

IMPLEMENTATION OF MENTAL HEALTH DATA MINING USING ASSOCIATION AND CLUSTERIZATION

IMPLEMENTASI DATA MINING KESEHATAN MENTAL MENGGUNAKAN ASOSIASI DAN KLASTERISASI

Johanes Fernandes Andry¹, Wandy Wandy², Francka Sakti Lee³, Honni⁴, Christian Ronaldo Yusup⁵

Program Studi Sistem Informasi, Universitas Bunda Mulia^{1,3,4,5}

Program Studi Sistem Informasi, Universitas Sampoerna²

jandry@bundamulia.ac.id¹

ABSTRACT

This study analyzed the data mining dataset from kaggle.com, with the aim of identifying cluster patterns and association relationships related to mental health disorders. In the clustering process, the optimal value for the number of clusters was found at $k = 3$ with a Davies-Bouldin Index (DBI) value of 0.470, indicating a clear and separate cluster division. Meanwhile, in the association analysis using the FP-Growth model, 20 premises were found that described different relationships, with age as a factor that often appears and plays a significant role in influencing mental health disorders. Based on evaluation metrics such as support, confidence, lift, and conviction, some of the values that often appear were the highest support at 0.742, confidence at 0.821, and lift at 1.064. In addition, the Laplace and p-s values also show significant contributions to this finding. Overall, the results of the association analysis indicate that age has a significant influence on mental health disorders, with various metrics showing consistency in the relationship. These findings can be a basis for further research in understanding the factors influencing mental health based on individual characteristics.

Keywords: Data Mining, Mental Health, Association, Clusterization

ABSTRAK

Penelitian ini menganalisis dataset hasil data mining dari kaggle.com, dengan tujuan untuk mengidentifikasi pola kluster dan hubungan asosiasi yang berkaitan dengan gangguan kesehatan mental. Dalam proses klusterisasi, nilai optimal untuk jumlah kluster ditemukan pada $k = 3$ dengan nilai Davies-Bouldin Index (DBI) sebesar 0,470, yang menunjukkan pembagian kluster yang jelas dan terpisah. Sementara itu, dalam analisis asosiasi menggunakan model FP-Growth, ditemukan 20 premis yang menggambarkan hubungan yang berbeda-beda, dengan usia (Age) sebagai faktor yang sering muncul dan berperan signifikan dalam mempengaruhi gangguan kesehatan mental. Berdasarkan metrik evaluasi seperti support, confidence, lift, dan conviction, beberapa nilai yang sering muncul adalah support tertinggi pada 0.742, confidence pada 0.821, dan lift pada 1.064. Selain itu, nilai Laplace dan p-s juga menunjukkan kontribusi penting pada temuan ini. Secara keseluruhan, hasil analisis asosiasi mengindikasikan bahwa usia memiliki pengaruh yang cukup besar terhadap gangguan kesehatan mental, dengan berbagai metrik menunjukkan konsistensi dalam hubungan tersebut. Temuan ini dapat menjadi dasar untuk penelitian lebih lanjut dalam memahami faktor-faktor yang mempengaruhi kesehatan mental berdasarkan karakteristik individu.

Kata Kunci: Data Mining, Kesehatan Mental, Asosiasi, Klusterisasi

PENDAHULUAN

Musik memiliki kekuatan untuk dapat merangsang emosi dan tanggapan, mengembangkan kreativitas, teknis, keterampilan sosial dan detail, meningkatkan ekspresi, komunikasi, kepercayaan diri dan harga diri, dan memfasilitasi perubahan positif dalam perilaku dan kesejahteraan suatu makhluk (WHO, 2019). Musik juga memainkan peran penting, khususnya untuk

mendukung manusia di semua tahap kehidupan: dari membantu bayi yang baru lahir, mengembangkan ikatan yang sehat dengan orang tuanya untuk menawarkan perawatan paliatif yang vital, sensitif, dan kasih sayang antara orang tua dan anak (Rebecchini, 2021), termasuk dalam kesehatan mental yang merupakan landasan dan memainkan peran penting dalam mendorong perkembangan positif pada manusia (Wang et al., 2023). Secara

signifikan mempengaruhi lingkungan, terlepas dari wilayah atau negara mana individu tersebut berasal (Qiaoyi & Chen, 2023). Hal tersebut fundamental bagi kesejahteraan keseluruhan individu dan mutlak penting untuk kehidupan yang produktif dan efisien (Hou et al., 2020).

Namun, mental dapat mengalami gangguan yang memiliki efek bagi seseorang dan keluarga mereka, serta bagi masyarakat secara keseluruhan (Rahman et al., 2020). Masalah kesehatan mental atau gangguan mental adalah masalah kesehatan yang mempengaruhi cara seseorang merasakan, berpikir, berperilaku, dan berkomunikasi dengan orang lain (Mutalib et al., 2021). Dalam situasi saat ini, berbagai aspek masyarakat sedang menghadapi peningkatan tekanan yang semakin terasa setiap waktunya, dan tekanan datang dari banyak pihak. Beberapa bagian masyarakat biasanya lebih peka terhadap situasi dan kondisi serta kemampuan kontrol psikologis yang lemah, yang mana mudah terserang serangkaian masalah psikologis (Hou, 2022). Beberapa gejala yang terjadi, seperti meningkatnya prevalensi gejala depresi, kecemasan, gangguan makan, dan penyakit mental lainnya di kalangan masyarakat menunjukkan krisis kesehatan mental. Oleh karena itu, lakukan studi secara aktif dan mencari solusi yang mungkin untuk memenuhi kebutuhan (Wang et al, 2022). Karya-karya terdahulu dalam bidang ini cenderung mengeksplorasi area-area stabil dalam penyelidikan psikologi perkembangan, yaitu perbedaan usia dan gender yang umum dalam merespons pemicu stres yang signifikan (Schwartz et al., 2021).

