Journal of Advances in Mathematical & Computational Sciences
An International Pan-African Multidisciplinary Journal of the SMART Research Group
International Centre for IT & Development (ICITD) USA

Available online at https://www.isteams.net/ mathematics-computationaljournal.info
CrossREF Member Listing - https://www.crossref.org/06members/50go-live.html

A Conceptual Model for Undergraduate Suicide Risk Data Labeling

Ezea, I. L, Udanor, C. N., Agu, N. M. & Bakpo, F.S.
[1,2,3,4]Department of Computer Science
Faculty of Physical Sciences
University of Nigeria, Nsukka.
[1]ezea.ikenna@funai.edu.ng, [2]collins.udanor@unn.edu.ng, [3]monica.agu@unn.edu.ng,
[4]francis.bakpo@unn.edu.ng
Phone: +2348025107142

ABSTRACT
Suicide risk assessment is a critical domain in mental health research and clinical practice, with the potential to save lives through early intervention. This study presents a novel approach, the Synchronized Model for Undergraduate Suicide Risk Dataset Labeling (SMUSRDLabel), aimed at improving the accuracy and ethical handling of datasets about undergraduate suicide risk. The primary objective of this research is to address the challenges associated with dataset labeling by developing a synchronized model that leverages multi-source data and cutting-edge machine-learning techniques. Using a synthetic dataset, SMUSRDLabel integrates demographic information, behavioral patterns, mental health history, and textual data from diverse sources, fostering a holistic understanding of suicide risk factors among undergraduate populations. Our methodology employs a combination of supervised and semi-supervised learning algorithms, allowing for effective labeling even when labeled data is limited. Ethical considerations remain paramount, with stringent privacy preservation mechanisms in place to protect sensitive information. Through extensive experimentation and validation on a comprehensive synthetic dataset, we demonstrate that SMUSRDLabel outperforms traditional labeling methods in terms of accuracy, sensitivity, and specificity. Moreover, our model excels in identifying nuanced risk factors and offers interpretable insights into the underlying factors contributing to suicide risk. This research contributes to the field of mental health and machine learning by offering a robust framework for responsible and accurate dataset labeling, facilitating advancements in suicide risk assessment. As an interdisciplinary collaboration between machine learning practitioners and mental health experts, this synchronized model bridges the gap between cutting-edge technology and ethical considerations, with the potential to impact suicide prevention strategies and support systems for at-risk undergraduate individuals.

Keywords: Conceptual Model, Undergraduates, Suicide, Risks, Data Labeling

Ezea, I. L, Udanor, C. N., Agu, N. M. & Bakpo, F.S. (2024): A Conceptual Model for Undergraduate Suicide Risk Data Labeling. Journal of Advances in Mathematical & Computational Science. Vol. 12, No. 1. Pp 23-36.
Available online at www.isteams.net/mathematics-computationaljournal.

1. INTRODUCTION

Suicide remains a profound public health concern, with tragic consequences for individuals, families, and communities. The alarming rise in suicide rates among undergraduate students has drawn heightened attention to the need for effective risk assessment and intervention strategies within this vulnerable population. To address this pressing issue, the field of mental health research has increasingly turned to the power of data-driven approaches and machine learning. These technologies offer the potential to identify nuanced risk factors, enhance prediction accuracy, and ultimately save lives through timely interventions.

However, the successful application of machine learning in suicide risk assessment hinges on the availability of high-quality labeled datasets, a crucial resource for training and evaluating predictive models. Yet, the process of dataset labeling in the context of suicide risk is rife with challenges. Privacy concerns, ethical considerations, and the complexity of capturing the multifaceted nature of suicide risk factors pose substantial obstacles. Moreover, the scarcity of labeled data, often limited in size and scope, further complicates the development of accurate and robust predictive models. This research addresses these challenges by introducing a groundbreaking approach, a Synchronized Model for Undergraduate Suicide Risk Dataset Labeling (SMUSRDLabel). The central aim of this study is to revolutionize the way undergraduate suicide risk datasets are labeled and utilized within the domain of machine learning. SMUSRDLabel endeavors to bridge the gap between the pressing need for accurate risk assessment and the ethical imperative of responsible data handling.

