Journal of Quality Measurement and Analysis
e-ISSN: 2600-8602
https://doi.org/10.17576/jqma.2103.2025.04

JQMA 21(3) 2025, 61-78
http://www.ukm.my/jqma

A COMPARISON OF DECISION TREE, LOGISTIC REGRESSION, ARTIFICIAL NEURAL NETWORK AND RANDOM FOREST ALGORITHMS TO PREDICT SUICIDAL IDEATION AMONG YOUNG ADULTS IN MALAYSIA

(*Penggunaan Algoritma Pepohon Keputusan, Regresi Logistik, Rangkaian Neural Buatan dan Hutan Rawak untuk Meramal Pemikiran Bunuh Diri di Kalangan Dewasa Muda di Malaysia*)

CHAN SIN YIN* & CH’NG CHEE KEONG

ABSTRACT

Suicide is a significant global public health issue, and Malaysia is no exception, with a high incidence rate. On average, approximately 10 suicide deaths occur daily in the country, alongside numerous attempted suicides. Hence, the key indicators for suicidal ideation should be identified so that communities can be aware of the characteristics of suicide attempters and assist them. Therefore, this study aims to develop predictive models using four predictive techniques, which are Decision Tree, Logistic Regression, Artificial Neural Network and Random Forest to anticipate suicidal ideation among young adults in Malaysia. By analysing key indicators, such as demographic, socio-economic, and psychological factors, the model seeks to enable proactive intervention and support for vulnerable individuals. A total of 33 predictive models are generated and evaluated based on their performance using the misclassification rate. Among these models, Gini decision tree models with 2 and 3 branches (80:20) showed superior performance, with the lowest misclassification rate recorded at 19.44%. Consequently, the model with 2 branches is selected for its practicality and accuracy in identifying vulnerable individuals. Early intervention is crucial in identifying and supporting young adults at risk of suicidal ideation. The developed predictive model offers valuable insights for proactive intervention and support, aiding in prevention efforts and reducing the prevalence of suicide. It carries significant policy implications for suicide prevention in Malaysia, enabling targeted intervention strategies for vulnerable young adults. By prioritising resources based on the identified risk factors, policymakers can enhance mental health support systems and prevent tragic outcomes.

Keywords: data mining; decision tree; prediction model; suicidal ideation among young adults; suicide

ABSTRAK

Bunuh diri adalah isu kesihatan awam global yang signifikan, dan Malaysia tidak terkecuali, dengan kadar kejadian yang tinggi. Secara purata, sekitar 10 kematian bunuh diri berlaku setiap hari di negara ini, di samping banyak yang cuba membunuh diri. Oleh itu, penunjuk utama bagi pemikiran bunuh diri perlu dikenal pasti supaya masyarakat dapat mengenali ciri-ciri individu yang cuba membunuh diri dan membantu mereka. Justeru itu, kajian ini bertujuan untuk membangunkan model ramalan menggunakan empat teknik ramalan iaitu Pepohon Keputusan, Regresi Logistik, Rangkaian Neural Buatan dan Hutan Rawak bagi meramalkan ideasi bunuh diri dalam kalangan dewasa muda di Malaysia. Dengan menganalisis penunjuk utama, seperti faktor demografi, sosio-ekonomi, dan psikologi, model ini bertujuan untuk membolehkan campur tangan proaktif yang lebih tepat dan memberi sokongan kepada individu yang berisiko. Sebanyak 33 model ramalan dihasilkan dan dinilai berdasarkan prestasi mereka menggunakan kadar ralat. Antara model-model ini, model Pepohon Keputusan indeks Gini dengan 2 dan 3 cabang (80:20) menunjukkan prestasi yang lebih baik, dengan kadar ralat terendah direkodkan pada 19.44%. Oleh itu, model dengan 2 cabang dipilih kerana kepraktisan dan ketepatannya dalam mengenal pasti individu yang berisiko. Pencegahan awal adalah sangat penting dalam mengenal pasti dan menyokong golongan dewasa muda yang berisiko mengalami pemikiran bunuh diri. Model ramalan yang dibangunkan ini menawarkan panduan yang berguna dalam usaha pencegahan dan mengurangkan prevalens bunuh diri. Ia membawa implikasi dasar yang penting untuk pencegahan bunuh diri di Malaysia, membolehkan pengambilan tindakan yang disasarkan untuk golongan dewasa muda yang berisiko. Dengan mengutamakan sumber berdasarkan faktor risiko yang dikenalpasti, penggubal dasar boleh memperkukuhkan sistem sokongan kesihatan mental dan mengelakkan akibat yang tragik.

Kata kunci: perlombongan data; pepohon keputusan; model ramalan; pemikiran bunuh diri dalam kalangan dewasa muda; bunuh diri

1. Introduction

Suicide persists as a significant global public health concern claiming the lives of 703,000 people end their lives every year, while many others attempt suicide (World Health Organization 2024). Each suicide further increases the number of individuals vulnerable to the risk of self-harm (Lew *et al.* 2022). Among young adults, suicide stands as the foremost cause of death, ranking within the top three primary causes of mortality for this age group (Centers for Disease Control and Prevention 2024). A study was constructed on people with suicidal ideation aged 18 and above, it was found that 17.7 per cent of them persisted in suicide plans and 10.6 per cent had ended their lives (Koh *et al.* 2023). Reducing the worldwide suicide death rate by one-third by 2030 is a target for both the World Health Organization (WHO) Mental Health Action Plan 2013-2030 and the United Nations Sustainable Development Goals (SDGs) (Lew *et al.* 2022).

Suicide is also a critical social problem in Malaysia (Mahathir 2021). The Ministry of Health in Malaysia reports that the suicide rate might reach 10–13 per 100,000 persons or 10 suicide deaths per day (Relate Malaysia n.d.). In Malaysia, there are at least 15 times more suicide attempts occur than completed suicides. Suicide by hanging, poisoning (by pesticides), jumping off a building and poisoning by car exhaust gas are the most common methods of suicide according to the Department of Statistics Malaysia (Lew *et al.* 2022). Besides, suicide rates are the highest among ethnic Indians, men, and those under the age of 40 (Relate Malaysia n.d.).

Suicide rates have been alarming since the beginning of the COVID-19 outbreak in 2020. Suicide cases in Malaysia rose from 609 in 2019 to 631 in 2020. The situation escalated further in 2021, witnessing an alarming 81% increase, with 1,142 recorded cases compared to the previous year (Mahathir 2021). Hospitals under the Ministry of Health claimed that 1,080 suicide attempt cases have been treated from March 2020 until May 2021 (Latfi *et al.* 2023). The number has decreased to 981 in 2022 but in 2023 it increased by 10% to 1,087 cases (Ova 2024).

The tendency of young adults to commit suicide is alarmingly high in Malaysia. Among 1,708 suicide cases that were reported in Malaysia between January 2019 and May 2021, 51% of them (872) were aged 15-18 (Aingaran 2023). Numerous suicide cases have been reported in recent years. For example, a Chinese male student aged 17 from Johor ended his life by jumping from the 4th floor of a mall due to constant bullying by his two classmates (Sheralyn 2020). Other than that, a 24-year-old law student attempted suicide by setting himself on fire and running to the oncoming road traffic after failing the exam 3 times (Lok 2023). A 40-year-old woman also took her life by jumping from the 21st floor of an apartment (Selvam 2024). These suicide cases highlight the need to address such avoidable deaths.