Penelitian yang melaporkan hubungan positif antara keterlibatan dalam musik dan indikator kesehatan mental, yaitu bahwa keterlibatan dalam musik dapat menjadi perlindungan terhadap masalah kejiwaan. Meskipun studi epidemiologi yang menyelidiki hubungan antara musik dan risiko masalah kesehatan mental masih jarang dilakukan

(Wesseldijk et al., 2019). Musik dapat mengirimkan rangsangan ke otak melalui pendengaran, seperti thalamus dan lenticular nucleus. Area otak ini mengandung dopaminergic neurons, dan dopamine yang disekresikannya adalah neurotransmitter penting di otak. Musik merangsang sekresi dopamine dan berpartisipasi dalam banyak aktivitas fisiologis, seperti pergerakan manusia, kognisi emosi, penguatan positif, dan pengaturan endokrin (Lu, 2023). Secara intuitif, keterlibatan musik (menyanyi atau memainkan alat musik) baik untuk kesehatan mental mereka, yaitu ada hubungan antara keterlibatan musik dan kesehatan mental realitas yang kompleks. Berbagai penelitian melaporkan bahwa orang-orang yang terlibat dalam musik melaporkan kesehatan fisik dan mental yang lebih baik, khususnya berkaitan dengan kesejahteraan dan keterhubungan sosial, keterlibatan musik tampaknya bermanfaat. Selain itu, terapi musik atau intervensi musik dalam pengaturan klinis dapat memberikan efek positif kesehatan mental pasien (Wesseldijk et al., 2023). Berbagai penelitian berkisar dari program berbasis komunitas untuk meningkatkan suasana hati atau mengurangi stres yang dirancang untuk mengobati *Serious Mental Illness* (SMI). Hasil menunjukkan bahwa musik dapat memberikan kesehatan manfaat seperti peningkatan respon imun dan peningkatan penanggulangan dan regulasi emosional. Mereka juga menyarankan agar musik dimasukkan ke dalam inisiatif kesehatan mental yang menawarkan beberapa manfaat, seperti peningkatan konektivitas sosial, peningkatan pemberian layanan kesehatan, dan peningkatan imunitas (Golden et al., 2022). Perkembangan teknologi musik digital telah memberikan dorongan pada pengembangan sistem rekomendasi musik yang dipersonalisasi yang merekomendasikan musik kepada pengguna. Memberikan rekomendasi berdasarkan big data yang tersedia di internet merupakan tantangan besar.

Namun dengan bantuan teknologi baru yaitu pembelajaran mesin, kecerdasan buatan, dan lain-lain, dapat menjadi mudah dalam mendeteksi suasana hati pengguna (Gupta, 2023).

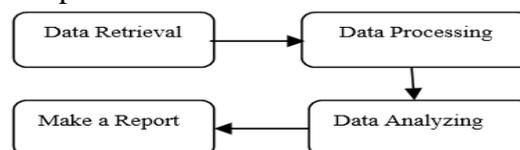
Tujuan dari penelitian ini adalah untuk mempelajari kecenderungan genre musik yang sering didengarkan dan efeknya terhadap kesehatan mental seseorang berdasarkan big data dari kaggle.com mengenai *Music & Mental Health Survey Results* menggunakan teknik data mining yaitu kelasterisasi dan asosiasi.

METODE

Berdasarkan Gambar 1. Tahapan Penelitian, yang dilakukan oleh peneliti secara teratur dan sistematis untuk melaksanakan penelitian ini. yang dilakukan adalah sebagai berikut:

1. *Data Retrieval*, peneliti mendapatkan data dari platform Kaggle yang dimana data tersebut akan digunakan oleh peneliti untuk dianalisis lebih lanjut (Andry et al., 2023). Peneliti memperoleh data tersebut dari tautan ini:
2. <https://www.kaggle.com/datasets/catherinerasgaitis/mxmh-survey-results>. Data bisa di lihat pada tabel 1, 2 dan 3 (karena data nya besar, hanya di tampilkan Sebagian)
3. *Data Processing*, peneliti melakukan pemrosesan dataset dengan melakukan filterisasi nilai-nilai kosong serta melakukan beberapa perubahan pada format dari data untuk dapat dilanjutkan pada tahapan berikutnya (Geasela, 2024).

4. *Analyzing Data*, peneliti melakukan analisis data pada dataset menggunakan RapidMiner dengan metode kelasterisasi dan asosiasi (Madyatmadja et al, 2021a).
5. *Make a report*, peneliti melakukan pelaporan dari hasil analisis data yang diperoleh.



Gambar 1. Tahapan Penelitian (Madyatmadja et al., 2021b)

Kelasterisasi adalah teknologi yang memeriksa hubungan logis atau fisik antara data dan membagi beberapa himpunan data menjadi banyak kluster, masing-masing terdiri dari himpunan data yang mirip dalam sifatnya (Pratama et al., 2022). Kelasterisasi juga digunakan dalam pendekatan analisis data umum untuk mengidentifikasi kelompok homogen dari objek berdasarkan nilai atribut, dan Kelasterisasi telah banyak dilakukan pada kehidupan nyata seperti segmentasi gambar, analisis data, pembelajaran mesin, mesin pencarian, pengambilan dokumen, pengenalan atau evaluasi objek, komputasi, ekonomi, studi perpustakaan, asuransi, dan lain-lain, pendekatan pengelompokan digunakan untuk mengelompokkan kelompok data terkait dalam set data multivariat serta metode pengelompokan termasuk *partitioning methods, fuzzy Kelasterisasi, hierarchical Kelasterisasi, density-based Kelasterisasi, dan model-based Kelasterisasi* (Gupta et al., 2021).

Tabel 1. Data Source (Bagian 1)

BPM	Frequency [Classical]	Frequency [Country]	Frequency [EDM]	Frequency [Folk]	Frequency [Gospel]	Frequency [Hip hop]	Frequency [Jazz]	Frequency [K pop]	Frequency [Latin]
103	Never	Never	Never	Never	Never	Very frequently	Never	Never	Never
120	Rarely	Never	Sometimes	Rarely	Never	Very frequently	Rarely	Never	Never

Asosiasi adalah metode analisis yang memeriksa status realisasi dari peristiwa lain, dalam penjelasan yang lebih rinci dan jelas aturan asosiasi memprediksi pekerjaan di masa depan berdasarkan analisis data yang ada, yaitu aturan asosiasi menentukan hubungan asosiasi antara himpunan big data, ini adalah algoritma yang sering dipilih dan banyak

digunakan saat ini serta dapat diandalkan dan berguna untuk menyoroti tren dalam himpunan data, di samping itu kriteria pendukung mengungkapkan frekuensi asosiasi antara item, sedangkan kriteria kepercayaan mengungkapkan akurasi dari asosiasi tersebut (Keleş et al., 2022).

