Decoding the Complexity of Adolescent Suicidal Behaviors with Exploratory Data Analysis Techniques

Amornvit Vatcharaphrueksadee[1], Rattikan Viboonpanich[2], Jirawin Deecharoenchitpong[3] and Soravit Chalermsaen[4]

[1,2,3,4]Faculty of Information Technology and Digital Innovation, North Bangkok University, Bangkok, 10220, Thailand

Abstract

Decoding the Complexity of Adolescent Suicidal Behaviors with Exploratory Data Analysis Techniques aims to explore the factors associated with suicidal behavior among adolescents using the Global School-based Student Health Survey dataset. Through exploratory data analysis and advanced analytical techniques, we investigated the relationship between suicidal behavior and various risk factors, including bullying, serious injuries, and social relationships. Our analysis also considered the potential moderating effects of protective factors, such as parental support and peer relationships, on the connection between risk factors and suicidal behavior. The results revealed that bullying, serious injuries, and a lack of close friends were the top three factors with the strongest correlation to attempted suicide among adolescents. We found that females exhibited higher suicide rates compared to males, and that the 13-15 age group showed higher suicide rates than the 16-17 age group. Additionally, our study highlighted the importance of data-driven methodologies in guiding future research and intervention efforts for adolescent suicide prevention. In conclusion, our findings contribute to the existing knowledge of factors influencing adolescent suicidal behavior and emphasize the need for targeted interventions and comprehensive prevention programs. Further research exploring the complex interplay between risk and protective factors is essential for developing effective strategies to mitigate suicide risk among adolescents and promote positive mental health outcomes.

***Keywords**:* Adolescent suicidal behavior, Exploratory data analysis, Risk factors, Protective factors, and Suicide prevention

Introduction

Suicide is a major public health issue that affects millions of people and families worldwide [1]. Suicide is the third-largest cause of mortality among teenagers, making it critical to investigate and comprehend the underlying risk factors and behaviors related to suicidal thoughts in this group [2]. Identifying these risk variables can help establish effective preventative measures and tailored treatments, eventually saving lives and decreasing the negative effects of suicide [3].

The current study examines suicide behavior in teenagers using data from the Global School-Based Student Health Survey (GSHS) [4]. This study seeks to add to our knowledge of the complex interaction of risk variables and give ideas for designing preventative methods by studying connections between numerous factors and suicidal behaviors among teenagers.

Exploratory data analysis (EDA) approaches are used in this research to find patterns, trends, and correlations in the dataset [5]. This study intends to utilize EDA to emphasize the significance of certain traits in predicting or explaining suicide behaviors among teenagers, as well as to improve public health policy and intervention development.

This study will evaluate the possible moderating impacts of protective variables such as parental support, peer connections, and school connectedness to gain a thorough knowledge of the factors impacting suicide behaviors [6]. Examining these protective variables is crucial for developing effective therapies that not only address risk factors but also boost resilience and promote positive mental health outcomes in adolescents [7].

Furthermore, recent breakthroughs in machine learning and data mining approaches have demonstrated promise in finding patterns and indicators of suicide behavior [7]. The combination of these approaches with EDA methodologies has the potential to provide deeper insights into the intricate interactions between many variables and suicidal behavior, guiding more focused and successful treatments [8].

In conclusion, this study aims to provide a comprehensive analysis of suicidal behaviors among adolescents using the GSHS dataset and EDA techniques. By examining the relationships between various risk factors, protective factors, and suicidal behaviors, the findings from this study will contribute to the existing knowledge base and inform the development of tailored prevention strategies and interventions to address the pressing issue of adolescent suicide.

Background and Statement of problem

The study of suicidal behaviors among adolescents is essential for understanding the factors that contribute to these behaviors and informing prevention strategies. The literature has identified several risk factors associated with suicidal ideation, attempts, and completed suicide, including psychiatric disorders, substance use, family history of suicide, and stressful life events [14], [15]. Additionally, social factors such as bullying, peer relationships, and family support have been shown to influence suicidal behaviors among young people [10], [11].