Aside from the loss of lives, suicide also may result in monetary loss and distress for friends and family (Koh *et al.* 2023). Therefore, critical factors that contribute to suicidal ideation should be identified to reduce suicide cases in Malaysia (Chan & Ch'ng 2022). According to the World Health Organization (2024), suicide is associated with several factors including mental disorders, previous suicide attempts, a sense of isolation and many else. Other risk factors for suicidal ideation are discussed in the following section.

1.1. *Factors contributing to suicide risk*

This section discusses the important risk factors that lead to suicidal intentions. Suicidal behaviour among youth is strongly correlated to family factors such as child abuse, parents with drug and alcohol addiction, divorced families, and poor relationships with family members. As a result of these factors, youth might engage in dangerous behaviours such as drug use, alcohol use, and even suicide (Abdu *et al.* 2020; Bilsen 2018; Costa *et al.* 2019).

Financial problem is also a key factor for suicide among young adults. They may feel stressed owing to the inability to cover their educational fees and living expenses (Embing *et al.* 2020). Stevenson and Wakefield (2021) found that a person who is faced with a financial crisis is more likely to have the intention to suicide.

Suicide is undeniably caused by several circumstances such as hopelessness (Embing *et al.* 2020). Hopelessness makes the individual feel that their life will not get better in the future (Primananda & Keliat 2019). People experiencing feelings of hopelessness are at an increased risk of losing interest in life and eventually, contemplating suicide (Lyu & Zhang 2019).

Low self-esteem also leads to suicidal thoughts as it perpetuates a continuous state of sadness and depression in individuals. People with low self-esteem tend to feel worthless and unable to cope with challenges. These feelings may make people more likely to develop suicidal thoughts when confronted with problems in life (Embing *et al.* 2020; Owusu-Ansah *et al.* 2020).

Suicidal thoughts are also associated with stress (Primananda & Keliat 2019). Academic stress, workload overload, interpersonal problems, financial problems, lack of leisure time, and other stressors for young adults are more likely to induce mental illness that might lead to suicidal ideation (Bilsen 2018; Primananda & Keliat 2019).

Costa *et al.* (2019) asserts that individuals who smoke are prone to commit suicide. Furthermore, long-term alcohol and drug use puts a person at a significant risk of suicide ideation (Junior *et al.* 2020). This statement is in line with the findings from Abdu *et al.* (2020) which conclude that alcohol use is closely related to suicidal behaviour.

Religion is crucial in helping people reduce stress and anxiety. As a result, participating in religious activities may lessen the likelihood of persons having suicidal thoughts (Abdu *et al.* 2020). Moreover, suicide is also highly prohibited in practically all religions (Gearing & Alonzo 2018; Nguyen *et al.* 2020). Hence, it can be concluded that lack of religion is one of the key factors that influence suicidal ideation among young adults.

People with poor social support are more likely to experience a lack of belonging. They are 1.66 times more likely to have suicidal thoughts than those who have good social relationships (Abdu *et al.* 2020). The majority of suicide attempters demonstrated social isolation, hence it can be said that poor social support increases the likelihood of suicide attempts (Pillay 2021).

Lyu and Zhang (2019) claimed that health problems are one of the key predictors of suicidal ideation. Individuals grappling with debilitating physical illnesses, obesity, disabilities, and other health-related problems appear to be more sensitive to suicidal thoughts and suicide attempts (Pillay 2021; Yu *et al.* 2021).

According to Bilsen (2018), 25-33% of suicides were committed by those with a history of self-harm or suicide. Hence, it can be said that individuals who have made previous suicide attempts are at a higher likelihood of inflicting self-harm again in the future (Bilsen 2018; Olfson *et al.* 2018).

According to Bilsen (2018), Pillay (2021) and Wasserman *et al.* (2021), some personality traits, such as the inability to manage an array of emotions, poor problem-solving skills and negative thinking are more likely to lead to feelings of insecurity, low self-esteem, emotional crises, and suicide. The scarcity of funds to meet academic and living expenses heightens the susceptibility of young adults to enter into a crisis leading to suicidal tendencies (Almaghrebi 2021; Berkelmans *et al.* 2021; Embing *et al.* 2020).

Social pressure is another important component that contributes to suicidal ideation. For instance, when students face societal pressures from sources like friends, social media, professors, family, and roommates at university, they are prone to experiencing heightened stress which might lead to suicidal ideation (Embing *et al.* 2020).

As stated by Bilsen (2018), stressful events such as marital problems, sexual abuse, cyberbullying, and the loss of friends or family members exert a profound negative influence on young people. Moreover, childhood trauma, including physical, sexual and emotional abuse might also contribute to suicidal ideation (Pillay 2021; Wasserman *et al.* 2021).

Numerous studies about suicide have been conducted in recent years. Most of the studies have discovered a range of biological, psychological, and sociocultural factors for suicide among different populations. Apart from that, past research also focused on developing the prediction model using various data mining techniques to identify the characteristics of the suicide attempters. However, the suicide rates among young adults in Malaysia are still high. Thus, this shows that more research should be done to discover more key factors that are highly associated with suicidal ideation among young adults.

Most of the previous studies only include a few factors in their studies. This paper will include more critical factors compared to previous studies when developing the prediction model using data mining techniques. This can help to identify the warning signs of the suicide attempters and assist them earlier. This paper also allows for the identification of key risk factors among youth specific to the Malaysian context, enabling a more targeted approach to suicide prevention.

Predictive modelling in data mining is indispensable for combating suicide ideation by facilitating early identification of individuals at risk. These models delve into historical data to discern patterns and risk factors associated with suicidal behaviour, enabling proactive intervention and support. By tailoring intervention strategies based on individual characteristics, predictive models optimise resource allocation and prioritise assistance for those most in need. Moreover, they play a crucial role in evaluating the effectiveness of intervention programs over time, informing evidence-based practices to reduce the incidence of suicidal behaviour and enhance mental health outcomes in communities.

1.2. *Data mining techniques*

Data mining is found critical in uncovering hidden patterns and usable insights from large amounts of data. Several data mining methodologies are used to comprehend the complexity and interplay of various aspects to construct the prediction model (Ishaq *et al.* 2021). Data mining methodologies have been widely used in a variety of industries, including education, healthcare, scientific applications, business, and advertising (Ramya & Rohini 2021). Below are some of the data mining approaches that are frequently employed in suicide topics.

1.2.1. *Decision tree*

Decision tree (DT) is widely used for identifying or classifying high-risk subjects (Amini *et al.* 2016). Each tree is formed by nodes and branches. Each node represents a characteristic in the category to be classified, and each subset determines a value that the node can take (Swain &

Hauska 1977). ID3, C4.5, and Classification and Regression Trees (CART) are among the methods often employed in decision tree construction. The difference between these methods is the splitting criteria. The two most common splitting criteria, Entropy and Gini index are used in this study (Zalar *et al.* 2018). Entropy is used to evaluate the impurity or uncertainty of the data. Entropy is always between zero and one. The closer the value gets to zero, the better (Charbuty & Abdulazeez 2021). The Gini index determines the level of impurity or class inequality in the presented dataset (Breiman *et al.* 1984). The value falls between 0 and 1, where 1 denotes inequality and 0 denotes equality. Eq. (1) shows the equation for Entropy index while Eq. (2) is the equation for Gini index.

$$Entropy = \sum_{j} p_j log_2 p_j \qquad (1)$$

p_j = ratio of the sample number of the subset,

j-th = attribute value.

$$Gini\ (t) = 1 - \sum_{j} [p(j|t)]^2 \qquad (2)$$

where,
$p(j|t)$ = relative frequency of class j at node t.

