



Computer Based Information System Journal

ISSN (Print): 2337-8794 | E- ISSN : 2621-5292
web jurnal : <http://ejournal.upbatam.ac.id/index.php/cbis>



CLUSTERING DALAM MENENTUKAN TINDAK LANJUT HASIL ANNUAL CHECK MENTAL HEALTH DENGAN ALGORITMA K-MEANS

Lorensius Bima Ade Prasetya

Universitas Putera Batam, Indonesia.

INFORMASI ARTIKEL

Diterima Redaksi: Januari 2025
Diterbitkan Online: Maret 2025

KATA KUNCI

Analysis, Clustering, Data Mining,
K-Means, Mental Health

KORESPONDENSI

E-mail: pb211510010@upbatam.ac.id

ABSTRACT

Addressing mental health issues has become a priority for the Indonesian government due to the rising prevalence of depression and anxiety, especially in Riau Archipelago Province, which reported as one of the provinces with the highest suicide rates in the country. The World Health Organization has pointed out the deficiencies in data collection, reporting, and knowledge management related to mental health. PT XYZ has collected mental health data through annual medical check-ups utilizing the DASS-21 scale, but this data has not been analyzed to extract useful and actionable insights for management. This research aims to profile the mental health of PT XYZ workers using annual medical check-up data from June to September 2023, involving a total of 3,699 workers. By implementing the K-Means clustering algorithm based on three variables—depression, anxiety, and stress—the analysis revealed that 242 workers were identified in a cluster with severe mental health conditions, 1,271 in a cluster with moderate conditions, and the remaining 2,186 with mild conditions in a cluster.

1. Latar Belakang

World Health Organization (WHO) pada tahun 2019 menyatakan bahwa sebanyak 970 juta orang di seluruh dunia hidup dengan gangguan mental dengan kecemasan dan depresi sebagai yang paling umum. Lebih lanjut, Kepulauan Riau merupakan salah satu provinsi dengan angka bunuh diri tertinggi di Indonesia [1]. Dengan begitu, urgensi mengenai kesehatan mental semakin menjadi sorotan. Sistem informasi dalam bidang kesehatan mental sangat penting untuk memantau dan menganalisis kondisi kesehatan mental individu, terlebih di lingkungan kerja. Kesehatan mental karyawan semakin diprioritaskan di Indonesia, tetapi masih terdapat

banyak tantangan dalam pengumpulan dan analisis data [2].

Pemerintah Indonesia telah menetapkan kesehatan mental sebagai prioritas dalam agenda pembangunan nasional. Namun, laporan WHO mengindikasikan bahwa kebijakan kesehatan mental yang komprehensif masih belum ada, dan data terkait kesehatan mental sangat minim [2]. Hal ini menyulitkan pemahaman yang lebih dalam mengenai kondisi kesehatan mental di berbagai sektor, termasuk industri. Sebagaimana yang terjadi pada salah satu perusahaan di Kota Batam yaitu PT XYZ yang meskipun telah dilakukan pemeriksaan kesehatan mental menggunakan DASS-21 (*Depression, Anxiety, Stress Scale – 21*), data yang dikumpulkan belum dianalisis secara mendalam. Data atau informasi

tidak dikelola lebih lanjut dengan algoritma yang sesuai dan mendukung sehingga informasi yang penting mungkin saja tidak didapatkan sehingga tidak dapat disajikan dan dipergunakan sebagaimana mestinya.

Dengan demikian, penelitian ini akan melakukan analisis terhadap data dengan menjalankan data mining dengan teknik clustering yaitu algoritma K-Means. Teknik ini dilakukan dengan mengolompokkan data-data dengan karakteristik serupa satu diantara lainnya dan kemudian melakukan pengujian terhadap kelompok-kelompok yang telah terbentuk. Penelitian dengan teknik serupa telah dilakukan dan berhasil menggunakan clustering K-Means dan analisis profil laten untuk mengidentifikasi berbagai profil kecemasan siswa, yang terbagi menjadi tiga kelompok berdasarkan tingkat risiko yaitu rendah, sedang, dan tinggi [3]. Oleh karena hasil penelitian yang baik itu, peneliti bermaksud menggunakan metode clustering K-Means yang pada lokasi penelitian yaitu PT. XYZ di Kota Batam dengan harapan peneliti dapat mengolah data kesehatan mental karyawan sehingga data tidak hanya disimpan namun akan dianalisa lebih lanjut dan menghasilkan profil-profil kesehatan mental karyawan berdasarkan tingkat *depression*, *anxiety* & *stress* secara bersamaan dan menjadi pertimbangan bagi PT. XYZ dalam hal menyusun langkah – langkah strategis & komprehensif untuk dapat membantu atau menangani kondisi kesehatan mental karyawannya. Berdasarkan latar belakang yang telah diuraikan di atas, maka peneliti tertarik melakukan penelitian dengan judul “*Clustering Dalam Menentukan Tindak Lanjut Hasil Annual Check Mental Health Dengan Algoritma K-Means*”.

