicidal ideation expressed by an at-risk individual is key to effective prevention, as it facilitates timely intervention by mental health professionals [10]. However, there are several challenges associated with suicide prevention. They include (1) social stigma, (2) limited access to professional help, and (3) inadequate training of clinicians in dealing with suicidal patients [11]. The combination of these factors creates a new challenge—(4) fragmented professional care, which entails having large time gaps between mental health assessments [11].

At the same time, an increasing number of at-risk individuals are turning to online communication channels to express their feelings and discuss their suicidal thoughts [12][13][14]. This tendency prompted research that focuses on detecting suicide risk and other mental health issues on social networks and online forums by applying machine learning (ML) and natural language processing (NLP) techniques [10][13][15]. The quantifiable signals in user-generated online data aid researchers in gaining insight into an individual's emotional state and detecting cues indicative of suicidality [16][17]. The feasibility of such an approach has been demonstrated by numerous studies on different mental health disorders. For examples, a study [18] used the textual data from Facebook posts of consenting study participants to predict depression diagnoses recorded in their electronic medical records with high accuracy, using a logistic regression model. In the study of [19], using pre-trained machine learning models, the researchers detected negative changes in Twitter users' sentiment, stress, anxiety, and loneliness measures after the declaration of emergency in the US due to the COVID-19 pandemic.

2. Leveraging Reddit for Suicidal Ideation Detection

2.1. Detection of Suicidal Ideation on Social Media

The social stigma related to having suicidal ideations has a particularly significant effect. The fear of social stigma has been shown to discourage individuals at risk of suicide from discussing their experiences in person and seeking support [20][21][22][23]. Further, it obstructs the extant suicide-risk screening methods, such as questionnaires and interviews, since they require patients to explicitly disclose their intentions to commit suicide [24]. According to a meta-analysis of 71 studies, on average, nearly 80% of people in non-psychiatric settings—primary healthcare patients, general population, military personnel, and incarcerated individuals—who died by suicide did not reveal their suicidal intentions when they were surveyed before their suicide attempt [25]. Thus, there is a need for novel suicidality detection methods that do not require face-to-face interactions [21]. In this case, detecting suicidal ideations on online platforms can be more effective since the anonymity of social media and forums enables people to openly share their struggles with suicidal thoughts without fear of judgment [11][16][26][27].

Although the Columbia-Suicide Severity Rating Scale (C-SSRS) has been widely used as a screening instrument, the administration of C-SSRS may place a burden on health-care providers [28]. Therefore, another motivation for detecting suicidal ideations on online platforms is to reduce the load on the health-care system. The goal is to create a tool that would automatically and instantaneously detect if a user is exhibiting any signs of suicidality based on their online activity before engagement with providers. Ideally, these screening tools should be highly scalable and adaptable so that they can be used with a variety of data sources and be readily integrated into existing health-care IT systems [10][28]. The adoption of such suicidal ideation detection tools can assist mental health professionals and even those without specialized training (e.g., primary-care physicians and social workers) in quickly identifying individuals at risk and making informed decisions [23].

Studying the online activity for suicidal ideation detection can also help address the challenges of fragmented care for existing patients [29]. Given that about 70% of psychiatric patients are active on social media, mental health professionals can monitor their online activity to obtain information relevant to patients' mental state during gaps in patient–clinician interactions [11]. In this scenario, suicidality detection tools can be employed to automatically detect signals of deteriorating mental condition and alert health-care providers, prompting them to attend to a patient under their care [28].

2.2. Reddit as a Source for Suicidal Ideation Detection

Reddit has generated particular interest among researchers due to its distinctive characteristics. Reddit is a popular online forum, covering a wide range of topics, with subcommunities called subreddits [30]. Currently, there are over 13 billion posts and comments distributed across more than 100,000 active communities [31]. More than 50 million active unique users interact with the platform in a single day. Researchers choose Reddit over other platforms as the source of data for several reasons.