1.2.2. *Logistic regression*

Logistic regression (LR) is utilised when categorical dependent variables are present (Niu 2020). The primary goal of this method is to determine the association between the dependent variable and one or more independent variables, which could be continuous or categorical (Ranganathan *et al.* 2017). The formula of the logistic regression model is as follows:

$$log\left[\frac{p}{1-p}\right] = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + \cdots + b_nx_n \qquad (3)$$

where,
j = the probability of the event occurring,
$p\ /1 - p$ = the odds ratio,
b = parameter of x,
x = independent variables.

LR employs various variable selection techniques, among which the most commonly utilised are forward selection, backward elimination and stepwise selection. Backward selection is the simplest variable selection method, beginning with a full model that includes all variables. The least significant variable is removed iteratively until only significant variables remain (Chowdhury & Turin 2020). In contrast, forward selection starts with no variables and adds the most significant one at each step, without eliminating any once included (Bursac *et al.* 2008). Stepwise selection combines both approaches, allowing variables to be added and removed in a bidirectional process. Following each addition, all variables are reviewed, and insignificant ones are excluded. The process ends when only significant variables remain in the model (Chowdhury & Turin 2020).

1.2.3. *Artificial neural network*

Artificial Neural Network (ANN) are made up of neurons stacked in layers that translate input into output. It includes a procedure for taking input, applying a function to it, and passing the result to the next layer. The nodes would receive outputs from preceding nodes and process them further, the output would be used to forecast future values (Kumar & Garg 2018).

1.2.4. *Random forest*

Random forest (RF) consists of an ensemble of independent decision trees built from randomly selected subsets of features. This robust modelling approach effectively captures the joint distribution of features while reducing the risk of overfitting (Bayramli *et al*. 2022). Generally, RF outperforms single DT classifiers, such as CART and C4.5 (Zhao & Zhang 2008).

1.3. ***Application of data mining techniques in suicide***

Previous studies have been done to identify the key factors associated with suicidal ideation among different population groups using the logistic regression technique. Based on the studies conducted by Costa *et al*. (2019), family history of mental disorder and childhood and adolescent abuse were the important predictors for suicide attempts among employees whereas educational level, childhood maltreatment and smoking habits were among the predictors for students. The risk factors associated with suicide in elderly people have also been found using the LR technique. Some of the factors discovered were staying alone, being unemployed, having no social support, being depressed, feeling hopeless and extremely lonely (Niu 2020). LR was found useful in detecting the risk factors for suicide.

Apart from that, Lyu and Zhang (2019) utilised ANN to develop a prediction model for suicide attempts. In this study, 12 risk factors for suicide were screened. Results showed that the suicide history of family members, mental illness, poor health condition, hopelessness, etc were among the risk predictors for suicide attempts. In addition, Lin *et al.* (2022) developed a prediction model for future multiple suicide attempts using the ANN approach. According to a study done by Morales-Rodríguez et al. (2023), ANN was used to anticipate suicide risk among young people. Findings indicated that the significant independent variables associated with suicidal risk are anxiety, coping skills, emotional intelligence, and perfectionist level.

Based on previous studies, DT is the most popular data mining technique that is often used in predicting suicide attempts. For instance, DT was used by Lin *et al.* (2022) and Pereira *et al.* (2024) to predict the tendency of individuals to reattempt suicide and the models showed a high accuracy rate. Apart from that, there were also several studies which employed DT in the topic of suicide (Lee *et al.* 2021; Shafiee & Mutalib 2020). Therefore, DT is selected in this paper since it is one of the most effective tools for data mining and it has been widely employed in numerous areas because it is easy to apply, free of ambiguity, and stable even when some missing values exist (Song & Lu 2015).

Although numerous researches have been done on the topic of suicide, a predominant focus has been on adolescents, university students, and the elderly population (Chin & Wu 2020; Cho *et al*. 2021; Islam *et al*. 2018). Hence, there is a notable gap in research models specifically addressing suicidal ideation among young adults in Malaysia. This paper aims to address this void by employing the widely used data mining techniques including DT, LR, RF and ANN, to predict the likelihood of suicidal ideation among young adults. This is critical in aiding government bodies, universities, schools, communities, and any other authorities in discerning the characteristics of potential suicide attempters. Such insight enables timely interventions to be implemented in the early stage, preventing potential tragedies.

2. Methodology

The target respondents for this survey are young adults in Malaysia, specifically those aged between 15 and 40 years old. The questionnaire seeks to gather the participants' perceptions regarding the risk factors associated with suicidal ideation. Prior to taking the questionnaire, the respondents are required to complete the Patient Health Questionnaire-9 (PHQ-9). PHQ-9 is a 9-item self-report questionnaire designed to evaluate the prevalence and severity of depression symptoms over the previous two weeks. Each item is rated on a 4-point scale ranging from 0 (not at all) to 3 (nearly every day), with higher scores reflecting higher severity. Only people with moderate to severe depression levels are used as samples.

A two-step approach is employed. Stratified sampling is initially used to determine the required sample size for each state, followed by convenient random sampling to select respondents within each state. The questionnaires utilise both online platforms through Google Forms link and physical delivery methods.

The study derives its insights into the risk factors for suicidal ideation from existing research. The 13 factors under examination in this study are family factor (F1), financial problems (F2), hopelessness (F3), low self-esteem (F4), stress (F5), substance abuse (F6), lack of religion (F7), poor social support (F8), health problem (F9), previous suicide attempts (F10), negative personality characteristics (F11), social pressure (F12) as well as negative life events that break down into several subgroups: life-threatening accident (F13), sexual harassment (F14), physical abuse (F15), family member being physically abused (F16), loss of loved one (F17), someone close committed suicide (F18) and personal loss or relationship problem (F19) while suicidal ideation (SI) as the target variable. Moreover, the demographics of the respondents including age (D1), gender (D2) and monthly income (D3), occupation (D4), race (D5), religion (D6) and state (D7) are also examined. Table 1 below illustrates the details of the variables included in this study.

Table 1: Model role, measurement level, and description of variables

Variable	Model roles	Measurement level	Description
D1	Input	Nominal	Age of the respondent
D2	Input	Binary	Gender of the respondent
D3	Input	Nominal	The monthly income of the respondent
D4	Input	Nominal	Occupation of the respondent
D5	Input	Nominal	Race of the respondent
D6	Input	Nominal	Religion of the respondent
D7	Input	Nominal	State of the respondent
F1	Input	Interval	This factor determines if the respondent has a family problem.
F2	Input	Interval	This factor determines if the respondent has financial problems.
F3	Input	Interval	This factor determines if the respondent feels hopeless.
F4	Input	Interval	This factor determines if the respondent lacks confidence.
F5	Input	Interval	This factor determines if the respondent feels stress.
F6	Input	Interval	This factor determines if the respondent takes drugs or alcohol.

