

Mental Health Prediction Models Using Machine Learning in Higher Education Institution

Sofianita Mutalib¹, Nor Safika Mohd Shafiee², Shuzlina Abdul-Rahman³

^{1,2}Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, 40450 Shah Alam, Selangor Darul Ehsan, Malaysia.

³Research Initiative Group Intelligent Systems, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia.

Article History: Received: 11 January 2021; Accepted: 27 February 2021; Published online: 5 April 2021

Abstract: Today, mental health problem has become a grave concern in Malaysia. According to the National Health and Morbidity Survey (NHMS) 2017, one in five people in Malaysia suffers from depression, two in five from anxiety, and one in ten from stress. Higher education students are also at risk of being part of the affected community. The increased data size without proper management and analysis, and the lack of counsellors, are compounding the issue. Therefore, this paper presents on identifying factors in mental health problems among selected higher education students. This study aims to classify students into different categories of mental health problems, which are stress, depression, and anxiety, using machine learning algorithms. The data is collected from students in a higher education institute in Kuala Terengganu. The algorithms applied are Decision Tree, Neural Network, Support Vector Machine, Naïve Bayes, and logistic regression. The most accurate model for stress, depression, and anxiety is Decision Tree, Support Vector Machine, and Neural Network, respectively.

Keywords: Anxiety, Depression, Higher Education, Stress, Mental Health.

Introduction

Identifying the factors of mental health problems among students has become a challenging task. The factors can be influenced by biological, psychological, and environmental issues. Diagnosis can be tricky and complex as symptoms and factors are often similar; this can lead the doctor to misdiagnose [1], and the wrong treatment being administered to the patient, thus jeopardizing the patient's psychological conditions, both emotionally and behaviourally. The World Health Organization (WHO) defines mental health problems or mental disorders as the combination of abnormal thoughts, emotions, behaviour in daily activities, and relationships with others [2]. The presence of machine learning would help in the extraction of knowledge and may improve the quality of medical practices [3].

Mental Health Problem in Higher Education System In Malaysia

The higher education system in Malaysia is the full responsibility of and operated by the higher education institutions (HEIs) under the authority of the Ministry of Higher Education (MOHE) [4]. They cover public institutions funded by the government and private institutions, with both offering certificate, diploma, undergraduate, and postgraduate programmes. The five levels of higher education qualification are Certificate, Diploma, Bachelor's Degree, Master's Degree, and Doctor of Philosophy (PhD). Based on statistics from the Malaysian Ministry of Education, 552,702 students enrolled at, and 119,345 students graduated from twenty Public Universities in Malaysia until 31 December 2018 [5]. These institutes for higher education help with human development and upward mobility out of poverty by producing high-quality graduates to work as high-income professionals who contribute to the economic sector and the social environment [5].

The university is a place to gain knowledge, although life at university can be challenging and filled with obstacles. However, students can still excel. Nowadays, the majority of students grumble about the high level of stress they experience in their university lives, including feelings of anxiety and depression, especially towards the end of the semester [6]. The level of stress increases as the learning process progresses due to the need to balance assessments, workload, and examinations [7]. Other factors may also affect students' mental health. Students may face a high risk of developing mental health problems due to family issues, uncertainties about their future careers, financial troubles and difficulties arising out of living away from home [8]. Balancing between life at university and other demands or needs can also lead the students to face the risk of developing mental health problems [7], [9].

Students experiencing symptoms of mental health problems have claimed that they are not receiving any treatments and would not seek help to address their emotional troubles. These students do not place any importance on their predicament as their peers also experience similar symptoms, and thus they see this as something common in their university lives [10]. However, some of them are aware of the need for proper treatment, but they lack the courage to seek help and worry too much about other people's perceptions [7],[9]-[10]. They fear that the stigma of being diagnosed with mental health problems may lead to discrimination or

prejudice by society, and they worry about the negative impact of being labelled sick, overly emotional or crazy [11]. Thus, universities need to consider new strategies to encourage students to get diagnosed and receive appropriate treatments for their mental health problems.

A mental health problem or mental illness is a health issue that affects the way a person feels, thinks, behaves, and communicates with others [11]-[12]. According to the American Psychiatric Association, a mental health problem or mental illness is a health condition that affects a person's emotional state, thought process and behaviour, or a combination of several health conditions associated with social, work, or family-related issues [12]. Thus, it can be concluded that mental health issues affect a person's emotions, thoughts, behaviour, connection with society, and daily activities. The types of mental health problems include anxiety disorders, depression, stress, and Schizophrenia [12]. The most common mental health problems in Malaysia are depression, anxiety disorders, and stress [13].

