

Klasifikasi Tingkat Stres Mahasiswa Dalam Penyelesaian Tugas Akhir Menggunakan Naïve Bayes Dan K-Nearest Neighbor

¹Lenni Pefrianti, ²Ibnu Rasyid Munthe, ³Irmayanti, ⁴Masrizal

^{1,2,3,4}Sistem Informasi, Fakultas Sains dan Teknologi, Universitas Labuhanbatu

Email : ¹lennifebrianti9@gmail.com, ²ibnurasyidmunthe@gmail.com,
³irmayantiritonga02@gmail.com, ⁴masrizal120405@gmail.com

Corresponding Author : lennifebrianti9@gmail.com

Abstract

This study aims to analyze the stress levels of final-year students and compare the performance of Naïve Bayes and K-Nearest Neighbor (KNN) algorithms in stress classification. Data were collected from 82 respondents through a questionnaire consisting of seven variables (S1–S7) measuring factors contributing to stress, which were classified into low, moderate, and high stress levels. The results show that both algorithms can classify student stress effectively, with Naïve Bayes achieving the highest accuracy (90.15%) compared to KNN (87.72%). Distribution analysis by study program indicates that Agrotechnology has the highest proportion of students with high stress (42.86%), followed by Information Systems (40.63%) and Information Technology (13.64%). This study provides insights for the university to offer targeted support through counseling or stress management workshops.

Keyword : Student Stress Levels, Naive Bayes Algorithm, K-Nearest Neighbor Algorithm.

1. Pendahuluan

Mahasiswa tingkat akhir rentan mengalami tekanan tinggi selama penyusunan tugas akhir akibat kompleksitas penelitian, tuntutan akademik, serta tanggung jawab menghasilkan karya ilmiah sesuai standar perguruan tinggi. Tekanan tersebut dipengaruhi oleh kesulitan memahami teori, keterbatasan referensi, intensitas bimbingan, serta faktor pribadi seperti kondisi ekonomi dan dukungan sosial (Djoar et al., 2024). Stres yang tidak terkelola dapat menurunkan konsentrasi, motivasi, dan kinerja akademik. Perkembangan *machine learning* memungkinkan analisis data psikologis secara lebih akurat untuk memprediksi tingkat stres. Penelitian ini menerapkan algoritma Naive Bayes dan K-Nearest Neighbors (KNN) pada data DASS-21 guna menghasilkan model klasifikasi tingkat stres mahasiswa. Hasil penelitian diharapkan membantu kampus dalam memetakan kondisi psikologis mahasiswa dan meningkatkan layanan pendampingan akademik.

2. Landasan Teori

Machine Learning

Machine Learning adalah bagian dari kecerdasan buatan yang memungkinkan komputer belajar dari data untuk mengenali pola yang sulit terlihat oleh manusia (Sarker, 2021). Dalam penelitian ini, Machine Learning digunakan untuk melakukan klasifikasi tingkat stres mahasiswa berdasarkan data kuesioner (Dewi Nasien et al., 2024).

Stres

Stres merupakan respons psikologis dan fisiologis individu terhadap tekanan atau tuntutan yang melebihi kemampuan adaptasi (Yuda et al., 2022). Dalam konteks akademik, stres sering dialami mahasiswa akibat beban tugas, tekanan waktu, dan tuntutan penyelesaian studi (Chai, 2025). Pengukuran tingkat stres dalam penelitian ini menggunakan instrumen adaptasi DASS-21 yang telah banyak digunakan dalam penelitian kesehatan mental (Thiyagarajan et al., 2022).

Algoritma Naïve Bayes

Naïve Bayes merupakan algoritma klasifikasi berbasis probabilitas yang menggunakan Teorema Bayes dengan asumsi independensi antar atribut (Susanti, 2024). Algoritma ini dikenal sederhana, cepat, dan efektif untuk dataset berukuran kecil hingga menengah. Dalam penelitian ini, Naïve Bayes digunakan untuk memprediksi kategori tingkat stres mahasiswa berdasarkan jawaban kuesioner (Wibisono et al., 2025).

$$P(C|X_1, X_2, \dots, X_n) = P(C) \prod_{i=1}^n P(X_i|C).$$

Keterangan:

$P(C|X)$ = Probabilitas kelas stres (Rendah, Sedang, Tinggi)