Table 1 (continued)

F7	Input	Interval	This factor determines if the respondent believes in religion.
F8	Input	Interval	This factor determines if the respondent lacks social support.
F9	Input	Interval	This factor determines if the respondent has a health problem.
F10	Input	Interval	This factor determines if the respondent has attempted suicide before.
F11	Input	Interval	This factor determines if the respondent has negative personality characteristics.
F12	Input	Interval	This factor determines if the respondent experiences societal pressure.
F13	Input	Binary	This factor determines if the respondent has been involved in a life-threatening accident.
F14	Input	Binary	This factor determines if the respondent has ever been a sexual harassment victim.
F15	Input	Binary	This factor determines if the respondent has been physically abused or hurt.
F16	Input	Binary	This factor determines if the respondent has seen family members being physically abused or hurt.
F17	Input	Binary	This factor determines if the respondent lost a loved one.
F18	Input	Binary	This factor determines if the family member or someone close to the respondent has committed suicide before.
F19	Input	Binary	This factor determines if the respondent experienced a significant personal loss or relationship break-up.
SI	Target	Binary	This factor determines if the respondent has suicidal ideation.

3. Data Analysis

Figure 1 shows the research process. During the data selection phase, only the respondents aged between 15 and 40 with a PHQ-9 score indicating moderate depression or above are chosen as the sample. 891 data points were accepted and underwent the research. The second phase involves data preprocessing to ensure the validity and feasibility of the data. This includes addressing missing values by replacing them with the mean value. In the data partition phase, the data are split into two sets, which are training and validation, with the ratios of 60:40, 70:30 and 80:20 respectively. Then, the next step involves the development of the predictive models including DT, LR, RF and ANN. In the final stage, the models are assessed using the misclassification rate, with the model exhibiting the lowest misclassification rate deemed the best and utilised for predicting suicidal ideation among young adults. Cross-validation is also implemented to evaluate the performance of the model. Microsoft Excel and SAS Enterprise Miner software are employed for data preprocessing and analysis.

The data is split into three groups - 60:40, 70:30 and 80:20. Within each data partition, four types of predictive models are generated, including DT, LR, ANN and RF. Multiple DT, LR and ANN models are created across the three data partition sets. There are five models for DT: Entropy with 2 branches, Entropy with 3 branches, Gini with 2 branches, Gini with 3 branches and interactive DT model, three models for LR: forward, backward and stepwise, two models for ANN: 2 hidden units and 3 hidden units while only one model for RF as shown in Figure 2. Since there are three data partition sets, a total of 33 models are developed.

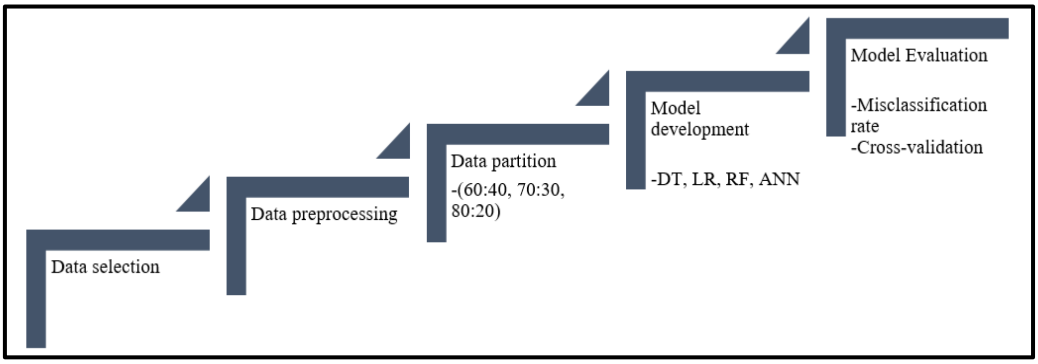

Figure 1: Research flow

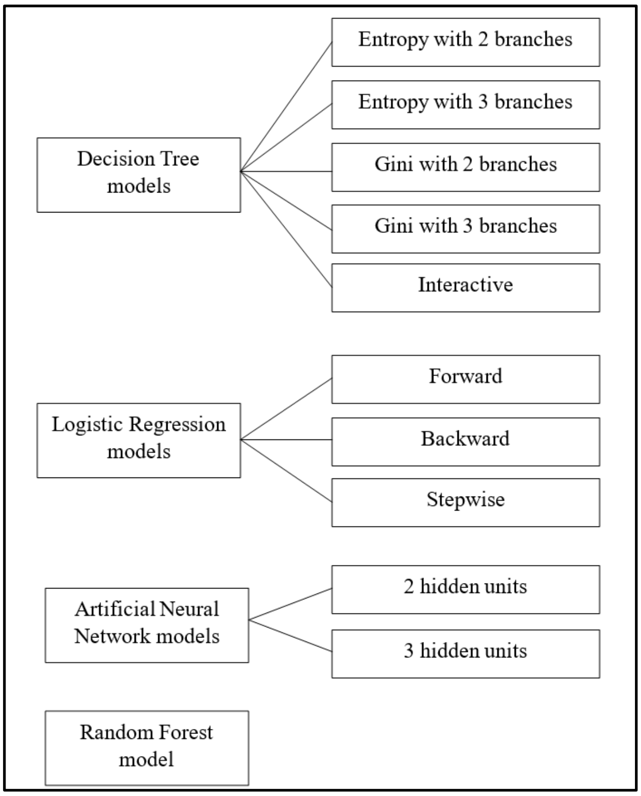

Figure 2: DT, LR, ANN and RF models

4. Analysis and Results

This section presents the characteristics of the data and the analysis results.

4.1. *Descriptive analysis*

Figure 3 shows the demographics of the respondents. A total of 891 data sets were collected for this study, with 501 identified as female and 390 as male. Most respondents fall within the age range of 20 to 24 which is 485 people. The distribution by ethnicity reveals that Malay constitutes the largest proportion (46.35%), followed by Chinese (40.07%), Indians (12.23%), and other ethnicities (1.35%). Additionally, the percentage of respondents with Islam religion (46.35%) surpasses those from other religious backgrounds.

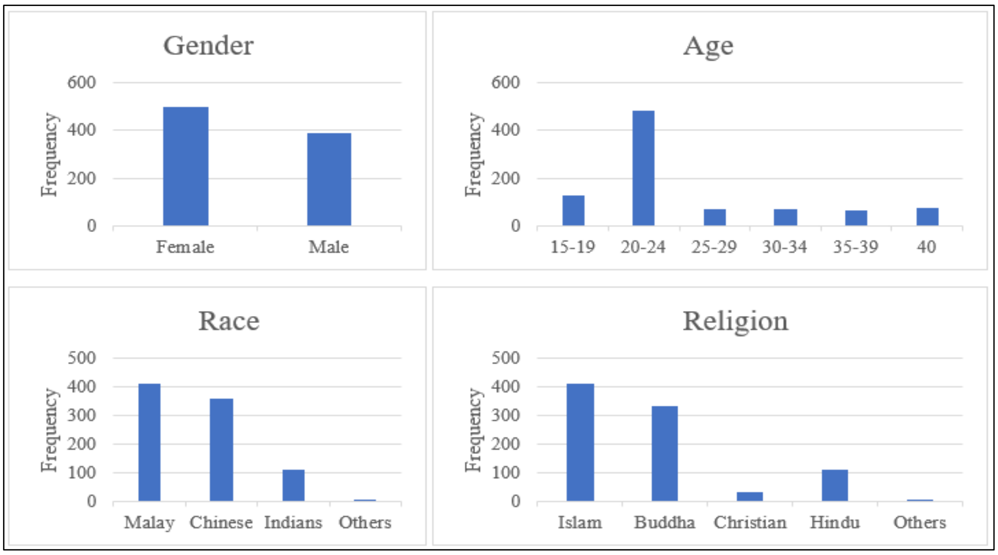

Figure 3: Demographics of respondents (gender, age, race, religion)

Figure 4 represents the demographics of respondents in the aspects of occupation, monthly income, state and suicidal ideation. Since most of the respondents are students (515 respondents), the group with no income exhibits the highest frequency (518 respondents) in the dataset. Kedah (21.44%) and Selangor (16.50%) had a higher proportion of responses than the other states.