2. Kajian Literatur

2.1. Knowledge Discovery in Databases

Knowledge Discovery in Databases (KDD) merupakan serangkaian proses yang kompleks yang mencakup langkah-langkah dengan tujuan untuk memperoleh pengetahuan dari data yang besar dan tidak terstruktur [4]. Adapun langkah-langkah utama dalam KDD meliputi pemilihan data, pra-proses data, transformasi data, pemodelan, evaluasi dan penyajian pengetahuan.

2.2. Data Mining

Proses mengubah data mentah menjadi informasi yang bernilai dengan cara menemukan pola atau *patterns* dalam proses pengekstrasian data besar disebut dengan *data mining*. Data mining mencakup pendekatan *supervised learning* dimana bertujuan memprediksi *output* yang tidak diketahui berdasarkan data dengan label tersedia dan *unsupervised learning* yang berfokus menemukan pola atau struktur tertentu yang tersirat dalam data yang tidak memiliki label sebelumnya. Terdapat dua fungsi utama dalam unsupervised learning yaitu klasterisasi dan asosiasi [5].

2.3. Algoritma Clustering

Clustering bertujuan menganalisis data kemudian mengidentifikasi pola/tren sehingga menunjukkan bahwa beberapa objek lebih mirip satu sama lain daripada objek lainnya sehingga dapat dibagi menjadi kelompok-kelompok yang sama [6].

2.4. K-Means

K-Means adalah algoritma *partitional clustering* yang populer, yang bertujuan untuk membagi kumpulan data menjadi k cluster dengan memaksimalkan kesamaan antar objek dalam satu cluster dan meminimalkan kesamaan antar cluster. K-Means menghitung jarak Euclidean untuk menentukan kedekatan antara objek dengan pusat cluster (*centroid*) secara berulang-ulang untuk mencari jarak terdekat dengan rumus sebagai berikut:

$$d(i, j) = \sqrt{\sum_{k=1}^n (X_{ik} - X_{jk})^2}$$

Rumus 1. Euclidean Distance

Sementara untuk mencari jarak antara objek i dan titik *centroid* j dilakukan perhitungan atribut berjumlah k dari hasil kuadrat hasil pengurangan i dan j .

2.5. Davies Bouldin Index (DBI)

Davies Bouldin Index (DBI) merupakan cara menilai kualitas atau performa hasil dari *clustering* dengan nilai yang diharapkan

adalah nilai terkecil dalam kondisi non-negatif [7]. Langkah pertama ialah mencari rasio atau perbandingan antara rata-rata jarak dalam cluster (*Sum of Square Within*) dengan rata-rata jarak antar-cluster (*Sum of Square Between*) yang dihitung dengan menggunakan Euclidean Distance. Rasio ini didapatkan melalui rumus berikut:

$$R_{ij} = \frac{SSW_i + SSW_j}{SSB_{i,j}}$$

Rumus 2. Rasio

Selanjutnya nilai rasio tersebut digunakan untuk dalam persamaan berikut untuk mendapatkan nilai DBI:

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} (R_{i,j})$$

Rumus 3. Davies Bouldin Index (DBI)

2.6. Altair AI Studio (RapidMiner)

Altair AI Studio (sebelumnya RapidMiner) adalah sebuah platform perangkat lunak *open-source* untuk analisis data dan *machine learning* dengan berbagai operator dan fungsi yang dapat digunakan untuk menyelesaikan berbagai jenis tugas analitik, seperti *pre-processing data*, klasifikasi, regresi, clustering, dan lain-lain [8].