Tabel 2. Data Source (Bagian 2)

<i>Frequency [Lofi]</i>	<i>Frequency [Metal]</i>	<i>Frequency [Pop]</i>	<i>Frequency [R&B]</i>	<i>Frequency [Rap]</i>	<i>Frequency [Rock]</i>	<i>Frequency [Video game music]</i>	<i>Anxiety</i>	<i>Depressi on</i>	<i>Insomnia</i>
<i>Never</i>	<i>Never</i>	<i>Never</i>	<i>Sometimes</i>	<i>Very frequently</i>	<i>Never</i>	<i>Rarely</i>	8	8	4
<i>Never</i>	<i>Sometimes</i>	<i>Sometimes</i>	<i>Rarely</i>	<i>Sometimes</i>	<i>Very frequently</i>	<i>Never</i>	5	7	10

HASIL DAN PEMBAHASAN

Kelasterisasi

Pada bagian ini, dilakukan proses Kelasterisasi terhadap dataset yang telah peneliti tentukan. Metode yang digunakan pada Kelasterisasi ini adalah K-Means. Proses Kelasterisasi pada penelitian ini ditujukan untuk memperoleh nilai k yang mampu mengklasterisasi data-data yang ada secara optimal. Nilai k tersebut ialah nilai k yang memiliki nilai DBI (Davies-Bouldin Index) yang paling kecil dari nilai k lain yang dibandingkan. Pada proses ini, nilai k yang digunakan ialah k = 2 sampai dengan k = 7.

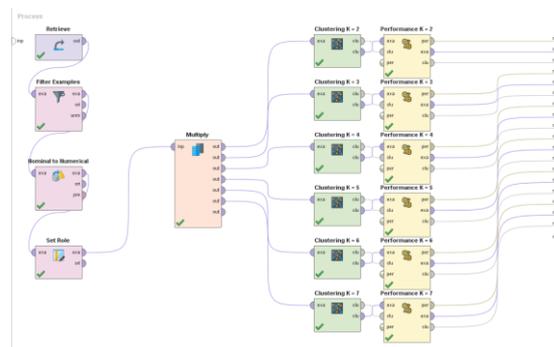
Proses Kelasterisasi yang dilakukan oleh peneliti melibatkan beberapa operator pada RapidMiner. Operator-operator tersebut meliputi:

- a. *Operator retrieve*
- b. *Operator filter examples*
- c. *Operator nominal to numerical*
- d. *Operator set role*
- e. *Operator multiply*
- f. *Operator Kelasterisasi, dan*
- g. *Operator cluster distance performance*

Keseluruhan proses kelasterisasi yang dilakukan oleh peneliti diilustrasikan pada Gambar 2.

Tabel 3. Data Source (Bagian 3)

<i>OCD</i>	<i>Music effects</i>	<i>Permissions</i>
3	<i>Improve</i>	<i>I understand.</i>
0	<i>Improve</i>	<i>I understand.</i>



Gambar 2. Model RapidMiner Kelasterisasi

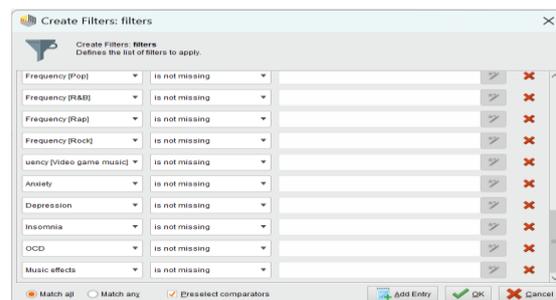
Berdasarkan pada gambar 2, tahapan yang dilakukan oleh peneliti dalam melakukan proses Kelasterisasi adalah sebagai berikut:

Melakukan impor data ke dalam RapidMiner dengan *operator retrieve*, melakukan data preparation dengan *operator filter examples* yang bertugas untuk membuang missing values yang terdapat pada dataset. Parameter yang

digunakan pada operator. Operator filter examples untuk Kelasterisasi dimana parameter filter yang digunakan adalah is not missing untuk atribut primary streaming service, hours per day, while working, instrumentalist, composer, fav genre, exploratory, foreign languages, BPM, frequency [Classical]. Operator filter examples untuk Kelasterisasi dimana parameter filter yang digunakan adalah is not missing untuk atribut frequency [Country], frequency [EDM], frequency [Folk], frequency [Gospel], frequency [Hip Hop], frequency [Jazz], frequency [K-pop], frequency [Latin], frequency [Lofi], frequency [Metal].

Gambar 3 menggambarkan operator filter examples untuk Kelasterisasi dimana parameter filter yang digunakan adalah is not missing untuk atribut frequency [Pop], frequency [R&B], frequency [Rap], frequency [Rock], frequency [Video Game Music], anxiety, depression, insomnia, OCD, music effects.

Pada tahap Operator Nominal to Numerical Untuk Kelasterisasi, atribut-atribut yang terkena dampak dari perubahan nominal menjadi numerical adalah atribut primary streaming service, while working, instrumentalist, composer, fav genre, exploratory, foreign languages, frequency [Classical], frequency [Country], frequency [EDM], frequency [Folk], frequency [Gospel], frequency [Hip Hop], frequency [Jazz], frequency [K-pop], frequency [Latin], frequency [Lofi], frequency [Metal], frequency [Pop], frequency [R&B], frequency [Rap], frequency [Rock], frequency [Video Game Music], music effects. Atribut-atribut tersebut semulanya bertipe data polynominal pada RapidMiner.



Gambar 3. Operator Filter Examples Untuk Kelasterisasi

Menggunakan operator set role untuk menandai atribut. Menggunakan operator multiply untuk menggandakan hasil dari proses yang telah dilakukan sebelumnya.

Operator multiply yang digunakan untuk proses Kelasterisasi pada penelitian ini. Pada gambar tersebut, terdapat satu input yang merupakan hasil dari proses sebelumnya dan enam output untuk nantinya digunakan pada enam operator berbeda berikutnya.

Menggunakan operator k-means untuk melakukan proses Kelasterisasi dengan parameter-parameter yang ditetapkan pada panel parameters. Proses ini dilakukan sebanyak 6 kali dengan nilai k yang digunakan adalah k = 2 sampai dengan k = 7. Dibawah ini adalah parameters yang digunakan untuk proses cluster k = 2.

Pada panel tersebut, beberapa hal berikut ini ditetapkan untuk proses Kelasterisasi untuk k = 2 sampai dengan k = 7:

- Add cluster attribute dicentang
- Max runs ditetapkan dengan nilai sebesar 10
- Measure types yang digunakan adalah BregmenDivergences
- Divergence yang digunakan adalah SquaredEuclideanDistance
- Max optimization steps ditetapkan dengan nilai sebesar 100

Menggunakan operator cluster distance performance untuk melakukan proses evaluasi terhadap cluster yang telah dibentuk oleh masing-masing operator k-means dengan parameter-parameter yang ditetapkan pada panel parameters.