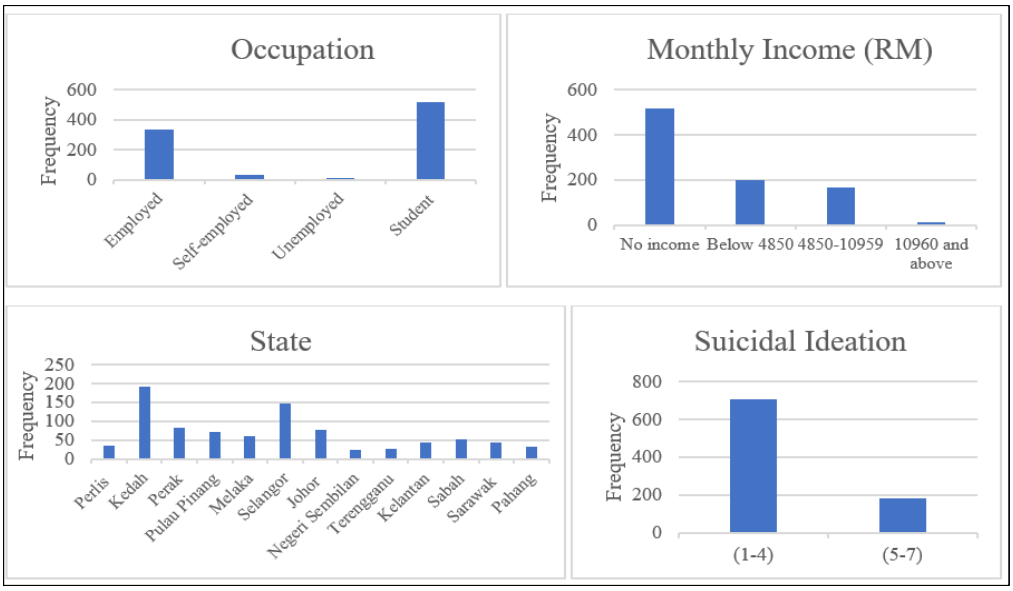

Figure 4: Demographics of respondents (occupation, monthly income, state, suicidal ideation)

The data have been classified into two groups, which are young adults who have a lower likelihood of experiencing suicidal ideation (1-4) and those who are more prone to suicidal ideation (5-7). According to the data, the number of individuals who are more vulnerable to suicidal ideation (182 respondents) is lower than that of those who are not (709 respondents). Predictive modelling will be carried out to predict the characteristics of individuals prone to having suicidal thoughts.

According to Table 2, the factor of negative personality characteristics (NPC) has the highest mean in both groups (1-4) and (5-7), while substance abuse has the lowest mean. This means that negative personality characteristics are the most significant factor in suicidal ideation. Standard deviation in both classes has the largest value in substance abuse as opposed to stress. Almost all factors have having normal range of skewness, which is between -2 to 2, except for substance abuse.

Table 2: Mean, std. deviation and skewness of the 12 factors that influence suicidal ideation

Class		**F1**	**F2**	**F3**	**F4**	**F5**	**F6**
(1-4)	Mean	2.05	3.87	3.08	3.83	3.13	1.82
	Std. Deviation	1.07	1.25	0.96	1.24	0.79	1.83
	Skewness	0.86	-0.02	0.17	-0.22	0.57	2.12
(5-7)	Mean	1.98	4.19	3.13	4.05	3.12	1.87
	Std. Deviation	0.99	1.22	0.89	1.10	0.85	1.88
	Skewness	1.02	-0.41	0.99	0.00	1.09	2.02
Class		**F7**	**F8**	**F9**	**F10**	**F11**	**F12**
(1-4)	Mean	2.77	3.35	3.08	2.48	4.19	2.70
	Std. Deviation	1.15	1.06	1.13	1.11	1.41	1.03
	Skewness	0.73	-0.08	0.28	0.60	-0.22	0.34
(5-7)	Mean	2.77	3.61	3.15	2.82	4.40	2.86
	Std. Deviation	1.09	0.93	1.10	1.26	1.27	0.99
	Skewness	0.55	0.42	0.31	0.78	0.12	0.59

4.2. *Model development and model evaluation*

Upon importing data to Microsoft Excel, preprocessing is initiated to prepare it for analysis, where missing values are replaced with the mean. Besides that, the average of each factor are calculated. Subsequently, the computed data are imported to SAS Enterprise Miner Workstation software. The skewed data is normalised using log transformation through Transform Variable node to enhance model performance. A total of 33 models were developed in this research. All these models are then compared using Model Comparison node and cross-validation is also implemented to evaluate the performance of the model. Figure 5 below depicts an overview of all the models created in this study.