Reddit posts have a higher character limit of 40,000 characters compared to Twitter, which only allows 280 characters [22]. It gives users more space to express their suicidal thoughts and describe their emotional state in more detail. The large posts provide a better insight into the author's mental state [23]. By analyzing long passages of text, the researchers capture and extract textual features that sufficiently indicate suicidal ideations [10][24].

Reddit facilitates better anonymity [22][23]. As per Reddit's privacy policy, users are not required to provide any identifying personal information or email address when creating an account [32]. The platform only requires a username and a password and the former does not have to relate to an actual name. This is unlike other social media sites. For instance, Facebook requires either a phone number or an email address during sign up, in addition to implementing a real-name policy that necessitates users to specify their real names on profiles [33]. Reddit users normally do not include their names and choose non-identifying ambiguous usernames. This level of anonymity allows people at risk of suicide to express themselves in an uninhibited fashion, without fear of social stigma [10][23][24]. This is valuable for researchers since unconstrained expounding of one's experiences and feelings builds a genuine picture of the user's psychological state.

Reddit has numerous specialized support forums dedicated to various mental health topics [23]. For example, the r/SuicideWatch subreddit is a subcommunity of 366,000 members where people share their suicidal thoughts, seek help, and provide support to others dealing with suicidal ideations [34][35]. This subreddit is extensively used by researchers as a source of suicidal posts to serve as positive samples in their datasets [10]. What further supports the validity of r/SuicideWatch as a source of genuine suicide-related posts is that this subreddit is monitored by moderators [22]. The moderators remove any irrelevant posts and posts that violate the community rules, e.g., abuse, criticism, and spam [35].

2.3. Machine Learning Approach for Suicidal Ideation Detection

2.3.1. Data Collection

The first step in the process of building a classifier is obtaining a dataset containing sufficient posts for each class label. Having an accurate dataset with labeled examples is critical for the success of the ML model. The dataset is used to train and then test the model. The model's predictive performance and its generalizability strongly depend on the quality and amount of training data. There are two broad data collection approaches adopted by the studies: collecting data directly from Reddit and using datasets created by other researchers.

2.3.2. Data Annotation

Supervised ML algorithms require annotated datasets. During the training stage, the algorithm generates a function that maps the relationship between the features and the target variables. To train the model to detect posts with suicidal ideations, the researchers need examples of posts annotated as suicidal and not suicidal. For the multiclass classification problem, posts with annotations for different suicide risk levels are required.

2.3.3. Data Preprocessing

The data collected from Reddit consist of raw, unstructured text and contain noise that can negatively impact the predictive performance of the model. The noise includes punctuation, special characters, URLs, emails, etc. The raw text needs to be converted into a numerical representation before it can be fed into a classifier. During the preprocessing stage, the input data are cleaned and standardized. Therefore, it is an important step that lays the foundation for feature extraction and classification.

2.3.4. Feature Engineering

To use ML algorithms, researchers need to extract features from the data. These features then serve as an input to a classifier algorithm. Therefore, the quality of extracted features is one of the factors that significantly affects the predictive performance of the model. Most studies combined techniques to extract different types of features. The researchers primarily focused on extracting features from the textual content of posts. However, several studies also considered statistical metadata, such as the number of posts per user, the frequency of posting, and the number of votes [13][23].

2.3.5. Model Development

All the studies in the corpus frame their contributions as building a predictive model that detects suicidal ideations from Reddit data. They tested multiple algorithms with different sets of features and proposed best-performing models. In total, 21 supervised ML algorithms were explored by the researchers. Most studies (18 out of 26 studies) included deep learning techniques. The researchers chose deep learning because, when used in conjunction with word embeddings, the deep-learning-based models can effectively detect suicidal ideations without the need for feature engineering.

2.3.6. Model Validation

Once the predictive model is trained, the performance of the model is evaluated. The most common evaluation metrics include accuracy, precision, recall, and F1-score. However, two studies also calculated the area under the curve (AUC) metric [11][23].