Our synchronized model capitalizes on the synergistic fusion of multi-source data, including demographic information, behavioral patterns, mental health history, and textual data. By leveraging a variety of data modalities, SMUSRDLabel aims to provide a holistic and comprehensive understanding of the intricate web of factors contributing to undergraduate suicide risk. Furthermore, our methodology incorporates both supervised and semi-supervised learning techniques, enabling effective labeling even when labeled data is scarce. In the following sections, we delineate the development and validation of the SMUSRDLabel model, emphasizing its ethical underpinnings and commitment to safeguarding individual privacy. Through a rigorous evaluation process, we demonstrate the model's superior performance in comparison to traditional labeling methods. Additionally, we delve into the interpretability of our model, shedding light on the underlying risk factors and their implications. This research stands at the intersection of cutting-edge machine learning and the pressing needs of mental health research and practice. With the potential to enhance our understanding of undergraduate suicide risk and improve prevention strategies, SMUSRDLabel offers a vital contribution to the ongoing efforts to address this critical public health concern.

2. LITERATURE REVIEW

The urgency of addressing suicide risk among undergraduate students calls for innovative methodologies that leverage online social networks' data. Over recent years, researchers have explored various approaches to utilizing social media data for suicide risk assessment, focusing on the context of undergraduate populations. This literature review delves into pivotal studies in this area, elucidating their objectives, methodologies, and contributions while illuminating gaps and constraints in current research.

Masuda et al. [1] endeavored to discern user characteristics linked to suicide ideation within online social networks. They utilized logistic regression, considering both social media-related and non-social media-related data. Nevertheless, this study's approach displayed limitations, including its reactive nature, single-point risk assessment, and a lack of specification regarding the user community. Moreover, it may be vulnerable to media contagion, potentially exposing at-risk individuals to further harm. Significantly, the study overlooked certain pivotal social demographic factors that could contribute to suicide ideation but may not be evident from users' profiles.

Jashinsky et al. [2] demonstrated the potential of tracking suicide risk factors among undergraduate students through Twitter, using tweets to identify the presence of suicide-related terms and keywords. Despite analyzing an extensive dataset of 1,659,274 tweets, the study faced limitations, including its reactive nature, single-point risk assessment, and susceptibility to media contagion. Additionally, the reliance on specific search terms might have led to incomplete detection of at-risk tweets, while challenges in identifying tweet locations in specific states impacted sample size and generalizability.
Varathan and Talib [3] proposed a suicide detection system based on Twitter to automatically detect suicide-related messages and facilitate timely assistance. However, their approach may generate false-positive alerts, as not all suicidal tweets explicitly indicate suicide ideation.

Coppersmith et al. [4] centered their research on quantifying suicidal ideation by examining language usage on social media. Although they aimed to provide rapid responses to suicide, they acknowledged limitations related to language usage, which may be influenced by an individual's interests and profession, potentially leading to biased results. Gunn and Lester [5] scrutinized Twitter postings in the 24 hours preceding a suicide, emphasizing linguistic patterns. This single-point analysis might not account for the diversity of suicidal behaviors across undergraduate individuals. Li et al.

developed a poison-based model to extract stressor events from social media posts among undergraduate students. However, this approach may not capture all stressful information, particularly among users with unique interests. Coppersmith et al. [7] conducted exploratory data analysis of social media preceding suicide attempts, primarily focusing on women aged 15-29. However, this narrow focus may not represent the broader diversity of at-risk groups within the undergraduate population.

Colombo et al. [8] analyzed the connectivity and communication patterns of suicidal users on Twitter, shedding light on the propagation of suicidal content. However, the study's findings may be hindered by limitations in the dataset, with only 10% representing actual suicide ideation cases. Vioul`es et al. [9] introduced an approach that identifies sudden changes in online behavior as indicators of suicide risk among undergraduate students. However, recognizing such changes may not always be straightforward, particularly when users maintain multiple social media accounts. Ji et al. [10] employed supervised learning to detect early signs of suicide ideation within user-generated online content. Nevertheless, their approach may overlook non-explicit indicators of suicide ideation. Cao et al.