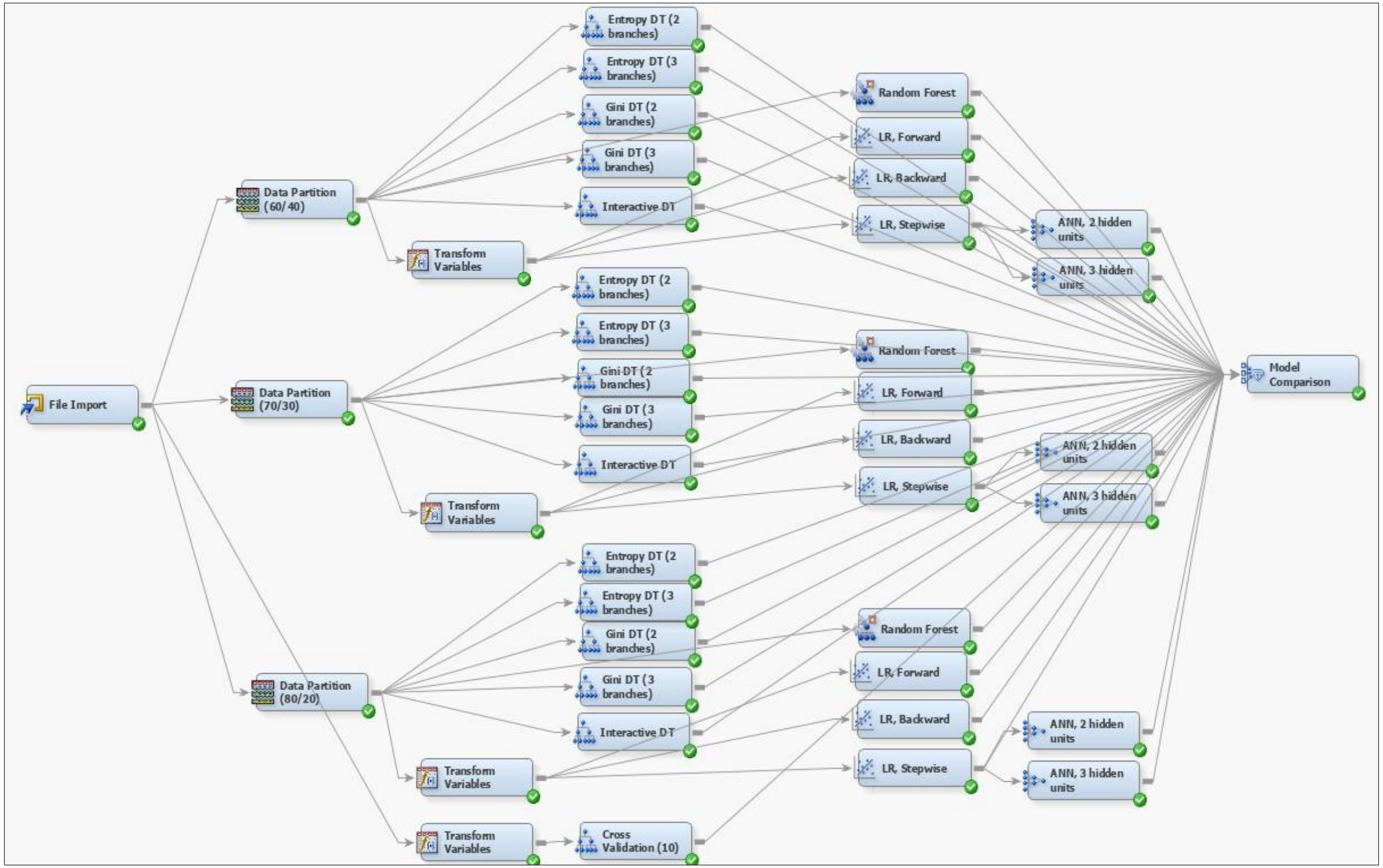

Figure 5: Decision Tree models

Table 3 displays the comparison of misclassification rates for all the models that are constructed. The results reveal that the decision tree models employing Gini with 2 branches and Gini with 3 branches (80:20) outperform other models, exhibiting a misclassification rate of 19.44%. As both models obtained the same misclassification rate, hence only one of them is chosen for predicting the suicidal attempts of young adults in Malaysia. The Gini decision tree model with 2 branches, with the splitting rules 80:20 is selected as the predictive model in this study.

Table 3: Comparison of misclassification rate for all Decision Tree models.

Percentage of data partition (%)	Criteria	Misclassification rate (%)
60:40	Entropy DT (2 branches)	20.06
	Entropy DT (3 branches)	19.50
	Gini DT (2 branches)	20.61
	Gini DT (3 branches)	19.50
	Interactive DT	26.46
	Random Forest	20.61
	LR (Forward)	22.28
	LR (Backward)	22.28
	LR (Stepwise)	22.28
	ANN, 2 hidden units	23.40
	ANN, 3 hidden units	22.56
70:30	Entropy DT (2 branches)	20.74
	Entropy DT (3 branches)	19.63
	Gini DT (2 branches)	20.74
	Gini DT (3 branches)	20.00
	Interactive DT	28.52
	Random Forest	20.74
	LR (Forward)	20.74
	LR (Backward)	20.74
	LR (Stepwise)	20.74
	ANN, 2 hidden units	21.48
	ANN, 3 hidden units	20.74
80:20	Entropy DT (2 branches)	20.56
	Entropy DT (3 branches)	20.00
	Gini DT (2 branches)	**19.44**
	Gini DT (3 branches)	**19.44**
	Interactive DT	20.56
	Random Forest	20.56
	LR (Forward)	22.22
	LR (Backward)	22.22
	LR (Stepwise)	22.22
	ANN, 2 hidden units	23.33
	ANN, 3 hidden units	23.89

To evaluate the model, a 10-fold cross validation model is performed. The results reveal a misclassification rate of 20.43%, corresponding to an overall accuracy of 79.57%, demonstrating the model's predictive capability. Hence, the models are suitable for predicting suicidal ideation risk in young adult populations.

Table 4 shows the variable importance from the selected model. The values represent the importance of each variable in the range of 0 to 1. The most important variable is previous suicide attempts (1.0000), followed by financial problems (0.7751) and age (0.6842).

Table 4: Variables importance

Variable	Importance
Previous suicide attempts	1.0000
Financial problems	0.7751
Age	0.6842

Figure 6 shows the comprehensive decision tree diagram of 2 branches with Gini as the target criterion (80:20). In this decision tree model, the root node is the Previous suicide attempts variable. The model consists of 5 levels, generating a total of 9 nodes. The results highlight the significance of factors such as previous suicide attempts, financial problems, and age in understanding and predicting suicidal ideation among young adults in this study.

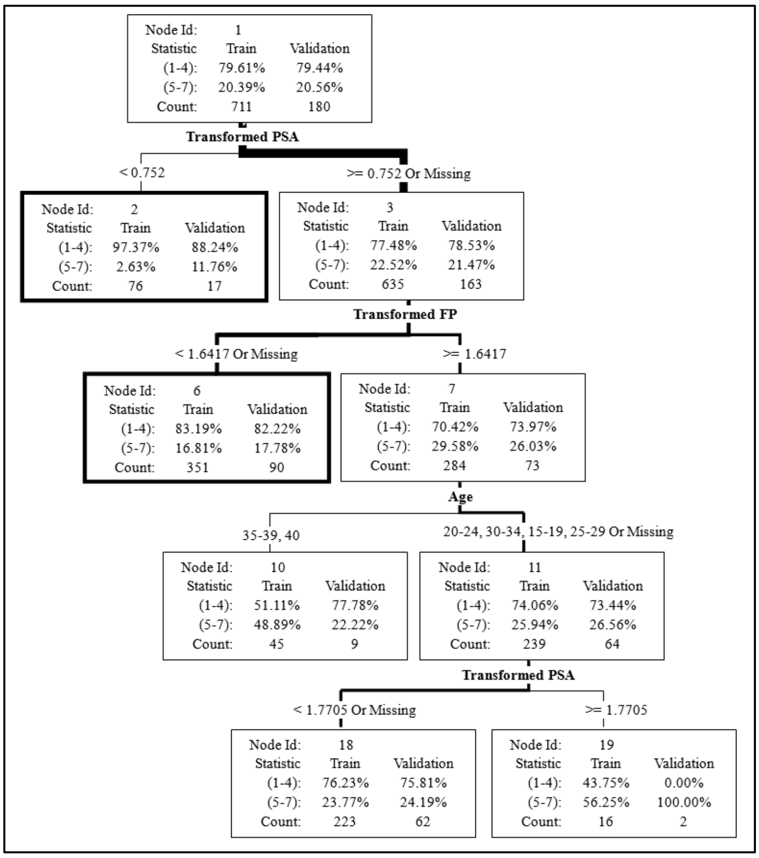

Figure 6: Gini Decision Tree with two branches

Table 5 illustrates the node rules of the decision tree model with an accuracy rate above 80%.