Previous research on adolescent mental health and suicidal behaviors has often relied on large-scale survey data, such as the GSHS or the Youth Risk Behavior Surveillance System (YRBSS) [12], [13]. These studies have provided valuable insights into the prevalence of suicidal behaviors, as well as the relationships between various risk factors and suicidal ideation, attempts, and completions. However, there is still a need for further exploration of these relationships, particularly in diverse populations and settings.

There has been a growing interest in the application of machine learning and data mining techniques to identify patterns and predictors of suicidal behaviors [7], [8]. These approaches have shown promise in detecting high-risk individuals and informing targeted interventions. For example, Passos et al. [7] used machine learning algorithms to identify important predictors of suicidal ideation in a large sample of U.S. adolescents, demonstrating the potential of these techniques for informing suicide prevention strategies.

Despite the progress made in understanding suicidal behaviors among adolescents, gaps remain in the literature. For instance, some studies have primarily focused on Western populations, limiting the generalizability of their findings to other cultural contexts [9]. Furthermore, there is a need for research that employs EDA techniques to provide a comprehensive understanding of the complex relationships between various risk factors and suicidal behaviors among adolescents. The present study aims to address these gaps by utilizing a dataset derived from the GSHS and employing EDA techniques to explore the relationships between diverse factors and suicidal behaviors in adolescents.

Additionally, gender differences in suicidal behaviors have been a topic of interest in the literature [16]. Some studies have found that females are more likely to report suicidal ideation and attempts, while males are more likely to die by suicide [17]. This highlights the importance of examining gender-specific risk factors and protective factors in understanding and addressing adolescent suicidal behaviors.

Several protective factors have also been identified in the literature, which may buffer against the negative effects of risk factors and reduce the likelihood of suicidal behaviors. For example, strong social support, positive family relationships, and school connectedness have been found to be associated with lower suicide risk among adolescents [18], [19]. Investigating the moderating role of these protective factors in the relationship between risk factors and suicidal behaviors may provide valuable insights for designing more effective prevention programs.

In summary, the current literature has made significant strides in identifying risk and protective factors for adolescent suicidal behaviors. However, gaps remain in terms of understanding the complex interplay between these factors in diverse populations and settings. Moreover, there is potential for leveraging advanced analytical techniques, such as EDA and machine learning, to uncover novel insights and inform more targeted interventions. The present study seeks to address these gaps by exploring relationships between various factors and suicidal behaviors among adolescents using the GSHS dataset and employing EDA techniques to provide a comprehensive understanding of this complex issue. The next section will discuss about the data and methods used in this research for analyzing the suicidal behavior in adolescents.

Research Methodology

a. Dataset Description

The dataset, titled "Suicidal Behaviors Among Adolescents," is derived from the Global School-based Student Health Survey (GSHS) for adolescents aged 13 to 17 years [9]. The GSHS is a collaborative surveillance project designed to help countries measure and assess behavioral risk factors and protective factors among young people. The dataset contains information on various factors that can potentially influence suicidal behavior, such as alcohol consumption, drug use, mental health, and violence [10] illustrated in Table 1.

Table 1 shows the Global School-based Student Health Survey (GSHS) dataset description.