$P(C)$ = Probabilitas prior kelas

$P(X_i|C)$ = Probabilitas atribut ke- i pada kelas C

n = jumlah atribut

Algoritma K-Nearest Neighbor

K-Nearest Neighbor (KNN) merupakan algoritma klasifikasi yang menentukan kelas suatu data berdasarkan mayoritas kelas dari K tetangga terdekatnya (Alamsyah & Wijaya, 2025). Metode ini menggunakan perhitungan jarak, seperti Euclidean Distance, untuk menentukan kedekatan antar data. Dalam penelitian ini, KNN digunakan sebagai metode pembandingan untuk mengetahui performa klasifikasi tingkat stres mahasiswa (Najjar & Majeed Breesam, 2023).

Rumus yang paling umum digunakan adalah Euclidean Distance:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

Keterangan:

$d(x, y)$ = jarak antara data uji dan data latih

x_i = nilai atribut ke- i data uji

y_i = nilai atribut ke- i data latih

n = jumlah atribut

Google Colab

Google Colaboratory (Google Colab) merupakan platform berbasis cloud yang menyediakan lingkungan pemrograman Python secara daring tanpa perlu instalasi lokal. Platform ini mendukung berbagai pustaka Machine Learning seperti Scikit-learn dan Pandas sehingga memudahkan proses pengolahan dan analisis data. (Dewi Nasien et al., 2024)

Microsoft Excel

Microsoft Excel merupakan perangkat lunak pengolahan data yang digunakan untuk input, pengkodean, dan praproses data sebelum dilakukan analisis lebih lanjut (Divisi et al., 2017).

3. Metodologi Penelitian

Jenis Penelitian

Penelitian ini menggunakan pendekatan kuantitatif karena data yang dikumpulkan berupa angka dari hasil kuesioner. Data tersebut kemudian diolah dan dianalisis menggunakan metode Machine Learning untuk mengklasifikasikan tingkat stres mahasiswa.

Populasi dan Sampel

Populasi penelitian berjumlah 452 mahasiswa tingkat akhir angkatan 2022 Fakultas Sains dan Teknologi Universitas Labuhanbatu (Sistem Informasi 174, Teknologi Informasi 1, Agroteknologi 153). Sampel sebanyak 82 mahasiswa ditentukan menggunakan rumus Slovin

$$n = \frac{N}{1 + N(e)^2}$$

n = jumlah sampel

N = jumlah populasi

e = tingkat kesalahan (margin of error)

1 = konstanta

(e = 10%) dan dibagi proporsional menjadi SI 32 orang, TI 22 orang, dan Agro 28 orang.

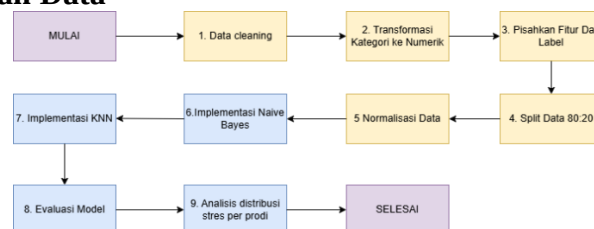
Instrumen Penelitian

Instrumen penelitian berupa kuesioner yang diadaptasi dari Depression Anxiety Stress Scales 21 (DASS-21). Dari 21 pernyataan yang mencakup aspek depresi, kecemasan, dan stres, penelitian ini hanya menggunakan 7 pernyataan khusus pada aspek stres. Penilaian menggunakan skala Likert 4 poin: (1) Pernah, (2) Jarang, (3) Sering, dan (4) Sangat Sering.

Teknik Pengumpulan Data

Data dikumpulkan melalui kuesioner adaptasi DASS-21 untuk mengukur tingkat stres mahasiswa tingkat akhir dengan skala Likert.

Tahapan Pengolahan Data



Gambar 1. Flowchart Pengolahan Data Tingkat Stres Mahasiswa

1. Data Cleaning

Data diperiksa untuk memastikan tidak ada nilai kosong, data ganda, atau kesalahan input. Data yang tidak lengkap dihapus agar hasil analisis lebih akurat.

