SITEKNIK
Sistem Informasi, Teknik dan Teknologi Terapan
E-ISSN: 3032-3991
Vol. 2. No. 2 April 2025. Pages. 76-87

Optimization of Stress Classification Among Students Using Random Forest Algorithm

Raffa Nur Listiawan Dhito Eka Santoso[1], Kusnawi[2]

[1,2] Informatics, Faculty of Computer Science, AMIKOM University Yogyakarta, Indonesia

Keywords: Stress, Students, Random Forest, Classification, Hyperparameter Tuning

Submitted: 15/03/2025
Revised: 25/02/2025
Accepted: 17/03/2025

Abstrak

Stress is a condition of physical and psychological discomfort experienced by students due to academic pressure, demands from parents and teachers, and schoolwork. This stress can lead to physical tension, behavioral changes, and mental health problems if not handled properly. Random Forest is a promising approach to analyze and classify student stress. The aim of this study is to classify stress among students to enable the development of targeted interventions to support student well-being and academic success. The dataset used was sourced from Kaggle and included 1100 datasets with 21 columns. The research stages included data preprocessing, exploratory data analysis, modeling, Decision tree classification and evaluation of the confusion matrix model and Deployment as a measure of stress level. Classification results were evaluated by calculating accuracy, precision, recall and f1-score for stress classes (low, medium and high). The results of this study resulted in an accuracy value before tuning of 87.27% and after tuning of 88.64%. This research can provide insights for schools, parents, and government to develop more effective strategies in addressing the problem of stress among students. The use of Random Forest algorithm is proven to be effective in analyzing and classifying stress, so that it can help in decision making and appropriate welfare interventions to tackle before stress reaches critical levels.

Corresponding Author:
Raffa Nur Listiawan Dhito Eka Santoso
Informatics, Faculty of Computer Science, AMIKOM University Yogyakarta
Jl.Ring Road Utara, Condong Catur, Sleman, Yogyakarta Telp : (0274) 884201 - 207
Email: raffadhitoo@students.amikom.ac.id

INTRODUCTION

Education in Indonesia has gone through an interesting journey from ancient times to the current digital era. Indonesia's education system has evolved from informal to formal in preparation for the next generation. In 1975, there was a reformation where education was used as a tool for national development and in 1998 the reformation era brought significant changes in education which began to focus on student empowerment and a more inclusive curriculum (Clarasatin Rera Owa, n.d.). Indonesia should learn from Finland's success and adopt a quality education model. Finland is ranked 3rd after China and Korea, while Indonesia is 57th out of 67 countries (Ananda et al., 2023). There are several indicators of the quality of education in Indonesia and Finland, namely the education system in Indonesia is characterized by competition while Finland prioritizes equality and learning hours in Indonesia +/- 40 hours per week while Finland is 30 hours per week (I. E. D. Putra et al., 2023).

Academic stress affects stress coping strategies among students. This can be seen, if a student experiences high academic stress they are more likely to use emotional problem-focused coping strategies (Che Bakar & Surat, 2022). This causes several factors for students to take bad risks such as harming themselves and can be desperate to commit suicide (R. & F., 2023). There are several factors that cause students to experience stress, namely First is the factor of the role of parents in instilling character values which is often caused by wrong parenting. Secondly, the role of schools should also be a place for good moral learning but some teachers do not support it because they are too harsh, indifferent, and often humiliate students which has a negative impact on character building (Faiz et al., 2021).

Random Forest Classifier is one of the classifiers that is often used in building ensemble models by combining many decision trees (Aldiansyah & Saputra, 2023). The advantages of Random Forest include its ability to produce accurate classifications with low error rates, handle large amounts of data, and effectively handle missing data (Husin, 2023). Random Forest forms many independent trees by using a randomly selected subset through the boostrap method of training samples and input variables at each node (P. H. Putra et al., 2023).

Research conducted by Dhea Fuji Astari, Yulison Herry Chrisnanto, and Melina in 2023 entitled “Classification of Stress Levels During Sleep Using Random Forest Algorithm”. In this study the main variables used include snoring, breathing frequency, body temperature, limb movement, oxygen levels, eye movement, sleep duration, and heart rate. This research resulted in an accuracy of 93.65%, with the advantage of the Random Forest algorithm being able to handle sensitive data through a combination of different variables in each decision tree. This result confirms that Random Forest is an effective method for classification of sleep stress levels (Fuji Astari et al., 2024).

Further research in 2024 applied the Random Forest Algorithm to predict the early stages of heart disease (RSIJ case study). Researchers concluded that implementing the Random Forest algorithm can predict heart disease with an accuracy value of 86.9%, with a sensitivity of 90.6% and a specificity of 82.7%. Based on Receiver Operating Characteristic (ROC) analysis, researchers obtained the diagnosis rate for heart disease prediction using Random Forest is 93.3%. Random Forest algorithm has proven to be the most efficient algorithm for heart disease classification (Nangon, 2024).