Features	**count**	**mean**	**std**	**min**	**max**
1. Year	106	2014.698113	2.089292	2010	2018
2. Currently_Drink_Alcohol	106	31.815094	53.454089	1.4	548
3. Really_Get_Drunk	106	22.496226	16.553129	0.8	80.2
4. Overwieght	106	23.69434	15.764075	3.3	70.6
5. Use_Marijuana	106	7.642453	8.713536	0	43.2
6. Have_Understanding_Parents	106	33.190566	11.559408	5.6	63.9
7. Missed_classes_without_permssion	106	29.996226	10.786728	6.5	62.2
8. Had_sexual_relation	106	26.679245	17.401318	2.5	73.9
9. Smoke_cig_currently	104	15.546154	10.748501	1.2	43.8
10. Had_fights	106	32.448113	15.297701	3.5	76.5
11. Bullied	102	31.109804	14.185479	9.9	78.6
12. Got_Seriously_injured	106	43.723679	14.808421	15.2	87.7
13. No_close_friends	106	7.74434	4.365254	1.5	24.8
14. Attempted_suicide	106	14.45283	9.273621	2.7	67.2
15. Country	106	-	-	-	-
16. Age Group	106	-	-	-	-
17. Sex	106	-	-	-	-

b. Data Preprocessing

Python programming language and its associated libraries, including NumPy, Pandas, Matplotlib, and Seaborn [19], were employed for data preprocessing. The initial preprocessing steps involve several important tasks including 1) Identifying and handling missing values in the dataset: Missing values can lead to biased or incorrect results during analysis. Depending on the nature and distribution of the data, appropriate imputation methods, such as median, mean, or mode imputation, are employed to replace the missing values; 2) Removing irrelevant or redundant features: To simplify the analysis and avoid multicollinearity issues, features that do not contribute significantly to the research objectives or those that are highly correlated with other features are removed from the dataset, and 3) Transforming or encoding categorical variables: To ensure compatibility with various statistical and machine learning algorithms, categorical variables are transformed into numerical representations, such as one-hot encoding or label encoding, depending on the specific requirements of the analysis. Based on Table 1, there are missing values in "*Smoke_cig_currently*" and "*Bullied*" features. These missing values were replaced using median imputation due to the benefit of unbiased imputation [20]. Additionally, irrelevant, or redundant features were dropped from the dataset to simplify the analysis [21].

c. Exploratory Data Analysis

With the cleaned dataset, exploratory data analysis (EDA) was conducted to uncover patterns, trends, and relationships among the variables [22]. EDA involved generating descriptive statistics and visualizations using Python libraries such as Matplotlib and Seaborn [23]. These techniques helped identify potential relationships between variables that informed further analysis and hypothesis generation.

d. Statistical and Machine Learning Methods

In addition to EDA, this study employed various statistical methods or machine learning algorithms to further explore the relationships between the risk factors and attempted suicide among adolescents [24]. These methods included correlation analysis, regression models, clustering, or classification techniques [25]. The choice of specific methods depended on the research questions, the nature of the data, and the findings from the EDA.

Throughout the data analysis process, Python programming language and its associated libraries (NumPy, Pandas, Matplotlib, and Seaborn) were utilized for data manipulation, visualization, and statistical modeling [19]. These tools allowed for a flexible and efficient approach to data analysis, ensuring the best possible understanding of the relationships between various factors and adolescent suicidal behavior.

Research results

The results section presents the key findings from the data analysis, exploratory data analysis (EDA), and the application of statistical methods. These findings provide insights into the relationships between various factors and adolescent suicidal behavior, as well as the predictive performance of the models.

a. Trends at the Country Level

The analysis of the relationship between "Country" and the count of "Attempted_suicide" revealed that Samoa has the highest prevalence of attempted suicide among adolescents (34.38%), while Indonesia has the lowest (3.6%). Thailand's suicide rate among adolescents stands at 12.7% [26] as shown in Figure 1. This finding suggests that unique regional or cultural factors may contribute to the elevated risk of suicidal behavior in certain countries [9, 10]. Further investigation into these factors could help inform targeted interventions and support strategies.