2.7. Pekerja

Pekerja/buruh adalah setiap orang yang bekerja dengan menerima upah atau imbalan dalam bentuk lain [9]. Aktivitas yang dimaksud sebagai "bekerja" mencakup berbagai jenis pekerjaan, baik formal (memiliki kontrak kerja) maupun informal (tanpa kontrak kerja), yang dilakukan untuk menghasilkan barang atau jasa guna memenuhi kebutuhan hidup.

2.8. Kesehatan Mental

Kesehatan mental mengacu pada kondisi mental yang sehat, di mana seseorang menunjukkan kesejahteraan emosional, dapat menyesuaikan perilakunya dengan baik, relatif bebas dari kecemasan atau gejala yang mengganggu, serta mampu

membangun hubungan yang positif dan mengatasi tantangan hidup secara efektif [10].

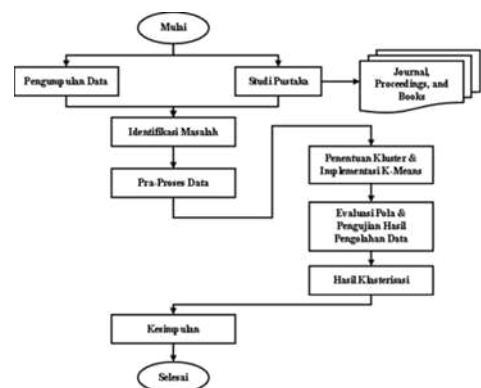
2.9. DASS-21

DASS-21 adalah singkatan untuk *Depression, Anxiety, Stress Scale* yang berisikan 21 item pertanyaan. DASS-21 merupakan versi pendek dari DASS-42 yang memiliki 42 item pertanyaan. DASS merupakan alat ukur psikologis yang dikembangkan oleh Peter F. Lovibond dan Shirley H. Lovibond pada tahun 1995 untuk mengukur atau menilai kondisi kesehatan mental seseorang berdasarkan skala *depression*, *anxiety* dan *stress* [11].

3. Metodologi

3.1. Desain Penelitian

Penelitian ini menerapkan pendekatan deskriptif kuantitatif dengan desain studi *cross-sectional* untuk menganalisis kondisi kesehatan mental pekerja PT XYZ pada satu titik waktu tertentu. Penelitian ini didesain untuk dapat memberikan gambaran awal mengenai tingkat depresi, kecemasan, dan stres secara bersamaan pada karyawan PT. XYZ.



Gambar 1. Desain Penelitian

- a) Pengumpulan Data & Studi Pustaka
Data primer yang digunakan dalam penelitian ini berupa informasi kesehatan mental yang diperoleh dari *annual medical check-up* yang dilakukan PT XYZ pada periode Juni hingga September tahun 2023. Peneliti juga

- melakukan studi pustaka juga mempelajari literatur-literatur relevan guna menunjang penelitian.
- b) **Identifikasi Masalah**
Identifikasi masalah dalam penelitian ini meliputi tidak adanya pengolahan data dan analisa lebih lanjut terhadap data kesehatan mental pekerja PT. XYZ meskipun data sudah didapatkan sehingga perusahaan tidak mengetahui kondisi atau profil kesehatan mental pekerjanya dan apakah diperlukan tindak lanjut terhadap kondisi ini.
 - c) **Pra-proses Data**
Pra-proses data dimaksudkan untuk mempersiapkan data agar dapat diolah dan dianalisa lebih lanjut dengan cara pembersihan serta transformasi data.
 - d) **Penentuan Kluster & Implementasi K-Means**
Pembentukan jumlah cluster merupakan tahapan awal K-Means dan peneliti memutuskan untuk membentuk 3 cluster serta menginisiasi *centroid* awal secara acak. Selanjutnya peneliti menghitung Euclidean Distance dan menghitung *centroid* secara berulang (iterasi) hingga tidak ada lagi perubahan nilai *centroid* maupun anggota pada setiap *cluster*.
 - e) **Evaluasi Pola & Pengujian Hasil Pengolahan Data**
Evaluasi pola dilakukan dengan mengukur *Davies Bouldin Index (DBI)* dengan nilai yang diharapkan ialah nilai terkecil mendekati 0 dalam kondisi non-negatif. Selanjutnya, *cluster* yang dihasilkan dari perhitungan algoritma K-Means akan diuji menggunakan aplikasi Altair AI Studio (RapidMiner) untuk memastikan bahwa data telah dikelompokkan dengan benar, membentuk kluster berdasarkan karakteristik berkaitan dengan tingkat depresi, kecemasan, dan stres secara bersamaan.
 - f) **Hasil Klasterisasi**
Hasil klasterisasi merupakan 3 cluster yang merepresentasikan profil kesehatan mental pekerja PT. XYZ dimana masing-masing memiliki karakteristik khusus berdasarkan tingkat depresi, kecemasan dan stress.