Furthermore, research conducted by Yoga Religia, Agung Nugroho, and Wahyu Hadikristanto in 2021 which Compares Optimization Algorithms on Random Forest (RF) in Bank Marketing Data Classification related to loan application acceptance. Classification is carried out with the RF algorithm to predict loan acceptance optimally. This study also compares Random Forest optimization with Bagging and Genetic Algorithm. The results show that the best performance is achieved by RF with 88.30% accuracy, AUC(+) 0.500, and AUC(-) 0.000. Optimization with Bagging and Genetic Algorithm did not improve the performance of Random Forest (Yoga Religia et al., 2021).

Another research by Arifin Yusuf Permana, et al, in 2023 discussed the Application of Data Mining for Lung Cancer Prediction Using Random Forest Algorithm. This study displays the exact number of lung cancer patients from the data analyzed, and uses Random Forest because of its ability to select the optimal data subset. With the right feature selection, this algorithm works efficiently on big data and helps classify lung cancer patients, both positive and negative. These observations formed the basis of the lung cancer prognosis analysis, and Random Forest's performance was tested on a lung cancer survey dataset using Rapidminer software. As a result, the algorithm achieved an accuracy of 90.61% with an AUC value of 0.941, proving its superiority in lung cancer prediction (Arifin Yusuf Permana et al., 2023).

This study aims to determine the classification of stress levels among students. Calculated and classified into three categories namely low, medium, and high. To classify stress levels, a Random Forest algorithm approach is used, so that countermeasures can be taken to help them overcome stress levels, before stress reaches a critical point and students take the wrong steps (Patil et al., 2021).

RESEARCH METHODS

This research method uses the Random Forest algorithm which is one of the classification techniques in machine learning applied to the data mining process by creating a decision tree presented as a rule. This algorithm uses two types of input, namely training samples to build a decision tree that is tested for accuracy, and samples as data to be classified (Rahmadhani et al., 2020). The sequence of this research flow is illustrated in the following figure.

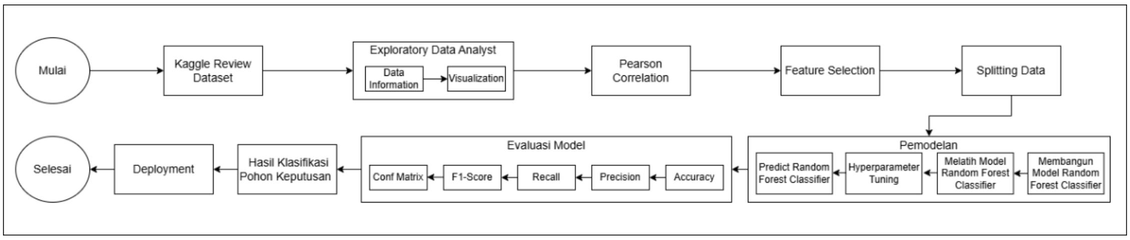


Figure 1. Research Flow

This research flow is explained as follows :

Data Collection

Data collection took the form of taking a dataset entitled Student Stress Factors A Comprehensive Analysis consisting of 1100 datasets and 21 features available on the kaggle platform for further analysis.

Exploratory Data Analyst (EDA)

The Exploratory Data Analyst process includes two stages, the first is to see an overview of the characteristics of the dataset content by displaying the basic information contained in the dataset to understand the dataset structure and find out each value whether there are NaN values, the second is data visualization using heatmaps to understand the most influential features and correlation analysis between each feature of the relationship between variables.Pearson Correlation

The Pearson Correlation process is used to filter the relationships of variable 1 and variable 2 that are positively correlated >0.6 to understand the inter-correlation value of each student stress feature.

Feature Selection

The feature selection process on features that are positively correlated >0.6 with the Pearson correlation method. The selected features include Anxiety level, Depression, Future career concerns, Headache, Mental health history and Bullying because these features are directly related to psychological factors and individual experiences that affect stress levels. After that, researchers divided the training data as much as 80% and test data as much as 20%.

Modelling

After feature selection is Random Forest algorithm machine learning modeling which includes building a model, training the model using training and test data sets that have been determined in the previous process, and using the random forest classifier model for prediction.

Hyperparameter Tuning

At this stage hyperparameter tuning using the grid search method is used to optimize model parameters to get the best performance. With parameters that have been determined in random forest.

Model Performance Evaluation

A thorough evaluation of the model was conducted starting from a simple metric of Accuracy to a more in-depth analysis of Confusion Matrix. to identify the strengths and weaknesses of the model, allowing for improvements to be made if needed.