[11] applied deep learning to detect latent suicide risk within hidden social media posts. However, their model's inability to consider diverse social demographic factors could limit its effectiveness in assessing suicide risk among undergraduate students.

Sawhney et al. [12] introduced a time-aware transformer-based model for suicide ideation detection. Nonetheless, its reliance on historical tweets from a specific platform and locality may hinder generalizability to diverse undergraduate populations.

Cao et al. [13] constructed a suicide-oriented knowledge graph for detecting suicide ideation within social media posts. However, the integrity of information shared on microblogging platforms

3. METHODOLOGY

Dataset Generation: To conduct this research, a synthetic dataset comprising 10,000 entries was generated using a custom Python code developed specifically for this study. The dataset revolves around attributes related to both Social Network and Student Performance.

Variables Used: The variables included in the dataset are crucial for capturing and analyzing the intricate relationship between social network dynamics and academic performance. The selected variables and their descriptions are as follows:

- sno: Serial number assigned to each entry in the dataset (This attribute is redundant and was not considered during the computation).
- regNumber: Registration number uniquely identifying each student.
- Friendship: Metric representing the extent of social connections or friendships.
- academicWorkload: Measure of the academic workload carried by the student.
- lectureAttendance: Attendance record in academic lectures.
- cgpa: Cumulative Grade Point Average, indicating academic performance.
- Level: Academic level or year of study.
- spillover: Indicates the number of years a student has stayed in excess to his or her expected year of graduation.
- suicideCount: Count of reported suicides.
- nonSuicideCount: Count of instances with no reported suicides.
- Suicide Risk Factor: Calculated factor assessing the risk of suicide based on various parameters.
- suicidePercistencyFactor: Factor indicating the persistence of suicidal tendencies.
- suicideResistencyScore: Score reflecting the resistance to suicidal thoughts.
- averageSRS: Average Suicide Resistency Score across the dataset.

Sample Data: A snippet of the sample data from the generated dataset is presented next:

Table 1: A Snippet of the Sample Data from the Generated Dataset

sno	regNumber	Friendship	Academic Workload	Lecture Attendance	cgpa	Level	spillover	Suicide Count	non Suicide Count	Suicide Risk Factor	Suicide Percistency Factor	Suicide Resistency Score	averageSRS
1	Funai/bsc/1	5	20	15	4.85	5	0	1	15	0.93	1.04	23.11	4.62
2	Funai/bsc/2	1	20	18	4.56	1	1	18	13	0.41	2.33	1.95	1.95
3	Funai/bsc/3	0	20	18	1.84	4	0	12	7	0.36	3.01	2.45	0.61
4	Funai/bsc/4	0	20	16	2.69	1	1	14	15	0.51	2.41	1.11	1.11
5	Funai/bsc/5	1	20	17	1.49	3	1	1	2	0.66	1.62	2.77	0.92
6	Funai/bsc/6	1	20	14	2.41	2	3	1	10	0.90	1.45	3.30	1.65

This dataset serves as the foundation for exploring the intricate interplay between social network dynamics and student performance, particularly focusing on factors related to suicide risk and resilience. The variables selected aim to provide a comprehensive understanding of the multifaceted aspects influencing the well-being and academic success of students.

4. MODEL DEVELOPMENT

The synchronization model for suicide risk assessment, as depicted in Figure 1, represents a mathematical framework illustrating the flow of information within the synchronization system. This model comprises three main components: the input, the processing stage, and the output.

Input Component: The input component encompasses a list of all students slated for suicide risk assessment. For each student within this array, specific inputs are generated through various mathematical computations (refer to equations 1, 2, and 3). These computed inputs are then fed into the Synchronization/Time Utilization layer.

Synchronization/Time Utilization Layer: Within this layer, the computed inputs are combined with additional data elements such as Level, GPA, β, and ε, sourced from the students. Together, they contribute to the assessment of the students' suicide risk scores.

Suicide Risk Ratio/Assessment Layer: The outcomes of the Synchronization/Time Utilization layer, along with other relevant inputs, are channeled into the Suicide Risk Ratio/Assessment layer. Here, the students' scores in relation to suicide risk assessment are determined.

