Implementation of a Decision Tree Algorithm to Analyze the Risk Factors of Stress Among University Students in Lampung Province

Fiqih Satria[1] and Rika Damayanti[2]

[1, 2]*UIN Raden Intan Lampung, Jl. Letkol H. Endro Suratmin, Bandar Lampung, Indonesia*

fiqihsatria@radenintan.ac.id, rikadamayanti@radenintan.ac.id

Kata kunci: Risk Factors, Stress, Students, Decision Tree

Abstrak: Stress among university students in Lampung Province has become a significant concern. This study aims to delve deeper into the unique and context-specific risk factors associated withstress by leveraging the decision tree algorithm. Through in-depth analysis of data encompassing academic pressure, social support,extracurricular involvement, adaptability, sleep patterns, the impact of campus green spaces, and spirituality, this research endeavors to develop an accurate predictive model. Consequently, it is expected to identify distinctive stress patterns and provide targeted intervention recommendations, such as data-driven counseling programs, the development of peer support communities, or the optimization of academic counseling services. The findings of this study are anticipated to make a substantial contribution to improving the mental well-being of students in Lampung Province and serve as a reference for similar studies in other regions.

1 INTRODUCTION

Stress is a significant concern among university students, impacting their academic performance, mental health, and overall well-being. The pressures of academic demands, social obligations, financial constraints, and future career uncertainties contribute to elevated stress levels in this demographic. Understanding the risk factors associated with stress is essential for developing effective intervention strategies and support systems.

In recent years, decision tree algorithms have emerged as a powerful tool in data analysis and predictive modeling. These algorithms offer a straightforward and interpretable method for identifying patterns and relationships within complex datasets.

By leveraging decision trees, researchers can uncover the key factors that contribute to stress and predict which students are at higher risk.

This study aims to analyze the risk factors of stress among university students in Lampung Province using a decision tree algorithm. Lampung Province, known for its diverse student population, provides a representative sample for this analysis. The variables examined in this study include academic pressure, social support, extracurricular involvement, adaptability, sleep patterns, the impact of campus green spaces, and spirituality. These factors are considered critical in understanding the stress levels among students and their overall mental health.

The study will identify the primary stressors and evaluate their relative importance, offering insights that can inform targeted interventions and policies to alleviate student stress. By focusing on these specific variables, the research seeks to provide a comprehensive understanding of the multidimensional aspects of student stress.

The paper is structured as follows: the next section reviews relevant literature on stress among university students and the application of decision tree algorithms in similar contexts. Following that, the methodology section details the data collection process, the decision tree algorithm used, and the criteria for evaluating the model's performance. The results section presents the findings, highlighting the most significant risk factors. Finally, the discussion and conclusion section interprets the results, discusses their implications, and suggests directions for future research and practical applications.

2 METHODOLOGY

2.1 Research Design

This study adopts a quantitative research approach using secondary data analysis. The data used is survey data collected from university students in Lampung Province through questionnaires. The study employs a

[1]https://orcid.org/0009-0006-3452-1287
[2]https://orcid.org/009-0001-3193-1495

decision tree algorithm to analyze the risk factors contributing to stress among students.

2.2 Population and Sample

The target population for this research is students registered at various universities in Lampung Province. The random sampling method uses an online questionnaire that is distributed randomly to students to ensure a representative sample. a total of 937 students completed the questionnaire to ensure adequate representation and reliability of the results.