Random Forest Classification

Based on the performance evaluation of the model, it presents the final results of the classification including the accuracy of the model and the interpretation of the classification results of the Random Forest model using decision trees. In random forest using the following calculation formula:

$$RFs(X) = argmax \sum_{s=1}^{S} I(RFs(x) = j), for\ j = 1, \dots, K \tag{1}$$

Where :

RFs(x) determines the final class for the test pattern x by selecting the class label j that is most predicted by the S classifiers. Each classifier assigns a label, and the indicator I (RFs(x) = j) is 1 if the selected label is j, or 0 otherwise. The total number of indicators indicates how many classifiers selected j, and argmax selects the label with the highest number as the final prediction.

Deployment

In the last stage, the deployment process is performed by implementing the best model into a website so that it can be used as a basis for designing more effective intervention strategies in managing stress in students.

RESULTS AND DISCUSSION

Data Collection

In the data collection process, the dataset used in this study came from Kaggle with the title “Student Stress Factors: A Comprehensive Analysis”, which was uploaded in November 2023. Data collection is done by reviewing datasets using google colab by utilizing the pandas library to analyze datasets from kaggle with the keyword “Student Stress Factors”. The dataset consists of 1100 data entries and 21 columns, which are then saved in CSV format. Details of the attributes contained in the dataset can be seen below:

DOI: https://doi.org/10.5281/zenodo.15130385

Table 1. Overall Attributes

No.	Column Name	Data Type	Description
1	Anxiety Level	Int	Tingkat kecemasan
2	Self Esteem	Int	Harga diri
3	Mental Health History	Int	Riwayat kesehatan mental
4	Depression	Int	Depresi
5	Headache	Int	Sakit kepala
6	Blood Pressure	Int	Tekanan darah
7	Sleep Quality	Int	Kualitas tidur
8	Breathing Problem	Int	Masalah pernafasan
9	Noise Level	Int	Tingkat kebisingan
10	Living Conditions	Int	Kondisi kehidupan
11	Safety	Int	Keamanan
12	Basic Needs	Int	Kebutuan dasar
13	Academic Performance	Int	Kinerja akademik
14	Study Load	Int	Beban belajar
15	Teacher Students Relationship	Int	Hubungan guru dan murid
16	Future Career Concerns	Int	Kekhawatiran karier masa depan
17	Social Support	Int	Dukungan sosial
18	Peer Pressure	Int	Tekanan teman sebaya
19	Extracurricular Activities	Int	Kegiatan ekstrakurikuler
20	Bullying	Int	Perundungan
21	Strees Level	Int	Tingkat stress

Exploratory Data Analyst (EDA)

The Exploratory Data Analyst stage is conducted to see patterns in the data and identify relationships between variables. The main purpose of the EDA process is to understand the structure of the dataset, find interesting patterns or anomalies. The following are the EDA stages carried out in this study.

Data Information : At this stage is to check the data whether there is a null value or the absence of undefined data values in a variable, column, and element.

Table 2. Data Infomation

Atributes	Total Non-Null	Data Type
Anxiety Level	1100 non-null	Int64
Self Esteem	1100 non-null	Int64
Mental Health History	1100 non-null	Int64
Depression	1100 non-null	Int64
Headache	1100 non-null	Int64
Blood Pressure	1100 non-null	Int64
Sleep Quality	1100 non-null	Int64
Breathing Problem	1100 non-null	Int64
Noise Level	1100 non-null	Int64
Living Conditions	1100 non-null	Int64
Safety	1100 non-null	Int64
Basic Needs	1100 non-null	Int64
Academic Performance	1100 non-null	Int64
Study Load	1100 non-null	Int64
Teacher Students Relationship	1100 non-null	Int64
Future Career Concerns	1100 non-null	Int64
Social Support	1100 non-null	Int64
Peer Pressure	1100 non-null	Int64
Extracurricular Activities	1100 non-null	Int64
Bullying	1100 non-null	Int64
Strees Level	1100 non-null	Int64

This table 2 the process of checking whether there are null values in all features in the dataset using the code above. The results of checking using python show that the dataset used by researchers in this study has good quality, without any missing values in all variables.

