Manchester Metropolitan University

e-space

Manchester Metropolitan University's Research Repository

Please cite the Published Version

Drousiotis, Efthyvoulos, Joyce, Dan W, Dempsey, Robert C, Haines-Delmont, Alina, Spirakis, Paul G, Shi, Lei and Maskell, Simon (2023) Probabilistic decision trees for predicting 12-month university students likely to experience suicidal ideation. In: 19th International Conference on Artificial Intelligence Applications and Innovations, 14 June 2023 - 17 June 2023, León, Spain.

DOI: https://doi.org/10.1007/978-3-031-34111-3_40

Publisher: Springer

Version: Accepted Version

Downloaded from: https://e-space.mmu.ac.uk/631811/

Usage rights: © In Copyright

Additional Information: This version of the conference paper has been accepted for publication, after peer review (when applicable) and is subject to Springer Nature's AM terms of use, but is not the Version of Record and does not reflect post-acceptance improvements, or any corrections. The Version of Record is available online at: http://dx.doi.org/10.1007/978-3-031-34111-3_40

Enquiries:
If you have questions about this document, contact openresearch@mmu.ac.uk. Please include the URL of the record in e-space. If you believe that your, or a third party's rights have been compromised through this document please see our Take Down policy (available from https://www.mmu.ac.uk/library/using-the-library/policies-and-guidelines)

Probabilistic Decision Trees for Predicting 12-Month University Students Likely to Experience Suicidal Ideation

Efthyvoulos Drousiotis[1], Dan W Joyce[2], Robert C Dempsey[3], Alina Haines[4], Paul G Spirakis[5], Lei Shi[6], and Simon Maskell[1]

[1] Department of Electrical Engineering and Electronics, University of Liverpool, Liverpool L69 3GJ, UK; {E.Drousiotis,S.Maskell}@liverpool.ac.uk
[2] Institute of Population Health, University of Liverpool, Liverpool L69 3GJ, UK; D.Joyce@liverpool.ac.uk
[3] Department of Psychology, Manchester Metropolitan University, Manchester M15 6BH, UK; r.dempsey@mmu.ac.uk
[4] Department of Nursing, Manchester Metropolitan University, Manchester M15 6BH, UK; a.haines@mmu.ac.uk
[5] Department of Computer Science, University of Liverpool, Liverpool L69 3BX, UK spirakis@liverpool.ac.uk
[6] School of Computing, Newcastle University, Newcastle NE4 5TG, UK; Lei.Shi@newcastle.ac.uk

Abstract. Environmental stressors combined with a predisposition to experience mental health problems increase the risk for SI (Suicidal Ideation) among college/university students. However, university health and wellbeing services know little about machine learning methods and techniques to identify as early as possible students with higher risk. We developed an algorithm to identify university students with suicidal thoughts and behaviours using features universities already collect. We used data collected in 2020 from the American College Health Association (ACHA), a cross-sectional population-based survey including 50,307 volunteer students. A state-of-the-art parallel Markov Chain Monte Carlo (MCMC) Decision tree was used to overcome overfitting problems and target classes with fewer representatives efficiently. Two models were fitted to the survey data featuring a range of demographic and clinical risk factors measured on the ACHA survey. The first model included variables universities would typically collect from their students (e.g., key demographics, residential status, and key health conditions). The second model included these same variables plus additional suicide-risk variables which universities would not typically measure as standard practice (e.g., students' sense of belonging at university). Models' performance was measured using precision, recall, F1 score, and accuracy metrics to identify any potential overfitting of the data efficiently.

Keywords: Bayesian Decision Trees · Suicidal Ideation (SI) · ACHA

1 Introduction and Related Work

Suicidal thoughts, intent, plans, and attempts are concepts with varying definitions [18], resulting in variability of measurement and consequently difficulty establishing robust conclusions about which populations or individual-level clinical factors merit either screening or, at least, surveillance. Prediction for an individual's absolute risk of dying by suicide is difficult, not least because tools claiming to demonstrate predictive performance have significant caveats – for example, even in robustly developed, validated and calibrated suicide risk prediction tools [16], for thresholds of a 1% risk of suicide they display very low positive predictive value (i.e. they are poor at discriminating those who will commit suicide) but a significantly higher negative predictive value (i.e. the tool can reliably predict those who will *not* commit suicide). This pattern of relatively poor positive predictive performance alongside neglect of *how* these tools could augment human actors' work in helping people at risk of suicide [40] has led to some national guidelines recommending *against* using any structured assessment tools for managing the risk of suicide. For example, in the United Kingdom, the National Institute for Health and Care Excellence guidelines advise against using risk stratification tools [26].