Anxiety Disorders

Anxiety disorders are characterized by overwhelming worry and fear, especially when confronted with problems or decision-making [13]-[14]. The lives of people suffering from anxiety disorders are affected by symptoms of extreme nervousness, anxiousness, and excessive fear [14]. When faced with unpleasant situations, other symptoms may appear, and these include heart palpitations, breathing difficulties, excessive sweating, tremors, or nausea [14]. Anxiety disorders are not restricted to specific conditions or age groups, so anyone can suffer from it [15], especially when coupled with adversities experienced during childhood. There are a few types of anxiety disorders, such as Generalized Anxiety Disorder (GAD), panic disorder, and social anxiety disorder.

People with GAD experience severe anxiety or stress about things like personal safety, jobs, social interactions, and everyday life events on most days for at least six months [15]. They prefer to avoid or seek reassurance in a situation where the result is unpredictable and are unnecessarily concerned about things that might go wrong. [14]. People with panic disorder can suffer panic attacks when they are assaulted by feelings of sudden fear or anxiety [14]-[15]. They become terrified, and may experience heart palpitations, excessive sweating, tremors, shortness of breath, and the sensation of losing control.

Phobia is a type of anxiety disorder where those affected have an intense fear of specific objects or situations [14]. People with phobias may have an irrational concern about a feared object or circumstance, which they also try to avoid [14]. Specific phobia is an excessive and persistent fear of a particular item, situation, or activity that are generally harmless, for example, heights, and animals or insects, such as dogs, spiders, and snakes [14]-[15]. People with social anxiety disorder have extreme fears about their attitude or behaviour being judged by others, causing them to feel embarrassed [14]. They avoid situations that they think might place them at the centre of attention [15]. People with agoraphobia are terrified when faced with any two or more of the following instances: using public transportation, being in open or enclosed spaces, standing in line, being in crowds, and being alone outside the house [14]-[15]. People with separation anxiety disorder are terrified of being apart from those they are emotionally attached to. If separation is happening or anticipated, they may have hallucinations or nightmares about the expected parting.

Depression

Depression is characterized by constant sadness, loss of interest or excitement, feelings of guilt or low self-worth, disturbed sleep, loss of appetite, fatigue, and inability to concentrate [15]. According to the National Institute of Mental Health, depression or clinical depression is a serious mood disorder that causes severe symptoms that affect the way one feels, thinks, and handles daily activities [13]. Depression can cause pain to the person suffering from the ailment and the people around them. It can be a serious health concern as it may lead to suicide [15]. The signs and symptoms of depression include perpetually feeling sad, empty, hopeless, lack of interest in hobbies and activities, and exhausted [13],[15].

There are several types of depression, and a few of them are persistent depressive disorder, postpartum depression and psychotic depression [13]. Persistent depressive disorder, also known as dysthymia, is a state of low mood that lasts for at least two years [15]. A person who is diagnosed with persistent depressive disorder may have major depressive episodes along with periods of less severe symptoms, but signs must last more than two years in order to be considered persistent depressive disorder [15]. People with psychotic depression experience severe depression with some form of psychosis, such as having disturbingly false beliefs, or hearing/seeing disturbing things that others cannot hear/see. The symptoms of psychotic depression typically have a grim "theme," such as delusions of guilt, poverty, or illness [15]. The development of another mood disorder, seasonal affective disorder (SAD), generally happens in winter months when less natural sunlight is available [15]. Winter depression, usually accompanied by social isolation, excessive sleep, and increased weight, emerges and dissipates at the same time every year. A person with bipolar disorder experiences intense mood episodes that shift from the extreme low (which meets the major depression characteristics) to the extreme high, also known as mania (when the person is either euphoric or irritable). A less severe form of mania is known as hypomania [15].

Factors behind Mental Health Problems

Generally, mental health problems are based on biological factors, as well as social and socioeconomic environments [15]. As shown in Table 1, the main factors leading to mental health problems among higher education students are lack of social support, financial troubles, and learning environment. The lack of social support is defined as insufficient support within the community that ultimately increases stress [10]. The support from family members and other people around a student can positively impact them. Otherwise, the student is bound to experience loneliness, hence increasing their stress level, which may lead to mental health problems such as depression and anxiety disorder.

Table 1 Comparison Factor of Mental Health among Higher Education Students

| Factor | Reference |
|------------------------|----------------------|
| Lack of Social Support | [10, 14, 17, 19, 20] |
| Financial Problem | [10, 14, 17, 18, 22] |
| Learning Environment | [10, 14, 17, 18, 21] |
| Being Female | [14, 16, 18, 19] |
| Family Problem | [10, 18, 22] |
| Peer Relationship | [10, 17, 20] |

Biological factor in mental health problems refers to the abnormal functioning of nerve cell circuits or pathways that connect the brain regions, which may be caused by genetics, infection due to brain damage, brain defect, prenatal damage, among other factors. Social environment refers to how a person interacts with their surroundings, culture, or way of life. It is about the person's relationship with their family, friends, colleagues, and local community [16], the lack of social support, and discrimination at the workplace. Socioeconomic environment reflects the person's financial status. Financial difficulty can become a major factor that causes mental health problems as people with low financial standing are prone to stress and anxiety [16]. Learning environment refers to daily life at university, assessments, and learning styles.