Suicide Risk Decision/Output Layer: The results obtained from the Suicide Risk Ratio/Assessment layer undergo regression analysis within the suicide risk decision/output layer. This analysis aids in identifying students who may be at risk of developing suicidal ideation, based on the risk assessment. In essence, this synchronization model facilitates a systematic evaluation of students' suicide risk by processing a range of inputs and utilizing mathematical computations, ultimately leading to the identification of individuals at potential risk.

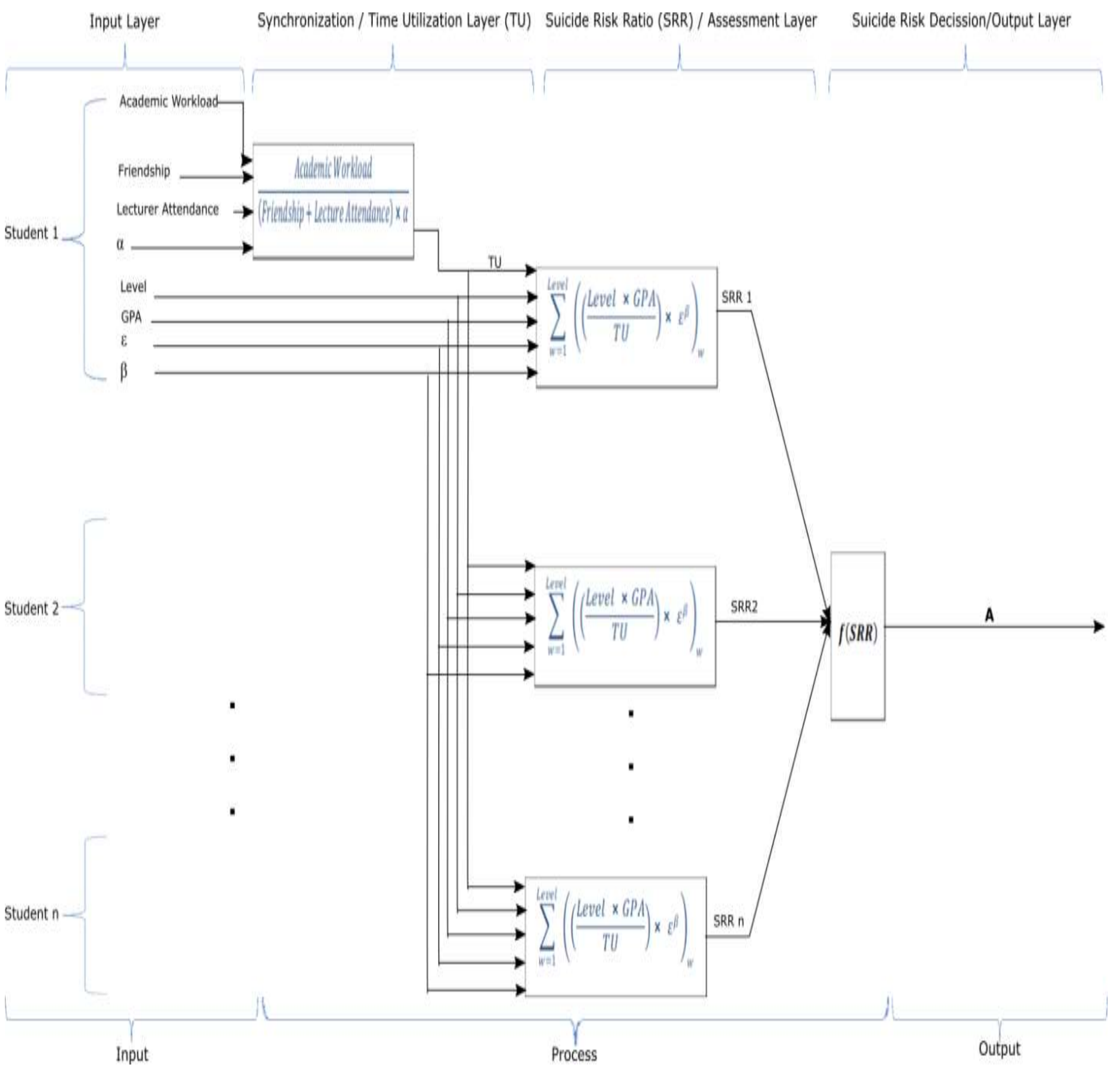

Figure 1: Synchronization Model for Suicide Risk Assessment

5. MATHEMATICAL FORMULATION AND DERIVATION

In Figure 1, the processing component of the model consists of two fundamental mathematical expressions, denoted as Equation 4 and Equation 5. These equations form the core of this article, and we will outline the derivation steps in the subsequent sections.

