Indonesian Journal of Engineering, Science and Technology
IJENSET
Volume 02 - Issue 02, December 2025
e-ISSN: 3089-2392
Journal Link: https://jurnal.umla.ac.id/index.php/ijenset/index

A DATA-DRIVEN MACHINE LEARNING FOR PREDICTING STUDENT STRESS LEVELS

Neni Purwati[1]*, Agus Priyono[1], Adam Rizky Al'insani[1], Darmawan[2]

[1]*Medical Informatics Department; Faculty Health Sciences; University Muhammadiyah Lamongan; Lamongan 62218, Indonesian*
[2]*Information Technology; Faculty of Business Management and Information Technology; Universitas Muhammadiyah Malaysia; Perlis 02100, Malaysia*

*Corresponding Author Email: nenipurwati@umla.ac.id

Article Information

Submitted	:	November 26, 2025
Revised	:	December 07, 2025
Accepted	:	December 15, 2025
Paper page	:	56-65
DOI	:	10.38040/ijenset.v2i2.1380

ABSTRACT

Academic stress significantly impacts students' psychological well-being and academic performance. This study focuses on predicting students' stress levels using a data-driven machine learning framework. The dataset was obtained from a questionnaire comprising 25 indicators encompassing emotional, psychological, academic, and environmental aspects of students. The research procedure involved data preprocessing, checking for missing values and redundancy, normalization, descriptive statistical analysis, model development, and performance evaluation using metrics such as recall, precision, sensitivity, specificity, F-measure, and accuracy. The implemented algorithm achieved excellent results, with an overall accuracy of 0.98. The model demonstrated high effectiveness in classifying Eustress and Distress, while its performance in detecting the No Stress category was limited, although precision and specificity indicate a strong capacity to differentiate between classes. These findings confirm that a machine learning approach can effectively capture patterns of student stress based on questionnaire responses and offers valuable guidance for developing early warning systems and targeted psychological intervention strategies. The study highlights the potential of data-driven predictive methods in supporting students' mental health through empirical data analysis.

***Keywords*—** *LibSVM; Machine Learning; Predicting; Stress Levels; Tree Ensemble*.

I. INTRODUCTION

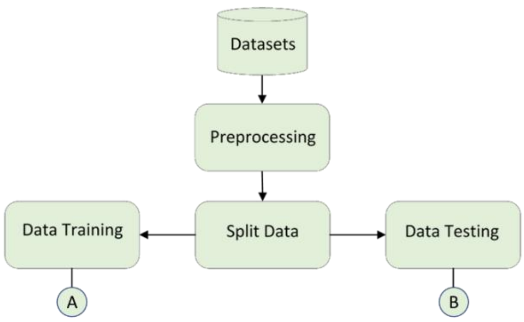
Student stress is increasingly recognized as a significant mental health issue in higher education. Academic pressure arises from beliefs about academic achievement and excessive study workload (Ruiz-Camacho & Gozalo, 2025). Stress can be experienced by anyone and can have negative effects (Siagian et al., 2025) which is a natural response to pressure (Amalia et al., 2025). The main stressors are excessive homework load, assessment pressure, and difficulties in balancing academic and personal life (Pérez-Jorge et al., 2025). This stress not only affects students' psychological well-being but can also influence their academic performance (Condroningtyas & Marsofiyati, 2024). Loneliness and academic stress simultaneously have a significant influence on psychological well-being (Putria et al., 2025). Academic stress can affect both psychological and physical aspects (Mufatihah et al., 2025). The impact of academic stress can lead to psychological and physiological changes, as well as affect learning motivation and learning quality (Bachtiar et al., 2023). The majority of students experience academic stress at a moderate level, and the increase in stress they face is due to poor coping strategies (Rahmawati et al., 2025). The integration of spiritual values into students' daily lives can enhance mental well-being and help them cope with academic challenges (Alviasari et al., 2025).

The use of machine learning technology has been widely applied to predict stress levels among students. The Support Vector Machine (SVM) algorithm is used in classifying student stress levels based on factors such as study patterns, sleep quality, social activities, and other psychological conditions (Fatah & Hasanah, 2025). A stress-level prediction model based on individual lifestyle patterns using a Machine Learning approach with the Random Forest algorithm (Anissa & Qoiriah, 2025). The applied XGBoost algorithm found that targeted intervention strategies at the school, family, and social levels can enhance students' psychological well-being and provide a reference for comprehensive psychological stressor analysis with high accuracy (Ma et al., 2025). Logistic Regression, Random Forest, Gradient Boosting, and Artificial Neural Network models used for the Perceived Stress Score (PSS) have the most significant impact on prediction (Agboro, 2025). Sentiment analysis using machine learning and deep learning algorithms, specifically BERT for sentiment classification, produced a trained model capable of detecting emotional states by analyzing stress or depression based on social interactions (Nijhawan et al., 2022). Early detection of anxiety and stress by combining physiological signals with machine learning (ML) methods (Liu et al., 2025).