For suicidality detection task, true positive (TP) represents the number of posts that were correctly classified as suicidal. True negative (TN) represents the number of posts that were correctly classified as non-suicidal. False positive (FP), also known as Type I error, represents the number of non-suicidal posts that were misclassified as suicidal. False negative (FN), also known as Type II error, represents the number of suicidal posts that were misclassified as non-suicidal.

Accuracy measures the overall portion of correct predictions [36]. It is a ratio of all correctly classified posts to the total number of posts:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{1}$$

Precision is a ratio of correctly classified suicidal posts to the total number of posts classified as suicidal (both correctly and incorrectly) [36]:

$$Precision = \frac{TP}{TP + FP} \qquad (2)$$

Recall, also called sensitivity or true-positive rate, is the ratio of correctly classified suicidal posts to the total number of suicidal posts, including both correctly classified posts and posts that should have been classified as suicidal [36]:

$$Recall = \frac{TP}{TP + FN} \qquad (3)$$

This metric is especially useful for selecting the best model where there is a high cost of false-negative predictions [37]. In the suicidal ideation detection model, false positives are more tolerable than false negatives [38]. In other words, it is better to raise a false alarm by incorrectly predicting someone as suicidal than to miss someone who is indeed at risk of suicide.

*F*1-score is the harmonic mean of precision and recall:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \qquad (4)$$

For multiclass classification problems, the macro-averaged *F*1-score can be determined by calculating individual *F*1-scores for each class and finding their unweighted mean.

The receiver operating characteristic curve is a graph that plots the true-positive rate (Equation (3)) against the false-positive rate (Equation (5)) at different classification thresholds [39]. It provides a graphical representation of the classifier's performance and a larger area under the curve indicates better performance.

$$False\ Positive\ Rate = \frac{FP}{FP + TN} \qquad (5)$$

References

1. World Health Organization. Suicide Worldwide in 2019: Global Health Estimates; World Health Organization: Geneva, Switzerland, 2021.

2. World Health Organization. Preventing Suicide: A Global Imperative; World Health Organization: Geneva, Switzerland, 2014.

3. O'Connor, R.C.; Nock, M.K. The Psychology of Suicidal Behaviour. Lancet Psychiatry 2014, 1, 73–85.

4. Risk Factors, Protective Factors, and Warning Signs. American Foundation for Suicide Prevention. Available online: https://afsp.org/risk-factors-protective-factors-and-warning-signs/ (accessed on 21 July 2022).

5. Franklin, J.C.; Ribeiro, J.D.; Fox, K.R.; Bentley, K.H.; Kleiman, E.M.; Huang, X.; Musacchio, K.M.; Jaroszewski, A.C.; Chang, B.P.; Nock, M.K. Risk Factors for Suicidal Thoughts and Behaviors: A Meta-Analysis of 50 Years of Research. Psychol. Bull. 2017, 143, 187–232.

6. Castillo-Sánchez, G.; Marques, G.; Dorronzoro, E.; Rivera-Romero, O.; Franco-Martín, M.; De la Torre-Díez, I. Suicide Risk Assessment Using Machine Learning and Social Networks: A Scoping Review. J. Med. Syst. 2020, 44, 205.

7. Aladağ, A.E.; Muderrisoglu, S.; Akbas, N.B.; Zahmacioglu, O.; Bingol, H.O. Detecting Suicidal Ideation on Forums: Proof-of-Concept Study. J. Med. Internet Res. 2018, 20, e215.

8. Harmer, B.; Lee, S.; Duong, T.v.H.; Saadabadi, A. Suicidal Ideation. In StatPearls; StatPearls Publishing: Treasure Island, FL, USA, 2022.

9. Simon, R.I. Passive Suicidal Ideation: Still a High-Risk Clinical Scenario. Curr. Psychiatry 2014, 13, 13–15.

10. Ji, S.; Pan, S.; Li, X.; Cambria, E.; Long, G.; Huang, Z. Suicidal Ideation Detection: A Review of Machine Learning Methods and Applications. IEEE Trans. Comput. Soc. Syst. 2021, 8, 214–226.