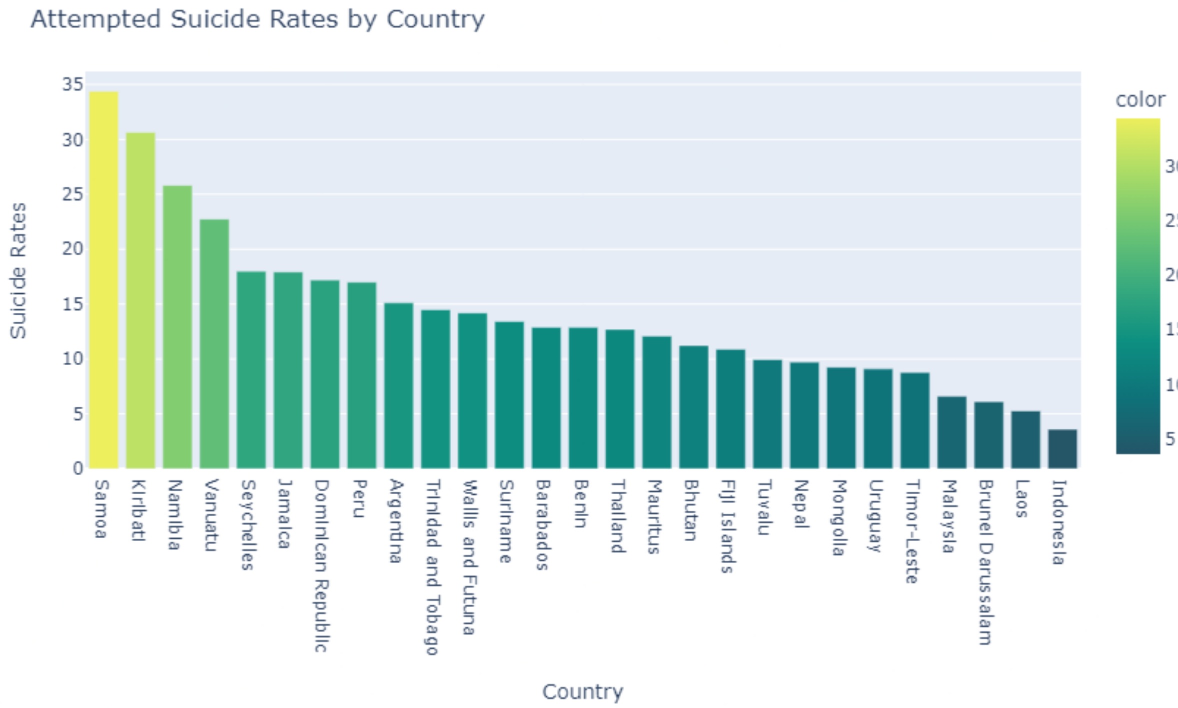

Figure 1 the analysis of the relationship between country and suicide rates

b. Suicidal Behavior and Demographics

Figure 2 illustrates the analysis of attempted suicide rates among different age groups and sexes. It has been found that females have a higher suicide rate (15.12%) compared to males (13.78%). Additionally, the suicide rate among adolescents is higher in the 13-15 age group (15.54%) than in the 16-17 age group (12.9%) [27] as shown in Figure 3. These findings align with previous research highlighting the importance of considering demographic factors when studying adolescent suicidal behavior [3, 4].