Previous studies have highlighted the causes and impacts of student stress and have used various algorithms to predict it; however, most still focus on single models such as SVM, Random Forest, or XGBoost without conducting in-depth model comparisons. No research has combined Tree Ensemble analysis and LibSVM to comprehensively predict stress categories Eustress, Distress, and No Stress particularly with class-wise evaluation using a confusion matrix. In addition, the use of the KNIME platform to build a structured predictive workflow has not been found in prior studies. Your research fills this gap by providing a comparison of two different models, complete evaluation, and a more comprehensive data-driven approach for detecting student stress.

II. METHOD

The research method is an essential section that serves to explain the systematic approach used to achieve the research objectives, providing a detailed description of the steps undertaken throughout the study, from the initial use of the dataset to the final process, which is model evaluation. Following research method steps are illustrated in Figure 1:



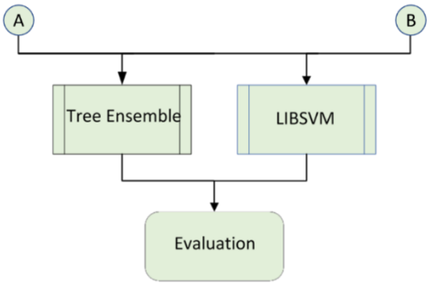


Figure 1. Research Method

A. Datasets

The dataset comes from the kaggle.com website which was collected via Google Forms, this dataset captures emotional, academic, and health-related stress indicators from college students aged 18–21. Datasets consist of rows 843, columns 25 as a predictor and the target column contains type of stress (eustress or distress or no stress).

B. Preprocessing

Preprocessing data has been conducted to ensure that the dataset is properly prepared and suitable for modeling, which includes verifying data types, identifying missing values, detecting duplicate entries, and examining skewness in the data distribution. A data duplication was found in rows 27, so that row was removed. This process has been done and explained on kaggle.com.

Preprocessing was carried out using Google Colab with the Python programming language using the following script:

a. Check Data Type

```
# Check data types
print("\nData Types:")
print(df.dtypes)
```

Output:

```
Data Types:
Gender                                                              int64
Age                                                                 int64
Have you recently experienced stress in your life?                  int64
Have you noticed a rapid heartbeat or palpitations?                 int64
Have you been dealing with anxiety or tension recently?             int64
Do you face any sleep problems or difficulties falling asleep?      int64
Have you been dealing with anxiety or tension recently?.1           int64
Have you been getting headaches more often than usual?              int64
Do you get irritated easily?                                        int64
Do you have trouble concentrating on your academic tasks?           int64
Have you been feeling sadness or low mood?                          int64
Have you been experiencing any illness or health issues?            int64
Do you often feel lonely or isolated?                               int64
Do you feel overwhelmed with your academic workload?                int64
Are you in competition with your peers, and does it affect you?     int64
Do you find that your relationship often causes you stress?         int64
Are you facing any difficulties with your professors or instructors?  int64
Is your working environment unpleasant or stressful?                int64
Do you struggle to find time for relaxation and leisure activities?   int64
Is your hostel or home environment causing you difficulties?        int64
Do you lack confidence in your academic performance?                int64
Do you lack confidence in your choice of academic subjects?         int64
Academic and extracurricular activities conflicting for you?        int64
Do you attend classes regularly?                                    int64
Have you gained/lost weight?                                        int64
Which type of stress do you primarily experience?                  object
dtype: object
```

b. Check for Missing Value

```
# Check for missing values
print("\nMissing Values:")
missing_values = df.isnull().sum()
if missing_values.sum() == 0:
    print("No missing values found in the dataset!")
else:
    print(missing_values[missing_values > 0])
```

Output:

```
Missing Values:
No missing values found in the dataset!
```

c. Check for Duplicate Rows

```
# Check for duplicate rows
duplicates = df.duplicated().sum()
print(f"\nDuplicate Rows: {duplicates}")
```

Output:

```
Duplicate Rows: 27
```

Berikut script menghapus duplikat data 27 rows tersebut:

```
# Menghapus baris duplikat
df_clean = df.drop_duplicates()
```

d. Check for Skewness

```
# Check for skewness
print("\nSkewness of Numeric Features:")
skewness = df[numeric_cols].skew()
print(skewness.sort_values(ascending=False))
```