- g) **Kesimpulan**
Kesimpulan diambil dengan menginterpretasikan hasil *clustering*, sehingga profil dan karakteristik masing-masing kelompok pekerja dapat dipahami.

3.2. Objek Penelitian

Objek penelitian ini adalah pekerja PT. XYZ yang sebagaimana dijabarkan dalam latar belakang ingin diidentifikasi profil kesehatan mentalnya.

3.3. Populasi dan Sampel

Populasi dalam penelitian ini meliputi pekerja PT. XYZ yang menjalankan annual medical check-up periode Juni 2023 hingga September 2023 berjumlah 3699 pekerja. Sampel penelitian menggunakan *saturation sampling* atau sampling jenuh sehingga semua anggota populasi dijadikan sampel penelitian dengan pertimbangan pengumpulan data yang komprehensif, mendalam, dan akurat dari seluruh karyawan.

3.4. Variabel Penelitian

Berdasarkan alat ukur psikologis bernama DASS-21 yang digunakan PT. XYZ dalam annual medical check-up, maka variabel dalam penelitian ini ialah:

- a) *Depression*
- b) *Anxiety*
- c) *Stress*

4. Pembahasan

4.1. Perhitungan Algoritma K-Means

Perhitungan algoritma K-Means dapat dilakukan setelah data sudah melalui tahapan pra-proses data yang meliputi pembersihan data dan transformasi data. Perhitungan algoritma K-Means dimulai dengan menentukan jumlah cluster yang ingin dibentuk. Selanjutnya menentukan nilai *centroid* secara acak pada iterasi pertama dan dilanjutkan dengan menghitung jarak dari setiap data terhadap nilai *centroid* acak tersebut kemudian mengelompokkan data-data dengan jarak terdekat pada *centroid* sebagai sebuah cluster. Perhitungan pada iterasi kedua dilakukan dengan menghitung kembali jarak terdekat

dengan nilai *centroid* baru yang didapatkan dari nilai rata-rata jarak pada iterasi pertama. Langkah ini terus dilakukan hingga nilai *centroid* maupun anggota pada setiap *cluster* tidak lagi berubah. Perhitungan algoritma K-Means terhadap 3699 data kesehatan mental pekerja PT XYZ selesai pada iterasi kesembilanbelas oleh karena hasil *clustering* dan nilai *centroid* pada iterasi kesembilanbelas dan iterasi kedua puluh adalah sama.

Tabel 1. Centroid Iterasi Kesembilanbelas

	Cluster A	Cluster B	Cluster C
Depression	9.351	3.172	0.542
Anxiety	8.694	4.148	1.062
Stress	10.463	5.415	1.112

Perhitungan algoritma K-Means pada iterasi kesembilan belas menghasilkan 3 cluster dengan detail sebagai berikut

- Cluster A* memiliki 242 anggota,
- Cluster B* memiliki 1271 anggota, dan
- Cluster C* memiliki 2186 anggota

4.2. Evaluasi Pola

Evaluasi perhitungan Algoritma K-Means dengan mencari Davies Bouldin Index melalui rasio SSW dan SSB menghasilkan nilai sebesar 0.815 yang menunjukkan bahwa clustering secara keseluruhan cukup baik.

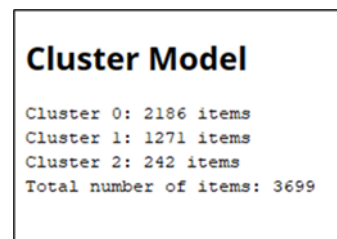
Tabel 2. Perbandingan DBI

Jumlah Cluster	3 Cluster	4 Cluster	5 Cluster	6 Cluster
DBI	0.815	0.913	1.062	1.173

Perbandingan DBI pada tabel di atas menunjukkan bahwa pengelompokan data ke dalam 3 *cluster* memiliki nilai DBI paling rendah dari antara jumlah cluster lainnya sehingga membuktikan bahwa penentuan 3 *cluster* dalam penelitian ini sudah sesuai.