Other factors include gender (being female), distance from home [10, 17], family problems, childhood trauma and sexual orientation (identifying as LGBT) [15], race (being non-white) [14], alcohol consumption [20], and internet addiction [17]-[23]. Maintaining a balance between the university and other demands in life is one of the factors contributing to a student's mental health problems. They may risk losing their scholarship, or having the amount reduced if their academic performance drops. The stress level will also increase towards the end of the semester, especially during the examination period [10].

Related Studies using Machine Learning Algorithms

Machine learning is a scientific discipline that focuses on how computers learn or gain knowledge from data. Machine learning is defined as a field of study that gives computers the capability to learn without being explicitly programmed [3, 27]. Machine learning can be divided into four categories: supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning [23]. Based on Table 2, supervised learning is the most selected data mining techniques to solve the problems in classifying mental health problems.

Table 2 Comparison Factor of Mental Health among Higher Education Students

| Author | Techniques | Variables | Significances |
|--------|--------------------------------------|--|--|
| [18] | Random forest, NB, SVM, KNN | Perceived Stress Scale (PSS) questionnaire | The accuracy of random forest 83.33%, NB 71.42%, SVM 85.71% and KNN 55.55% |
| [24] | DT, SVM, ANN, BN | Outgoing activity, toileting activity, sleeping activity, disease, mental status (GDS) | DT 95.1%, SVM 75.6%, ANN 96.7%, BN 92.7% |
| [25] | KNN, SVM BN | Depression level, sex, grade, major, technical, novel, amusing, psychological | Mild depression: DT 94.3%, SVM 75.6%, ANN 96.2%, BN 91.3% |
| [26] | Random forest, random tree, MLP, SVM | Age, MMSE score, neurological condition, depression (GDS), MoCA test | Severe depression: DT 99.4%, SVM 99.5%, ANN & BN 100% |

The most commonly applied algorithms are Support Vector Machine (SVM), followed by Decision Tree and Neural Network. These three models are highly accurate, above 70%, with good generalization capabilities that prevent overfitting [27]-[28].

Methods

The project framework is adapted from the Cross-Industry Standard Process for Data Mining (CRISP-DM). The adapted cycles from the CRISP-DM framework are problem and data understanding, modelling, and model evaluation. The principal sources are websites that focus on the topic of mental health problems: World Health Organization (WHO) [15], National Institute of Mental Health [13], and American Psychiatric Association [12]. Data collection is performed using surveys with important segments, Depression Anxiety Stress Scale (DASS-21), and World Health Organization 40 Quality of Life survey (WHOQOL). The DASS-21 survey is used to determine the student’s level of stress, depression, and anxiety. Stress, depression, and anxiety levels will be the target attributes in the modelling stage. The WHOQOL is utilized to factor the mental health problems if the student experiences them.

Selected Factors

There are 15 factors listed in WHOQOL that have been selected: positive feeling, memory, self-esteem, appearance, negative feeling, personal relationship, social support, safety, home environment, financial, leisure and religion. The survey is constructed in three sections: demographic profile in section 1, DASS-21 segment in section 2, and section 3 consists of questions for WHOQOL in five segments: Psychological, Social Relationship, Environment, University Life, and Spirituality/Religion/Personal Beliefs. There are 629 total respondents from a higher education institute in Terengganu. Each respondent provides answers to each question rated on a scale of 1 to 5 where 1 = not at all, 2 = a little, 3 = a moderate amount, 4 = very much, and 5 = an extreme amount (indicating high, positive perceptions).

Table 3 Attributes for data collection

| Group | Attribute | Attribute value | Number of attributes |
|--|--|-----------------|----------------------|
| Demographic | Gender, Age, Program, Part, CGPA [3.50 - 4.00, 3.00 – 3.49, 2.20 – 2.99, 2.00-2.19 and below 2.00], Financial Support | Nominal | 7 |
| DASS-21 (Target/Output) | S1 – S21 | Numeric | 21 |
| WHOQOL Section A: Psychological | A1.1, A1.2, A1.3 A2.1, A2.2, A2.3, A2.4 A3.1, A3.2, A3.3, A3.4 A4.1, A4.2, A4.3, A4.4 A5.1, A5.2, A5.3, A5.4 | Numeric | 19 |
| Section B: Social Relationship | B1.1, B1.2, B1.3, B1.4 B2.1, B2.2, B2.3, B2.4 | Numeric | |
| Environment | C1.1, C1.2, C1.3, C1.4 C2.1, C2.2, C2.3, C2.4 C3.1, C3.2, C3.3, C3.4 C4.1, C4.2, C4.3, C4.4 | Numeric | 12 |
| University Life | D1.2, D1.3, D1.4 D2.1, D2.2, D2.3, D2.4 D3.1, D3.2, D3.3, D3.4 | Numeric | 11 |
| Spirituality/Religion/ Personal Beliefs | E1.1, E1.2, E1.3, E1.4 | Numeric | 4 |
| Total of attributes | | | 82 |