Output:

```
Skewness of Numeric Features:
Age                                                                  11.863526
Gender                                                                0.630366
Is your hostel or home environment causing you difficulties?          0.577519
Have you gained/lost weight?                                          0.572894
Are you facing any difficulties with your professors or instructors?  0.530070
Are you in competition with your peers, and does it affect you?       0.519131
Is your working environment unpleasant or stressful?                  0.458688
Do you find that your relationship often causes you stress?           0.435945
Do you struggle to find time for relaxation and leisure activities?   0.431621
Do you often feel lonely or isolated?                                 0.421002
Have you been feeling sadness or low mood?                            0.413460
Have you been dealing with anxiety or tension recently?               0.394281
Have you been experiencing any illness or health issues?              0.387678
Do you lack confidence in your academic performance?                  0.377648
Do you feel overwhelmed with your academic workload?                  0.361278
Have you been getting headaches more often than usual?                0.357832
Do you lack confidence in your choice of academic subjects?           0.344224
Have you been dealing with anxiety or tension recently?.1             0.325314
Do you have trouble concentrating on your academic tasks?             0.266796
Academic and extracurricular activities conflicting for you?          0.241671
Have you noticed a rapid heartbeat or palpitations?                   0.231942
Do you get irritated easily?                                          0.221462
Do you face any sleep problems or difficulties falling asleep?        0.188477
Have you recently experienced stress in your life?                    0.063336
Do you attend classes regularly?                                     -0.289718
dtype: float64
```

C. Split Data

Dataset splitting is performed to divide the data into training and testing sets so that the model can be trained and evaluated separately. The purpose is to ensure the model can generalize well, prevent overfitting, and produce objective and accurate performance evaluation. In this study, the dataset was split into 80% training data and 20% testing data, with 674 rows used for training and 169 rows for testing out of a total of 843 rows.

D. Tree Ensemble

Tree Ensemble is a machine learning method that combines multiple decision trees to produce more accurate, stable, and overfitting-resistant predictions, enabling it to capture more complex data patterns compared to a single decision tree (Blanco et al., 2024). Tree Ensemble combines multiple decision trees using specific techniques.

E. LibSVM

LibSVM is an implementation of the Support Vector Machine (SVM) algorithm that searches for the optimal hyperplane to separate data classes and supports various kernels, enabling it to handle non-linear patterns and provide accurate and stable classification performance (Somantri et al., 2023).

F. Evaluasi

This stage is carried out to assess the performance of the machine learning model in predicting students' stress levels, ensuring that the model can generalize well to new data. Evaluation is conducted using confusion metrics and performance measures such as accuracy, precision, recall, F1-score, sensitivity, and specificity to identify the strengths and weaknesses of the model, allowing the selection of the most optimal algorithm.

III. RESULT AND DISCUSSION

The modeling workflow was built using KNIME Analytics Platform version 5.3.2 through a structured sequence of nodes designed to manage data, build models, and evaluate prediction results. The node arrangement begins with the data loading stage using the CSV Reader node to import the dataset into the workflow. The Partitioning node is used to split the dataset. The Tree Ensemble Learner node serves as the training model and the Tree Ensemble Predictor node as the testing model. Similarly, the LibSVM Learner node is used for training and the LibSVM Predictor node for testing. The Scorer node is used to evaluate the dataset. A more detailed illustration is shown in Figure 2:

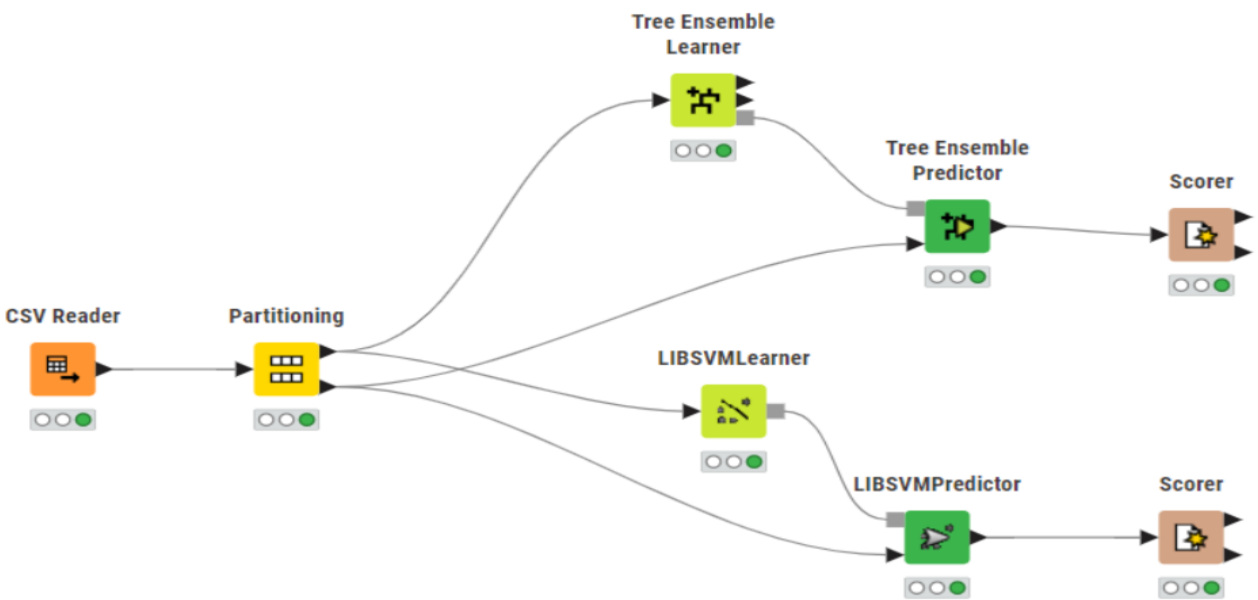


Figure 2. Research Models

The parameters required by the model include:

Model Tree Ensemble	Model LibSVM
Limit number of levels (tree depth) = 10	Kernel Parameters Degree = 3
Minimum split node size = 1	SVM Parameters Cost = 1.0
Minimum child node size = 1	SVM Parameters Nu = 0.5
Number of models = 100	SVM Parameters Lost-Epsilon = 0.1
	SVM Parameters Cachesize (in MB) = 601
	SVM Parameters Epsilon = 0.001

Table 1. Data Prediction

Name	Min	Max	Mean	Standar Deviasi
Gender	0	1	0.32	0.468
Age	14	37	19.781	2.649
Have you recently experienced stress in your life?	1	5	2.953	1.122
Have you noticed a rapid heartbeat or palpitations?	1	5	2.751	1.09
Have you been dealing with anxiety or tension recently?	1	5	2.503	1.23
Do you face any sleep problems or difficulties falling asleep?	1	5	2.781	1.284
Have you been dealing with anxiety or tension recently?	1	5	2.586	1.312
Have you been getting headaches more often than usual?	1	5	2.562	1.322
Do you get irritated easily?	1	5	2.728	1.392
Do you have trouble concentrating on your academic tasks?	1	5	2.657	1.345
Have you been feeling sadness or low mood?	1	5	2.527	1.196
Have you been experiencing any illness or health issues?	1	5	2.479	1.235
Do you often feel lonely or isolated?	1	5	2.361	1.198
Do you feel overwhelmed with your academic workload?	1	5	2.385	1.239
Are you in competition with your peers, and does it affect you?	1	5	2.432	1.194
Do you find that your relationship often causes you stress?	1	5	2.639	1.312
Are you facing any difficulties with your professors or instructors?	1	5	2.42	1.153
Is your working environment unpleasant or stressful?	1	5	2.396	1.151
Do you struggle to find time for relaxation and leisure activities?	1	5	2.308	1.205
Is your hostel or home environment causing you difficulties?	1	5	2.402	1.217
Do you lack confidence in your academic performance?	1	5	2.657	1.406
Do you lack confidence in your choice of academic subjects?	1	5	2.633	1.308
Academic and extracurricular activities conflicting for you?	1	5	2.645	1.265
Do you attend classes regularly?	1	5	3.237	1.278
Have you gained/lost weight?	1	5	2.467	1.17
Prediction (Confidence)	0.52	1	0.913	0.097

A. *Tree Ensemble*

The Tree Ensemble model produced prediction data as shown in Table 1.

The Min (minimum) value indicates whether any respondents rated an item very low and helps identify deviations, such as extremely low responses. The Max (maximum) value shows whether any respondents gave the highest rating and reflects the extent of variation in perception levels within the respondent group. The Mean value represents the general tendency of respondents toward a statement and is useful for determining whether responses tend to be low, moderate, or high. Meanwhile, the Standard Deviation measures the degree of variation or dispersion of responses from the mean.