The relationship between Suicidal Ideation (SI) to either attempting or committing suicide is complex. For example, the assumption of a straightforward 'linear progression' from SI to action is not empirically supported. Moreover, several key models of suicide, such as the Interpersonal Theory of Suicide [38], the Integrated Motivational-Volitional model [27], and the Three Step Theory of Suicide [21] separate the development of SI from the progression of ideation to an attempt. Given this complexity, locating individuals who might benefit from enhanced support and early intervention for both mental health problems and features associated with suicide may be logistically and practically more appropriate.

A study of UK university students designed to test the central tenets of the Three Step Theory [11] identified that SI develops in the presence of self-reported psychological pain and hopelessness cognitions. Further, "connectedness" (to others) appears to play a role in preventing the escalation of ideation to action, but also, a lack of social connectedness was independently associated with SI. University or college can be stressful for students, especially for those who are psychologically vulnerable and have poor support. University students are at an elevated risk for SI, planning, and attempts [15] potentially arising due to increased academic pressure, psycho-social stress, heavy workloads, and difficulties adapting to a new environment [5]. Around $16\% - 25\%$ of students have experienced some form of SI during university, and approximately 40% and 20% of students with SI reported that they have considered or attempted suicide respectively [25].

Early identification of at-risk students is important for effective suicide prevention and allows universities to refer students to appropriate well-being, counseling, and pastoral support. Various influences and predictors on the experience of suicidal ideation amongst university students have been identified in the existing literature. In terms of mental health-related variables, the experience of more severe depressive symptoms, psychosis, and greater perceived stress have been associated with higher suicide risk [1, 32], with similar associations between increased mental health symptoms and past 12-month suicidal ideation reported [3, 33]. Whilst there have been limited prospective studies, there is evidence that increased suicide risk over 12 months amongst students is associated with the experience of clinically significant mood and other common mental health conditions (e.g. generalised anxiety), stressful life events, childhood adversities, reporting of physical or sexual assault [6]. Various other risk factors for self-harm and suicidality amongst university students have been identified, including sleep difficulties [36], alcohol use problems [9, 29], being of a younger age and/or status as an undergraduate student [32], as well as interpersonal difficulties, such as unsupportive family environments [42].

Theoretical models of suicide [21, 38] suggest addressing SI as a potentially modifiable factor in preventing suicide. Identifying students with SI is, however, challenging as most of the studies [34] rely on information about an individual's mental health obtained from (e.g.) interactions with skilled pastoral support workers. Most of the previous studies [2, 8, 10], including those aiming to predict SI among different populations, focus on identifying predictors for SI using post-hoc, inferential statistical analyses, which may be helpful to inform predictive models. For example, inferential statistical methods also require *a priori* models describing the relationships between predictors and outcomes to test a specific hypothesis. As a result, there may be a number of different pathways to suicidality amongst at-risk university students considering the various risk factors that may interact to increase the vulnerability to suicidal thoughts and behaviours (i.e., suicide-related events).

The majority of the literature [5, 15, 20, 22, 23, 37, 41] uses traditional statistical and linear methods to analyse and identify putative factors that may be predictors of SI. Linear models and inferential statistical hypothesis testing often assume linearity of individual predictor's contribution to an outcome and very often, the models assume additivity of risk. Given the complexity of suicide in general, such approaches could hamper the ability to locate variables (or their combinations) that inform decision-making in a meaningful way [19, 39]. Melissa et al. [24] conducted the only study that uses a non-linear machine learning (Random Forest) algorithm to identify possible predictors for students with SI considering 70 features. However, identifying factors associated with SI does not necessarily imply that they could help predict future students with SI [7].

The aforementioned studies were mainly conducted as explanatory analyses to help inform universities about the possible factors leading to SI in student populations. In contrast, our study attempts to efficiently and effectively predict students with SI using data universities and colleges possess before a student starts their studies. Universities need understandable algorithms that are easily applied to data they already have, for example, to stratify students at risk of SI and to target resources appropriately to those who might most benefit.