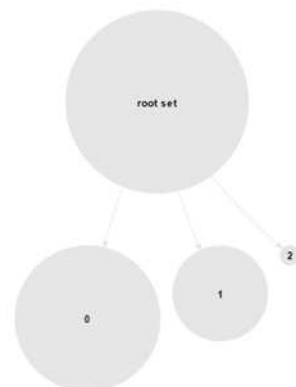
4.3. Pengujian dengan Aplikasi Altair AI Studio (RapidMiner)

Data-data yang telah dihitung dengan algoritma K-Means secara manual akan diuji dengan menggunakan aplikasi Altair AI Studio (RapidMiner) untuk mengetahui apakah hasilnya telah sesuai dengan berdasarkan jarak data dengan anggota cluster berdasarkan kaidah *clustering* K-Means. Pengujian dengan aplikasi Aplikasi Altair AI Studio (RapidMiner) pada menu *Cluster Model* menunjukkan hasil *clustering* pada *Cluster 0* berjumlah 2186 data, *Cluster 1* berjumlah 1271 data dan *Cluster 2* berjumlah 242 data sebagaimana terlihat pada Gambar 3 di bawah ini.



Gambar 2. Cluster Model

Terdapat perbedaan urutan pada cluster yang terbentuk antara perhitungan manual dengan pengujian dengan aplikasi Altair AI Studio (RapidMiner) oleh karena titik centroid yang dipilih secara acak. Pada grafik cluster model tergambar bahwa dari 3 cluster yang terbentuk, Cluster 0 memiliki data paling banyak dan Cluster 2 memiliki data paling sedikit.



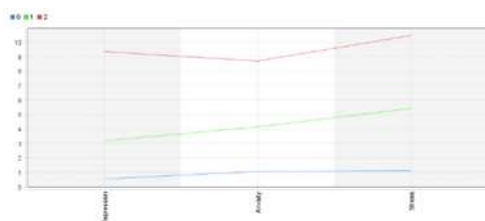
Gambar 3. Grafik Cluster Model

Nilai *centroid* yang didapatkan pada pengujian dengan aplikasi terlihat pada **Gambar 4** dan sama dengan nilai *centroid* pada perhitungan algoritma K-Means secara manual. Nilai *centroid* untuk Cluster 0 pada variabel *depression*, *anxiety* dan *stress* secara berurutan ialah 0.542, 1.062, dan 1.112. Pada Cluster 1 nilai *centroid* untuk variabel *depression* adalah 3.172, *anxiety* adalah 4.148 dan *stress* adalah 5.415. Pada Cluster 2 nilai *centroid* untuk variabel *depression*, *anxiety* dan *stress* secara berurutan ialah 9.351, 8.694 dan 10.463.

Attribute	cluster_0	cluster_1	cluster_2
Depression	0.542	3.172	9.351
Anxiety	1.062	4.148	8.694
Stress	1.112	5.415	10.463

Gambar 4. Centroid Table

Sementara itu, **Gambar 5** memberikan visualisasi terhadap nilai atau titik *centroid* yang terlihat pada *centroid table*. Dengan demikian terlihat bahwa *Cluster 0* merupakan kelompok pekerja dengan tingkat keparahan kondisi mental *mild*, *Cluster 1* ialah kelompok pekerja dengan tingkat keparahan kondisi mental *moderate*, dan yang terakhir *Cluster 2* adalah kelompok pekerja dengan tingkat keparahan kondisi mental *severe*.



Gambar 5. Centroid Plot View

Pada **Gambar 6** terlihat *performance vector* yang merupakan hasil evaluasi terhadap clustering K-Means yang telah dibuat. Diketahui bahwa rata-rata jarak dalam centroid sebesar 7.618. Rata-rata jarak centroid dengan data pada Cluster 0 sebesar 3.935, pada Cluster 1 sebesar 9.073 dan pada Cluster 2 sebesar 33.242. Sementara itu nilai Davies Bouldin sebesar 0.815 menyatakan kualitas pengelompokan secara keseluruhan cukup baik.