Table 1 above shows that the age range of respondents is 14–37 years, with an average of 19.78 years, and the gender composition is dominated by females with a mean of 0.32. The mean values range from 2.3 to 2.9, indicating that students' stress levels tend to be moderate. Most standard deviations are above 1, suggesting considerable variation in responses, so stress perceptions across the data are not homogeneous. Sleep disturbances, anxiety, irritability, loss of focus, and low self-confidence have relatively higher mean values compared to other items, indicating that these symptoms are more frequently experienced by students. Meanwhile, attendance has the highest mean value of 3.23, indicating a generally positive trend. The Prediction (Confidence) value with a mean of 0.91 and a low standard deviation of 0.10 demonstrates that the predictive model has a fairly high and stable level of confidence in providing classification results.

Table 2. Prediction Output

Question																									T	P
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**		
0	20	3	4	3	2	3	3	4	1	3	1	2	2	3	2	2	3	1	2	3	2	4	4	4	E	E
0	20	3	2	1	5	3	5	1	1	2	3	2	3	3	4	4	4	2	4	4	3	1	5	1	E	E
0	19	4	5	4	4	2	3	2	2	1	1	1	1	1	2	3	3	1	2	3	1	3	5	3	E	E
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
0	18	5	5	5	4	3	3	5	4	4	4	3	3	2	2	3	2	2	3	5	5	5	4	5	D	E
1	17	4	3	4	4	4	3	4	4	4	4	4	3	4	3	4	4	3	3	3	4	4	2	4	D	E
0	19	1	1	4	1	2	3	2	3	3	3	2	1	3	3	2	1	1	1	1	1	1	1	1	N	E
0	19	2	2	3	2	1	1	1	1	1	5	2	2	3	2	2	1	1	1	1	4	3	4	3	E	E
0	19	3	4	3	3	1	1	2	3	2	5	5	5	5	5	5	5	5	5	3	3	3	3	3	D	E
0	14	2	2	1	1	1	1	1	1	2	1	1	2	3	3	3	3	3	4	3	4	1	5	1	E	E
1	18	1	3	3	1	3	4	1	1	2	2	1	2	1	2	1	2	1	2	3	1	1	1	1	N	E

Description:

T: Target E: Eustress

P: Prediction D: Distress

N: No Stress

The table 2 shows that most predictions fall into the 'E' category, while the actual targets are 'D', 'E', or 'N'. This indicates some discrepancies between the T (target) and P (prediction) values, suggesting that the model still misclassifies certain cases. For example, when the Target = D but the Prediction = E, it means the model has not yet fully distinguished stress patterns across categories. The table provides an initial overview of the model's prediction accuracy and consistency and serves as a basis for further evaluation using the confusion matrix and other performance metrics.

B. LibSVM

The LibSVM model produced classification data as shown in Table 3 below:

Table 3. Data Classification

Name	Min	Max	Mean	Standar Deviasi
Gender	0	1	0.32	0.468
Age	14	37	19.781	2.649
Have you recently experienced stress in your life?	1	5	2.953	1.122
Have you noticed a rapid heartbeat or palpitations?	1	5	2.751	1.09
Have you been dealing with anxiety or tension recently?	1	5	2.503	1.23
Do you face any sleep problems or difficulties falling asleep?	1	5	2.781	1.284
Have you been dealing with anxiety or tension recently?	1	5	2.586	1.312
Have you been getting headaches more often than usual?	1	5	2.562	1.322
Do you get irritated easily?	1	5	2.728	1.392
Do you have trouble concentrating on your academic tasks?	1	5	2.657	1.345
Have you been feeling sadness or low mood?	1	5	2.527	1.196
Have you been experiencing any illness or health issues?	1	5	2.479	1.235
Do you often feel lonely or isolated?	1	5	2.361	1.198
Do you feel overwhelmed with your academic workload?	1	5	2.385	1.239
Are you in competition with your peers, and does it affect you?	1	5	2.432	1.194
Do you find that your relationship often causes you stress?	1	5	2.639	1.312
Are you facing any difficulties with your professors or instructors?	1	5	2.42	1.153

Is your working environment unpleasant or stressful?	1	5	2.396	1.151
Do you struggle to find time for relaxation and leisure activities?	1	5	2.308	1.205
Is your hostel or home environment causing you difficulties?	1	5	2.402	1.217
Do you lack confidence in your academic performance?	1	5	2.657	1.406
Do you lack confidence in your choice of academic subjects?	1	5	2.633	1.308
Academic and extracurricular activities conflicting for you?	1	5	2.645	1.265
Do you attend classes regularly?	1	5	3.237	1.278
Have you gained/lost weight?	1	5	2.467	1.17
Prob_Eustress	0	1	0.9	0.248
Prob_No Stress	0	0.99	0.069	0.215
Prob_Distress	0	1	0.031	0.132

The data in Table 3 above is dominated by females (mean 0.32) with an average age of 19.78 years. The mean values across all items range from 2.3 to 2.9, indicating that students generally experience moderate stress levels. Common stress symptoms include sleep disturbances, irritability, lack of focus, and low self-confidence, as reflected by relatively higher means for these items. Most variables have a standard deviation greater than 1, indicating considerable variation in responses across the data. For attendance behavior, the mean is highest at 3.23, reflecting relatively good class attendance habits. The model's probability results show that the Eustress category has the highest average probability at 0.90, while No Stress and Distress are much lower at 0.069 and 0.031, respectively, indicating that the model tends to classify the majority of respondents in the Eustress condition.