2.3 Data Collection

Data is collected using a validated questionnaire that includes various variables assumed to influence students' stress levels. The questionnaire covers questions related to:

1. Academic pressure (e.g., Too much coursework, Worried about not meeting grade targets)
2. Social support (e.g., Feeling lonely without close friends on campus, Missing family and friends at home, Receiving support from friends and family)
3. Life changes and adaptation (e.g., Struggling to balance study and other activities, Experiencing significant life changes, Able to adapt to changes and challenges)
4. Mental Health History (e.g., Diagnosed with a mental disorder, Currently receiving treatment for a mental health issues, Have a family member with mental illness)

2.4 Data Processing

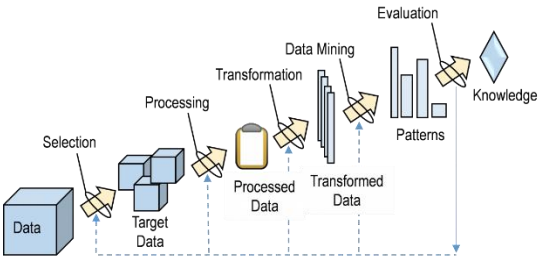
The collected data undergoes several preprocessing steps, including:

1. Data Cleaning: Removing or correcting missing or invalid data.
2. Data Transformation: Converting qualitative variables into numerical forms when necessary.
3. Data Normalization: Adjusting the data scale to ensure that each variable has a balanced influence in the analysis.

2.5 Decision Tree Algorithm

The decision tree algorithm is trained using the training set to build a predictive model that identifies the most influential factors contributing to stress.

Data mining process with decision tree algorithm in this research using orange data mining ver. 3.36.



2.6 Model Evaluation

The decision tree model is evaluated using several metrics, including:

1. **Accuracy**: The percentage of correct predictions out of the total predictions.
2. **Precision and Recall**: These metrics evaluate the model's performance in predicting specific classes.
3. **Confusion Matrix**: A matrix providing a detailed breakdown of the model's correct and incorrect predictions.

2.7 Interpretation and Analysis of Results

The results of the decision tree model will be analyzed to identify the key factors contributing to student stress. The importance of each factor will be evaluated based on its position in the decision tree and its frequency as a primary splitting point. Additional analysis will explore the interactions between these factors and how they collectively influence stress levels among students.

3 DISCUSSION

3.1 Survey Instrument

Questionnaire consists of two parts:

1. **DASS-21**

Consists of 21 questions that measure students' level of mental health, including depression, anxiety, and stress.

DASS-21

No	Questions	D	A	S
1	I find it difficult to calm down		☐	

2	I feel like my mouth is dry			☐
3	I am unable to feel any positive emotions at all	☐		
4	I have difficulty breathing (e.g., breathing too quickly, panting without physical exertion)			☐
5	I find it difficult to get things started		☐	
6	I tend to overreact to situations		☐	
7	I experience trembling (for example, of the hands)			☐
8	I feel anxious		☐	
9	I feel completely hopeless	☐		
10	I feel very irritable		☐	
11	I find it difficult to relax	☐		
12	I feel sad and depressed	☐		
13	I feel intolerant of interruptions to my activities		☐	
14	I feel very nervous		☐	
15	I feel that nothing can make me happy	☐		
16	I feel worthless	☐		
17	I feel irritable			☐
18	I feel heart rate changes (for example, faster heart rate)			☐
19	I feel afraid for no apparent reason		☐	
20	I feel like life is meaningless	☐		
21	I find it very difficult to stay calm	☐		

Answer	**Score**
Never	1
Sometimes	2
Often	3
Very Often	4

2. Stress Risk Factor Questions

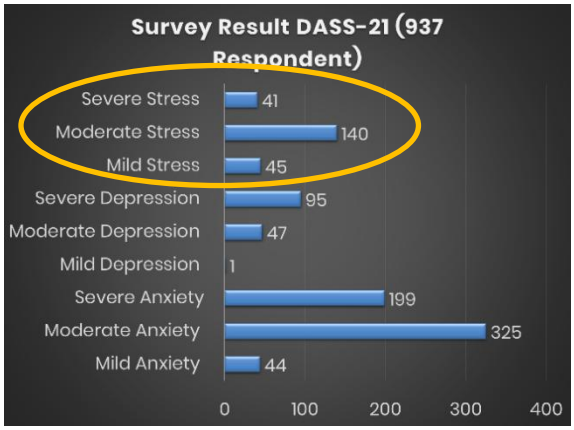
Consists of 11 questions that evaluate factors that could potentially cause stress in college students