PerformanceVector

```
PerformanceVector:
Avg. within centroid distance: 7.618
Avg. within centroid distance_cluster_0: 3.935
Avg. within centroid distance_cluster_1: 9.073
Avg. within centroid distance_cluster_2: 33.242
Davies Bouldin: 0.815
```

Gambar 6. Performance Vector

5. Kesimpulan

Setelah hasil penelitian diperoleh maka peneliti menarik kesimpulan sebagai berikut:

1. Implementasi *Data Mining* yaitu Algoritma K-Means dapat digunakan untuk mengidentifikasi profil kesehatan mental pekerja dan menunjukkan hasil *clustering* yang baik dan efisien dimana hasil pengujian menunjukkan cluster pekerja dengan tingkat keparahan kondisi mental *mild*, *moderate* dan *severe*.
2. Cluster yang terbentuk dari perhitungan algoritma K-Means dan dari pengujian dengan aplikasi Altair AI Studio (RapidMiner) menunjukkan Cluster C atau Cluster 0 yang merupakan kelompok karyawan dengan tingkat keparahan kondisi mental *mild* berjumlah 2186 pekerja, Cluster B atau Cluster 1 merupakan tingkat keparahan kondisi mental *moderate* sejumlah 1271 pekerja dan Cluster A atau Cluster 2 merupakan kelompok pekerja dengan tingkat keparahan kondisi mental *severe* berjumlah 242 pekerja.

Ucapan Terima Kasih

Peneliti mengucapkan syukur dan terima kasih kepada Tuhan Yang Maha Esa atas berkat dan kekuatan yang diberikan selama proses penelitian ini. Penulis juga mengucapkan terima kasih kepada Bapak Sasa Ani Arnomo yang telah memberikan bimbingan, dukungan serta pengetahuan sepanjang penelitian.

Daftar Pustaka

- [1] S. Onie *et al.*, "Indonesia's first suicide statistics profile: an analysis of suicide and attempt rates, underreporting,

- geographic distribution, gender, method, and rurality,” *The Lancet Regional Health - Southeast Asia*, vol. 22, p. 100368, 2024, doi: 10.1016/j.lansea.2024.100368.
- [2] Asia World Health Organization - Regional Office for South-East, *Addressing Mental Health in Indonesia*. 2022. doi: 978-92-9021-018-4.
- [3] F. Liu *et al.*, “Use of latent profile analysis and k-means clustering to identify student anxiety profiles,” *BMC Psychiatry*, vol. 22, no. 1, pp. 1–11, 2022, doi: 10.1186/s12888-021-03648-7.
- [4] M. Oded and L. Rokach, *Data Mining and Knowledge Discovery Handbook*. Boston, MA: Springer US, 2010. doi: 10.1007/978-0-387-09823-4.
- [5] C. C. Aggarwal, *Data Mining*. Cham: Springer International Publishing, 2015. doi: 10.1007/978-3-319-14142-8.
- [6] I. A. Nur Afifah and H. Nurdyanto, “DATA MINING CLUSTERING DALAM PENGELOMPOKAN BUKU PERPUSTAKAAN MENGGUNAKAN ALGORITMA K-MEANS,” *JIPi (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 8, no. 3, pp. 802–814, Aug. 2023, doi: 10.29100/jipi.v8i3.3891.
- [7] S. A. Arnomo and Y. Yulia, “Clustering the potential bandwidth upgrade of FTTH broadband subscribers,” *ILKOM Jurnal Ilmiah*, vol. 13, no. 1, pp. 51–57, 2021, doi: 10.33096/ilkom.v13i1.805.51-57.
- [8] D. Aprilla, D. A. Baskoro, L. Ambarwati, and I. W. S. Wicaksana, *Belajar Data Mining dengan RapidMiner*, vol. 5, no. 4. 2013.
- [9] *Undang-Undang Republik Indonesia Nomor 13 Tahun 2003 tentang Ketenagakerjaan*.
- [10] Association American Psychological, “Mental Health.” [Online]. Available: <https://www.apa.org/topics/mental-health>
- [11] I. Ifdil *et al.*, “The depression anxiety stress scales (DASS-21): an Indonesian validation measure of the depression anxiety stress,” *COUNS-EDU: The International Journal of Counseling and Education*, vol. 5, no. 4, pp. 205–215, 2022, doi: 10.23916/0020200536840.