Table 4. Data Classification

Pertanyaan																									T	P
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**		
0	20	3	4	3	2	3	3	4	1	3	1	2	2	3	2	2	3	1	2	3	2	4	4	4	E	E
0	20	3	2	1	5	3	5	1	1	2	3	2	3	3	4	4	4	2	4	4	3	1	5	1	E	E
0	19	4	5	4	4	2	3	2	2	1	1	1	1	1	2	3	3	1	2	3	1	3	5	3	E	E
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	..
0	18	5	5	5	4	3	3	5	4	4	4	3	3	2	2	3	2	2	3	5	5	5	4	5	D	D
1	17	4	3	4	4	4	3	4	4	4	4	4	3	4	3	4	4	3	3	3	4	4	2	4	D	D
1	19	2	3	2	1	5	2	3	1	1	4	4	2	4	1	3	3	2	1	1	3	3	3	3	E	E
0	19	2	2	3	2	1	1	1	1	1	5	2	2	3	2	2	1	1	1	1	4	3	4	3	E	E
0	19	3	4	3	3	1	1	2	3	2	5	5	5	5	5	5	5	5	5	3	3	3	3	3	D	E
0	14	2	2	1	1	1	1	1	1	2	1	1	2	3	3	3	3	3	4	3	4	1	5	1	E	E
1	18	1	3	3	1	3	4	1	1	2	2	1	2	1	2	1	2	1	2	3	1	1	1	1	N	N

Description:

T: Target	E: Eustress
P: Prediction	D: Distress
	N: No Stress

Most of the data in Table 4 above fall within the range of 2–4, indicating that the majority of students experience stress symptoms at mild to moderate levels. The variation in responses across the data appears quite diverse, reflecting differences in perception and stress experiences. The comparison between T (target) and P (prediction) also shows that most model predictions fall into the same category as the target, indicating that the model demonstrates fairly good classification performance.

C. *Evaluasi*

1. Confusion Matrix

This matrix provides information on the number of correct predictions (True Positive and True Negative) as well as prediction errors (False Positive and False Negative). By examining the distribution of values in the confusion matrix, it is possible to identify which classes are well-predicted and which classes still pose challenges for the model. Table 5 presents the confusion matrix for the Tree Ensemble model:

Table 5. Confusion Matrix Tree Ensemble

	TP	FP	TN	FN
Eustress	154	14	1	0
Distress	1	0	164	4
No Stress	0	0	159	10

The model demonstrates excellent ability in recognizing Eustress (no FN), although there is a slight tendency to over-predict (as indicated by the remaining FP). The model tends to be less sensitive to the Distress class because most Distress data are not correctly detected (relatively high FN); however, it does not misclassify other classes as Distress (FP = 0), making the Distress predictions very "strict." The model fails to recognize the No Stress class, as there are no correct predictions (TP = 0), indicating that the model cannot distinguish the distinct patterns of truly stress-free data. Additionally, the high FN value (FN = 10) shows that all No Stress respondents were classified into other categories.

Table 6 below presents the confusion matrix results for the LibSVM model:

Table 6. Confusion Matrix LibSVM

	TP	FP	TN	FN
Eustress	153	3	12	1
Distress	9	1	158	1
No Stress	3	0	164	2

Table 6 above shows that the model performs very well in recognizing Eustress, with a high TP and very low FN indicating strong model sensitivity, while the small FP suggests the model rarely misclassifies other classes as Eustress. The model shows fairly good performance for the Distress class, with both FP and FN being low, indicating a balanced prediction for this class, although TP is still relatively low and could be improved. The model can recognize No Stress, but sensitivity remains low as TP is only 3, while FP = 0 indicates that the model is strict and does not easily misclassify other data as No Stress. However, the FN shows that there are still errors in detecting stress-free respondents.