Pada panel tersebut, dua hal berikut ini ditetapkan untuk operator cluster distance performance:

- Main criterion yang digunakan adalah Davies Bouldin
- Parameter maximize dicentang

Berdasarkan pada tabel 4, didapati bahwa nilai Davies-Bouldin Index (DBI) yang paling kecil adalah nilai DBI yang dihasilkan dari proses klasterisasi dengan $k = 3$, yakni sebesar 0.470. Artinya, klaster yang paling cocok untuk mengelompokkan data-data pada dataset googleplaystore pada penelitian ini adalah 3 klaster. Nilai DBI yang dihasilkan oleh proses klaster $k = 2$ tidak dipilih sebagai nilai k yang cocok karena terdapat kesalahan pada proses

klaster yang mengakibatkan nilai DBI yang dihasilkan menjadi 0.

Berikut ini ditampilkan deskripsi secara lengkap dari hasil operator performance untuk proses klaster beserta dengan chart dari hasil proses klasterisasi dengan nilai k tertentu yang telah dilakukan.

Tabel 4. Hasil Proses Klaster

Nilai K	Klaster (Banyaknya data)							Total Data	DBI
	0	1	2	3	4	5	6		
2	1	615	-	-	-	-	-	616	0,000
3	388	1	227	-	-	-	-	616	0,470
4	177	1	131	307	-	-	-	616	0,525
5	172	136	1	69	238	-	-	616	0,609
6	94	165	136	1	219	1	-	616	0,518
7	11	1	146	227	157	1	73	616	0,513

PerformanceVector

```
PerformanceVector:
Avg. within centroid distance: 1621.358
Avg. within centroid distance_cluster_0: 0.000
Avg. within centroid distance_cluster_1: 1623.995
Davies Bouldin: 0.000
```

Gambar 4. Hasil Proses Klaster K = 2

Gambar 4 hasil operator performance dari proses klaster dengan nilai $k = 2$. Didapati hasil average within centroid distance atau rata-rata jarak dalam centroid, average within centroid distance_cluster_0 atau rata-rata jarak dalam klaster 0, average within centroid distance_cluster_1 atau rata-rata jarak dalam klaster 1, dan nilai Davies-Bouldin Index (DBI) sebagai berikut:

a. Average within centroid distance atau rata-rata jarak dalam centroid sebesar 1.621,358. Nilai tersebut adalah nilai yang menggambarkan jarak rata-rata semua titik data dalam dataset ke centroid klaster mereka. Karena nilai perolehan tersebut cukup tinggi, maka dapat dikatakan bahwa klaster yang

terbentuk tidak terlalu rapat atau tidak terdefinisi dengan baik.

- Average within centroid distance_cluster_0 atau rata-rata jarak dalam klaster 0 sebesar 0, yang artinya tidak ada jarak sama sekali. Hal ini diakibatkan karena klaster 0 hanya memiliki satu titik data saja pada klaster tersebut.
- Average within centroid distance_cluster_1 atau rata-rata jarak dalam klaster 1 sebesar 1.623,995. Nilai tersebut adalah jarak rata-rata titik data dalam klaster 1 ke centroid klaster 1.
- Nilai Davies-Bouldin Index (DBI) sebesar 0.

Berikut ini adalah chart dari hasil proses klasterisasi dengan nilai $k = 2$.

Pada Gambar 5 hasil operator performance dari proses klaster dengan nilai $k = 3$. Didapati hasil average within centroid distance atau rata-rata jarak dalam centroid, average within centroid distance_cluster_0 atau rata-rata jarak

dalam kluster 0, average within centroid distance_cluster_1 atau rata-rata jarak dalam kluster 1, average within centroid distance_cluster_2 atau rata-rata jarak dalam kluster 2, dan nilai Davies-Bouldin Index (DBI) sebagai berikut:

- a. Average within centroid distance atau rata-rata jarak dalam centroid sebesar 861,392. Nilai tersebut adalah nilai yang menggambarkan jarak rata-rata semua titik data dalam dataset ke centroid kluster mereka. Nilai tersebut lebih rendah daripada proses kluster dengan nilai $k = 2$ yang dilakukan sebelumnya.
- b. Average within centroid distance_cluster_0 atau rata-rata jarak dalam kluster 0 sebesar 537,026. Nilai tersebut adalah jarak rata-rata titik data dalam kluster 0 ke centroid kluster 0.
- c. Average within centroid distance_cluster_1 atau rata-rata jarak dalam kluster 1 sebesar 0, yang artinya tidak ada jarak sama sekali. Hal ini diakibatkan karena kluster 1 hanya memiliki satu titik data saja pada kluster tersebut.
- d. Average within centroid distance_cluster_2 atau rata-rata jarak dalam kluster 2 sebesar 1.419,609. Nilai tersebut adalah jarak rata-rata titik data dalam kluster 2 ke centroid kluster 2.
- e. Nilai Davies-Bouldin Index (DBI) sebesar 0,470.

Berikut ini adalah chart dari hasil proses klusterisasi dengan nilai $k = 3$

PerformanceVector

```
PerformanceVector:
Avg. within centroid distance: 861.392
Avg. within centroid distance_cluster_0: 537.026
Avg. within centroid distance_cluster_1: 0.000
Avg. within centroid distance_cluster_2: 1419.609
Davies Bouldin: 0.470
```

Gambar 5. Hasil Proses Kluster $K = 3$

Hasil operator performance dari proses kluster dengan nilai $k = 4$. Didapati hasil average within centroid distance atau rata-rata jarak dalam centroid, average within centroid distance_cluster_0 atau rata-rata jarak dalam kluster 0, average

within centroid distance_cluster_1 atau rata-rata jarak dalam kluster 1, average within centroid distance_cluster_2 atau rata-rata jarak dalam kluster 2, average within centroid distance_cluster_3 atau rata-rata jarak dalam kluster 3, dan nilai Davies-Bouldin Index (DBI) sebagai berikut:

- a. Average within centroid distance atau rata-rata jarak dalam centroid sebesar 627,939. Nilai tersebut adalah nilai yang menggambarkan jarak rata-rata semua titik data dalam dataset ke centroid kluster mereka. Nilai tersebut lebih rendah daripada proses kluster dengan nilai $k = 2$ dan nilai $k = 3$ yang dilakukan sebelumnya.
- b. Average within centroid distance_cluster_0 atau rata-rata jarak dalam kluster 0 sebesar 421,852. Nilai tersebut adalah jarak rata-rata titik data dalam kluster 0 ke centroid kluster 0.
- c. Average within centroid distance_cluster_1 atau rata-rata jarak dalam kluster 1 sebesar 0, yang artinya tidak ada jarak sama sekali. Hal ini diakibatkan karena kluster 1 hanya memiliki satu titik data saja pada kluster tersebut.
- d. Average within centroid distance_cluster_2 atau rata-rata jarak dalam kluster 2 sebesar 1.898,758. Nilai tersebut adalah jarak rata-rata titik data dalam kluster 2 ke centroid kluster 2.
- e. Average within centroid distance_cluster_3 atau rata-rata jarak dalam kluster 3 sebesar 206,351. Nilai tersebut adalah jarak rata-rata titik data dalam kluster 3 ke centroid kluster 3.
- f. Nilai Davies-Bouldin Index (DBI) sebesar 0,525.

Hasil proses kluster dengan nilai $k = 4$ yang menggambarkan empat kluster, yakni kluster 0 yang berwarna biru toska, kluster 1 yang berwarna merah, kluster 2 yang berwarna hijau muda, dan kluster 3 yang berwarna biru. Gambar tersebut sesuai dengan tabel 4.1 dimana kluster 3 memiliki lebih banyak titik-titik data di dalamnya sehingga membuatnya lebih

padat ketimbang klaster 0, klaster 2, dan terutama klaster 1 yang hanya memiliki satu titik data saja didalamnya.

Hasil operator performance dari proses klaster dengan nilai $k = 5$. Didapati hasil average within centroid distance atau rata-rata jarak dalam centroid, average within centroid distance_cluster_0 atau rata-rata jarak dalam klaster 0, average within centroid distance_cluster_1 atau rata-rata jarak dalam klaster 1, average within centroid distance_cluster_2 atau rata-rata jarak dalam klaster 2, average within centroid distance_cluster_3 atau rata-rata jarak dalam klaster 3, average within centroid distance_cluster_4 atau rata-rata jarak dalam klaster 4 dan nilai Davies-Bouldin Index (DBI) sebagai berikut:

- a. Average within centroid distance atau rata-rata jarak dalam centroid sebesar 531,087. Nilai tersebut adalah nilai yang menggambarkan jarak rata-rata semua titik data dalam dataset ke centroid klaster mereka. Nilai tersebut lebih rendah daripada proses klaster dengan nilai $k = 2$, nilai $k = 3$, dan nilai $k = 4$ yang dilakukan sebelumnya.
- b. Average within centroid distance_cluster_0 atau rata-rata jarak dalam klaster 0 sebesar 165,460. Nilai tersebut adalah jarak rata-rata titik data dalam klaster 0 ke centroid klaster 0.
- c. Average within centroid distance_cluster_1 atau rata-rata jarak dalam klaster 1 sebesar 425,718. Nilai tersebut adalah jarak rata-rata titik data dalam klaster 1 ke centroid klaster 1.
- d. Average within centroid distance_cluster_2 atau rata-rata jarak dalam klaster 2 sebesar 0, yang artinya tidak ada jarak sama sekali. Hal ini diakibatkan karena klaster 2 hanya memiliki satu titik data saja pada klaster tersebut.
- e. Average within centroid distance_cluster_3 atau rata-rata jarak dalam klaster 3 sebesar 3.011,334. Nilai tersebut adalah jarak rata-rata titik data dalam klaster 3 ke centroid klaster 3.

- f. Average within centroid distance_cluster_4 atau rata-rata jarak dalam klaster 4 sebesar 138,700. Nilai tersebut adalah jarak rata-rata titik data dalam klaster 4 ke centroid klaster 4.
- g. Nilai Davies-Bouldin Index (DBI) sebesar 0,609.

Hasil proses klaster dengan nilai $k = 5$ yang menggambarkan lima klaster, yakni klaster 0 yang berwarna biru, klaster 1 yang berwarna biru toska, klaster 2 yang berwarna merah, klaster 3 yang berwarna kuning, dan klaster 4 yang berwarna hijau. Gambar tersebut sesuai dengan tabel 4.1 dimana klaster 4 memiliki lebih banyak titik-titik data di dalamnya sehingga membuatnya lebih padat ketimbang klaster 0, klaster 1, klaster 3 dan terutama klaster 2 yang hanya memiliki satu titik data saja didalamnya.

Hasil operator performance dari proses klaster dengan nilai $k = 5$. Didapati hasil average within centroid distance atau rata-rata jarak dalam centroid, average within centroid distance_cluster_0 atau rata-rata jarak dalam klaster 0, average within centroid distance_cluster_1 atau rata-rata jarak dalam klaster 1, average within centroid distance_cluster_2 atau rata-rata jarak dalam klaster 2, average within centroid distance_cluster_3 atau rata-rata jarak dalam klaster 3, average within centroid distance_cluster_4 atau rata-rata jarak dalam klaster 4, average within centroid distance_cluster_5 atau rata-rata jarak dalam klaster 5, dan nilai Davies-Bouldin Index (DBI) sebagai berikut:

- a. Average within centroid distance atau rata-rata jarak dalam centroid sebesar 219,839. Nilai tersebut adalah nilai yang menggambarkan jarak rata-rata semua titik data dalam dataset ke centroid klaster mereka. Nilai tersebut lebih rendah daripada proses klaster dengan nilai $k = 2$, nilai $k = 3$, nilai $k = 4$, dan nilai $k = 5$ yang dilakukan sebelumnya.
- b. Average within centroid distance_cluster_0 atau rata-rata jarak

- dalam kluster 0 sebesar 284,116. Nilai tersebut adalah jarak rata-rata titik data dalam kluster 0 ke centroid kluster 0.
- c. Average within centroid distance_cluster_1 atau rata-rata jarak dalam kluster 1 sebesar 138,281. Nilai tersebut adalah jarak rata-rata titik data dalam kluster 1 ke centroid kluster 1.
 - d. Average within centroid distance_cluster_2 atau rata-rata jarak dalam kluster 2 sebesar 425,718. Nilai tersebut adalah jarak rata-rata titik data dalam kluster 2 ke centroid kluster 2.
 - e. Average within centroid distance_cluster_3 atau rata-rata jarak dalam kluster 3 sebesar 0, yang artinya tidak ada jarak sama sekali. Hal ini diakibatkan karena kluster 3 hanya memiliki satu titik data saja pada kluster tersebut.
 - f. Average within centroid distance_cluster_4 atau rata-rata jarak dalam kluster 4 sebesar 127,853. Nilai tersebut adalah jarak rata-rata titik data dalam kluster 4 ke centroid kluster 4.
 - g. Average within centroid distance_cluster_5 atau rata-rata jarak dalam kluster 5 sebesar 0, yang artinya tidak ada jarak sama sekali. Hal ini diakibatkan karena kluster 5 hanya memiliki satu titik data saja pada kluster tersebut.
 - h. Nilai Davies-Bouldin Index (DBI) sebesar 0,518.