Visualization : After checking the null values, the next step is to display the correlations in the data set "Student Stress Factors: A Comprehensive Analysis". Visualization was done using python by utilizing the matplotlib.pyplot and seaborn libraries. The visualization details are available in the image below :

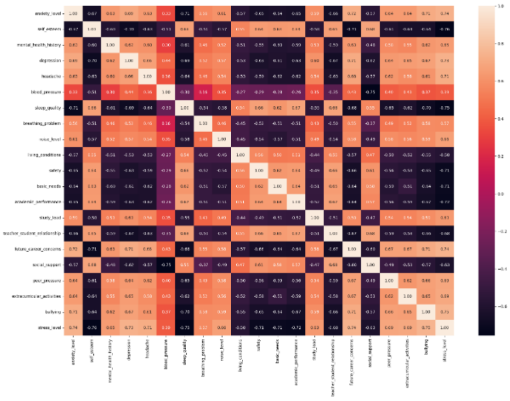


Figure 2. Heatmap Data Correlation Visualization

Figure 2. Heatmap Data Correlation Visualization Using the heatmap before going into the preprocessing stage, it displays a correlation calculation that measures the extent to which two variables have a linear relationship, with values varying from -1 which is a perfect negative relationship to 1 which is a perfect positive relationship. From the visualization patterns and relationships between variables can be observed with a strong negative correlation between anxiety and sleep quality of -0.71, and a strong positive correlation between anxiety and depression of 0.70.

Pearson Correlation

In this study, the pearson correlation method was used to measure the strength and direction of the linear relationship between variable 1 and variable 2. The results of this analysis give researchers an idea of how strong and significant the relationship between the two variables is.

	bullying	stress_level
anxiety_level	0.709982	0.736795
self_esteem	NaN	NaN
mental_health_history	0.624366	0.648644
depression	0.665790	0.734379
headache	0.609775	0.713484
blood_pressure	NaN	NaN
sleep_quality	NaN	NaN
breathing_problem	NaN	NaN
noise_level	NaN	0.663371
living_conditions	NaN	NaN
safety	NaN	NaN
basic_needs	NaN	NaN
academic_performance	NaN	NaN
study_load	NaN	0.634156
teacher_student_relationship	NaN	NaN
future_career_concerns	0.711278	0.742619
social_support	NaN	NaN
peer_pressure	0.661058	0.690684
extracurricular_activities	0.651979	0.692977
bullying	NaN	0.751162
stress_level	0.751162	NaN

[21 rows x 21 columns]

Pasangan Fitur dengan Korelasi Positif > 0.6:

	Feature 1	Feature 2	Correlation
0	anxiety_level	mental_health_history	0.634450
1	anxiety_level	depression	0.694340
2	anxiety_level	headache	0.632738
3	anxiety_level	noise_level	0.608624
4	anxiety_level	future_career_concerns	0.717016
..	...	...	...
105	stress_level	study_load	0.634156
106	stress_level	future_career_concerns	0.742619
107	stress_level	peer_pressure	0.690684
108	stress_level	extracurricular_activities	0.692977
109	stress_level	bullying	0.751162

[110 rows x 3 columns]

Figure 3. Pearson Correlation and Pearson Correlation Positive > 0.6

Figure 3. Pearson Correlation and Figure 4. Pearson Correlation Positive > 0.6 Pearson correlation analysis provides an initial overview of the relationship between various variables related to student stress factors and displays pairs of features in a dataset that have a very strong positive relationship. The process begins by calculating the pearson correlation matrix for all feature pairs, then filtering only those pairs that have a

correlation value above 0.6. This pearson correlation method is important for further analysis, such as selecting relevant features for machine learning models.

Feature Selection

The feature selection process explains how to select features using the random forest classifier algorithm. In the feature selection process using the Pearson correlation method, it succeeds in filtering features that have a strong relationship with the target stress level variable. By selecting features that have a significant effect on student stress levels, such as Anxiety level, Depression, Future career concerns, Headache, Mental health history, and Bullying, researchers can improve model performance, reduce model complexity, and improve model interpretation.

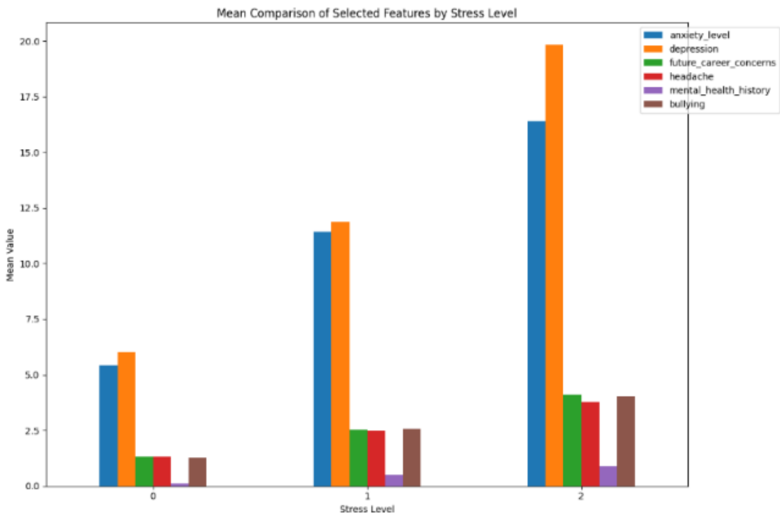


Figure 5. Bar Graph 6 Features of Stress in Students

In the picture 5. Displaying the average bar graph regarding the classification of the six features selected in the feature selection process above by stress level. Depression and Anxiety Level are the most dominant features at all stress levels. At low stress level Level 0, the average values of Anxiety and Depression are still low but higher than other features while Headache, Future Career Concerns, Bullying, and Mental Health History tend to be stable and low. At moderate stress Level 1, there is a significant increase in Anxiety and Depression, followed by an increase in Future Career Concerns and Headache. At high stress Level 2, Depression had the highest mean score, followed by Anxiety Level, with additional contributions from Future Career Concerns, Headache, and Bullying, while Mental Health History remained stable.