2. Classification Report

A summary of the classification model evaluation, presenting key metrics to assess how well the model predicts each class, is shown in table 7 below:

Table 7. Classification Tree Ensemble

	R	P	S	SP	FM	A
Eustress	0.99	0.98	0.99	0.8	0.99	-
Distress	0.9	0.9	0.9	0.99	0.9	-
No Stress	0.6	1	0.6	1	0.75	-
Overall	-	-	-	-	-	0.98

Based on the evaluation results in table 7 above, the model demonstrates excellent performance for the Eustress class, indicated by perfect recall and sensitivity values (1) and a high F-measure of 0.96, although the low specificity suggests that many non-Eustress cases were misclassified as Eustress. For the Distress class, precision reaches 1, meaning every Distress prediction is correct, but the low recall of 0.20 indicates that most Distress cases are not detected by the model. Meanwhile, for the No Stress class, the model fails to detect any cases (recall = 0), although it can correctly recognize non-No Stress data (specificity = 1). Overall, the accuracy reaches 91.7%, but the imbalance in performance across classes indicates that the model is still biased toward the majority class and requires improvement to better recognize all stress categories more evenly.

The classification results from the LibSVM model are shown in Table 8 below:

Table 8. Classification LibSVM

	R	P	S	SP	FM	A
Eustress	1	0.92	1	0.07	0.96	-
Distress	0.2	1	0.2	1	0.33	-
No Stress	0	-	0	1	-	-
Overall	-	-	-	-	-	0.917

Based on the model evaluation results in Table 8, the classification performance shows very good results for most classes. The Eustress class achieved a recall of 0.99, precision of 0.98, and sensitivity of 0.99, indicating that the model can accurately recognize nearly all Eustress data, supported by a specificity of 0.80 and an F-measure of 0.99, reflecting a well-balanced positive prediction. For the Distress class, recall, precision, and sensitivity are all 0.90, demonstrating the model's consistent ability to detect this class, with specificity of 0.99 and F-measure of 0.90 indicating stable performance. Meanwhile, the No Stress class has a recall and sensitivity of 0.60, meaning the model still struggles to identify some No Stress data, although precision is very high at 1 and specificity is perfect at 1. Overall, the model achieves an accuracy of 0.98, indicating highly accurate predictions, though improvements are still needed for the No Stress class.

IV. CONCLUSION

This study successfully developed a data-driven machine learning model to predict student stress levels, targeting Eustress, Distress, and No Stress. The entire process, from preprocessing, statistical exploration, modeling, to evaluation, demonstrated that the data used were sufficiently representative of students' psychological conditions. Model evaluation results showed that the applied algorithms could deliver excellent predictive performance, with an overall accuracy of 0.98. The model exhibited very high capability in recognizing the Eustress and Distress classes, as indicated by high values of recall, precision, sensitivity, and F-measure, although challenges remain in optimally detecting the No Stress class. Overall, this study confirms that a data-driven machine learning approach is highly effective in analyzing student stress patterns and has great potential to be utilized as an early detection system, helping educational institutions design more targeted psychological interventions and support services. Such a predictive system also adds value in understanding stress-inducing factors, providing a basis for decision-making aimed at enhancing students' mental well-being.

REFERENCES

Agboro, D. (2025). Leveraging Explainable AI and Multimodal Data for Stress Level Prediction in Mental Health Diagnostics. *International Journal of Research and Scientific Innovation*. https://philpapers.org/archive/AGBLEA.pdf

Alviasari, A., Fauzi, N. H. P., Aldi, M., Sofa, M. A. N., Nayara, N. S., Nabila, S., & Suresman, E. (2025). Peran Pendidikan Islam dalam Mengatasi Stres karena Tekanan Akademik Mahasiswa di Universitas Pendidikan Indonesia. *Jurnal Pendidikan Islam (IHSAN)*, *3*(2), 271–278. https://doi.org/10.61104/ihsan.v3i2.923

Amalia, S. D., Barata, M. A., & Yuwita, P. E. (2025). Optimization of Random Forest Algorithm with Backward Elimination Method in Classification of Academic Stress Levels. *Journal of Applied Informatics and Computing (JAIC)*, *9*(3), 633–641. https://doi.org/10.30871/jaic.v9i3.9280

Anissa, A. I., & Qoiriah, A. (2025). Prediksi Tingkat Stres Berdasarkan Pola Hidup Menggunakan Machine Learning. *Ournal of Informatics and Computer Science (JINACS)*, *7*(1), 292–300. https://ejournal.unesa.ac.id/index.php/jinacs/article/view/72508/52035

Bachtiar, N., Hasan, M. I., & Andodo, C. (2023). Stres Akademik Mahasiswa Kesehatan Terhadap Perubahan Psikologis, Fisiologis, Perubahan Motivasi Dan Kualitas Belajar Mahasiswa. *Jurnal Ilmu Dan Teknologi Kesehatan (BHAMADA)*, *14*(2). https://doi.org/10.36308/jik.v14i2.539