Most of the papers in the literature tend to report analyses conducted on samples that contain relatively small amounts of data, with some large inconsistencies in sample sizes across previous studies. This problem is previously addressed [35], where the median number of participants across studies related to suicidal ideation was 79 and the mean 710. The features sample size in past studies have varied from 1085 to 5572 [15, 20, 22, 24, 37, 41]. To address this issue of inconsistent and relatively small sample sizes used in past research, we conducted machine learning-based analyses on a sample of 50,307 students from the American College Health Association (ACHA) survey to provide a more robust and reliable understanding of the predictors of suicidal ideation in university students. To achieve this aim, we conducted two separate analyses. Firstly, a machine learning model featuring data that universities are likely to collect as part of their normal operations was fitted to the data (e.g., information about the student's general health). Secondly, we tested a more comprehensive model featuring this information plus a variety of variables from the ACHA survey implicated in suicidality (e.g., social connectedness variables) which universities may not normally collect from their students. We ran these two models to investigate which offers the more robust and accurate predictive model of suicidal ideation in university students and identify possible important vulnerability factors for suicidality in students that universities may need to collect in the future. A scalable and transparent state-of-the-art machine learning algorithm, a parallel MCMC Decision Tree, processed the models.

2 Methods

2.1 Dataset and Data Preparation

The dataset used in this study is provided by ACHA[7](American College Health Association), a nationally recognised research survey that provides precise data about students' health habits, behaviours, and perceptions. The dataset contains information from 2020 of 50,307 students (68% $females$, 31% $males$, $<$ 1% $intersex$) and 694 features including demographics, Likert scale questions, and multiple choice questions (MCQ).

The mean age of the sample was 22.5 years. Most students reported living either on campus or in university housing (39.6%) or off-campus or in non-university housing(46.4%). 12.8% of the students lived in their parents'/guardian' or other family member's house, 0.2% temporarily stayed with a relative, friend, or "couch surfing" until they found housing, $<$ 0.1% did not (currently) have a place to live, and 0.9% lived in other types of accommodation. 52% of the students were not in a relationship, 39.3% were in a relationship, and 8.7% were married or partnered. Of the 50,307 participants in the sample, 53.8% never thought about killing themselves, 25.7% had a brief passing thought about suicide/planning suicide, 15.3% had a plan at least once to kill themselves, and 5.2% had attempted to kill themselves. Specifically, 55.7% of girls, 57.8% of boys, and

[7] https://www.acha.org/ACHA/Resources/Survey$_Data/ACHA/Resources/Survey_Data.aspx$

44.4% of intersex students never thought about killing themselves, 27.3% of girls, 25.9% of boys, and 22.2% of intersex students had a brief passing thought about suicide/planning suicide, 11.1% of girls, 12.9% of boys, and 27.3% of intersex students had a plan at least one to kill themselves, and 5.9% of girls, 3.4% of boys, and 5.6% of intersex students had attempted to kill themselves. The main baseline characteristics did not significantly differ based on gender.

From the 694 available features, we used only the 16 features that universities already had (information asked during their UCAS application) to predict students' SI. For the secondary analysis, we used 182 features implicated in suicide risk which universities did not typically collect from students. Due to the large number of features we used for the secondary model, please find the table containing the features and the associated features' importance on the supplementary material. Our target variable asks the students if they thought about suicide in the past year, meaning that we predict students with suicidal ideation. For this particular question, there are five possible answers, never (1), rarely (2), sometimes (3), often (4), and very often (5). Participants who chose answers 2 to 5 were coded as positive for SI. We avoided any data pre-processing techniques (Oversampling, Undersampling, Principal Component Analysis, etc.), as we wanted our data to be as original as possible and keep the transparent nature of our algorithm, and allow practical conclusions to be drawn based on the raw data as collected on the ACHA survey. Table1 demonstrates the imbalanced nature of our data, showing the number of students who reported SI compared to those who did not report SI. Specifically, 34,626 students didn't report any SI during the last year, while 15,681 students reported at least one SI during the last year.

Number of students	Percentage	SI
34626	68.3%	No(0)
15681	31.17%	Yes(1)

Table 1. Percentage of students reporting SI

2.2 Markov Chain Monte Carlo Decision Tree

A decision tree typically starts with a root node, which branches into possible outcomes. Each of those outcomes leads to additional decision nodes, which branch off into other possibilities ending up in leaf nodes. This gives it a tree-like shape.

Our model describes the conditional distribution of y given x, where x is a vector of predictors $[x = (x_1, x_2, ..., x_p)]$. The main components of the $tree(T)$ include the depth of the tree, $(d(T))$, the features, $(k(T))$, and the thresholds, $(c(T))$, for each node where $\theta = [k(T), c(T)]$, and the conditional probabilities $p(Y|T, \theta, x)$ for each leaf node, $(L(T))$. If x lies in the region corresponding to the i_{th} terminal node, then $y|x$ has distribution $f(y|\theta_i)$, where f represents a parametric family indexed by θ_i. The model is called a probabilistic classification tree, according to the quantitative response y.