Splitting Data

The data splitting stage divides the dataset into two parts, namely training data and testing data. This division aims to train the prediction model using part of the data and test the performance of the model on data that has never been used in the training process.

Tabel 3. Splitting Data

Pembagian	Jumlah Baris	Rasio (%)	Keterangan
Training Set	770	80 %	Untuk melatih model Random Forest
Test Set	330	20 %	Untuk evaluasi kinerja model

Table 3. Division of Training Data and Test Data in the source code snippet above, the data division process is carried out using the train_test_split function from the scikit-learn library in python with a ratio of 80:20 to divide the dataset. The 80:20 proportion was chosen because it provides a balance between data for training and model evaluation. The split was done by random sampling to ensure an even distribution of data between

training and test data. This technique avoids bias that may arise if the data is divided non-randomly.

Modelling

After dividing the data into training and test data, the next step is to build a model to learn the patterns in the data and make predictions. Modeling is used to train and test the prepared dataset using the Random Forest algorithm.

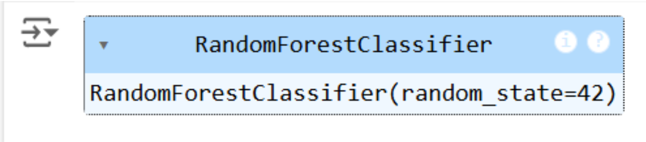


Figure 6. Modelling Random Forest Classifier

Figure 6. The stages of modelling a random forest classifier model to produce a classification model that is able to reduce the risk of overfitting. In this research, the parameter n_estimators = 100 the model will build 100 decision trees to make predictions and random_state parameter = 42 to ensure the results obtained are consistent and reproducible in each execution.

Hyperparameter Tuning

At this stage hyperparameter tuning is performed to improve accuracy before evaluation. This process involves setting parameters that were not optimized during training. The tuning method used is grid search which includes a thorough search through hyperparameter combinations.

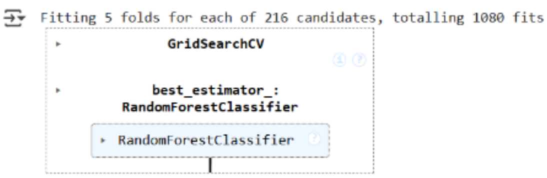


Figure 7. Grid Search Cross Validation

The figure 7. shows the results of the GridSearchCV process that searches for the best parameters for the RandomForestClassifier model. This process uses 5-fold cross-validation, where each combination of 216 parameters is tested by dividing the training data into 5 parts, training the model on 4 parts, and testing on the remaining part, resulting in a total of 1080 iterations.

Model Performance Evaluation

The model evaluation stage is to measure the performance of the classification model that has been built and trained using the Random Forest algorithm. Evaluation is done by comparing the predicted results of the y_pred model with the actual value of the test data. In this research, several evaluation metrics are used to get a comprehensive picture of model performance, including accuracy, precision, recall, F1-Score, and confusion matrix.