Data Labelling and Representation

The target value of stress, depression, and anxiety are constructed based on the score in DASS-21 in equation 1 for Stress_Score, 2 for Depress_Score, and 3 for Anxiety_Score:

$$Stress_{score} = S1 + S6 + S8 + S11 + S14 + S18 \tag{1}$$

$$Depress_{score} = S3 + S5 + S10 + S13 + S16 + S17 + S21 \tag{2}$$

$$Anxiety_{score} = S2 + S4 + S7 + S9 + S15 + S19 + S20 \tag{3}$$

The scores are then mapped into different levels of stress, depression, and anxiety as shown in Table 4. The target variables are nominal attributes for each Stress_Level, Depression_Level, and Anxiety_Level, with labels for normal, mild, moderate, severe, and extremely severe.

Table 4 Levels of Mental Health Types

| Score/No of Samples/Level | Normal | Mild | Moderate | Severe | Extremely severe |
|---------------------------|--------|-------|----------|---------|------------------|
| Stress | 0 – 7 | 8 – 9 | 10 – 13 | 14 – 17 | Above 18 |
| No. of Samples | 377 | 87 | 113 | 43 | 9 |
| Depression | 0 – 5 | 6 – 7 | 8 – 10 | 11 – 14 | Above 15 |
| No. of Samples | 373 | 100 | 54 | 64 | 38 |
| Anxiety | 0 – 4 | 5 – 6 | 7 – 8 | 9 – 10 | Above 11 |
| No. of Samples | 261 | 119 | 84 | 66 | 99 |

There are two main processes in preparing datasets for modelling. First, as presented in Table 3, the exact value for each attribute in the survey are captured. The dataset holds the numerical values of attributes S1 to S21, numerical values of attributes A1.1 to E1.4, and nominal values of demographic attributes. Next, feature extraction is performed to get the average score for factors in WHOQOL. Below are three examples of factors from specific attributes and the formula for the average score in WHOQOL using equations 4, 5, and 6.

$$positive_feeling = \frac{A1.1 + A1.2 + A1.3}{3} \quad (4)$$

$$personal_relationship = \frac{B1.1 + B1.2 + B1.3 + B1.4}{4} \quad (5)$$

$$safety = \frac{C1.1 + C1.2 + C1.3 + C1.4}{4} \quad (6)$$

Machine Learning Algorithms

In this study, the modelling phase is repeated for several experiments with the following machine learning algorithms: Decision Tree, Neural Network, Support Vector Machine (SVM), Naïve Bayes, and Logistic Regression using SPSS Modeler.

A Chi-squared Automation Interaction Detection (CHAID) decision tree was developed for the prediction model based on chi-square statistic, as shown in equation (7), where y is actual and y' is expected, with a probability between 0 and 1. A chi-square value closer to 0 indicates that there is a significant difference between the two classes which are being compared. Similarly, a value closer to 1 indicates that there is not any significant difference between the 2 classes. The predictor variable with the smallest adjusted p -value, i.e., the predictor variable that will yield the most significant split will be considered for the next split in the tree.

$$Chi - squared\ value = \sqrt{\frac{(y - y')^2}{y'}} \quad (7)$$

Neural network is a brain model architecture with the elements of input layer, hidden layer, connection weight, and output layer. Hidden layer is a processing layer that concert input into output. Connection weights is expressed the relative strength of the input. It has summation function and transformation function in each node in the input and hidden layer. The input layer consists of nodes that represent the input variables. Meanwhile the output layer is presenting the output variable of a prediction problem, the figure 1 is the neural network for predicting anxiety level.

Naïve Bayes algorithm is using probabilistic theory to perform classification task. It based on Bayes Theorem in equation (8) which to find the probability of A is happening, given that B has occurred. Here, B is the evidence and A is the hypothesis with an assumption that the predictors/features are independent.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{8}$$

Above,

- $P(A|B)$ is the posterior probability of class (A is target variable) given predictor (B is attributes).
- $P(A)$ is the prior probability of class.
- $P(B|A)$ is the likelihood which is the probability of predictor given class.
- $P(B)$ is the prior probability of predictor.

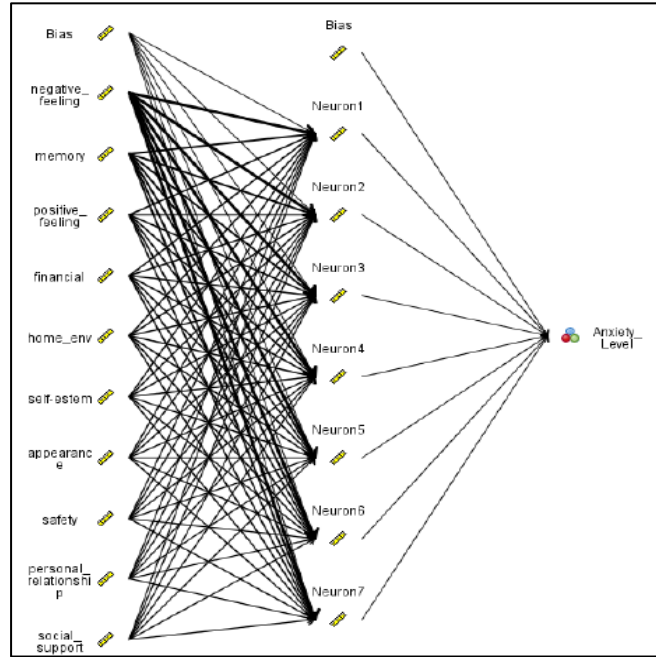


Figure 1 Neural network for anxiety level.