Academic Workload, a crucial metric in our analysis, represents the total lecture time (excluding days when lecturers were absent) within a day, further accumulated over a week or month(s).

It is computed by aggregating the time each lecturer spent in class on a given day, considering the following variables:

i) **Lecturer Time**: The amount of time each lecturer dedicates to teaching on a specific day.
ii) **Number of Lecturers**: The total count of lecturers who conducted classes on a given day.
iii) **Week(s):** The duration, in terms of weeks, during which lectures were held.
iv) **Lecturers**: Refers to the specific lecturers responsible for the courses being taught.

This composite measure provides insights into the intensity of academic engagements and forms an integral part of our analysis.

$$AcademicWorkload = \sum_{Day=1}^{Weeks} \left(\sum_{lecturer=1}^{Number of Lecuters} LectureTime_{Lecturer} \right) \tag{1}$$

Lecture Attendance represents the cumulative count of a student's presence in lectures for a course or multiple courses within a week. It is calculated by summing the daily instances of a student attending lectures for a specific course(s) over the course of a week.

This metric encompasses the following components:
i) **Time in Lecture:** The total duration a student spends actively participating in classroom lectures.
ii) **Course(s):** The number of courses in which a student is enrolled or participates in during the lecture sessions.
iii) **Week(s):** The timeframe typically measured in weeks, during which the lectures are conducted.

Lecture Attendance is a comprehensive measure that reflects a student's commitment and engagement with their academic coursework throughout the specified week(s).

$$LectureAttendance = \sum_{Day=1}^{Weeks} \left(\sum_{Course=1}^{Courses} TimeinLecture_{Course} \right) \tag{2}$$

Total Suicide Risk Factor (α) is the outcome of a machine learning assessment that gauges students' depression ratios, derived from their engagement in social media activities. This measure is obtained by aggregating the collective depression ratios (referred to as Suicide Risk Factor) over a designated week. The underlying assumption is grounded in the belief that an average individual should engage in at least one non-suicidal conversational interaction within a week or a semester, depending on the chosen evaluation interval for α.

$$\alpha = \begin{cases} \sum_{Day=1}^{week} \left(\frac{Non\ Suicide\ Treat\ in\ conversation\ (NSTC) + 1}{Total\ number\ of\ conversations} \right) \ if\ NSTC = 0 \\ \sum_{Day=1}^{week} \left(\frac{Non\ Suicide\ Treat\ in\ conversation\ (NSTC)}{Total\ number\ of\ conversations} \right) \ if\ NSTC > 0 \end{cases} \quad 3$$

$$SPF = \begin{cases} \frac{Academic\ Workload}{(Friendship + Lecture\ Attendance) \times \alpha} \ if\ friendship > 0 \\ \frac{Academic\ Workload}{(1 + Lecture\ Attendance) \times \alpha} \ if\ friendship = 0 \end{cases} \quad 4$$

The assumption is that when you don't have a friend you are alone.

Grade Point Average: this is the ratio of the sum product of a student's credit unit and grade, and the total credit unit taken over the sessions. It is used to access students' general academic performance. It can be derived using the equation 5 and the following variables associated with courses: Credit Unit, Grade, Courses, and Total Credit Unit Taken.

$$GradePointAverage(GPA) = \frac{\sum_{Course=1}^{courses} CreditUnit \times Grade_{Course}}{TotalCreditUnitTaken} \quad 5$$

Level: This is the total academic duration spent by a student as he or she pursues a degree program. It is the total number of academic sessions spent by a student. It can be derived using equation 6.