As Decision Trees are identified by (θ, T), a Bayesian analysis of the problem proceeds by specifying a prior probability distribution, $p(\theta, T)$. Because θ indexes the parametric model for each T, it will usually be convenient to use the relationship

$$p(Y_1 :_N, T, \theta | x_1 :_N) = p(Y|T, \theta, x) p(\theta|T) p(T) \tag{1}$$

The Metropolis-Hastings (MH) algorithm for simulating the Markov Chain in Decision Trees (see equation 2) is defined as follows. Starting with an initial tree T_0, iteratively simulate the transitions from T_i to $T_i + 1$ by these two steps:

1. Generate a candidate value T' with probability distribution $q(T_i, T')$.
2. Set $T_{i+1} = T'$ with probability

$$a(T_i, T') = \min(1, \frac{\pi(Y_1 :_N, T', \theta' | x_1 :_N)}{\pi(Y_1 :_N, T, \theta | x_1 :_N)} \frac{q(T, \theta | T', \theta')}{q(T', \theta' | T, \theta)}) \tag{2}$$

Otherwise set $T_{i+1} = T_i$.

More information and a detailed explanation of the algorithm can be found here [14]

To evaluate our predictive model's performance, we used the following four metrics.

- Precision: the ratio of the correctly predicted positive observations to the total predicted positive observations.
- Recall: the ratio of correctly predicted positive observations to all observations in the actual positive class.
- F1-score: the weighted average of Precision row and Recall row. Therefore, this score takes both false positives and false negatives into account.
- Accuracy: the most intuitive performance measure and it is simply a ratio of correctly predicted observations to the total observations.

We also performed cross-validation by randomly splitting the initial dataset into 10 folds, where each fold was used as a test, and we repeated this process until all the folds were used as test sets.

The main advantage of the probabilistic machine learning (ML) models over conventional ML is that they are known to generalise better on imbalanced data and are less overfitting prone [13] [12], allowing us to avoid data pre-processing techniques. We have further modified the algorithm to produce even more accurate results by adding special weights to the students with SI. Specifically, we instructed our algorithms to classify a student as a 0 (people with no SI) only if it is more than 69% confident. This practice enabled us to further increase our performance metrics (accuracy, F1-score) and fight off overfitting due to the imbalanced nature of the dataset. In general, we believe that applying heavy data pre-processing techniques alters the dataset nature, leading us to solve a different problem, ending up having algorithms working only for the specific dataset and not being able to generalise. Our philosophy is to change and modify the algorithms to fit the problem, not vice versa.

3 Results

3.1 Predicting

For the students with SI using the features universities normally collect (e.g., through university application forms), the model had an out-of-bag error of $29.2\% \pm 0.9$, leading to an accuracy of $70.8\% \pm 0.9$. The predictive (0) values were 0.77, whereas the predictive (1) values were 0.55, meaning that 77% and 55% of predicted cases were actually cases. Table 2 shows the scores analytically for predicting students with SI. Specifically precision score is $65.33\% \pm 0.7$, recall score is $64.16\% \pm 0.4$, and f1-score $64.8\% \pm 0.4$ Analyses of the importance of the variables for the prediction, measured by the times each variable is used on the predictive model, revealed that the following four variables were the most predictive, as shown in Table 3: Depression, Eating Disorder, Approximate Cumulative Average Grade, and Attending Classes, Discussion sections, or Labs.

3.2 Secondary Analysis

For the students with SI using the features universities might not ordinarily collect, the model had an out-of-bag error of $27.1\% \pm 3.96$, leading to an accuracy of $72.9\% \pm 3.96$. The predictive (0) values were 0.83, whereas the predictive (1) values were 0.58, meaning that 83% and 58% of predicted cases were actually cases. Table 2 shows the scores analytically for predicting students with SI. Specifically precision score is $70.30\% \pm 2.28$, recall score is $71.3\% \pm 2.23$, and f1-score $69.8\% \pm 2.99$ Analyses of the importance of the variables for the prediction, measured by the times each variable is used on the predictive model, revealed that the following four variables were the most predictive, as shown in Table 3: Financial Problems, Bullying, Allergies to pets/animals, Gastroesophageal Reflux Disease/Acid Reflux.