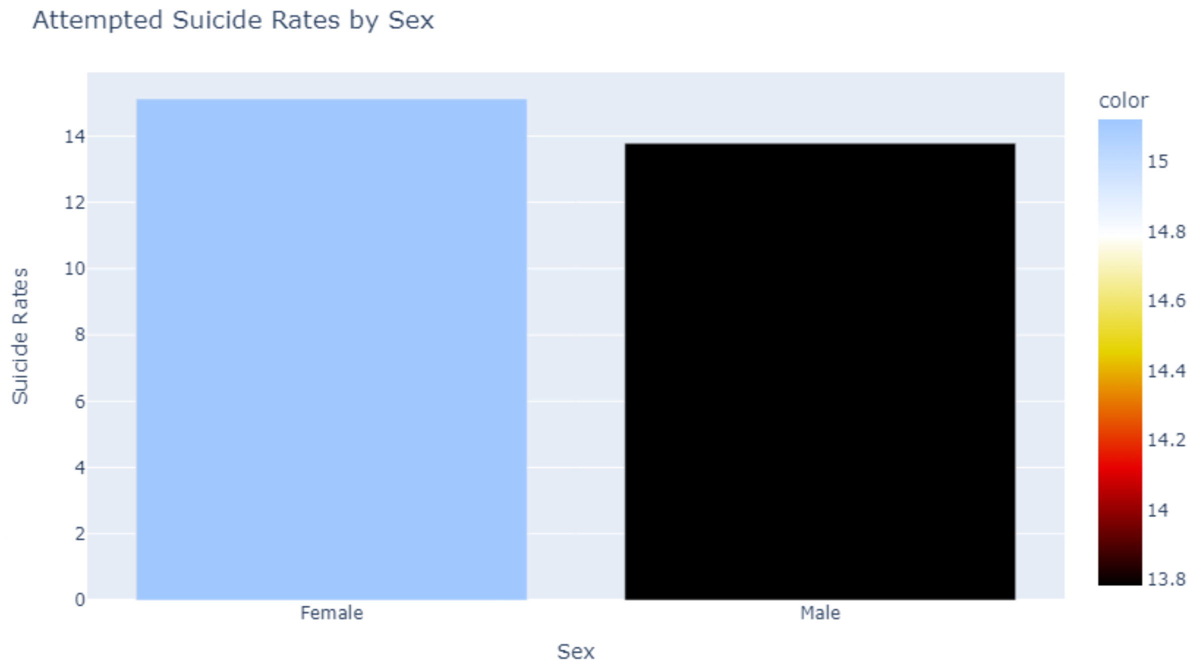

Figure 2 the analysis of the relationship between sex and suicide rates

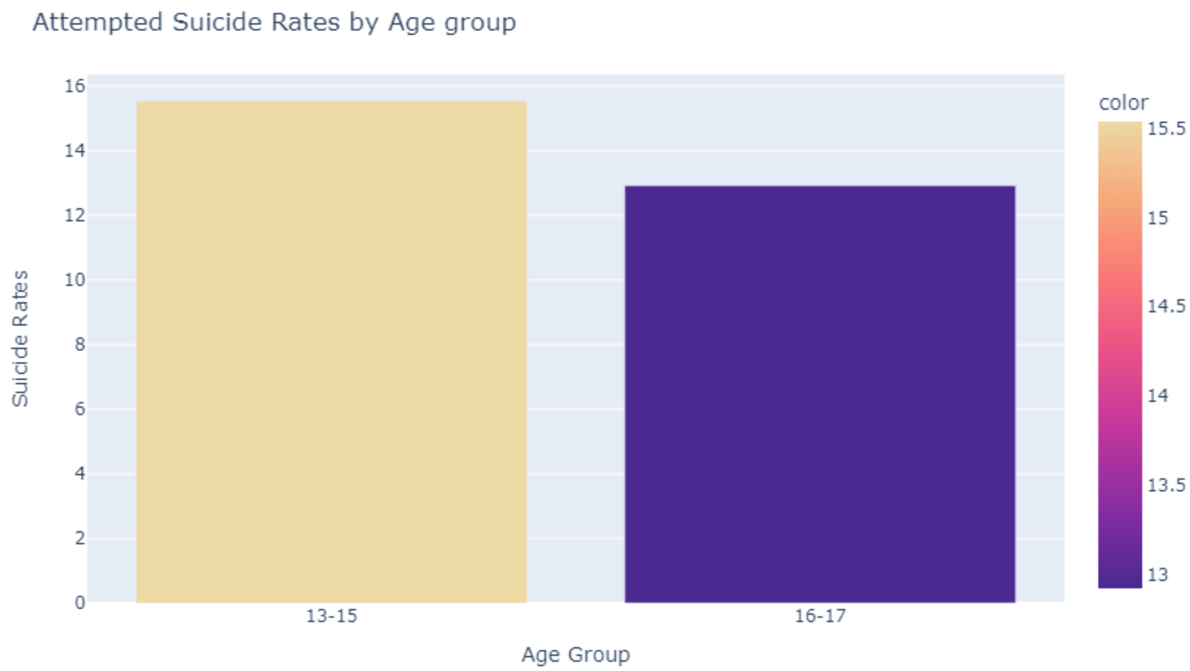

Figure 3 the analysis of the relationship between age group and suicide rates

c. The Relationships Between Suicidal Behavior and Risk Factors

Correlation analysis of behavioral factors and "Attempted_suicide" revealed that "Bullied" (0.64), "Got_Seriously_injured" (0.49), and "No_close_friends" (0.45) had the highest correlation values with suicidal behavior [28]. These findings are consistent with previous research that has identified bullying [10], serious injuries [6], and lack of social support [7] as risk factors for suicidal behavior among adolescents. In contrast, having understanding parents showed a very weak correlation, suggesting that parental support may serve as a protective factor against suicidal behavior, as has been suggested in prior studies [8].

Figure 4 the correlation between risk factors toward suicide rates

Furthermore, the study employed the median imputation technique [20] to handle missing values in the "Smoke_cig_currently" and "Bullied" features. This approach is considered effective in avoiding bias in the dataset, ensuring a more accurate analysis of the relationships between variables.

The results of this study offer valuable insights into the complex relationships between various factors and adolescent suicidal behavior. The identification of significant risk factors, such as bullying, serious injuries, and lack of close friends, can help inform targeted interventions and support strategies for at-risk individuals. Moreover, the study highlights the importance of considering demographic factors like age and sex, as well as regional and cultural factors, in understanding suicidal behavior and designing appropriate interventions [29].

Summary of the study

Expanding upon the findings from our analysis, several directions for future research and interventions can be proposed. First, longitudinal studies examining the long-term impact of bullying and serious injuries on suicidal behavior can shed light on how these risk factors evolve over time and inform the development of targeted prevention strategies [10]. For example, by exploring how the severity and duration of bullying experiences influence suicidal behavior, interventions can be tailored to address the specific needs of individuals at various stages of exposure to bullying [1].