Chart hasil proses kluster dengan nilai $k = 6$ yang menggambarkan enam kluster, yakni kluster 0 yang berwarna hijau muda, kluster 1 yang berwarna biru, kluster 2 yang berwarna biru toska, kluster 3 yang berwarna merah, kluster 4 yang berwarna hijau, dan kluster 5 yang berwarna kuning. Gambar tersebut sesuai dengan tabel 4.1 dimana kluster 4 memiliki lebih banyak titik-titik data di dalamnya sehingga membuatnya lebih padat ketimbang kluster 0, kluster 1, kluster 2, kluster 4, terutama kluster 3 dan kluster 5 yang hanya memiliki satu titik data saja didalamnya.

Hasil operator performance dari proses kluster dengan nilai $k = 5$. Didapati hasil average within centroid distance atau rata-rata jarak dalam centroid, average within centroid distance_cluster_0 atau rata-rata jarak dalam kluster 0, average within centroid distance_cluster_1 atau rata-rata jarak dalam kluster 1, average within centroid distance_cluster_2 atau rata-rata jarak dalam kluster 2, average within centroid distance_cluster_3 atau rata-rata jarak dalam kluster 3, average within centroid distance_cluster_4 atau rata-rata jarak dalam kluster 4, average within centroid distance_cluster_5 atau rata-rata jarak dalam kluster 5, average within centroid distance_cluster_6 atau rata-rata jarak dalam kluster 6 dan nilai Davies-Bouldin Index (DBI) sebagai berikut:

- a. Average within centroid distance atau rata-rata jarak dalam centroid sebesar 160,358. Nilai tersebut adalah nilai yang menggambarkan jarak rata-rata semua titik data dalam dataset ke centroid kluster mereka. Nilai tersebut lebih rendah daripada proses kluster dengan nilai $k = 2$, nilai $k = 3$, nilai $k = 4$, nilai $k = 5$, dan nilai $k = 6$ yang dilakukan sebelumnya.
- b. Average within centroid distance_cluster_0 atau rata-rata jarak dalam kluster 0 sebesar 636,612. Nilai tersebut adalah jarak rata-rata titik data dalam kluster 0 ke centroid kluster 0.
- c. Average within centroid distance_cluster_1 atau rata-rata jarak dalam kluster 1 sebesar 0, yang artinya tidak ada jarak sama sekali. Hal ini diakibatkan karena kluster 3 hanya memiliki satu titik data saja pada kluster tersebut.
- d. Average within centroid distance_cluster_2 atau rata-rata jarak dalam kluster 2 sebesar 138,401. Nilai tersebut adalah jarak rata-rata titik data dalam kluster 2 ke centroid kluster 2.
- e. Average within centroid distance_cluster_3 atau rata-rata jarak dalam kluster 3 sebesar 121,520. Nilai

- tersebut adalah jarak rata-rata titik data dalam kluster 3 ke centroid kluster 3.
- f. Average within centroid distance_cluster_4 atau rata-rata jarak dalam kluster 4 sebesar 164,022. Nilai tersebut adalah jarak rata-rata titik data dalam kluster 4 ke centroid kluster 4.
 - g. Average within centroid distance_cluster_5 atau rata-rata jarak dalam kluster 5 sebesar 0, yang artinya tidak ada jarak sama sekali. Hal ini diakibatkan karena kluster 5 hanya memiliki satu titik data saja pada kluster tersebut.
 - h. Average within centroid distance_cluster_6 atau rata-rata jarak dalam kluster 6 sebesar 249,793. Nilai tersebut adalah jarak rata-rata titik data dalam kluster 6 ke centroid kluster 6.
 - i. Nilai Davies-Bouldin Index (DBI) sebesar 0,513.

Chart hasil proses kluster dengan nilai $k = 7$ yang menggambarkan tujuh kluster, yakni kluster 0 yang berwarna hijau muda, kluster 1 yang berwarna kuning, kluster 2 yang berwarna hijau toska, kluster 3 yang berwarna biru, kluster 4 yang berwarna biru muda, kluster 5 yang berwarna merah, dan kluster 6 yang berwarna hijau. Gambar tersebut sesuai dengan tabel 4.1 dimana kluster 3 memiliki lebih banyak titik-titik data di dalamnya sehingga membuatnya lebih padat ketimbang kluster 1, kluster 2, kluster 4, kluster 6, terutama kluster 1 dan kluster 5 yang hanya memiliki satu titik data saja didalamnya.

Asosiasi

Dalam menganalisis berbagai genre musik yang memberikan efek terhadap pendengar berdasarkan association yang menggunakan metode FP-Growth, harus dilakukan dengan pengumpulan data awal agar dapat menemukan masalah pada data atau dapat mendeteksi subset unik dalam membentuk suatu hipotesis data yang dimiliki. Hal tersebut dapat dilakukan dengan menganalisis data set yang akan kelompok kami gunakan untuk

memperoleh data awal. Dataset yang digunakan pada association Age, Anxiety, BPM (Beats per Minute), Depression, Hour per day, Insomnia, OCD (obsessive compulsive disorder). Beberapa hal berikut ini dilakukan untuk menyiapkan dataset sebelum dilakukan proses asosiasi lebih lanjut:

1. Melakukan filter atribut dari data set yang kami miliki agar dapat memperoleh pengetahuan dari atribut tersebut. Dalam studi kasus ini kami melakukan filter terhadap atribut frequency (jazz).
2. Melakukan pemilihan atribut yang akan digunakan untuk menganalisis data set, dengan menggunakan operator select attributes. Dalam studi kasus ini, kami telah menentukan atribut yang akan digunakan, yaitu Age, Anxiety, BPM (Beats per Minute), Depression, Hour per day, Insomnia, OCD (obsessive compulsive disorder).
3. Melakukan perubahan data yang memiliki jenis numerik menjadi binominal.
4. Melakukan perubahan data yang memiliki jenis nominal menjadi binominal.