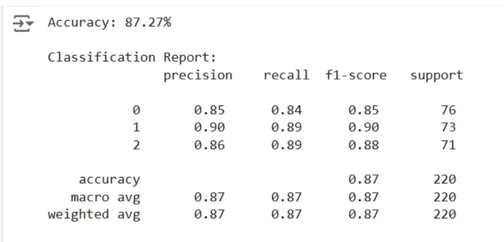
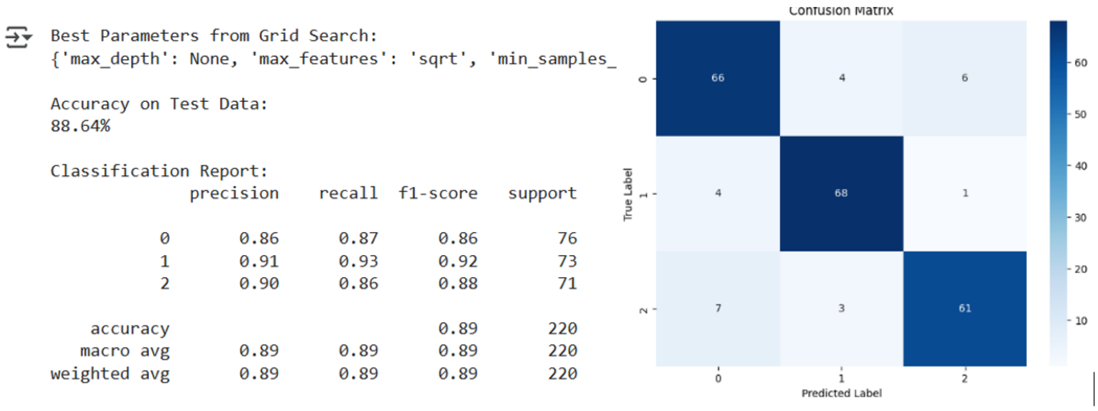

```
Accuracy: 87.27%

Classification Report:
              precision    recall  f1-score   support

           0       0.85      0.84      0.85        76
           1       0.90      0.89      0.90        73
           2       0.86      0.89      0.88        71

    accuracy                           0.87       220
   macro avg       0.87      0.87      0.87       220
weighted avg       0.87      0.87      0.87       220
```

Figure 8. Evaluation of Classification Report Model Before Tuning

This figure 8. Evaluation Model Classification Report before tuning displays the results of the evaluation of the model that has been built using the Random Forest algorithm before tuning. Based on the evaluation results, the model has achieved an accuracy of 87.27% in classifying the test data.



```
Best Parameters from Grid Search:
{'max_depth': None, 'max_features': 'sqrt', 'min_samples_

Accuracy on Test Data:
88.64%

Classification Report:
              precision    recall  f1-score   support

           0       0.86      0.87      0.86        76
           1       0.91      0.93      0.92        73
           2       0.90      0.86      0.88        71

    accuracy                           0.89       220
   macro avg       0.89      0.89      0.89       220
weighted avg       0.89      0.89      0.89       220
```


Figure 9. Evaluation of Classification Report Model After Tuning and Confusion Matrix

Figure 9 displays the evaluation of the Random Forest Classifier model after tuning using GridSearchCV, which achieves 88.64% accuracy on the test data with the best parameters of max_depth, max_features, min_samples_leaf 4, and min_samples_split 5. Metrics such as precision, recall, and F1-score show good performance in detecting each class, ensuring balanced predictions. Displays the Confusion Matrix after tuning, where rows represent the actual classes and columns represent the predicted classes. The main diagonal values indicate the number of correctly predicted samples, while the off-diagonal values indicate misclassification. The model shows high accuracy for classes 0 and 1, although there are still some misclassifications for class 2.

Random Forest Classification

In the classification stage using decision trees, a decision tree is created based on the results of the stress classification on students who have been trained and modeled. The decision tree was chosen because of its ability to handle numerical and categorical data, easy model interpretation.