Second, future research should investigate the potential moderating effects of protective factors on the relationship between risk factors and suicidal behavior. For instance, examining the role of social support and coping strategies in buffering the effects of bullying and serious injuries on suicidal behavior can provide valuable insights for the design of comprehensive intervention programs [19]. Additionally, understanding the relative contributions of different protective factors, such as parental support, peer relationships, and school connectedness, can help optimize the allocation of resources for suicide prevention efforts [18].

Third, the integration of advanced data analysis techniques, such as machine learning and network analysis, can further enhance our understanding of the complex interplay between risk and protective factors for suicidal behavior among adolescents [7]. These approaches can help identify subgroups of adolescents with distinct patterns of risk and resilience, enabling the development of targeted interventions that address the unique needs of each group [5].

In conclusion, the findings from our analysis offer valuable insights into the factors contributing to adolescent suicidal behavior and emphasize the importance of data-driven methodologies in guiding future research and intervention efforts [22]. By building on these insights and incorporating advanced analytical techniques, researchers and practitioners can work together to develop effective strategies for preventing adolescent suicide and promoting positive mental health outcomes for this vulnerable population.

Discussions

In conclusion, our study sheds light on the factors associated with suicidal behavior in adolescents, emphasizing the significance of bullying, serious injuries, and social relationships. The research highlights the necessity of utilizing data-driven methodologies and advanced analytical techniques in understanding the multifaceted dynamics underpinning adolescent suicidal behavior.