Koneksi antar operator yang berawal dari penggunaan operator replace missing value untuk menggantikan nilai data yang hilang, kemudian penggunaan operator filter examples untuk menentukan atribut patokan yang ingin dianalisis, lalu melakukan select atribut untuk memilih atribut yang akan digunakan untuk menganalisis atribut yang telah di set pada filter examples, penggunaan operator numerical to binominal untuk melakukan perubahan jenis data numerik menjadi binominal, dan penggunaan operator nominal to binominal untuk melakukan perubahan jenis data nominal menjadi binominal.

Menggunakan model operator FP-Growth untuk melakukan metode association ini. Dimana model ini digunakan untuk mengkalkulasi semua item set yang sering muncul pada dataset

menggunakan struktur data FP tree secara efisien.

Hasil dari pemodelan association rules dengan menggunakan model FP-Growth yang dapat kita lihat bahwa ditemukan 20 premis berbeda-beda yang memiliki kesimpulan yang berbeda-beda juga, tetapi diantaranya terdapat kesimpulan yang sering tampil, yaitu Age serta efek yang mempengaruhi gangguan kesehatan mental. Pada support terdapat jumlah yang sering tampil dan memiliki nilai berjumlah 0.728, 0.729, 0.731, 0.734, 0.741, dan 0.742. Lalu pada confidence terdapat jumlah yang sering tampil dan memiliki nilai tertinggi berjumlah 0.821. Pada laplace terdapat jumlah yang sering tampil dan memiliki nilai tertinggi berjumlah 0.916, pada gain terdapat nilai tertinggi berjumlah -1.045, pada p-s memiliki sebuah nilai yang sering tampil dan memiliki nilai 0.039, 0.040, 0.044, dan 0.045. Pada lift nilai yang sering tampil dan tertinggi adalah 1.064. Serta pada conviction terdapat beberapa nilai 1.254, 1.258, 1.267, 1.276, 1.278, 1.280, dan 1.285.

Pada metrik support, yang menunjukkan frekuensi kemunculan premis dalam dataset, terdapat beberapa nilai yang sering muncul, yaitu 0.728, 0.729, 0.731, 0.734, 0.741, dan 0.742. Nilai ini menunjukkan bahwa premis-premis tersebut muncul dengan frekuensi yang cukup tinggi dalam dataset. Pada metrik confidence, yang mengukur kekuatan dari hubungan antara premis dan kesimpulan, terdapat nilai tertinggi sebesar 0.821. Ini menunjukkan bahwa aturan dengan confidence ini memiliki tingkat kepastian yang tinggi bahwa premis akan menghasilkan kesimpulan yang sesuai. Pada metrik laplace, yang digunakan untuk mengatasi masalah ketidakseimbangan data, terdapat nilai tertinggi sebesar 0.916. Ini menunjukkan bahwa aturan tersebut cukup stabil meskipun ada variasi dalam data. Pada metrik gain, yang mengukur peningkatan informasi dari aturan tersebut dibandingkan dengan ekspektasi acak,

terdapat nilai tertinggi sebesar -1.045. Nilai negatif menunjukkan bahwa aturan tersebut mungkin tidak memberikan peningkatan informasi yang signifikan dibandingkan dengan ekspektasi acak. Pada metrik p-s, yang mengukur signifikansi statistik dari aturan, terdapat beberapa nilai yang sering muncul, yaitu 0.039, 0.040, 0.044, dan 0.045. Nilai-nilai ini menunjukkan bahwa aturan-aturan tersebut cukup signifikan secara statistik. Pada metrik lift, yang mengukur sejauh mana premis dan kesimpulan terjadi bersamaan dibandingkan dengan jika mereka independen, terdapat nilai tertinggi sebesar 1.064. Nilai ini menunjukkan bahwa premis dan kesimpulan memiliki hubungan yang sedikit lebih kuat daripada yang diharapkan jika mereka independen. Pada metrik conviction, yang mengukur tingkat kepercayaan dari aturan tersebut, terdapat beberapa nilai yang sering muncul, yaitu 1.254, 1.258, 1.267, 1.276, 1.278, 1.280, dan 1.285. Nilai-nilai ini menunjukkan bahwa aturan-aturan tersebut cukup meyakinkan dalam prediksi kesimpulan berdasarkan premis. Secara keseluruhan, hasil dari pemodelan association rules dengan model FP-Growth ini menunjukkan bahwa usia (Age) sering menjadi faktor yang signifikan dan memiliki pengaruh yang cukup kuat terhadap gangguan kesehatan mental, seperti yang ditunjukkan oleh berbagai metrik yang digunakan untuk mengevaluasi aturan-aturan yang dihasilkan.

SIMPULAN

Untuk klasterisasi, berdasarkan pada proses klastering yang telah dilakukan, didapatkan nilai $k = 3$ merupakan nilai k yang paling cocok dibandingkan dengan nilai k lainnya untuk mengklasterisasi dataset yang digunakan dengan nilai DBI sebesar 0,470.

Untuk Asosiasi, model asosiasi yang telah dibuat, ditemukan 20 premis berbeda-beda yang memiliki kesimpulan yang berbeda-beda juga, tetapi diantaranya

terdapat kesimpulan yang sering tampil, yaitu Age serta efek yang mempengaruhi gangguan kesehatan mental, yaitu Pada support terdapat jumlah yang sering tampil dan memiliki nilai berjumlah 0.728, 0.729, 0.731, 0.734, 0.741, dan 0.742. Pada confidence terdapat jumlah yang sering tampil dan memiliki nilai tertinggi berjumlah 0.821. Pada laplace terdapat jumlah yang sering tampil dan memiliki nilai tertinggi berjumlah 0.916, pada gain terdapat nilai tertinggi berjumlah -1.045, pada p-s memiliki sebuah nilai yang sering tampil dan memiliki nilai 0.039, 0.040, 0.044, dan 0.045. Pada lift nilai yang sering tampil dan tertinggi adalah 1.064. Pada conviction terdapat beberapa nilai 1.254, 1.258, 1.267, 1.276, 1.278, 1.280, dan 1.285.

Secara keseluruhan, hasil dari pemodelan association rules dengan model FP-Growth ini menunjukkan bahwa usia (Age) sering menjadi faktor yang signifikan dan memiliki pengaruh yang cukup kuat terhadap gangguan kesehatan mental, seperti yang ditunjukkan oleh berbagai metrik yang digunakan untuk mengevaluasi aturan-aturan yang dihasilkan.