Logistic Regression is one of algorithms to solve classification problems by implement the concept of probability, but it uses more complex cost function, which is called as Sigmoid function. Logistic regression uses an equation as the model representation in equation (9). Independent values (x) are combined linearly using weights or coefficient values to predict an output value (y). Example of logistic regression equation is:

$$y = \frac{e^{b_0+b_1*x}}{1 + e^{b_0+b_1*x}} \tag{9}$$

where y is the predicted output, b_0 is the bias or intercept term and b_1 is the coefficient for the single input value (x).

Support vector machines (SVM) are a class of linear algorithms that can be used for classification, regression, density estimation, novelty detection, and other applications. SVM is using classification techniques to build a predictive model. SVM algorithm is purposely to find a hyperplane in an N -dimensional space that distinctly classifies the data points. Separating two classes of data point may lead to many possible hyperplanes to be the choices. Hyperplane equation is below in (10):

$$w^T x + b = 0 \tag{10}$$

where w is a weight vector, x is input vector and b is bias.

SVM maximizes the margin of the classifier in order to separated two class of data points. In order to maximize the margin, we thus need to minimize $\|w\|$ equation (11) and (12), with the condition that there are no data points between two lines.

$$w^T x + b \geq 0 \text{ for } d_i = +1 \tag{11}$$

$$w^T x + b \leq 0 \text{ for } d_i = -1 \tag{12}$$

where, d is margin of separation. It is a separation between the hyperplane and the closest data point for a given weight vector w and bias b .

Performance Measures

Each mental health problem (stress, depression, and anxiety) is tested in different models by each algorithm with feature selection process. Each model is evaluated based on accuracy (13), sensitivity (14), specificity (15), and precision (16), and the best prediction model is then selected.

$$Accuracy = \frac{TP + TN}{Total\ Sample} \quad (13)$$

$$Sensitivity = \frac{TP}{TP + FN} \quad (14)$$

$$Specificity = \frac{TN}{TN + FP} \quad (15)$$

$$Precision = \frac{TP}{TP + FP} \quad (16)$$

Result and discussion

This section presents visualization of the descriptive analysis results via the dashboard and the modelling results of the algorithms. The visualization of the results on the dashboard provides the distribution of students. During the modelling phase, the models are constructed and fine-tuned until the highest accuracy is obtained.

Descriptive Analysis of the Factors

The graph presents the percentage of students according to gender in figure 2; the dominant gender in this data collection is female at 71%. The dashboard shows the number of records in the data collection that answered the DASS-21 survey. The second, third, and fourth graphs show the number of students based on gender and level of stress, depression, and anxiety, respectively. From the data, 50% of the students have normal levels of stress, depression, and anxiety. The number of students with extremely severe level of stress and depression is small. However, the number of students with severe and extremely severe levels of anxiety is more than 100 students. It is assumed that the students are perhaps uncertain about their future as they are still unfamiliar with their new life at university. Figure 2 shows the distribution of anxiety status for normal, mild, moderate, extremely severe and severe, according to gender for female in pink bar and male in blue bar.