The results of our analysis call for the development of targeted interventions that cater to the specific requirements of adolescents facing risk factors, such as bullying and serious injuries. Moreover, the study accentuates the value of investigating protective factors and their potential in moderating the impact of risk factors on suicidal behavior. This knowledge can be instrumental in creating well-rounded prevention programs.

To sum up, our research enriches the existing literature on the determinants of adolescent suicidal behavior and sets the stage for future investigations aimed at comprehending and preventing suicide within this susceptible group. By delving deeper into the intricate interplay between risk and protective factors, researchers and practitioners can join forces to devise and implement efficient strategies to reduce suicide risk among adolescents and foster positive mental health outcomes.

Acknowledgment

We would like to express our sincere gratitude to North Bangkok University for their generous financial support and resources provided throughout this research project. We are also grateful to the faculty, staff, and fellow researchers for their invaluable expertise, encouragement, and constructive feedback, which have significantly contributed to our understanding of adolescent suicidal behaviors and the application of exploratory data analysis techniques. Their support has played a crucial role in the successful completion of our study and the advancement of knowledge in this field.

Reference

[1] World Health Organization (WHO). (2021). Suicide worldwide in 2019: Global health estimates. Geneva: WHO.
[2] Hawton, K., Saunders, K. E. A., & O'Connor, R. C. (2012). Self-harm and suicide in adolescents. The Lancet, 379(9834), 2373-2382. DOI: 10.1016/S0140-6736(12)60322-5.
[3] Klonsky, E. D., May, A. M., & Saffer, B. Y. (2016). Suicide, Suicide Attempts, and Suicidal Ideation. Annual Review of Clinical Psychology, 12, 307-330. DOI: 10.1146/annurev-clinpsy-021815-093204.
[4] World Health Organization (WHO). (n.d.). Global School-based Student Health Survey (GSHS). Retrieved from https://www.who.int/teams/noncommunicable-diseases/surveillance/systems-tools/global-school-based-student-health-survey.
[5] Tukey, J. W. (1977). Exploratory Data Analysis. Reading, MA: Addison-Wesley.