No	**Questions**	**Answer**
1	Burdened with many college assignments	○ Never ○ Sometimes ○ Often ○ Very Often
2	Anxious and worried about not being able to meet the desired grade targets	○ Never ○ Sometimes ○ Often ○ Very Often
3	Difficulty managing time between study and other activities	○ Never ○ Sometimes ○ Often ○ Very Often
4	Being lonely and not having close friends on campus	○ Never ○ Sometimes ○ Often ○ Very Often
5	Missing family and friends at home	○ Never ○ Sometimes ○ Often ○ Very Often
6	experienced significant life changes	○ Yes ○ No
7	have been diagnosed with a mental disorder	○ Yes ○ No
8	is undergoing treatment for a mental health problem	○ Yes ○ No
9	have a family member with mental illness	○ Yes ○ No
10	Support from friends and family in dealing with problems	○ Never ○ Sometimes ○ Often ○ Very Often
11	Able to adapt to changes and difficult situations	○ Never ○ Sometimes ○ Often ○ Very Often

3.2 Survey Result

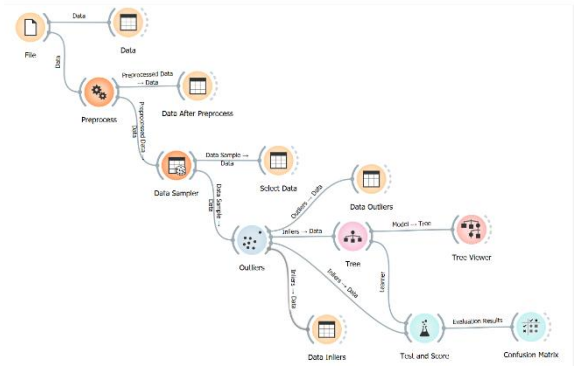
The target variable is the mental health condition of students who are categorized into 9 categories, namely: mild depression, moderate depression, severe depression, mild anxiety, moderate anxiety, severe anxiety, mild stress, moderate stress and severe stress. the categorization of mental health uses the results of the calculation of scores on the

DASS 21 questionnaire. based on the calculation of the DASS-21 score, the following results are obtained:

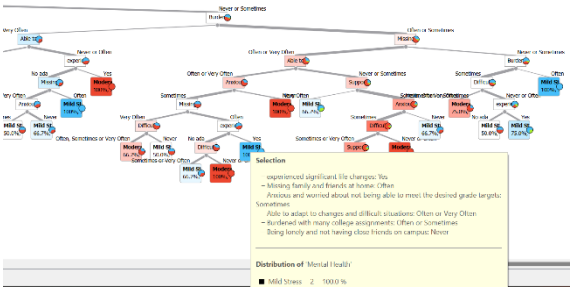


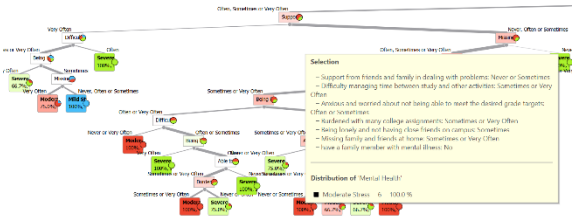
This research focuses on analyzing stress factors in students so that the data to be analyzed using the decision tree algorithm is student data with stress categories with a total of 226 data consisting of 45 mild stress, 140 moderate stress and 41 severe stress.