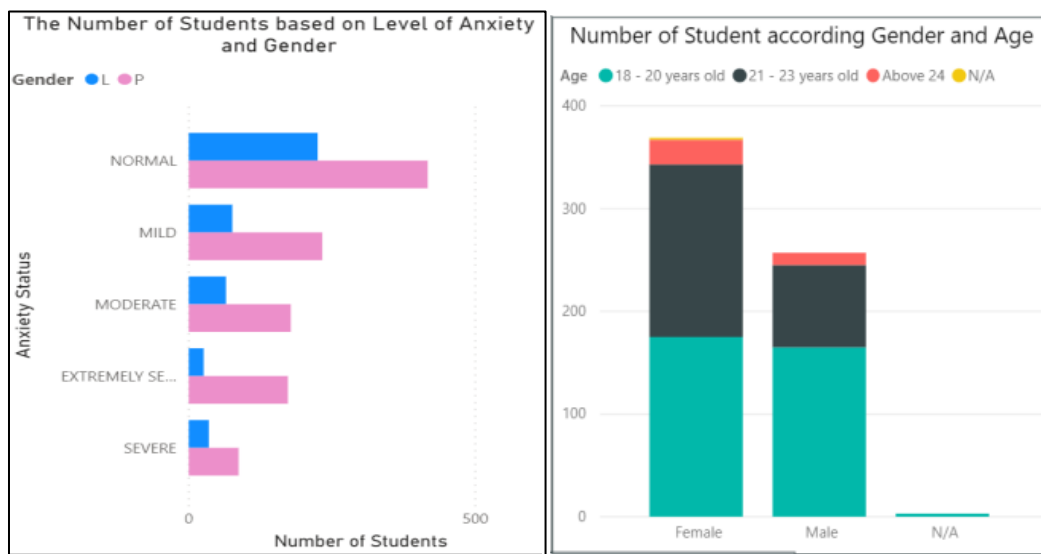


Figure 2 Distribution of sample

Comparison of Algorithms in Prediction of Stress Level

The best algorithm is developed by CHAID Decision Tree (DT) to classify stress. The initial target variable has five levels: normal, mild, moderate, severe, and extremely severe. Those levels are further transformed into two categories of stress. Normal and mild levels are labelled 0, indicating that the students do not have mental health problems in terms of stress. Moderate, severe, and extremely severe levels are labelled 1, indicating that the students have the probability of facing mental health problems in terms of stress. The target attribute value is Stress_Level (1 = Stressed with 464 samples, and 0 = Not Stressed with 156 samples). Feature selection is used for ranking, and six attributes were selected out of 18. The attributes represent factors of positive_feeling, memory, negative_feeling, personal_relationship, leisure, and religion. The depth of the decision tree is four levels. The total number of rules is 11 from the tree in figure 3 with the negative_feeling as the root node and {memory, positive feeling, leisure, religion and personal_relationship} as the intermediate nodes in the tree.

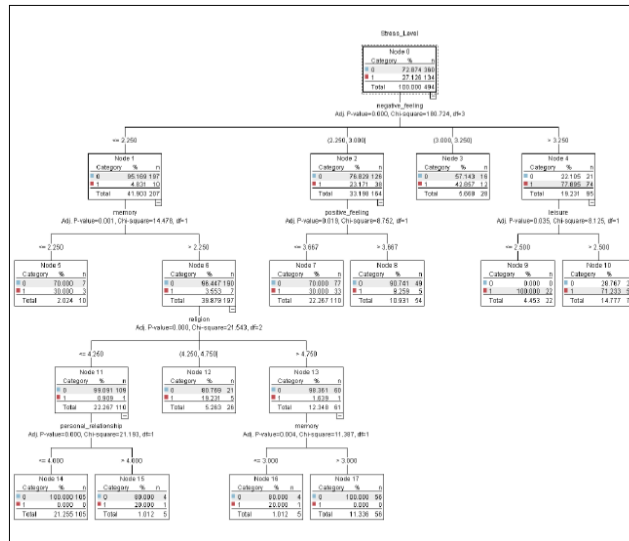


Figure 3The CHAID decision tree model

For testing, as shown in Table 5, the highest accuracy gained is 84.44%, sensitivity is 54.84%, specificity is 93.27%, and precision is 70.83%, by the decision tree model.

Table 5Comparison of different algorithms

| Algo | Accuracy | Sensitivity | Specificity | Precision |
|------|----------|-------------|-------------|-----------|
| DT | 84.44 | 54.84 | 93.27 | 70.83 |
| MLP | 80.00 | 51.61 | 88.46 | 57.14 |
| NB | 74.81 | 41.94 | 84.62 | 44.83 |
| SVM | 82.22 | 51.61 | 91.35 | 64.00 |
| LR | 82.96 | 51.61 | 92.31 | 66.67 |

Therefore, the decision tree model is selected as the best model for stress prediction with six attributes: positive_feeling, memory, negative_feeling (the highest rank for predictor importance), personal_relationship, leisure, and religion. A small change in the decision tree for each factor is detected and leads to decision making.

Comparison of Algorithms in Prediction of Depression Level

Lastly, models are developed to predict depression with feature selection. Eleven attributes are selected for the modelling activities. The eleven attributes are negative_feeling, self-esteem, positive_feeling, social_support, memory, religion, safety, leisure, home_env, personal_relationship, and programme. The value of the target variable is Depression_Level (1 = Depressed with 332 samples, and 0 = Not Depressed with 342 samples). The initial target variable has five levels: normal, mild, moderate, severe, and extremely severe. The target value is transformed into two categories of depression. Normal and mild levels are labelled 0, indicating that the students do not have depression. Moderate, severe, and extremely severe levels are labelled 1, indicating that the students have the probability of facing mental health problems in terms of depression. The target attribute for depression

modelling is Depression_Level, and the linear SVM algorithm is applied. For testing, as shown in Table 6, the accuracy is 88.15%, sensitivity is 64.52%, specificity is 95.19%, and precision is 80.00%. The SVM model produces the highest accuracy and precision value compared to other models. The performances of the same experiment without feature selection in testing: accuracy is 82.96%, sensitivity is 61.29%, specificity is 89.42%, and precision is 63.3%.