Table 5: Node rules

Node	Node Rules
2	If previous suicide attempts > 1.125 then predicted: suicidal ideation = (5-7) = 97%
6	If previous suicide attempts <= 1.125 or missing and financial problem > 4.1667 or missing then predicted: suicidal ideation= (5-7) = 83%

5. Conclusions

Suicide is a critical global health concern, and in Malaysia, the suicide attempts issue is escalating. Nonetheless, the prospect of suicide prevention is attainable through a comprehensive understanding of the characteristics exhibited by individuals at risk. This understanding facilitates the implementation of timely and appropriate measures to assist them before reaching the point of contemplating or committing suicide. There are several factors that influence a person to possess suicidal ideation such as hopelessness, previous suicide attempts, etc. In this study, various predictive models are constructed to identify the specific characteristics that make individuals more prone to experiencing suicidal ideation. The results indicated that the best model for this case is Gini with 2 branches (80:20) with the lowest misclassification rate of 19.44%. Previous suicide attempts, age, and financial problems are significant indicators for the prediction of suicidal ideation among young adults in Malaysia. Research has consistently demonstrated that those who have attempted suicide in the past are far more likely to make another attempt. Besides that, age is associated with a higher risk of suicidal thoughts, especially in young adults who are transitioning from adolescence to adulthood, and experiencing social, financial, academic or career pressures, etc. Financial instability including struggling with debt, unemployment, or the inability to cover basic needs, can also lead to immense stress and suicidal thoughts. Communities should be aware of these risk factors and pay heightened attention to the individuals exhibiting these characteristics. By doing so, they can actively work towards diminishing suicide risks and contribute to fostering a healthy and harmonious environment for all. These findings enable communities to become aware of critical characteristics of persons who are at a higher risk of suicide at an early stage, allowing for appropriate actions to aid them and prevent suicide from occurring. The limitation of this study is the respondents are selected randomly from each state; thus, the number of respondents is imbalanced in some aspects such as age, gender, occupation, etc. Moreover, since the data utilized in this study was based on self-report measures, it is important to acknowledge the potential for biases and inaccuracies in participants' responses especially when discussing suicide, such a sensitive topic, which may affect the reliability of the findings. Future work can be done to explore more factors and add them to the prediction model to enhance the model. Apart from that, other data mining techniques can be employed and the performance of all the techniques can be compared to obtain the best model for this study.

6. Contribution of Study

Even though the research on suicide issues is on the rise, there are substantial gaps in our understanding of suicidal ideation, particularly in Malaysia. This study contributes significantly to our understanding of suicidal ideation among young adults in Malaysia. Besides that, the findings provide important insights for suicide prevention, allowing for early detection of individuals who are vulnerable to suicidal thoughts and prompt interventions. Furthermore, the study may promote public health policy, assist healthcare practitioners in detecting warning signals, raise awareness through educational initiatives, and contribute to lowering suicide rates and boosting community resilience. Other than that, the findings provide valuable, actionable insights for practitioners. By prioritizing the key predictors highlighted by the decision tree, organizations can make data-driven decisions that optimize resource allocation and guide effective strategic planning.

Acknowledgments

We express sincere gratitude to the School of Quantitative Sciences at Universiti Utara Malaysia (UUM) for their invaluable support and encouragement during this research project. Additionally, we extend heartfelt thanks to the editor and reviewers for their valuable feedback and contributions to this article. Lastly, our sincere appreciation goes to all respondents who generously shared their time and insights, enriching the study with valuable input. This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

References

Abdu Z., Hajure M. & Desalegn D. 2020. Suicidal behavior and associated factors among students in Mettu University, South West Ethiopia, 2019: An institutional based cross-sectional study. *Psychology Research and Behavior Management* **13**: 233–243.

Aingaran R. 2023. Expert: Rise in suicide cases recorded may be 'statistical.' *The Sun*. https://thesun.my/malaysia-news/expert-rise-in-suicide-cases-recorded-may-be-statistical-EG11607737 (3 November 2024).

Almaghrebi A.H. 2021. Risk factors for attempting suicide during the COVID-19 lockdown: Identification of the high-risk groups. *Journal of Taibah University Medical Sciences* **16**(4): 605–611.

Amini P., Ahmadinia H., Poorolajal J. & Amiri M.M. 2016. Evaluating the high risk groups for suicide: A comparison of logistic regression, support vector machine, decision tree and artificial neural network. *Iranian Journal of Public Health* **45**(9): 1179–1187.

Bayramli I., Castro V., Barak-Corren Y., Madsen E.M., Nock M.K., Smoller J.W. & Reis B.Y. 2022. Temporally informed random forests for suicide risk prediction. *Journal of the American Medical Informatics Association* **29**(1): 62-71.

Berkelmans G., van der Mei R., Bhulai S. & Gilissen R. 2021. Identifying socio-demographic risk factors for suicide using data on an individual level. *BMC Public Health* **21**(1): 1702.

Bilsen J. 2018. Suicide and youth: Risk factors. *Frontiers in Psychiatry* **9**: 540.

Breiman L., Friedman J., Olshen R.A. & Stone C.J. 1984. *Classification and Regression Trees*. 1st Ed. Monterey, CA: Wadsworth International Group.

Bursac Z., Gauss C.H., Williams D.K. & Hosmer D.W. 2008. Purposeful selection of variables in logistic regression. *Source Code for Biology and Medicine* **3**: 17.

Centers for Disease Control and Prevention. 2024. Facts About Suicide. Suicide Prevention. https://www.cdc.gov/suicide/facts/index.html (2 November 2024).

Chan S.Y. & Ch'ng C.K. 2022. Factors associated with suicidal ideation among university students in Malaysia. *E-Proceedings of the 4th Young Researchers Quantitative Symposium 2022*.

Charbuty B. & Abdulazeez A.M. 2021. Classification based on Decision Tree Algorithm for machine learning. *Journal of Applied Science and Technology Trends* **2**(01): 20–28.

Chin W.C. & Wu S.L. 2020. The predicting effects of depression and self- esteem on suicidal ideation among adolescents in Kuala Lumpur, Malaysia. *Journal of Health and Translational Medicine* **23**(1): 60–66.

Cho S.E., Geem Z.W. & Na K.S. 2021. Development of a suicide prediction model for the elderly using health screening data. *International Journal of Environmental Research and Public Health* **18(**19): 10150.

Chowdhury M.Z.I. & Turin T.C. 2020. Variable selection strategies and its importance in clinical prediction modelling. *Family Medicine and Community Health* **8**(1): e000262.

Costa A.C.B., Mariusso L.M., Canassa T.C., Previdelli I.T.S. & Porcu M. 2019. Risk factors for suicidal behavior in a university population in Brazil: A retrospective study. *Psychiatry Research* **278**: 129–134.

Embing J., Yusoff S.M. & Othman M.R. 2020. Exploration on perception of suicidal ideation among students of higher institutions. *Journal of Cognitive Sciences and Human Development* **6**(2): 37–51.

Gearing R.E. & Alonzo D. 2018. Religion and suicide: new findings. *Journal of Religion and Health* **57**(6): 2478-2499.

Ishaq A., Sadiq S., Umer M., Ullah S., Mirjalili S., Rupapara V. & Nappi M. 2021. Improving the prediction of heart failure patients' survival using SMOTE and effective data mining techniques. *IEEE Access* **9**: 39707–39716.

Islam M.A., Low W.Y., Tong W.T., Yuen C.C.W. & Abdullah A. 2018. Factors associated with depression among university students in Malaysia: A cross-sectional study. *KnE Life Sciences* **2018**: 415–427.

Junior A.C., Fletes Lemos, T., Teixeira E. & Souza M. deLourdes D. 2020. Risk factors for suicide: systematic review. *Saudi Journal for Health Sciences* **9**(3): 183-193.

Koh Y.S., Shahwan S., Jeyagurunathan A., Abdin E., Vaingankar J.A., Chow W.L., Chong S.A. & Subramaniam M. 2023. Prevalence and correlates of suicide planning and attempt among individuals with suicidal ideation: Results from a nationwide cross-sectional survey. *Journal of Affective Disorders* **328**: 87–94.