2. Transformasi Kategori ke Numerik

Jawaban kuesioner berbentuk kategori diubah menjadi skala numerik 1-4 agar dapat diproses oleh algoritma Machine Learning. Dan kategori Rendah=1, Sedang=2, Tinggi=3

3. Pemisahan Fitur dan Label
Data dipisahkan menjadi fitur (X) sebagai variabel input dan label (y) sebagai target klasifikasi tingkat stres.
4. Split Data
Dataset dibagi menjadi 80% data training dan 20% data testing menggunakan `train_test_split`.
5. Normalisasi Data
Dilakukan normalisasi menggunakan Min-Max Scaling untuk menyamakan rentang nilai, terutama untuk meningkatkan kinerja algoritma KNN.
6. Implementasi Naïve Bayes
Model Naive Bayes diterapkan untuk mengklasifikasikan tingkat stres ke dalam kategori rendah, sedang, dan tinggi berdasarkan pendekatan probabilitas.
7. Implementasi K-Nearest Neighbor
Algoritma KNN menentukan kategori stres berdasarkan kedekatan jarak antar data menggunakan metode Euclidean Distance.
8. Evaluasi Model
Kinerja model diukur menggunakan Accuracy, Precision, Recall, dan F1-Score pada data testing.
9. Analisis Distribusi Tingkat Stres Berdasarkan Prodi
Dilakukan analisis untuk melihat perbandingan tingkat stres antar program studi berdasarkan hasil klasifikasi.

4. Hasil dan Pembahasan

Deskripsi Dataset

Dataset penelitian diperoleh melalui kuesioner kepada 82 mahasiswa tingkat akhir Universitas Labuhanbatu. Terdapat 10 variabel independen yang merepresentasikan faktor penyebab stres dan 1 variabel dependen, yaitu Tingkat Stres.

Tabel 1. Variabel Kuesioner

Variabel	Pertanyaan / Arti	Skala
S1	Saya merasa sulit untuk bersantai selama mengerjakan tugas akhir	1–4
S2	Saya merasa tegang atau gelisah ketika memikirkan tugas akhir	1–4
S3	Saya merasa mudah tersinggung akibat tekanan dalam penyelesaian tugas akhir	1–4
S4	Saya merasa tertekan oleh tuntutan akademik yang berkaitan dengan tugas akhir	1–4
S5	Saya merasa kesulitan mengendalikan emosi ketika menghadapi kendala dalam tugas akhir	1–4
S6	Saya merasa tidak sabar dalam menjalani proses penyelesaian tugas akhir	1–4
S7	Saya merasa lelah secara mental karena memikirkan dan mengerjakan tugas akhir	1–4

Tabel 2. Contoh 5 Data Pertama

ID	Prodi	S1	S2	S3	S4	S5	S6	S7	SKOR	Kategori
R01	Sistem Informasi	2	3	3	3	3	2	3	19	2

R02	Sistem Informasi	2	3	2	2	2	2	2	15	2
R03	Sistem Informasi	4	4	3	3	4	4	4	26	3
R04	Sistem Informasi	3	4	2	3	3	3	3	21	3
R05	Teknologi Informasi	1	1	1	1	1	1	1	7	1

Tabel 3. Distribusi Tingkat Stres

Kategori Stres	Jumlah	Persentase (%)
Rendah	14	17,1%
Sedang	40	48,8%
Tinggi	28	34,1%
Total	82	100%

Terlihat bahwa kategori stres sedang mendominasi, diikuti tinggi dan rendah. Dataset sudah bersih dari missing value maupun duplikasi, sehingga siap dianalisis menggunakan algoritma klasifikasi.

Hasil Implementasi Naïve Bayes

Hasil pengujian model Naïve Bayes menggunakan Google Colab, yang menunjukkan nilai akurasi serta confusion matrix pada data testing. Akurasi model: 1,0 (100%).

```

print("== HASIL NAIVE BAYES ==")
print("Akurasi:", accuracy_score(y_test, y_pred_nb))
print("Confusion Matrix:")
print(confusion_matrix(y_test, y_pred_nb))
print("Classification Report:")
print(classification_report(y_test, y_pred_nb))

== HASIL NAIVE BAYES ==
Akurasi: 1.0
Confusion Matrix:
[[3 0 0]
 [0 7 0]
 [0 0 7]]
Classification Report:
              precision    recall  f1-score   support

    1         1.00        1.00        1.00         3
    2         1.00        1.00        1.00         7
    3         1.00        1.00        1.00         7

 accuracy         1.00        1.00        1.00        17
  macro avg         1.00        1.00        1.00        17
 weighted avg         1.00        1.00        1.00        17
    
```