	Precision	Recall	F1-score	Accuracy
With Uni Features	65.33 ± 0.74	64.16 ± 0.37	65.83 ± 0.37	70.83 ± 0.89
Without Uni Features	70.30 ± 2.28	71.30 ± 2.23	69.80 ± 2.99	72.9 ± 3.96

Table 2. Metrics for predicting students with SI

Features	**Feature Importance**(%)
Age	1.6
Attending Classes, Discussion sections, or Labs	8.0
Sex assigned at birth	5.6
Enrollment Status	3.60
Black or African American	4.2
Middle Eastern/North African or Arab Origin	1.3
Biracial or Multiracial	5.5
Approximate Cumulative Average Grade	9.1
Blind/Low Vision	5.2
Parent or Guardian of a Child	7.6
Anxiety(Generalized Anxiety, Social Anxiety, anic Disorder,	6.8
Bipolar and Related Conditions (Bipolar I, II, Hypomanic Episode)	7.4
Diabetes or pre-diabetes/insulin resistance	2.7
Depression(Major Depression, Persistent Depressive	15.0
Eating Disorder(Anorexia Nervosa, Bulimia Nervosa, Binge - Eating)	10.0
Insomnia	5.4

Table 3. Features Importance for Variables Universities Collect from Students

4 Discussion

Predicting which university students are at higher risk of SI and potentially suicidal behaviours is a difficult task. Past studies in this area have been limited by using relatively small and unrepresentative samples of data and relying on researcher-led choices of which data to use in predictive models of suicidal ideation, meaning a lack of consistency and comprehensiveness in previous research [15, 22, 24]. To address these issues, we applied machine learning approaches to understand the variables associated with suicidal ideation amongst students based on an existing large survey of US students (over 50,000 participants) and the testing of predictive variables that universities routinely and seldom collect from incoming students.

Using a parallel MCMC Decision Tree model for the features universities already have, we found that four main baseline variables predicted SI: Depression, Eating Disorder, Approximate Cumulative Average Grade, and Attending Classes(Discussion sections, Labs) with a significance level of 15%, 10%, 9.1%, 8% accordingly. The model including those variables showed a good predictive performance (accuracy $= 70.83 \pm 0.89$) estimated using cross-validation. In secondary analyses in a wider sample of (number of) features, the main predictive variables differed from the main analyses. Having Financial Problems contributed to a 4.13%, Bullying contributed to a 3.03%, Allergies to pets/animals contributed to a 4.13%, and Gastroesophageal Reflux Disease or acid reflux contributed to a 3.87%. The model, including the dataset with bigger feature space, had an improved predictive accuracy than the one using only the universities' typical features, with a predictive accuracy of 72.9 ± 3.96. We also achieved better results for both test cases (predicting students with SI, secondary analyses) compared to other studies predicting students with SI utilising a machine learning model(Random Forest). Specifically, [24] achieved 0.4 and 0.36 predicted (1) values for girls and boys, respectively, meaning that 40% and 36% of predicted cases were actual cases. In comparison, we achieved 0.55 and 0.58, which leads to a significant increase of 15% for girls and 19% for boys when we use features universities have, while 18% and 22% improvement achieved for our secondary analyses.

Machine learning approaches offer universities potentially powerful means of understanding the risk factors for suicidality amongst their student populations. There may be the potential for universities to use similar models at a local level considering risk factors that may be unique to their campuses, location, and student population. Understanding more local-level risk factors may be important for university health and wellbeing services to better identify those at risk for suicidal ideation and to provide more targeted early intervention support for students at a heightened risk.

There are some strengths and limitations to consider with the present study. As discussed earlier, previous studies of the risk factors associated with suicidal ideation amongst university students have been limited by their analysis of relatively small samples [20, 37]. In contrast, our study drew on data from a large national sample of US university students (over 50,000 students) and applied machine learning approaches to develop predictive models of students' suicidal ideation. Based on a prior call [17], the present machine learning study has allowed for the modelling of numerous variables in a predictive model of suicidal ideation. Such models offer a more detailed understanding of university student suicidal ideation and accommodate the modelling of potentially hundreds of predictors and their complex inter-relationships, compared to the dominance of regression-based analyses in the literature, which only accommodate the testing of relatively simplistic models of suicidality [17]. In addition, the use of an existing, large, representative national survey to model potential predictors of suicidality amongst university students reduced potential research ethical issues associated with collecting suicide-related data from at-risk individuals.