11. Gaur, M.; Aribandi, V.; Alambo, A.; Kursuncu, U.; Thirunarayan, K.; Beich, J.; Pathak, J.; Sheth, A. Characterization of Time-Variant and Time-Invariant Assessment of Suicidality on Reddit Using C-SSRS. PLoS ONE 2021, 16, e0250448.

12. Grant, R.N.; Kucher, D.; León, A.M.; Gemmell, J.F.; Raicu, D.S.; Fodeh, S.J. Automatic Extraction of Informal Topics from Online Suicidal Ideation. BMC Bioinform. 2018, 19, 211.

13. Ji, S.; Yu, C.P.; Fung, S.; Pan, S.; Long, G. Supervised Learning for Suicidal Ideation Detection in Online User Content. Complexity 2018, 2018, 6157249.

14. Vioules, M.J.; Moulahi, B.; Aze, J.; Bringay, S. Detection of Suicide-Related Posts in Twitter Data Streams. IBM J. Res. Dev. 2018, 62, 7:1–7:12.

15. Matero, M.; Idnani, A.; Son, Y.; Giorgi, S.; Vu, H.; Zamani, M.; Limbachiya, P.; Guntuku, S.C.; Schwartz, H.A. Suicide Risk Assessment with Multi-Level Dual-Context Language and BERT. In Proceedings of the Sixth Workshop on Computational Linguistics and Clinical Psychology, Minneapolis, MN, USA, 6 June 2019; Association for Computational Linguistics: Stroudsburg, PA, USA, 2019; pp. 39–44.

16. Tadesse, M.M.; Lin, H.; Xu, B.; Yang, L. Detection of Suicide Ideation in Social Media Forums Using Deep Learning. Algorithms 2019, 13, 7.

17. Jones, N.; Jaques, N.; Pataranutaporn, P.; Ghandeharioun, A.; Picard, R. Analysis of Online Suicide Risk with Document Embeddings and Latent Dirichlet Allocation. In Proceedings of the 2019 8th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW), Cambridge, UK, 3–6 September 2019; pp. 1–5.

18. Eichstaedt, J.C.; Smith, R.J.; Merchant, R.M.; Ungar, L.H.; Crutchley, P.; Preoţiuc-Pietro, D.; Asch, D.A.; Schwartz, H.A. Facebook Language Predicts Depression in Medical Records. Proc. Natl. Acad. Sci. USA 2018, 115, 11203–11208.

19. Guntuku, S.C.; Sherman, G.; Stokes, D.C.; Agarwal, A.K.; Seltzer, E.; Merchant, R.M.; Ungar, L.H. Tracking Mental Health and Symptom Mentions on Twitter During COVID-19. J. Gen. Intern. Med. 2020, 35, 2798–2800.

20. 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 August 2021; pp. 1–8.

21. Allen, K.; Bagroy, S.; Davis, A.; Krishnamurti, T. ConvSent at CLPsych 2019 Task A: Using Post-Level Sentiment Features for Suicide Risk Prediction on Reddit. In Proceedings of the Sixth Workshop on Computational Linguistics and Clinical Psychology, Minneapolis, MN, USA, 6 June 2019; Association for Computational Linguistics: Stroudsburg, PA, USA, 2019; pp. 182–187.

22. Yao, H.; Rashidian, S.; Dong, X.; Duanmu, H.; Rosenthal, R.N.; Wang, F. Detection of Suicidality Among Opioid Users on Reddit: Machine Learning-Based Approach. J. Med. Internet Res. 2020, 22, e15293.

23. Gaur, M.; Alambo, A.; Sain, J.P.; Kursuncu, U.; Thirunarayan, K.; Kavuluru, R.; Sheth, A.; Welton, R.; Pathak, J. Knowledge-Aware Assessment of Severity of Suicide Risk for Early Intervention. In Proceedings of the The World Wide Web Conference—WWW '19, San Francisco, CA, USA, 13–17 May 2019; pp. 514–525.