$$Level = CurrentSession - yearofadmission \quad 6$$

Spillover Value: This is the total number of years students spend in exclusion to the normal duration of a given programme. It can be calculated using the equation 7.

$$SpilloverValue(\beta) = \begin{cases} 0 \quad ifDurationofProgramme - Level \geq 0 \\ (Level - DurationofProgramme) ifDurationofProgramme - Level \leq 0 \end{cases} \qquad 7$$

Suicide Resistance Score (SRS): This is used to determine the likelihood that a student will exhibit suicidal behavior in the nearest future. It can be derived by substituting equation 4, 5 and 6 in equation 8. In addition to the mentioned equations it is also made up of three more components:

i) **Year:** the number of session to be used for the suicide assessment
ii) **Spillover value (β):** The number of sessions the student has spent in excess after the stipulated programme duration. It is usually less than the total academic duration of a given programme of study.

$$Suicide\ ResistanceScore\ (SRS)\ = \begin{cases} \sum_{year=}^{Level} \left(\left(\frac{(Level)\ \times GPA}{SuicidePercistencyFactor}\right)\right) \quad if\beta = 0 \\ \sum_{year=1}^{Level-\beta} \left(\left(\frac{(Level-\beta)\ \times GPA}{SuicidePercistencyFactor}\right)\right) \quad if\beta > 0 \end{cases} \qquad 8$$

$SRS_{Average}$:This is the average of overtime successive suicide resistance scores (SRS). It is derived by substituting equation 8 in equation 9.

$$SRS_{Average} = \begin{cases} \frac{\sum_{Year=1}^{Level}(Suicide\ \ ResistenceScore(SRS))}{Level} if\beta = 0 \\ \frac{\sum_{Year}^{Level}\ \ (Suicide\ \ ResistenceScore(SRS))}{Level-\beta} if\beta > 0\ and\beta\ != level \end{cases} \qquad 9$$

Suicide Resistance Score (Suicide Synchronization) Algorithm

Suicide Resistance Score algorithm is an algorithm that was generated using the mathematical expressions given in the section above.

The algorithm is as shown in figure 2.

1.1. Get all the courses a student offered
1.2. For each of the courses compute the total number of hours spent by each lecturer teaching the course
  1.2.1. While courses <=courses offered
    1.2.1.1. Lecturer hours = Previous hours + hours spent
1.3. Compute the total number of hours a student spends in lecture for all the courses
  1.3.1. While courses <=courses offered
    1.3.1.1. Student attendance = previous attendance + hours spent
1.4. Compute the suicide risk factor
  1.4.1. Suicide Risk Factor = Non Suicide Treat in conversation / Total number of conversations
1.5. Compute the total number of friends a student has maintained for the given period
  1.5.1. While period < = given period
    1.5.1.1. Total Friends = number of friends + previous count
1.6. Compute the persistency factor
  1.6.1. Suicide Persistency Factor = Lecturer hours / Total Friends + (student attendance * Suicide Risk Factor)
1.7. Compute Grade Point Average (GPA)
  1.7.1. Grade point average = credit unit * grade/total credit unit taken
1.8. Compute students level
  1.8.1. Student level = current session – year of admission
1.9. Compute the spillover value
  1.9.1. If ((duration of programme – level) >=0){
    1.9.1.1. Spillover value = 0
    1.9.1.2. }Else{
    1.9.1.3. Spillover value = level – duration of programme
    1.9.1.4. }
1.10. Compute Suicide Resistance Score (SRS)
  1.10.1. If (spillover value ==0){
    1.10.1.1. Suicide Resistance Score = ((Level) x GPA)/Suicide Persistency factor
    1.10.1.2. }else{
    1.10.1.3. ((Level - β) x GPA)/Suicide Persistency Factor
    1.10.1.4. }

Figure 2: Algorithm for Suicide Resistance Score

Ethical considerations were integrated into the model development process to ensure responsible and ethical usage of the data. This framework encompassed:

Privacy-Preserving Model: Developing models that prioritize user privacy by not requiring access to personally identifiable information (PII) during prediction.

Bias Mitigation: Implementing strategies to detect and mitigate potential biases in the dataset or model predictions to ensure fairness.