[6] Renshaw, K. L., Bolognino, S. J., Fletcher, S. E., Benshoof, C. M., & Searcy, S. J. (2019). Understanding the relationship between protective factors and suicidal ideation among high school students. School Mental Health, 11(3), 574-586.
[7] Passos, I. C., Mwangi, B., Cao, B., Hamilton, J. E., Wu, M.-J., Zhang, X. Y., Zunta-Soares, G. B., Quevedo, J., Kauer-Sant'Anna, M., Kapczinski, F., & Soares, J. C. (2016). Identifying a clinical signature of suicidality among patients with mood disorders: A pilot study using a machine learning approach. Journal of Affective Disorders, 193, 109-116.
[8] Kessler, R. C., Warner, C. H., Ivany, C., Petukhova, M. V., Rose, S., Bromet, E. J., Brown, M., Cai, T., Colpe, L. J., Cox, K. L., Fullerton, C. S., Gilman, S. E., Gruber, M. J., Heeringa, S. G., Lewandowski-Romps, L., Li, J., Millikan-Bell, A. M., Naifeh, J. A., Nock, M. K., Rosellini, A. J., Sampson, N. A., Schoenbaum, M., Stein, M. B., Wessely, S., Zaslavsky, A. M., & Ursano, R. J. (2015). Predicting suicides after psychiatric hospitalization in US Army soldiers: The Army Study To Assess Risk and Resilience in Service members (Army STARRS). JAMA Psychiatry, 72(1), 49-57.
[9] Nock, M. K., Borges, G., Bromet, E. J., Cha, C. B., Kessler, R. C., & Lee, S. (2008). Suicide and suicidal behavior. Epidemiologic Reviews, 30(1), 133-154.
[10] Brunstein Klomek, A., Sourander, A., & Elonheimo, H. (2015). Bullying by peers in childhood and effects on psychopathology, suicidality, and criminality in adulthood. The Lancet Psychiatry, 2(10), 930-941.
[11] Fergusson, D. M., Woodward, L. J., & Horwood, L. J. (2000). Risk factors and life processes associated with the onset of suicidal behaviour during adolescence and early adulthood. Psychological Medicine, 30(1), 23-39.
[12] Kann, L., McManus, T., Harris, W. A., Shanklin, S. L., Flint, K. H., Queen, B., Lowry, R., Chyen, D., Whittle, L., Thornton, J., Lim, C., Bradford, D., Yamakawa, Y., Leon, M., Brener, N., & Ethier, K. A. (2018). Youth risk behavior surveillance—United States, 2017. MMWR Surveillance Summaries, 67(8), 1-114.
[13] Centers for Disease Control and Prevention (CDC). (n.d.). Youth Risk Behavior Surveillance System (YRBSS). Retrieved from
[14] Nock, M. K., Green, J. G., Hwang, I., McLaughlin, K. A., Sampson, N. A., Zaslavsky, A. M., & Kessler, R. C. (2013). Prevalence, correlates, and treatment of lifetime suicidal behavior among adolescents: Results from the National Comorbidity Survey Replication Adolescent Supplement. JAMA Psychiatry, 70(3), 300-310.
[15] Beautrais, A. L. (2000). Risk factors for suicide and attempted suicide among young people. Australian & New Zealand Journal of Psychiatry, 34(3), 420-436.
[16] Canetto, S. S., & Sakinofsky, I. (1998). The gender paradox in suicide. Suicide and Life-Threatening Behavior, 28(1), 1-23.
[17] Centers for Disease Control and Prevention (CDC). (2020). Fatal injury data. Web-based Injury Statistics Query and Reporting System (WISQARS). Retrieved from https://www.cdc.gov/injury/wisqars/fatal.html.
[18] Resnick, M. D., Bearman, P. S., Blum, R. W., Bauman, K. E., Harris, K. M., Jones, J., Tabor, J., Beuhring, T., Sieving, R. E., Shew, M., Ireland, M., Bearinger, L. H., & Udry, J. R. (1997). Protecting adolescents from harm: Findings from the National Longitudinal Study on Adolescent Health. JAMA, 278(10), 823-832.
[19] Borowsky, I. W., Ireland, M., & Resnick, M. D. (2001). Adolescent suicide attempts: Risks and protectors. Pediatrics, 107(3), 485-493.
[20] van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate Imputation by Chained Equations in R. Journal of Statistical Software, 45(3), 1-67.
[21] Kelleher, J. D., Mac Namee, B., & D'Arcy, A. (2015). Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples, and Case Studies. Cambridge, MA: MIT Press.
[22] Tukey, J. W. (1977). Exploratory Data Analysis. Reading, MA: Addison-Wesley.
[23] Hunter, J. D. (2007). Matplotlib: A 2D Graphics Environment. Computing in Science & Engineering, 9(3), 90-95.
[24] James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An Introduction to Statistical Learning: with Applications in R. New York, NY: Springer.

[25] Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction (2nd ed.). New York, NY: Springer.
[26] World Health Organization. (2021). Suicide in the world: Global Health Estimates. Geneva: World Health Organization.
[27] Cash, S. J., & Bridge, J. A. (2009). Epidemiology of youth suicide and suicidal behavior. Current Opinion in Pediatrics, 21(5), 613-619.
[28] Klonsky, E. D., May, A. M., & Saffer, B. Y. (2016). Suicide, Suicide Attempts, and Suicidal Ideation. Annual Review of Clinical Psychology, 12, 307-330. DOI: 10.1146/annurev-clinpsy-021815-093204.
[29] O'Connor, R. C., & Nock, M. K. (2014). The psychology of suicidal behavior. The Lancet Psychiatry, 1(1), 73-85.