Table 6 Comparison of different algorithms

| Algo | Accuracy | Sensitivity | Specificity | Precision |
|------------|----------|-------------|-------------|-----------|
| DT | 83.70 | 61.29 | 90.38 | 65.52 |
| MLP | 84.44 | 48.39 | 95.19 | 75.00 |
| NB | 77.04 | 32.26 | 90.38 | 50.00 |
| SVM | 88.15 | 64.52 | 95.19 | 80.00 |
| LR | 87.41 | 61.29 | 95.19 | 79.17 |

Therefore, the Support Vector Machine (SVM) model is selected as the best model for depression prediction. The SVM model clearly separates between those with and without depression.

Comparison of Algorithms in Prediction of Anxiety Level

Modelling is performed to model anxiety. The rank of the predictor importance for Anxiety_Level are memory, positive_feeling, financial, home_env, self-esteem, appearance, safety, personal_relationship, social_support, and CGPA. The target variable for anxiety prediction is Anxiety_Level (1 = Having Anxiety with 300 samples, and 0 = Not Having Anxiety with 194 samples). The initial target variable has five levels: normal, mild, moderate, severe, and extremely severe. The anxiety level is further transformed into two categories: 0 and 1. Normal and mild levels are labelled 0, indicating that the students do not have mental health problems in terms of anxiety. Moderate, severe, and extremely severe levels are labelled 1, indicating that the students have the probability of facing mental health problems in terms of anxiety. Table 7 shows the comparative evaluation of the different models.

Table 7 Comparison of different algorithms

| Algo | Accuracy | Sensitivity | Specificity | Precision |
|------------|----------|-------------|-------------|-----------|
| DT | 65.93 | 45.45 | 85.00 | 67.57 |
| MLP | 68.89 | 60.00 | 75.00 | 62.26 |
| NB | 60.00 | 43.64 | 71.25 | 53.33 |
| SVM | 63.70 | 45.45 | 76.25 | 56.82 |
| LR | 71.85 | 47.27 | 88.75 | 74.29 |

The Logistic Regression (LR) model has the highest accuracy, specification, and precision value compared to other models, but has low sensitivity value. Meanwhile, a MultiLayer Perceptron (MLP) for ANN has lower accuracy at 68.89%, but higher sensitivity value at 60.00%, as well as specificity at 75.00% and precision at 62.26%.

Conclusion

This paper presents mental health prediction models using machine learning in higher education institution. We start by elaborating on the mental health problems and contributing factors among higher education students, and predicts the issues in three categories, namely stress, depression, and anxiety. We present how the data from DASS-21 can be used for modelling by using the attributes' score to label the individuals in the dataset. Meanwhile the factors of WHOQOL were used as input variables in modelling the health problems with feature selection approach. This is achieved by modelling the health problem using different machine learning algorithms. The most common factors identified in this study are the lack of social support, financial difficulties, and learning environment. The best models with the highest accuracy are Decision Tree for stress, and Support Vector Machine for depression. Linear Regression and Neural Networks are the two models that give fair results for anxiety with an accuracy range between 68% to 88%. In the future, more data can be collected for the algorithms to learn the pattern of mental health problems and improve performance.

Acknowledgements

The authors would like to thank the Senior Psychology Officer (Kuala Terengganu), Madam Raja Hasnahbinti Raja Mat, and Senior Psychology Officer (Dungun), Madam AzurabintiMamat, as well as Madam NorulHidayahbinti Isa for supporting this research project.