Kumar V. & Garg M.L. 2018. Predictive analytics: A review of trends and techniques. *International Journal of Computer Applications* **182**(1): 31–37.

Latfi N.S.I., Mahmud H.H., Kadir A.H.A., Yusof Z.M. & Misiran M. 2023. Factors affecting mental health among workers in Malaysia during Covid-19 pandemic. *AIP Conference Proceedings* **2896**(1): 050020.

Lee Y., Kim H., Lee Y. & Jeong H. 2021. Comparison of the prediction model of adolescents' suicide attempts using logistic regression and decision tree: Secondary data analysis of the 2019 youth health risk behavior web-based survey. *Journal of Korean Academy of Nursing* **51**(1): 40-53.

Lew B., Kõlves K., Lester D., Chen W.S., Ibrahim N.B., Khamal N.R.B., Mustapha F., Chan C.M.H., Ibrahim N., Siau C.S. & Chan L.F. 2022. Looking into recent suicide rates and trends in Malaysia: A comparative analysis. *Frontiers in Psychiatry* **12**: 770252.

Lin I.L., Tseng J.Y.C., Tung H.T., Hu Y.H. & You Z.H. 2022. Predicting the risk of future multiple suicide attempt among first-time suicide attempters: Implications for suicide prevention policy. *Healthcare* **10**(4): 667.

Lok J.W. 2023. Malaysian law student who failed exams 3 times sets himself on fire, tries to get hit by moving truck. *The Straits Times*. https://www.straitstimes.com/asia/se-asia/malaysian-law-student-who-failed-exams-3-times-sets-himself-on-fire-tries-to-get-hit-by-moving-truck (23 November 2024).

Lyu J. & Zhang J. 2019. BP neural network prediction model for suicide attempt among Chinese rural residents. *Journal of Affective Disorders* **246**: 465–473.

Mahathir M. 2021. Rapidly rising suicides should jolt Malaysia's leaders into action. *Nikkei Asia*. https://asia.nikkei.com/Opinion/Rapidly-rising-suicides-should-jolt-Malaysia-s-leaders-into-action (5 October 2024).

Morales-Rodríguez F.M., Martínez-Ramón J.P., Giménez-Lozano J.M. & Morales Rodríguez A.M. 2023. Suicide risk analysis and psycho-emotional risk factors using an artificial neural network system. *Healthcare* **11**(16): 2337.

Nguyen M., Cabral M.D. & Patel D.R. 2020. Suicide in adolescents: exploring the role of religion. *International Journal of Child Health and Human Development* **13**(4): 379–382.

Niu L. 2020. A review of the application of logistic regression in educational research: Common issues, implications, and suggestions. *Educational Review* **72**(1): 41–67.

Olfson M., Wall M., Wang S., Crystal S., Bridge J.A., Liu S.M. & Blanco C. 2018. Suicide after deliberate self-harm in adolescents and young adults. *Pediatrics* **141**(4): e20173517.

Ova. 2024. Suicide cases in Malaysia increase by 10% from 2022 to 2023. *Ova*. https://ova.galencentre.org/suicide-cases-in-malaysia-increase-by-10-from-2022-to-2023/ (5 November 2024).

Owusu-Ansah F.E., Addae A.A., Peasah B.O., Oppong Asante K. & Osafo J. 2020. Suicide among university students: prevalence, risks and protective factors. *Health Psychology and Behavioral Medicine* **8**(1): 220–233.

Pereira C.A., Peixoto R.C., Kaster M.P., Grellert M. & Carvalho J.T. 2024. Using data mining techniques to understand patterns of suicide and reattempt rates in Southern Brazil. *Proceedings of the 2nd International Conference on Biomedical Electronics and Biomedical Informatics*, pp. 385–392.

Pillay J. 2021. Suicidal behaviour among university students: A systematic review. *South African Journal of Psychology* **51**(1): 54–66.

Primananda M. & Keliat B.A. 2019. Risk and protective factors of suicidal ideation in adolescents. *Comprehensive Child and Adolescent Nursing* **42**(sup1): 179–188.

Ramya A. & Rohini K. 2021. A survey about role of data mining techniques and its applications in healthcare sector. *2021 2nd International Conference on Intelligent Engineering and Management*, pp. 277–281.

Ranganathan P., Pramesh C. & Aggarwal R. 2017. Common pitfalls in statistical analysis: Logistic regression. *Perspectives in Clinical Research* **8**(3): 148–151.

Relate Malaysia. n.d. *Suicide in Malaysia*. https://relate.com.my/why-get-help-2/suicide-in-malaysia/ (10 October 2024).

Selvam K. 2024. Woman falls to death from 21st floor of apartment. *Sinar Daily*. https://www.sinardaily.my/article/215625/focus/woman-falls-to-death-from-21st-floor-of-apartment#google_vignette (15 November 2024).

Shafiee N.S.M. & Mutalib S. 2020. Prediction of mental health problems among higher education student using machine learning. *International Journal of Education and Management Engineering* **10**(6): 1-9.

Sheralyn. 2020. 17yo M'sian boy commits suicide after his classmates bullied him for not joining their secret gang. *World Of Buzz*. https://worldofbuzz.com/msian-boy-commits-suicide-after-his-classmates-bullied-him-for-not-joining-their-secret-gang/ (15 November 2024).

Song Y.Y. & Lu Y. 2015. Decision tree methods: applications for classification and prediction. *Shanghai Archives of Psychiatry* **27**(2): 130–135.

Stevenson C. & Wakefield J.R.H. 2021. Financial distress and suicidal behaviour during COVID-19: Family identification attenuates the negative relationship between COVID-related financial distress and mental ill-health. *Journal of Health Psychology* **26**(14): 2665–2675.

Swain P.H. & Hauska H. 1977. The decision tree classifier: Design and potential. *IEEE Transactions on Geoscience Electronics* **15**(3): 142–147.

Wasserman D., Carli V., Iosue M., Javed A. & Herrman H. 2021. Suicide prevention in childhood and adolescence: a narrative review of current knowledge on risk and protective factors and effectiveness of interventions. *Asia-Pacific Psychiatry* **13**(3): e12452.

World Health Organization. 2024. *Suicide*. https://www.who.int/news-room/fact-sheets/detail/suicide (5 October 2024).

Yu R., Chen Y., Li L., Chen J., Guo Y., Bian Z., Lv J., Yu C., Xie X., Huang D., Chen Z. & Fazel S. 2021. Factors associated with suicide risk among Chinese adults: A prospective cohort study of 0.5 million individuals. *PLOS Medicine* **18**(3): e1003545.

Zalar B., Plesnicar B.K., Zalar I. & Mertik M. 2018. Suicide and suicide attempt descriptors by multimethod approach. *Psychiatria Danubina* **30**(3): 317-322.

Zhao Y. & Zhang Y. 2008. Comparison of decision tree methods for finding active objects. *Advances in Space Research* **41**(12): 1955-1959.

Department of Decision Science
School of Quantitative Sciences
Universiti Utara Malaysia
06010 Sintok
Kedah, MALAYSIA
E-mail: chan_sin_yin2@ahsgs.uum.edu.my,chee@uum.edu.my*

Received: 18 January 2025
Accepted: 7 May 2025

*Corresponding author