Gambar 2. Hasil Pengujian Naïve Bayes Pada Google Colab

Tabel 4. Confusion Matrix

Kelas Asli / Prediksi	Rendah (1)	Sedang (2)	Tinggi (3)
Rendah (1)	3	0	0
Sedang (2)	0	7	0
Tinggi (3)	0	0	7

Tabel 5. Classification Report

Kelas	Precision	Recall	F1-Score	Support
Rendah (1)	1.00	1.00	1.00	3
Sedang (2)	1.00	1.00	1.00	7
Tinggi (3)	1.00	1.00	1.00	7
Accuracy	-	-	1.00	17
Macro Avg	1.00	1.00	1.00	17
Weighted Avg	1.00	1.00	1.00	17

Model Naïve Bayes berhasil memprediksi seluruh data testing dengan benar, menghasilkan akurasi 100%. Confusion matrix menunjukkan tidak ada kesalahan prediksi pada semua kategori. Classification report menegaskan precision, recall, dan F1-score semuanya 1,00, menandakan model Naïve Bayes sangat efektif dalam mengklasifikasikan tingkat stres mahasiswa pada dataset ini.

Hasil Implementasi K-Nearest Neighbor

Hasil pengujian model KNN dengan parameter $k=5$, termasuk nilai akurasi dan confusion matrix pada data testing. Akurasi model: 0,824 (82,35%).

```

print("==== HASIL KNN ===")
print("Akurasi:", accuracy_score(y_test, y_pred_knn))
print("Confusion Matrix:")
print(confusion_matrix(y_test, y_pred_knn))
print("Classification Report:")
print(classification_report(y_test, y_pred_knn))

==== HASIL KNN ====
Akurasi: 0.8235294117647058
Confusion Matrix:
[[3 0]
 [2 5]
 [0 1]]
Classification Report:
              precision    recall  f1-score   support

    1         0.60         1.00         0.75         3
    2         0.83         0.71         0.77         7
    3         1.00         0.86         0.92         7

 accuracy          0.81          0.86          0.81         17
 macro avg         0.81          0.86          0.81         17
 weighted avg      0.86          0.82          0.83         17
    
```

Gambar 3. Hasil Pengujian K-Nearest Neighbor pada Google Colab

Tabel 6. Confusion Matrix

Kelas Asli \ Prediksi	Rendah (1)	Sedang (2)	Tinggi (3)
Rendah (1)	3	0	0
Sedang (2)	2	5	0
Tinggi (3)	0	1	6

Berdasarkan data testing, model KNN memiliki akurasi 82,35%. Confusion matrix menunjukkan bahwa kategori Rendah dan Tinggi sebagian besar diprediksi dengan benar, sedangkan kategori Sedang masih terdapat beberapa kesalahan prediksi. Classification report memperlihatkan nilai precision, recall, dan F1-score yang cukup tinggi, menunjukkan model KNN efektif untuk klasifikasi tingkat stres mahasiswa.

Tabel 7. Classification Report

Kelas	Precision	Recall	F1-Score	Support
Rendah (1)	0.60	1.00	0.75	3
Sedang (2)	0.83	0.71	0.77	7
Tinggi (3)	1.00	0.86	0.92	7
Accuracy	-	-	0.82	17
Macro Avg	0.81	0.86	0.81	17
Weighted Avg	0.86	0.82	0.83	17

Perbandingan Kinerja Algoritma

Hasil 5-Fold Cross Validation, di mana Naïve Bayes memperoleh rata-rata akurasi lebih tinggi dibandingkan KNN.

```

from sklearn.naive_bayes import GaussianNB
from sklearn.neighbors import KNeighborsClassifier
from sklearn.model_selection import cross_val_score

# Inisialisasi model
knn = KNeighborsClassifier(n_neighbors=5)
nb = GaussianNB()

# Cross validation
knn_cv = cross_val_score(knn, X, y, cv=5)
nb_cv = cross_val_score(nb, X, y, cv=5)

print("Rata-rata Akurasi KNN (5-fold):", knn_cv.mean())
print("Rata-rata Akurasi Naive Bayes (5-fold):", nb_cv.mean())

Rata-rata Akurasi KNN (5-fold): 0.877285823529412
Rata-rata Akurasi Naive Bayes (5-fold): 0.9814785882352942
    
```