References

1. K Shahidul, I Ariful, H Akther, Z Taiyeb Z and LH Abu. Supporting the Treatment of Mental Diseases using Data Mining. *International Conference on Innovations in Science, Engineering and Technology (ICISSET), Chittagong, Bangladesh*. 2018; 339–344.
2. E Kvarnstrom. The Dangers of Mental Health Misdiagnosis: Why Accuracy Matters. Bridges to Recovery. Available at: <https://www.bridgestorecovery.com/blog/the-dangers-of-mental-health-misdiagnosis-why-accuracy-matters/>, accessed 2017.
3. RC Deo. Machine learning in medicine. *Circulation*. 2015; **132**, 1920–1930.
4. Jabatan Pendidikan Tinggi. Direktori Universiti Awam. Available at: <http://jpt.mohe.gov.my/portal/ipta/institusi-pendidikan-tinggi-awam/direktori-universiti-awam>, accessed 2017.
5. HS Akareem and SS Hossain. Determinants of education quality: what makes students' perception different?. *Open Review of Educational Research*. 2016; **3**, 1, 52–67.
6. A Pitt, F Oprescu, G Tapia and M Gray. An exploratory study of students' weekly stress levels and sources of stress during the semester. *Active Learning in Higher Education*. 2018; **19**, 1, 61–75.
7. M McLafferty, CR Lapsley, E Ennis, C Armour, S Murphy et al. Mental health, behavioural problems and treatment seeking among students commencing university in Northern Ireland. *PLOS ONE*. 2017; **12**, 12, 1–14.
8. RA Vidourek and M Burbage. Positive mental health and mental health stigma: A qualitative study assessing student attitudes. *Mental Health and Prevention*. 2019; **13**, 1–6.
9. R Parekh. What Is Mental Illness?. Patients & Families, Psychiatry.org. Available at: <https://www.psychiatry.org/patients-families/what-is-mental-illness>, accessed 2019.
10. S Shannon, G Breslin, T Haughey, N Sarju, D Neill, M Lawlor and G Leavey. Predicting Student-Athlete and Non-Athletes' Intentions to Self-Manage Mental Health: Testing an Integrated Behaviour Change Model. *Mental Health and Prevention*, 2019; **13**, 92–99.
11. RA Vidourek and M Burbage. Positive mental health and mental health stigma: A qualitative study assessing student attitudes. *Mental Health and Prevention*. 2019; **13**, 1–6.
12. American Psychiatric Association (APA) (n.d.). What Is Mental Illness?. Available at: <https://www.psychiatry.org/patients-families/what-is-mental-illness>.
13. National Institute of Mental Health (NIMH) (n.d.). Anxiety Disorder. Available at: <https://www.nimh.nih.gov/health/topics/anxiety-disorders/index.shtml>.
14. G Andrews, C Bell, P Boyce, C Gale, L Lampe, O Marwat et al. Royal Australian and New Zealand College of Psychiatrists clinical practice guidelines for the treatment of panic disorder, social anxiety disorder and generalised anxiety disorder. *Australian and New Zealand Journal of Psychiatry*. 2018; **52**, 12, 1109–1172.
15. World Health Organization (WHO). Depression and Other Common Mental Disorders - Global Health Estimation. *Obstetrics and Gynecology*. 2017; **48**, 1, 56–60.
16. ILD Moutinho, ALG Lucchetti, ODSilver Ezequiel and G Lucchetti. Mental health and quality of life of Brazilian medical students: Incidence, prevalence, and associated factors within two years of follow-up. *Psychiatry Research*. 2019; **274**, 306–312.
17. D Guofeng and X Yuming. Internet addiction and mental health status of Chinese college freshmen. *2009 3rd International Conference on Bioinformatics and Biomedical Engineering, ICBBE 2009*. 2009; 1–4.
18. AA Sabourin, JC Prater and NA Mason. Assessment of mental health in doctor of pharmacy students. *Currents in Pharmacy Teaching and Learning*. 2019; **11**, 3, 243–250.
19. SM Sæther, M Knapstad, KG Askeland and JC Skogen. *Alcohol consumption, life satisfaction and mental health among Norwegian college and university students*. Addictive Behaviors Reports, 2019, 10, 100216.
20. A ZahirIzuan, S ShamsulAzhar, MKS Tan and SAR Syed-Sharizman. Neighbourhood influences and its association with the mental health of adolescents in Kuala Lumpur, Malaysia. *Asian Journal of Psychiatry*. 2018; **38**, 35–41.
21. R Ahuja and A Banga. Mental stress detection in university students using machine learning algorithms. *Procedia Computer Science*. 2019; **152**, 349–353.
22. M Mohssen, K Muhammad Badruddin and MB EihadBashier. *Machine Learning:*

- Algorithms and Applications*, 1st Edition. In CRC Press, ch. 1, 2017, pp. 37-44.
23. S Shalev-Shwartz and S Ben-David. *Understanding machine learning: From theory to algorithms*. Cambridge University Press, ch 1, 2013; 1-8.
 24. Y Hou, J Xu, Y Huang and X Ma. A big data application to predict depression in the university based on the reading habits. *2016 3rd International Conference on Systems and Informatics, ICSAI 2016*, 2017; 1085–1089.
 25. IM Spyrou, C Frantzidis, C Bratsas, I Antoniou and PD Bamidis. Geriatric depression symptoms coexisting with cognitive decline: A comparison of classification methodologies. *Biomedical Signal Processing and Control*. 2016; **25**, 118–129.
 26. F Ge, Y Li, M Yuan, J Zhang and W Zhang. Identifying predictors of probable posttraumatic stress disorder in children and adolescents with earthquake exposure: A longitudinal study using a machine learning approach. *Journal of Affective Disorders*. 2020; 264, 483-493.
 27. N Mohammad Suhaimi, S Abdul-Rahman, S Mutalib, NH Abdul Hamid and A Md Ab Malik. Predictive Model of Graduate-On-Time Using Machine Learning Algorithms. *Communications in Computer and Information Science*. 2019, **1100**, 130-141.
 28. SS Ahmad Tarmizi, S Mutalib, NH Abdul Hamid, S Abdul-Rahman and A Md Ab Malik. A Case Study on Student Attrition Prediction in Higher Education Using Data Mining Techniques. *Communications in Computer and Information Science*. 2019; **1100**, 181-192.