6. EVALUATION METRICS

Model Performance Evaluation

The performance of SMUSRD Label was rigorously assessed using standard evaluation metrics like:

Accuracy measures the ratio of correctly classified instance over the total number of instances classified. It can be performed using the following formula (TP + TN) / (TP + TN + FP + FN), where TP is the number of True Positive, TN is the number of True Negative, FP is the number of False Positive and FN is the number of False Negative.

Precision is the ratio of True Positive Prediction over the sum of True Positive and False Positive. It can be calculated using the following formula: TP/(TP + FP).

Recall otherwise known as the Sensitivity or True Positive Rate is the measure of True Positive predictions over the sum of the actual positive instances. It can be calculated using the following formula: TP/(TP + FN).
F1-Score measures the balance between precision and recall by measuring their harmonic mean. It can be calculated using the following formula: 2 * (Precision * Recall) / (Precision + Recall).

Specificity is the measure of the ratio of true negative predictions over the sum of the number of true negative and false positive instances. The formula is TN / (TN + FP).

Table 1: The Confusion Matrix for SMUSRDLabel and Manual Labeling

Model	Suicide Status	Precision	Recall	Specificity	F1-Score
SMUSRDLabel	non-suicide	0.96	0.97	0.96	0.96
	suicide	0.95	0.94	0.95	0.95
Manual Labeling	non-suicide	0.89	0.84	0.85	0.86
	suicide	0.84	0.89	0.88	0.86

7. VALIDATION AND CROSS-VALIDATION

To validate the robustness of SMUSRDLabel, extensive cross-validation procedures were conducted. The dataset was partitioned into training, validation, and testing subsets, with cross-validation performed iteratively to validate the model's performance across diverse subsets of data.

References

[1] N. Masuda, I. Kurahashi and H. Onari, "Suicide ideation of individuals in online social networks," *PloS one,* vol. 8, no. 4, pp. 1-8, 2013.

[2] J. Jashinsky, S. H. Burton, C. L. Hanson, J. West, C. Giraud-Carrier, M. D. Barnes and T. Argyle, "Tracking suicide risk factors through twitter in the us," *Crisis,* vol. 35, no. 1, pp. 51-59, 2014.

[3] K. D. Varathan and N. Talib, "Suicide detection system based on twitter," in *Science and Information Conference (SAI)*, London, 2014.

[4] G. Coppersmith, R. Leary, E. Whyne and T. Wood, "Quantifying suicidal ideation via language usage on social media," 2015.

[5] J. F. Gunn and D. Lester, "Twitter postings and suicide: An analysis of the postings of a fatal suicide in the 24 hours prior to death," *Suicidologi,* vol. 17, no. 3, 2015.

[6] Q. Li, Y. Xue, L. Zhao, J. Jia and L. Feng, "Analyzing and Identifying Teens Stressful Periods and Stressor Events from a Microblog," *IEEE Journal of Biomedical and Health Informatics,* pp. 2168-219
4, 2016.

[7] G. Coppersmith, K. Ngo, R. Leary and A. Wood, "Exploratory analysis of social media prior to a suicide attempt," in *Proceedings of the Third Workshop on Computational Linguistics and Clinical Psychology*, 2016.

[8] G. B. Colombo, P. Burnap, A. Hodorog and J. Scourfield, "Analysing the connectivity and communication of suicidal users on twitter," *Computer communications,* vol. 73, pp. 291-300, 2016.

[9] M. J. Vioul`es, J. A. B. Moulahi and S. Bringay, "Detection of suicide-related posts in twitter data streams," *IBM Journal of Research and Development,* vol. 62, no. 1, pp. 1-12, 2018.

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

[11] L. Cao, H. Zhang, L. Feng, ZihanWei, X. Wang, N. Li and X. He, "Latent Suicide Risk Detection on Microblog via Suicide-OrientedWord Embeddings and Layered Attention," in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, Hong Kong, 2019.

[12] R. Sawhney, H. Joshi, S. Gandhi and R. R. Shah, "A Time-Aware Transformer Based Model for Suicide Ideation Detection on Social Media," in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*, Online, 2020.