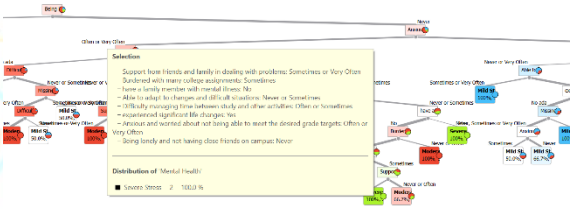
3.3 Data Mining Process



Picture Decission Tree Algorithm with Orange data mining analysis using decision trees using the entire population of 226 student stress data with mild, moderate, and severe stress categories.







from the resulting decision tree, the main factor patterns that cause stress in students are obtained, which differ between mild, moderate and severe stress. these factors can be seen in the following table

	Mild Stress	
1	Being lonely and not having close friends on campus	Never
2	Burdened with many college assignments	Often or Sometimes
3	Able to adapt to changes and difficult situations	Often or Very Often
4	Anxious and worried about not being able to meet the desired grade targets	Sometimes
5	Missing family and friends at home	Often
6	experienced significant life changes	Yes

	Moderate Stress	
1	have a family member with mental illness	No
2	Missing family and friends at home	Sometimes or Very Often
3	Being lonely and not having close friends on campus	Sometimes
4	Burdened with many college assignments	Sometimes or Very Often
5	Anxious and worried about not being able to meet the desired grade targets	Often or Sometimes
6	Difficulty managing time between study and other activities	Sometimes or Very Often
7	Support from friends and family in dealing with problems	Never or Sometimes
	Severe Stress	
1	Being lonely and not having close friends on campus	Never
2	Anxious and worried about not being able to meet the desired grade targets	Often or Very Often
3	experienced significant life changes	Yes
4	Difficulty managing time between study and other activities	Often or Sometimes
5	Able to adapt to changes and difficult situations	Never or Sometimes
6	have a family member with mental illness	No
7	Burdened with many college assignments	Sometimes
	Support from friends and family in dealing with problems	Sometimes or Very Often

3.4 Model Evaluation

Test and Score Thu Aug 08 24, 12:17:04

Settings

Sampling type: No sampling, test on training data
Target class: None, show average over classes

Scores

Model	AUC	CA	F1	Prec	Recall	MCC
Tree	0.960	0.837	0.838	0.853	0.837	0.715

Confusion Matrix Thu Aug 08 24, 12:17:46

Confusion matrix for Tree (showing number of instances)

		Predicted			
		Mild Stress	Moderate Stress	Severe Stress	Σ
Actual	Mild Stress	41	3	0	44
	Moderate Stress	15	110	3	128
	Severe Stress	4	9	24	37
	Σ	60	122	27	209

1. Area Under the Curve (AUC) close to 1 indicates that the Decision Tree model has an excellent ability to distinguish between positive and negative classes. This model is very effective in identifying different classes.
2. The accuracy value of 83.9% shows that the decision tree model is able to classify most of the data correctly. The high accuracy indicates that the decision tree model has an overall good performance.
3. The F1 score of 0.839 shows a good balance between precision and recall. This value indicates that the model has a good ability to classify data correctly, without too many false positives or false negatives.
4. Precision of 84.8% indicates that of all the positive predictions made by the model, 84.8% of them are actually positive. This means that the model has a low false positive error rate.
5. Recall of 83.9% indicates that out of all the positive cases, the model correctly identified 83.9% of them. This shows that the model has a low false negative error rate.
6. The Matthews Correlation Coefficient (MCC) of 0.718 shows a good correlation between the predicted and actual labels. The high MCC value indicates that the model has a good balance between true positive, true negative, false positive, and false negative.

4 CONCLUSIONS

The study explored various factors that contributed to the level of stress in students, focusing on three categories of stress: mild stress, moderate stress, and severe stress. (Severe Stress). Based on the results of analysis using the decision tree method, it was found that the following factors play an important role in determining the level of stress experienced by students:

Mild stress: Students who experience mild stress tend to still be able to adapt to changes and difficult situations, even though they often feel burdened with schoolwork and miss family and friends at home. They seldom feel lonely on campus, and anxiety about values usually happens only occasionally.

Moderate Stress: Stress is more common in students who experience anxiety about values, difficulty in organizing time between studies and other activities, and lack of support from friends and family. Besides, feelings of loneliness and missing family and friends at home also contribute to this level of stress.