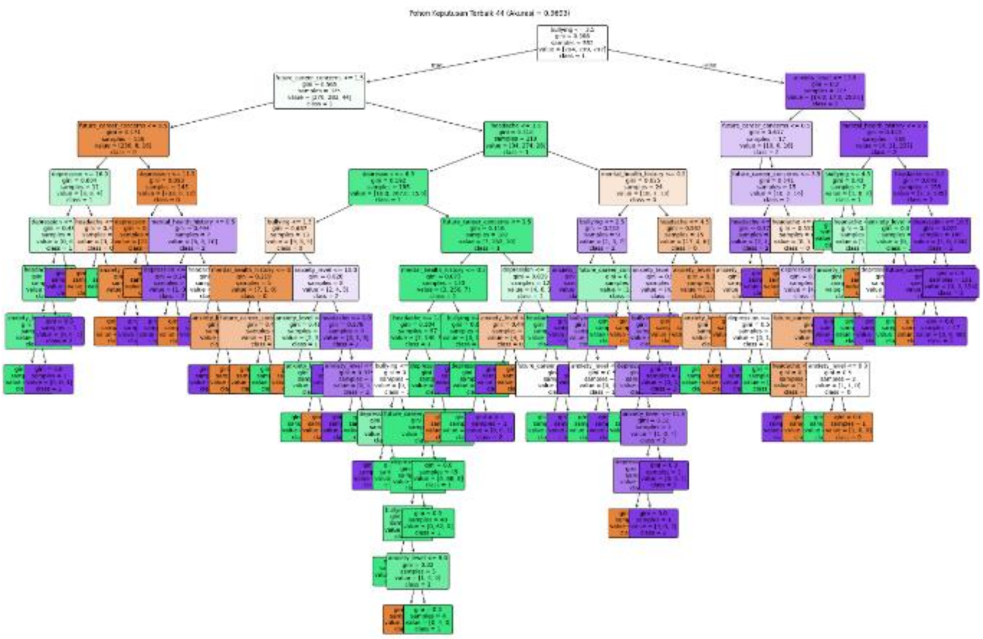


Figure 11. Best Random Forest for Decision Tree Classification

The image 11. Displays the Best Decision Tree by checking the bullying value If it is more than 3.5, the data is classified to class 2 with an accuracy of 0.9693. Otherwise, checks future career concerns. Values less than equal to 1.5 lead to class 1 with an accuracy of 0.8565, while higher values proceed to anxiety level. If the anxiety level is more than 13.5, the data goes to class 2 with an accuracy of 0.9253, otherwise headache is checked. Values less than equal to 3.5 remain in class 1, while larger values check mental health history. If depression is less than equal to 1.5 the data remains in class 1, if the value is more than bullying and anxiety level are checked again until the data is classified into class 1 or 2 based on the last value.

Deployment

The deployment stage is done by implementing the best model into a web application. The functional requirements of this system are that the website can classify the causes of stress in students and display the results of the stress level.

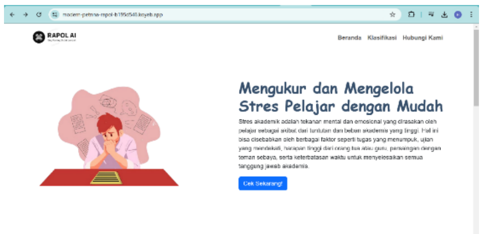


Figure 12. Display Deployment

Picture 12 The Deployment shows that this application is designed to address the problem of academic stress that is often experienced by students caused by various factors, such as piling up assignments, approaching exams, and high expectations from parents or teachers. There is a Check Now button that invites users to use the app immediately.

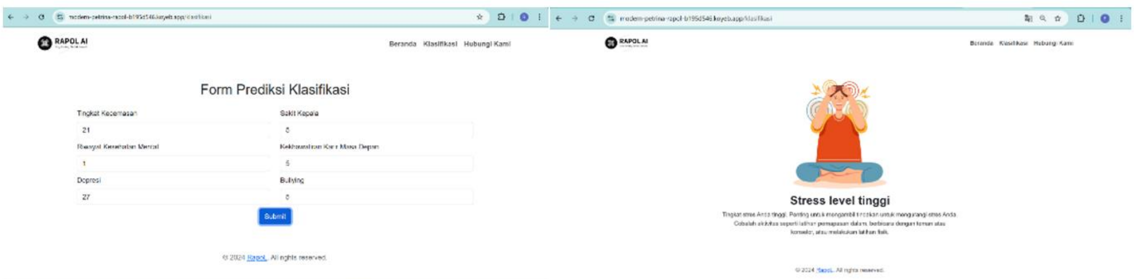


Figure 13. Display of Stress Cause Classification Form and Stress Level Result Form

The figure 13. displays the Stress Cause Classification Form by filling in information related to various features such as anxiety level between 0-30, mental health history between 0 and 1, depression between 0-30, headaches between 0-5, concerns about future career between 0-5, and bullying level between 0-5. After filling in all the features, you can press the submit button to predict the classification based on the data entered. Figure displaying the Stress Level Results Form shows that the learner is experiencing high levels of stress, as illustrated by an individual who appears anxious and frustrated. To help overcome this condition, the website provides recommendations for activities that can be done, such as breathing exercises, talking to friends or counselors, or doing physical activity. This website can be accessed publicly by accessing the following link https://modern-petrina-rapol-b195d546.koyeb.app/.

CONCLUSIONS AND SUGGESTIONS

Conclusion

This research successfully developed a stress classification model among students using the Random Forest algorithm through the stages of data collection, exploratory data analysis, data preprocessing, data splitting, model training and model performance evaluation and deployment as a measure of stress level.

The application of the Random Forest method produces high accuracy, before tuning it is 87.27% and after tuning it is 88.64%. This shows that the model is able to classify student stress based on factors that have a significant effect, including anxiety level, depression, future career concerns, headaches, mental health history, and bullying.

Suggestion

Future research that wants to explore similar topics can add other factors from the causes of stress, such as from the family environment and lifestyle in order to get a more comprehensive picture.

Increase the quality of the dataset for better generalization of the model and perform comparisons with other algorithms, such as Gradient Boosting or XGBoost to see if the accuracy can be improved further.