Gambar 4. Proses 5-Fold Cross-Validation pada Google Colab

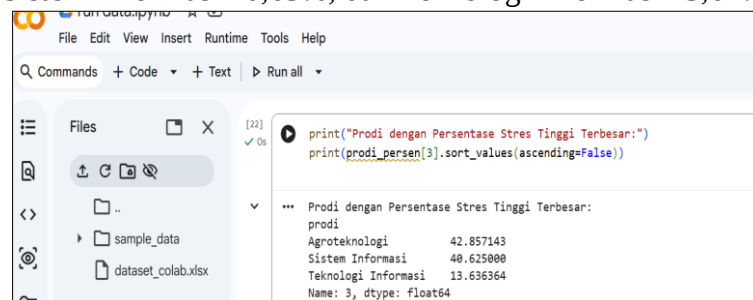
Tabel 8. Perbandingan Kinerja Algoritma

Algoritma	Rata-rata Akurasi (5-Fold CV)
Naïve Bayes	0.901 (90,15%)
K-Nearest Neighbor (k=5)	0.877 (87,72%)

Berdasarkan hasil 5-fold cross-validation, Naïve Bayes memiliki rata-rata akurasi 90,15%, sedikit lebih tinggi dibandingkan KNN dengan k=5 yang mencapai 87,72%. Hal ini menunjukkan bahwa kedua algoritma mampu mengklasifikasikan tingkat stres mahasiswa dengan baik, namun Naïve Bayes sedikit lebih akurat pada dataset ini

Analisis Tingkat Stres Berdasarkan Prodi

Hasil perhitungan persentase mahasiswa dengan kategori stres tinggi pada masing-masing program studi. Agroteknologi memiliki persentase tertinggi sebesar 42,86%, diikuti Sistem Informasi 40,63%, dan Teknologi Informasi 13,64%.



Gambar 5. Hasil Analisis Persentase Stres Tinggi Berdasarkan Program Studi pada Google Colab

Tabel 9. Presentasi Stres Tertinggi Berdasarkan Prodi

Program Studi	Persentase Stres Tinggi (%)
Agroteknologi	42,86 %
Sistem Informasi	40,63%
Teknologi Informasi	13,64%

Berdasarkan hasil klasifikasi tingkat stres, prodi Agroteknologi memiliki persentase mahasiswa dengan tingkat stres tinggi tertinggi, yaitu 42,86%, diikuti oleh Sistem Informasi (40,63%) dan Teknologi Informasi (13,64%). Hal ini menunjukkan bahwa sebagian besar mahasiswa yang paling tertekan berada di prodi Agroteknologi, sedangkan mahasiswa Teknologi Informasi relatif lebih rendah tingkat stresnya. Data ini dapat menjadi acuan bagi pihak kampus untuk memberikan perhatian khusus terhadap mahasiswa di prodi dengan persentase stres tinggi.

5. Kesimpulan dan saran

Penelitian ini menunjukkan bahwa algoritma Naïve Bayes dan K-Nearest Neighbor (KNN) berhasil mengklasifikasikan tingkat stres mahasiswa menjadi kategori rendah, sedang, dan tinggi. Dari hasil 5-fold cross-validation, Naïve Bayes memiliki akurasi tertinggi sebesar 90,15%, sedikit lebih unggul dibanding KNN yang mencapai 87,72%. Berdasarkan distribusi data, prodi Agroteknologi memiliki persentase mahasiswa dengan stres tinggi paling besar (42,86%), diikuti oleh Sistem Informasi (40,63%) dan Teknologi Informasi (13,64%). Untuk penelitian selanjutnya, disarankan menambah jumlah responden dan variabel lain agar analisis lebih mendalam. Sedangkan untuk pihak kampus, sebaiknya memberikan perhatian khusus kepada mahasiswa di prodi dengan stres tinggi melalui konseling, bimbingan tugas akhir, atau workshop manajemen stres.