Severe stress: Severe stress in students is associated with a combination of high anxiety about achievement of values, significant life changes, low adaptability, and difficulty in time management. Lack of social support from friends and family is also an important factor that exacerbates this level of stress.

From the results of this study, it can be concluded that to reduce high levels of stress among students, special attention is needed to manage academic burdens, develop adaptive skills, as well as increase social support from the surrounding environment. Interventions designed to improve time management and provide emotional support can be effective strategies in helping students manage their stress better.

REFERENCES

Auerbach, R. P. . M. P. . B. R. . A. J. . B. C. . C. P. . D. K. . E. D. . G. J. . H. P. . M. R. . N. M. K. ., & Pinder-Amaker, S. . S. N. A. . V. G. . Z. A. M. . K. R. C. . W. W.-I. C. (2018). t: Prevalence and distribution of mental disorder. WHO World Mental Health Surveys International College Student Project.

Prasetio, C. E., & Triwahyuni, A. (2022). Gangguan Psikologis pada Mahasiswa Jenjang Sarjana: Faktor-Faktor Risiko dan Protektif. Gadjah Mada Journal of Psychology (GamaJoP), 8(1), 56. https://doi.org/10.22146/gamajop.68205

Ahmad Saleh, 2015, Klasifikasi Gejala Depresi Pada Manusia dengan Metode Naïve Bayes Menggunakan Java, Yogyakarta.

Boone, Louis E dan Kurtz, David L. 2010. Pengantar Bisnis Kontemporer. Terjemahan Anwar Fadriansyah. Jakarta: Erlangga.

Davies. B., “Database Systems 3rd Edition”, Basingstoke, UK: Palgrave, 2004.

Griffin, Ricky W., and Moorhead, Gregory., 2014. Organizational Behavior: Managing People and Organizations. Eleventh Edition. USA: South Western.

Han, J. dan M. Kamber. 2006. Data Mining Concepts and Techniques Second Edition. San Francisco: Morgan Kaufmann.

Hermawati. 2013. Data Mining. Yogyakarta: Penerbit ANDI, Jl. Beo.

Ian H. Witten, E. F. (2005). Data mining: Practical Machine Learning Tools and Techniques, Second Edition. San Francisco: Elsevier Inc.

Jananto, A. 2013. Algoritma Naive Bayes untuk Mencari Perkiraan Waktu Studi Mahasiswa. Teknologi Informasi DINAMIK, Vol.18 No.1, 9–16.

Laksana, Fajar. 2008. Manajemen Pemasaran. Edisi 1. Yogyakarta: Grama Ilmu

Moleong, Lexy J. (2007) Metodologi Penelitian Kualitatif, Penerbit PT Remaja Rosdakarya Offset, Bandung

Murtopo, A. A. (2015). Prediksi Kelulusan Tepat Waktu Mahasiswa STMIK YMI Tegal Menggunakan Algoritma Naïve Bayes. CSRID Journal, 7(3), 145–154.

Pramudiono. I., (26 Mei 2017), “Pengantar Data mining: Menambang Permata Penxgetahuan di Gunung Data.” [online], available: http://www.ilmu computer.org/wp-content/uploads/ 2006/08/iko-datamining.

Pratiwi. 2013. Buku Ajar Sistem Pendukung Keputusan. Jakarta: Penerbit, Media.

Ridwan, M., Suyono, H., & Sarosa, M. (2013). Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier. Eeccis, 7(1), 59–64.

Santoso. Heroe, Hariyadi. I Putu, Prayitno, "Data Mining Analisa Pola Pembelian Produk Dengan Menggunakan Metode Algoritma Apriori", 2016

Sugiyono.2017. Metode Penelitian Kuantitatif, Kualitatif dan R&D. Bandung: Penerbit Alfabeta.

Tan, Pang Ning., Michael Steinbach, Vipin Kumar (2004). Introduction to Data Mining.