There are, however, some limitations to the present study to consider. The study analysed data from an ongoing national US student survey from only one-time point. Given the complex and dynamic nature of suicidality, how these factors identified in the machine learning model influence suicidality over the longer term requires further exploration. It should also be noted that the specific predictors identified here may not be generalisable to students in other countries, where there may be more local and unique pressures on students implicated in suicidal ideation. We also focused on the experience of suicidal ideation as a broad outcome in the machine learning models and did not include detailed assessments of the types of suicidal thoughts students experienced (such as active planning versus more passive ideation), and so care should be taken in assuming that these factors are similarly implicated in suicidal behaviours amongst students. Although, it is important to identify the potential factors implicated in the suicidal ideation-to-enaction process, particularly those associated with earlier suicidal thoughts where targeted interventions may be particularly effective. In addition, the survey data used in the machine learning approach here did not feature many key psychological variables implicated in the suicidality pathways, such as feelings of defeat and entrapment [17, 30, 31]. The factors in the models also tended to be more risk-focused in nature rather than encapsulate more protective factors against suicidality, and did not explicitly test existing models of suicidality which attempt to outline the ideation-to-enaction pathway [4, 27, 28]. Integrating factors associated with reduced suicidality, including more factors of a bio-psycho-social basis, may be promising for future machine learning approaches focusing on understanding suicidality in at-risk populations.

5 Conclusion

University students are a high-risk group for suicidal thoughts, feelings, and behaviours, but predicting which specific students are at higher risk is a difficult endeavour. Machine learning-based approaches offer a unique way of understanding suicide risk based on their ability to model a large and complex range of factors at the same time. Still, few machine learning approaches have been used to understand suicide risk in university students. Our study differs from most of the literature, as we discussed in section1. We trained a state-of-the-art MCMC Decision Tree with a large sample (over 50,000 participants) for the first time. We showed that such a machine learning-based approach could significantly contribute towards identifying and predicting suicidal ideation among university students. Unlike the other studies, we focused on the actual predictive model and how to produce optimal solutions instead of only identifying possible factors leading students to SI. Our approach can potentially help universities quickly identify and provide early interventions targeting students with these suicide-risk factors. Moreover, our model outperforms significantly any other similar implementation by an average of 17%, and 20% when a wider sample of features is used. This study, though, has some limitations. The study focused only on SI and not suicidal behaviours, as we should note that the results may not be generalisable to students in other countries.

References

1. Umair Akram, Antonia Ypsilanti, Maria Gardani, Kamila Irvine, Sarah Allen, Asha Akram, Jennifer Drabble, Eleanor Bickle, Lauren Kaye, Damian Lipinski, et al. Prevalence and psychiatric correlates of suicidal ideation in uk university students. *Journal of affective disorders*, 272:191–197, 2020.
2. Zakiah Mohamad Ashari, Yek Er Liow, and Nurul Farhana Binti Zainudin. Psychological risk factors and suicidal ideation among undergraduate students of a malaysian public university•. *Jurnal Kemanusiaan*, pages 33–40, 2022.