6. Daftar Pustaka

- Alamsyah, M. F., & Wijaya, A. (2025). Perbandingan Metode KNN dan Naïve Bayes dalam Deteksi Tingkat Stres Berdasarkan Ekspresi Wajah. *Jurnal Informatika: Jurnal Pengembangan IT*, 10(2), 359–369. <https://doi.org/10.30591/jpit.v10i2.8513>
- Chai, S. (2025). The Impact of Adolescent Academic Stress on Academic Achievement. *Lecture Notes in Education Psychology and Public Media*, 109(1), 202–208. <https://doi.org/10.54254/2753-7048/2025.ND26080>
- Dafitrii, H., & Sundari stth, S. (2022). SISTEM PAKAR MENDETEKSI TINGKAT STRES MAHASISWA HARAPAN DALAM PENYUSUNAN SKRIPSI TEKNIK INFORMATIKA DENGAN MENGGUNAKAN METODE TEOREMA BAYES. *Syntax : Journal of Software Engineering, Computer Science and Information Technology*, 2(2), 165–171. <https://doi.org/10.46576/syntax.v2i2.1678>
- Dewi Nasien, Darwin, R., Cia, A., Leo Winata, A., Go, J., M.C, R., Charles Wijaya, R., & Charles Lo, K. (2024). Perbandingan Implementasi Machine Learning Menggunakan Metode KNN, Naive Bayes, dan Logistik Regression Untuk Mengklasifikasi Penyakit Diabetes. *JEKIN - Jurnal Teknik Informatika*, 4(1), 10–17. <https://doi.org/10.58794/jekin.v4i1.640>
- Divisi, D., Leonardo, G. Di, Zaccagna, G., & Crisci, R. (2017). *Basic statistics with Microsoft Excel : a review*. 9(6), 1734–1740. <https://doi.org/10.21037/jtd.2017.05.81>
- Djoar, R. K., Putu, A., Anggarani, M., Airlangga, U., & Fisioterapi, P. S. (2024). *FAKTOR - FAKTOR YANG MEMPENGARUHI STRESS AKADEMIK*. 6(1), 52–59.
- Najjar, E., & Majeed Breesam, A. (2023). Supervised Machine Learning a Brief Survey of Approaches. *Al-Iraqia Journal for Scientific Engineering Research*, 2(4), 71–82. <https://doi.org/10.58564/IJSER.2.4.2023.121>
- Sarker, I. H. (2021). Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN Computer Science*, 2(3), 160. <https://doi.org/10.1007/s42979-021-00592-x>
- Singh, A., Singh, K., Kumar, A., Shrivastava, A., & Kumar, S. (2024). Machine Learning Algorithms for Detecting Mental Stress in College Students. *2024 IEEE 9th International Conference for Convergence in Technology (I2CT)*, 1–5. <https://doi.org/10.1109/I2CT61223.2024.10544243>
- Susanti, L. (2024). Klasifikasi Tingkat Stress pada Mahasiswa Teknik Informatika dalam Melakukan Perkuliahan Metode Hybrid Menggunakan Algoritma Naive Bayes. *STRING (Satuan Tulisan Riset Dan Inovasi Teknologi)*, 8(3), 243. <https://doi.org/10.30998/string.v8i3.17096>
- Tapidingan, Y. C., & Paseru, D. (2020). Comparative Analysis of Classification Methods of KNN and Naïve Bayes to Determine Stress Level of Junior High School Students. *Indonesian Journal of Information Systems*, 2(2), 80–89. <https://doi.org/10.24002/ijis.v2i2.3035>
- Thiyagarajan, A., James, T. G., & Marzo, R. R. (2022). Psychometric properties of the 21-item Depression, Anxiety, and Stress Scale (DASS-21) among Malaysians during COVID-19: a methodological study. *Humanities and Social Sciences*

- Communications*, 9(1), 220. <https://doi.org/10.1057/s41599-022-01229-x>
- Wibisono, B., Machmud, A., Suryani, N., & Yunita, Y. (2025). Analisis Sentimen Cyberbullying Pada Komentar X Menggunakan Metode Naïve Bayes. *Computer Science (CO-SCIENCE)*, 5(1), 8–16. <https://doi.org/10.31294/coscience.v5i1.5152>
- Yuda, A. K., Resita, C., Nurwansyah, R., Gani, R. A., Németh, Z., & Setiawan, E. (2022). Confidence, Academic Stress, Coping Strategies as Predictors of Student Academic Achievement in Physical Education Classes During Covid-19. *Physical Education Theory and Methodology*, 22(2), 180–187. <https://doi.org/10.17309/tmfv.2022.2.05>