REFERENCE

Aldiansyah, S., & Saputra, R. A. (2023). Comparison of Machine Learning Algorithms for Land Use and Land Cover Analysis Using Google Earth Engine (Case Study: Wanggu Watershed). *International Journal of Remote Sensing and Earth Sciences (IJReSES)*, *19*(2), 197. https://doi.org/10.30536/j.ijreses.2022.v19.a3803

Ananda, R., Syaputri, W. I., Suhesni, T., & Rossadah, N. (2023). Perbandingan Pendidikan di Indonesia dan Pendidikan di Finlandia. *JIIP - Jurnal Ilmiah Ilmu Pendidikan*, *6*(9), 6689–6694. https://doi.org/10.54371/jiip.v6i9.2812

Arifin Yusuf Permana, Hari Noer Fazri, M.Fakhrizal Nur Athoilah, Mohammad Robi, & Ricky Firmansyah. (2023). Penerapan Data Mining Dalam Analisis Prediksi Kanker Paru Menggunakan Algoritma Random Forest. *Jurnal Ilmiah Teknik Informatika Dan Komunikasi*, *3*(2), 27–41. https://doi.org/10.55606/juitik.v3i2.472

Che Bakar, W. M. A., & Surat, S. (2022). Stres Akademik dan Strategi Daya Tindak dalam kalangan Pelajar: Satu Kajian Sistematik. *Malaysian Journal of Social Sciences and Humanities (MJSSH)*, *7*(3), e001330. https://doi.org/10.47405/mjssh.v7i3.1330

Clarasatin Rera Owa. (n.d.). *https://unair.ac.id/post_fetcher/fakultas-ilmu-budaya-perkembangan-pendidikan-di-indonesia-dar i-masa-ke-masa/*. Februari 20,2024. https://unair.ac.id/post_fetcher/fakultas-ilmu-budaya-perkembangan-pendidikan-di-indonesia-dari-masa-ke-masa/

Faiz, A., Soleh, B., Kurniawaty, I., & Purwati, P. (2021). Tinjauan Analisis Kritis Terhadap Faktor Penghambat Pendidikan Karakter di Indonesia. *Jurnal Basicedu*, *5*(4), 1766–1777. https://doi.org/10.31004/basicedu.v5i4.1014

Fuji Astari, D., Herry Chrisnanto, Y., & Melina, M. (2024). Klasifikasi Tingkat Stres Saat Tidur Menggunakan Algoritma Random Forest. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *7*(5), 3676–3684. https://doi.org/10.36040/jati.v7i5.7750

Husin, N. (2023). Komparasi Algoritma Random Forest, Naïve Bayes, dan Bert Untuk Multi-Class Classification Pada Artikel Cable News Network (CNN). *Jurnal Esensi Infokom : Jurnal Esensi Sistem Informasi Dan Sistem Komputer*, *7*(1), 75–84. https://doi.org/10.55886/infokom.v7i1.608

Nangon, T. (2024). Prediksi Tahap Awal Penyakit Jantung Menggunakan Algoritma Random Forest (Studi Kasus RSIJ). *Da'watuna: Journal of Communication and Islamic Broadcasting*, *4*(4), 1561–1567. https://doi.org/10.47467/dawatuna.v4i4.1882

Patil, S., Sultana, H. P., Gregorio, A. J., & ... (2021). Mind Stress Analysis of Students using Classification Techniques. *Linguistica ...*, *February*. https://www.researchgate.net/profile/Abner-Fonseca-Livias/publication/358781143_Mind_Stress_Analysis_of_Students_using_Classification_Techniques/links/6214f9d46c472329dcfe1241/Mind-Stress-Analysis-of-Students-using-Classification-Techniques.pdf

Putra, I. E. D., Rusdinal, R., Ananda, A., & Gistituati, N. (2023). Perbandingan Kurikulum Pendidikan Indonesia dan Finlandia. *Journal on Education*, *06*(01), 7437–7448. https://journal.uinsi.ac.id/index.php/bjie/article/download/7346/2511/

Putra, P. H., Azanuddin, A., Purba, B., & Dalimunthe, Y. A. (2023). Random forest and decision tree algorithms for car price prediction. *Jurnal Matematika Dan Ilmu Pengetahuan Alam LLDikti Wilayah 1 (JUMPA)*, *4*(1), 81–89. https://doi.org/10.54076/jumpa.v3i2.305

R., A., & F., L. M. (2023). Stress Among Secondary School Students: Exploring Stress Causes and Stress Measurement Instruments. *Online Journal for TVET Practitioners*, *8*(2), 69–79. https://doi.org/10.30880/ojtp.2023.08.02.007

Rahmadhani, S. A., Asih, M. S., & Wulan, N. (2020). Implementasi Algoritma Decision Tree C.45 Pada Klasifikasi Penyakit Tuberkulosis. *Seminar Nasional Teknologi Informasi & Komunikasi*, 230–238.

Yoga Religia, Agung Nugroho, & Wahyu Hadikristanto. (2021). Klasifikasi Analisis Perbandingan Algoritma Optimasi pada Random Forest untuk Klasifikasi Data Bank Marketing. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, *5*(1), 187–192. https://doi.org/10.29207/resti.v5i1.2813