3. Jason R Bantjes, Ashraf Kagee, Taryn McGowan, and Henry Steel. Symptoms of posttraumatic stress, depression, and anxiety as predictors of suicidal ideation among south african university students. *Journal of American college health*, 64(6):429–437, 2016.
4. Shira Barzilay and Alan Apter. Psychological models of suicide. *Archives of suicide research*, 18(4):295–312, 2014.
5. Maria Jesús Blasco, Pere Castellví, José Almenara, Carolina Lagares, Miquel Roca, Albert Sesé, José Antonio Piqueras, Victoria Soto-Sanz, Jesús Rodríguez-Marín, Enrique Echeburúa, et al. Predictive models for suicidal thoughts and behaviors among spanish university students: rationale and methods of the universal (university & mental health) project. *BMC psychiatry*, 16(1):1–13, 2016.
6. Maria Jesús Blasco, Gemma Vilagut, Itxaso Alayo, José Almenara, Ana Isabel Cebrià, Enrique Echeburúa, Andrea Gabilondo, Margalida Gili, Carolina Lagares, José Antonio Piqueras, et al. First-onset and persistence of suicidal ideation in university students: A one-year follow-up study. *Journal of Affective Disorders*, 256:192–204, 2019.
7. Danilo Bzdok, Gael Varoquaux, and Ewout W Steyerberg. Prediction, not association, paves the road to precision medicine. *JAMA psychiatry*, 78(2):127–128, 2021.
8. Chin Wen Cong and Wu Shin Ling. The predicting effects of depression and selfesteem on suicidal ideation among adolescents in kuala lumpur, malaysia: Received 2019-10-10; accepted 2020-01-06; published 2020-04-17. *Journal of Health and Translational Medicine*, 23(1):60–66, 2020.
9. William Coryell, Adam Horwitz, Ronald Albucher, Kai Zheng, Jacqueline Pistorello, Daniel Eisenberg, Todd Favorite, and Cheryl King. Alcohol intake in relation to suicidal ideation and behavior among university students. *Journal of American college health*, pages 1–5, 2021.
10. Munmun De Choudhury, Emre Kiciman, Mark Dredze, Glen Coppersmith, and Mrinal Kumar. Discovering shifts to suicidal ideation from mental health content in social media. In *Proceedings of the 2016 CHI conference on human factors in computing systems*, pages 2098–2110, 2016.
11. Katie Dhingra, E David Klonsky, and Vojna Tapola. An empirical test of the three-step theory of suicide in uk university students. *Suicide and Life-Threatening Behavior*, 49(2):478–487, 2019.
12. Efthyvoulos Drousiotis, Panagiotis Pentaliotis, Lei Shi, and Alexandra I Cristea. Capturing fairness and uncertainty in student dropout prediction–a comparison study. In *International Conference on Artificial Intelligence in Education*, pages 139–144. Springer, 2021.
13. Efthyvoulos Drousiotis, Panagiotis Pentaliotis, Lei Shi, and Alexandra Cristea I. Balancing fined-tuned machine learning models between continuous and discrete variables-a comprehensive analysis using educational data. 2022.
14. Efthyvoulos Drousiotis and Paul G. Spirakis. Single mcmc chain parallelisation on decision trees. In Dimitris E. Simos, Varvara A. Rasskazova, Francesco Archetti, Ilias S. Kotsireas, and Panos M. Pardalos, editors, *Learning and Intelligent Optimization*, pages 191–204, Cham, 2022. Springer International Publishing.
15. Mehmet Eskin, Jian-Min Sun, Jamila Abuidhail, Kouichi Yoshimasu, Omar Kujan, Mohsen Janghorbani, Chris Flood, Mauro Giovanni Carta, Ulrich S Tran, Anwar Mechri, et al. Suicidal behavior and psychological distress in university students: a 12-nation study. *Archives of suicide research*, 20(3):369–388, 2016.

16. Seena Fazel, Achim Wolf, Henrik Larsson, Susan Mallett, and Thomas R Fanshawe. The prediction of suicide in severe mental illness: development and validation of a clinical prediction rule (oxmis). *Translational psychiatry*, 9(1):1–10, 2019.
17. Joseph C Franklin, Jessica D Ribeiro, Kathryn R Fox, Kate H Bentley, Evan M Kleiman, Xieyining Huang, Katherine M Musacchio, Adam C Jaroszewski, Bernard P Chang, and Matthew K Nock. Risk factors for suicidal thoughts and behaviors: A meta-analysis of 50 years of research. *Psychological bulletin*, 143(2):187, 2017.
18. Benjamin Goodfellow, Kairi Kolves, and Diego De Leo. Contemporary nomenclatures of suicidal behaviors: a systematic literature review. *Suicide and Life-Threatening Behavior*, 48(3):353–366, 2018.
19. Holly Hedegaard and Margaret Warner. Suicide mortality in the united states, 1999-2019. 2021.
20. Corey LM Keyes, Daniel Eisenberg, Geraldine S Perry, Shanta R Dube, Kurt Kroenke, and Satvinder S Dhingra. The relationship of level of positive mental health with current mental disorders in predicting suicidal behavior and academic impairment in college students. *Journal of American college health*, 60(2):126–133, 2012.
21. E David Klonsky and Alexis M May. The three-step theory (3st): A new theory of suicide rooted in the "ideation-to-action" framework. *International Journal of Cognitive Therapy*, 8(2):114–129, 2015.
22. Anne C Knorr, Brooke A Ammerman, Alexander J Hamilton, and Michael S McCloskey. Predicting status along the continuum of suicidal thoughts and behavior among those with a history of nonsuicidal self-injury. *Psychiatry research*, 273:514–522, 2019.
23. Cindy H Liu, Courtney Stevens, Sylvia HM Wong, Miwa Yasui, and Justin A Chen. The prevalence and predictors of mental health diagnoses and suicide among us college students: Implications for addressing disparities in service use. *Depression and anxiety*, 36(1):8–17, 2019.
24. Melissa Macalli, Marie Navarro, Massimiliano Orri, Marie Tournier, Rodolphe Thiébaut, Sylvana M Cote, and Christophe Tzourio. A machine learning approach for predicting suicidal thoughts and behaviours among college students. *Scientific reports*, 11(1):1–8, 2021.
25. Philippe Mortier, P Cuijpers, Glenn Kiekens, RP Auerbach, Koen Demyttenaere, JG Green, RC Kessler, MK Nock, and R Bruffaerts. The prevalence of suicidal thoughts and behaviours among college students: a meta-analysis. *Psychological medicine*, 48(4):554–565, 2018.
26. NICE. Self-harm: assessment, management and preventing recurrence. URL: https://www.nice.org.uk/guidance/ng225.
27. Rory C O'Connor and Olivia J Kirtley. The integrated motivational–volitional model of suicidal behaviour. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373(1754):20170268, 2018.
28. Rory C O'Connor and Matthew K Nock. The psychology of suicidal behaviour. *The Lancet Psychiatry*, 1(1):73–85, 2014.
29. Siobhan O'Neill, Margaret McLafferty, Edel Ennis, Coral Lapsley, Tony Bjourson, Cherie Armour, Sam Murphy, Brendan Bunting, and Elaine Murray. Sociodemographic, mental health and childhood adversity risk factors for self-harm and suicidal behaviour in college students in northern ireland. *Journal of affective disorders*, 239:58–65, 2018.