DAFTAR PUSTAKA

Andry, J. F., Sibarani, R., & Yefta, V. N. (2023). Analysis of Big Data Football Club Market Value Using K-Means and Linear Regression Mining Methods. *Journal of Computer Science*, vol. 19, no. 2, pp. 286-294, 2023, [https://doi: 10.3844/jcssp.2023.286.294](https://doi.org/10.3844/jcssp.2023.286.294).

Geasela, Y. M., Bernanda, D. Y., Andry, J. F., Jusuf, C. K., Winata, S., Lydia, & Everlin, S. (2024). Analysis of Student Mental Health Dataset Using Mining Techniques," *Journal of Computer Science*, vol. 20, no. 1, pp. 121-128, 2024, [https://doi: 10.3844/jcssp.2024.121.128](https://doi.org/10.3844/jcssp.2024.121.128).

Golden, T. L., Tetreault, L., Ray, C. E., Kuge, M. N., Tiedemann, A., & Magsamen, S. (2022). The State of

Music-Based Interventions for Mental Illness: Thought Leaders on Barriers, Opportunities, and the Value of Interdisciplinarity. *Community Mental Health Journal*, 58(3), 487-498. <https://doi.org/10.1007/s10597-021-00843-4>

Gupta, A., Sharma, H., & Akhtar, A. (2021). A comparative analysis of k-means and hierarchical clustering. *EPR International Journal of Multidisciplinary Research (IJMR)*, 7(8). <https://doi.org/10.36713/epira8308>

Gupta, G., Wayal, A., Jain, R., Dr. Waghmare, K., Kadale, P. (2023). Mood Based Music Recommendation System. *International Research Journal of Modernization in Engineering Technology and Science*. <https://doi.org/10.56726/irjmet39298>

Hou, J. (2022). Effective Ways for College Students' Mental Health Education Based on Music Therapy. *Journal of Healthcare Engineering*. Hindawi Limited. <https://doi.org/10.1155/2022/3031064>

Hou, T., Zhang, T., Cai, W., Song, X., Chen, A., Deng, G., & Ni, C. (2020). Social support and mental health among health care workers during Coronavirus Disease 2019 outbreak: A moderated mediation model. *Plos one*, 15(5). <https://doi.org/10.1371/journal.pone.0233831>

Keleş, A. E., Önen, E., & Górecki, J. (2022). Determination of green building awareness: a study in Turkey. *Sustainability*, 14(19), 11943. <https://doi.org/10.3390/su141911943>

Lu, D. (2023). Evaluation Model of Music Therapy's Auxiliary Effect on Mental Health Based on Artificial Intelligence Technology. *Journal of*

- Environmental and Public Health, 2022.
<https://doi.org/10.1155/2022/9960589>
- Madyatmaja, E. D., Jordan, S. I., & Andry, J. F. (2021a). Big Data Analysis Using RapidMiner Studio To Predict Suicide Rate In Several Countries. *ICIC Express Letters Part B: Applications*, 12(8), 757-764. <https://doi.org/10.24507/icicelb.12.08.757>
- Madyatmadja, E. D., Sembiring, D. J. M., Angin, S. M. B. P., Ferdy, D. and Andry, J. F. (2021b). Big data in educational institutions using RapidMiner to predict learning effectiveness," *Journal of Computer Science*, vol. 17, no. 4, pp. 403-413. [https://doi: 10.3844/jcssp.2021.403.413](https://doi.org/10.3844/jcssp.2021.403.413)
- Mutalib, S., Shafiee, N. S. M., & Abdul-Rahman, S. (2021). Mental health prediction models using machine learning in higher education institution. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, 12(5), 1782-1792. <http://dx.doi.org/10.17762/turcomat.v12i5.2181>
- Qiaoyi, Chen. (2023). Mental Health as a Mediator of the Transformational Leadership to Better Performance. *Journal of Education, Humanities and Social Sciences*. <https://doi.org/10.54097/ehss.v8i.4660>
- Pratama, Y., Pramudya, Y., Albert, E., Ferdinand, R., Juniansyah, V., & Rasywir, E. (2022). Klasterisasi Data Pertanian di Tingkat Provinsi Jambi Tahun 2021 Menggunakan Algoritma K-Means. *Bulletin of Computer Science Research*, 3(1), 57-63. <https://doi.org/10.47065/bulletincsr.v3i1.205>
- Rahman, R. A., Omar, K., Noah, S. A. M., Danuri, M. S. N. M., & Al-Garadi, M. A. (2020). Application of machine learning methods in mental health detection: a systematic review. *Ieee Access*, 8, 183952-183964. <https://doi.org/10.1109/ACCESS.2020.3029154>
- Rebecchini, L. (2021). Music, mental health, and immunity. *Brain, Behavior, and Immunity - Health*, 18. <https://doi.org/10.1016/j.bbih.2021.100374>
- Schwartz, K. D., Exner-Cortens, D., McMorris, C. A., Makarenko, E., Arnold, P., Van Bavel, M., ... & Canfield, R. (2021). COVID-19 and student well-being: Stress and mental health during return-to-school. *Canadian journal of school psychology*, 36(2), 166-185. <https://doi.org/10.1177/08295735211001653>
- Wang, K., Gao, S., & Huang, J. (2022). Learning About Your Mental Health From Your Playlist? Investigating the Correlation Between Music Preference and Mental Health of College Students. *Frontiers in Psychology*, 13. <https://doi.org/10.3389/fpsyg.2022.824789>
- Wang, Q., Zhang, W., & An, S. (2023). A systematic review and meta-analysis of Internet-based self-help interventions for mental health among adolescents and college students. *Internet Interventions*, 100690. <https://doi.org/10.1016/j.invent.2023.100690>
- Wesseldijk, L. W., Ullén, F., & Mosing, M. A. (2019). The effects of playing music on mental health outcomes. *Scientific Reports*, 9(1). <https://doi.org/10.1038/s41598-019-49099-9>

- Wesseldijk, L. W., Lu, Y., Karlsson, R., Ullén, F., & Mosing, M. A. (2023). A comprehensive investigation into the genetic relationship between music engagement and mental health. *Translational Psychiatry*, 13(1).
<https://doi.org/10.1038/s41398-023-02308-6>
- WHO (World Health Organization), 2019. What is the evidence on the role of the arts in improving health and well-being? A scoping review. World Health Organization. Regional Office for Europe;