24. Alambo, A.; Gaur, M.; Lokala, U.; Kursuncu, U.; Thirunarayan, K.; Gyrard, A.; Sheth, A.; Welton, R.S.; Pathak, J. Question Answering for Suicide Risk Assessment Using Reddit. In Proceedings of the 2019 IEEE 13th International Conference on Semantic Computing (ICSC), Newport Beach, CA, USA, 30 January–1 February 2019; pp. 468–473.

25. McHugh, C.M.; Corderoy, A.; Ryan, C.J.; Hickie, I.B.; Large, M.M. Association between Suicidal Ideation and Suicide: Meta-Analyses of Odds Ratios, Sensitivity, Specificity and Positive Predictive Value. BJPsych Open 2019, 5, e18.

26. Iavarone, B.; Monreale, A. From Depression to Suicidal Discourse on Reddit. In Proceedings of the 2021 IEEE International Conference on Big Data (Big Data), Orlando, FL, USA, 15–18 December 2021; pp. 437–445.

27. Rabani, S.T.; Khan, Q.R.; Khanday, A. A Novel Approach to Predict the Level of Suicidal Ideation on Social Networks Using Machine and Ensemble Learning. ICTACT J. Soft Comput. 2021, 11, 7.

28. Coppersmith, G.; Leary, R.; Crutchley, P.; Fine, A. Natural Language Processing of Social Media as Screening for Suicide Risk. Biomed. Inform. Insights 2018, 10, 117822261879286.

29. Zirikly, A.; Resnik, P.; Uzuner, Ö.; Hollingshead, K. 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, MN, USA, 6 June 2019; Association for Computational Linguistics: Stroudsburg, PA, USA, 2019; pp. 24–33.

30. Skaik, R.; Inkpen, D. Using Social Media for Mental Health Surveillance: A Review. ACM Comput. Surv. 2021, 53, 1–31.

31. Reddit by the Numbers. Available online: https://www.redditinc.com/press (accessed on 23 July 2022).

32. Reddit Privacy Policy. Available online: https://www.reddit.com/policies/privacy-policy (accessed on 23 July 2022).

33. Meta Privacy Policy—How Meta Collects and Uses User Data. Available online: https://www.facebook.com/privacy/policy/?entry_point=data_policy_redirect&entry=0 (accessed on 23 July 2022).

34. Peer Support for Anyone Struggling with Suicidal Thoughts. Available online: https://www.reddit.com/r/SuicideWatch/ (accessed on 23 July 2022).

35. Dutta, R.; Gkotsis, G.; Velupillai, S.; Bakolis, I.; Stewart, R. Temporal and Diurnal Variation in Social Media Posts to a Suicide Support Forum. BMC Psychiatry 2021, 21, 259.

36. Gasparetto, A.; Marcuzzo, M.; Zangari, A.; Albarelli, A. A Survey on Text Classification Algorithms: From Text to Predictions. Information 2022, 13, 83.

37. Khan, A.R. Facial Emotion Recognition Using Conventional Machine Learning and Deep Learning Methods: Current Achievements, Analysis and Remaining Challenges. Information 2022, 13, 268.

38. Roy, A.; Nikolitch, K.; McGinn, R.; Jinah, S.; Klement, W.; Kaminsky, Z.A. A Machine Learning Approach Predicts Future Risk to Suicidal Ideation from Social Media Data. NPJ Digit. Med. 2020, 3, 78.

39. De Oliveira, N.R.; Pisa, P.S.; Lopez, M.A.; de Medeiros, D.S.V.; Mattos, D.M.F. Identifying Fake News on Social Networks Based on Natural Language Processing: Trends and Challenges. Information 2021, 12, 38.

Retrieved from https://encyclopedia.pub/entry/history/show/75827