30. Rebecca Owen, Robert Dempsey, Steven Jones, and Patricia Gooding. Defeat and entrapment in bipolar disorder: exploring the relationship with suicidal ideation from a psychological theoretical perspective. *Suicide and Life-Threatening Behavior*, 48(1):116–128, 2018.
31. Rory C O'Connor and Gwendolyn Portzky. The relationship between entrapment and suicidal behavior through the lens of the integrated motivational–volitional model of suicidal behavior. *Current opinion in psychology*, 22:12–17, 2018.
32. Myra Parker, Bonnie Duran, Isaac Rhew, Maya Magarati, Leo Egashira, Mary Larimer, and Dennis Donovan. Prevalence of moderate and acute suicidal ideation among a national sample of tribal college and university students 2014–2015. *Archives of suicide research*, 25(3):406–423, 2021.
33. Md Estiar Rahman, Md Saiful Islam, Mohammed A Mamun, Mst Sabrina Moonajilin, and Siyan Yi. Prevalence and factors associated with suicidal ideation among university students in bangladesh. *Archives of suicide research*, 26(2):975–984, 2022.
34. Geoffrey L Ream. The interpersonal–psychological theory of suicide in college student suicide screening. *Suicide and Life-Threatening Behavior*, 46(2):239–247, 2016.
35. Jessica D Ribeiro, Joseph C Franklin, Kathryn Rebecca Fox, Kate H Bentley, Evan M Kleiman, Bernard P Chang, and Matthew K Nock. Self-injurious thoughts and behaviors as risk factors for future suicide ideation, attempts, and death: a meta-analysis of longitudinal studies. *Psychological medicine*, 46(2):225–236, 2016.
36. Kirsten Russell, Stephanie Allan, Louise Beattie, Jason Bohan, Kenneth MacMahon, and Susan Rasmussen. Sleep problem, suicide and self-harm in university students: A systematic review. *Sleep medicine reviews*, 44:58–69, 2019.
37. Geumsook Shim and Bumseok Jeong. Predicting suicidal ideation in college students with mental health screening questionnaires. *Psychiatry investigation*, 15(11):1037, 2018.
38. Kimberly A Van Orden, Tracy K Witte, Kelly C Cukrowicz, Scott R Braithwaite, Edward A Selby, and Thomas E Joiner Jr. The interpersonal theory of suicide. *Psychological review*, 117(2):575, 2010.
39. Colin G Walsh, Jessica D Ribeiro, and Joseph C Franklin. Predicting risk of suicide attempts over time through machine learning. *Clinical Psychological Science*, 5(3):457–469, 2017.
40. Daniel Whiting and Seena Fazel. How accurate are suicide risk prediction models? asking the right questions for clinical practice. *Evidence-based mental health*, 22(3):125–128, 2019.
41. Holly C Wilcox, Amelia M Arria, Kimberly M Caldeira, Kathryn B Vincent, Gillian M Pinchevsky, and Kevin E O'Grady. Prevalence and predictors of persistent suicide ideation, plans, and attempts during college. *Journal of affective disorders*, 127(1-3):287–294, 2010.
42. Hui Zhai, Bing Bai, Lu Chen, Dong Han, Lin Wang, Zhengxue Qiao, Xiaohui Qiu, Xiuxian Yang, and Yanjie Yang. Correlation between family environment and suicidal ideation in university students in china. *International journal of environmental research and public health*, 12(2):1412–1424, 2015.

View publication stats