[13] L. Cao, H. Zhang and L. Feng, "Building and Using Personal Knowledge Graph to Improve Suicidal Ideation Detection on Social Media," *IEEE Transactions on Multimedia,* pp. 1-16, 2020.

[14] J. L. Oliffe, J. S. Ogrodniczuk, J. L. Bottorff, J. L. Johnson and K. Hoyak, "“You feel like you can’t live anymore”: Suicide from the perspectives of Canadian men who experience depression," *Social Science & Medicine,* vol. 74, no. 2012, pp. 506-514, 2012.

[15] R. Sawhney, H. Joshi, L. Flek and R. R. Shah, "PHASE: Learning Emotional Phase-aware Representations for Suicide Ideation Detection on Social Media," in *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics*, Online, 2021.

[16] S. Chattopadhyay, "A study on suicidal risk analysis," in *International Conference on e-Health Networking, Application and Services*, 2007.

[17] S. Chattopadhyay, "A mathematical model of suicidal-intentestimation in adults," *American Journal of Biomedical Engineering,* vol. 2, no. 6, p. 251–262, 2012.

[18] S. Chattopadhyay, "A mathematical model of suicidal-intentestimation in adults," *American Journal of Biomedical Engineering,* vol. 2, no. 6, pp. 251-262, 2012.

[19] W. Wang, L. Chen, M. Tan, S. Wang and A. P. Sheth, "Discovering fine-grained sentiment in suicide notes," *Biomedical informatics insights,* vol. 5, no. 1, p. 137, 2012.

[20] M. Liakata, J. H. Kim, S. Saha, J. Hastings and D. Rebholzschuhmann, "Three hybrid classifiers for the detection of emotions in suicide notes," *Biomedical Informatics Insights,* vol. 2012, no. Suppl. 1, p. 175–184, 2012.

[21] J. Pestian, H. Nasrallah, P. Matykiewicz, A. Bennett and A. Leenaars, "Suicide note classification using natural language processing: A content analysis," *Biomedical informatics insights,* vol. 2010, no. 3, p. 19, 2010.

[22] A. Abboute, Y. Boudjeriou, G. Entringer, J. Az´e, S. Bringay and P. Poncelet, "Mining twitter for suicide prevention," in *International Conference on Applications of Natural Language to Data Bases/Information Systems*, 2014.

[23] S. R. Braithwaite, C. Giraud-Carrier, J. West, M. D. Barnes and C. L. Hanson, "Validating machine learning algorithms for twitter data against established measures of suicidality," *JMIR mental health,* vol. 3, no. 2, p. e21, 2016.

[24] E. Okhapkina, V. Okhapkin and O. Kazarin, "Adaptation of information retrieval methods for identifying of destructive informational influence in social networks," in *31st International Conference on Advanced Information Networking and Applications Workshops (WAINA)*, 2017.

[25] X. Huang, L. Zhang, D. Chiu, T. Liu, X. Li and T. Zhu, "Detecting suicidal ideation in chinese microblogs with psychological lexicons," in *IEEE 11th International Conference on Ubiquitous Intelligence and Computing and Autonomic and Trusted Computing and IEEE 14th International Conference on Scalable Computing and Communications and Its Associated Workshops (UTC-ATC-ScalCom).* , 2014.

[26] H.-C. Shing, S. Nair, A. Zirikly, M. Friedenberg, H. D. III and P. Resnik, "Expert; crowdsourced; and machine assessment of suicide risk via online postings," in *Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic*, 2018.

[27] M. M. a. J. Quinn, "Suicidal tendencies: The automatic classification of suicidal and non-suicidal lyricists using nlp.," in *IJCNLP*, 2013.

[28] X. Le, I. Lancashire and G. Hirst, "Regina Jokel Longitudinal detection of dementia through lexical and syntactic changes in writing: A case study of three British novelists," *Literary and linguistic computing,* vol. 26, no. 4, pp. 435-461, 2011.

[29] F. Ren, X. Kang and C. Quan, "Examining accumulated emotional traits in suicide blogs with an emotion topic model," *IEEE journal of biomedical and health informatics,* vol. 20, no. 5, pp. 11384-1396, 2016.