Blanco, V., Japón, A., Puerto, J., & Zhang, P. (2024). Optimal Probabilistic Feature Shifts for Reclassification in Tree Ensembles. *Mathematics: Optimization and Control*, 1–29. https://doi.org/10.48550/arXiv.2412.03722

Condroningtyas, S. A., & Marsofiyati, M. (2024). Pengaruh Tekanan Atau Stres Akademik Terhadap Prestasi Belajar Mahasiswa. *Cendikia: Jurnal Pendidikan Dan Pengajaran*, *2*(6), 205–212. https://jurnal.kolibi.org/index.php/cendikia/article/view/1697

Fatah, Z., & Hasanah, U. (2025). Prediksi Tingkat Stress dan Kesehatan Mental Mahasiswa Menggunakan Algoritma SVM. *Jurnal Mahasiswa Teknik Informatika (JAMASTIKA)*, *4*(2), 200–207. https://doi.org/10.35473/jamastika.v4i2.4542

Liu, Y., Palacio, M.-I., Bikki, T., Toledo, C., Ouyang, Y., Li, Z., Li, Z., Toledo, F., Zeng, H., & Herrero, M.-T. (2025). Machine Learning, Physiological Signals, and Emotional Stress/Anxiety: Pitfalls and Challenges. *MDPI Applied Sciences*, *15*(21). https://doi.org/10.3390/app152111777

Ma, G., Xing, A., Cui, J., Cai, C., & Guo, M. (2025). Analyzing the Psychological Stressors of College Students Based on XGBoost Algorithm: Precise Insights and Intervention Strategies. *MLNN '25: Proceedings of the 2025 International Conference on Machine Learning and Neural Networks*, 14–20. https://doi.org/10.1145/3747227.3747230

Mufatihah, Z., Zelya, A. P., Puriani, R. A., & Putri, R. M. (2025). Fenomena Stres Akademik Pada Mahasiswa. *EDU Research*, *6*(1), 573–580. https://doi.org/10.47827/jer.v6i1.564

Nijhawan, T., Attigeri, G., & Ananthakrishna, T. (2022). Stress Detection using Natural Language Processing and Machine Learning Over Social Interactions. *Journal Big Data*, *9*(33). https://doi.org/10.1186/s40537-022-00575-6

Pérez-Jorge, D., Boutaba-Alehyan, M., González-Contreras, A. I., & Pérez-Pérez, I. (2025). Examining the Effects of Academic Stress on Student Well-Being in Higher Education. *Humanities and Social Sciences Communications*, *12*(449). https://doi.org/10.1057/s41599-025-04698-y

Putria, W. T. A., Setyawati, R., Dewi, D. S. E., & Mildaeni, I. N. (2025). Dampak Kesepian Dan Stres Akademik Terhadap Kesejahteraan Psikologis Mahasiswa Luar Jawa di Perguruan Tinggi di Purwokerto (Universitas Muhammadiyah Purwokerto, Universitas Jenderal Soedirman, dan UIN Prof. K.H. Saifuddin Zuhri Purwokerto). *Jurnal Ilmiah Multidisiplin Ilmu (JIMI)*, *2*(2), 191–209. https://doi.org/10.69714/6sq7pj51

Rahmawati, R. D., Lestari, D. R., & Rahmayanti, D. (2025). Overview of Academic Stress Levels in Undergraduate Nursing and Medical Students. *Journal of Intan Nursing (JOIN)*, *4*(2), 1–6. https://doi.org/10.54004/join.v4i1.266

Ruiz-Camacho, C., & Gozalo, M. (2025). Predicting University Students' Stress Responses: The Role of Academic Stressors and Sociodemographic Variables. *European Journal of Investigation in Health, Psychology and Education (EJIHPE)*, *15*(8), 1–18. https://doi.org/10.3390/ejihpe15080163

Siagian, I. O., Linda, S., & Melia, S. (2025). Relationship Between Stress Levels and Students' Academic Achievement. *Jurnal Keperawatan Muhammadiyah (JKM)*, *10*(1), 114–119. https://doi.org/10.30651/jkm.v10i1.24537

Somantri, O., Maharrani, R. H., & Purwaningrum, S. (2023). Model Library Support Vector Machine (LibSVM) untuk Sentiment Review Penilaian Pesisir Pantai. *Jurnal Nasional Teknik Elektro Dan Teknologi Informasi*, *12*(2), 110–116. https://doi.org/10.22146/jnteti.v12i2.6367