*p-ISSN: 2460-173X*
*e-ISSN: 2598-5841*

# Analisis *Cluster* Atribut *Audio* pada Lagu Terpopuler Aplikasi TikTok

**Karina Auliasari[1)], Mariza Kertaningtyas[2)]**

[1)]Program Studi Teknik Informatika, Institut Teknologi Nasional Malang
[2)]Program Studi Teknik Industri, Institut Teknologi Nasional Malang
Jl. Raya Karanglo KM. 2 Tasikmadu, Kota Malang
[1)]karina.auliasari@lecturer.itn.ac.id
[2)]mariza_kertaningtyas@lecturer.itn.ac.id

**Abstrak**

Layanan *streaming* musik seperti TikTok telah merubah cara konsumen mendengarkan musik. Memahami apa atribut yang membuat lagu tertentu menjadi populer dapat memberikan informasi tertentu untuk menciptakan pengalaman pelanggan yang lebih baik serta lebih banyak lagi upaya pemasaran yang efektif bagi pengembang aplikasi TikTok. Pada penelitian ini dilakukan *cluster analisis* pada dataset yang berjumlah 6746 *track* yang paling populer pada aplikasi TikTok dari tahun 2004 hingga tahun 2021. Pada penelitian ini dilakukan empat proses dalam metode penelitian yaitu pengumpulan data, pra-pemrosesan data, penerapan algoritma K-Means dan analisis hasil klasterisasi. Dataset dari proses pengumpulan data memiliki sebelas atribut yaitu *danceability*, *key*, *energy*, *loudness*, *speechiness*, *acousticness*, *instrumentalness*, *liveness*, *valence*, *tempo*, dan *duration*. Dari hasil *cluster* menunjukkan menunjukkan ada dua kelompok data dimana data yang dikelompokkan merupakan data lagu atau musik dengan nilai popularitas lebih dari 50. Klaster pertama berisi 1846 data sedangkan pada klaster kedua ada 2876 data. Dari hasil klaster dapat diketahui bahwa terdapat beberapa atribut yang membuat lagu atau *track* musik pada aplikasi TikTok menjadi *trending* yaitu diantaranya atribut *instrumentalness* dengan nilai yang tinggi, durasi pemutaran yang lama, *danceability, loudness, speechiness, valence* dan *tempo* yang juga memiliki nilai yang tinggi.

**Kata kunci:** TikTok, K-Means, Klastering data, Atribut, Lagu

***Abstract***

*Music streaming services like TikTok have changed the way consumers listen to music. Understanding what attributes make a particular song popular can provide certain information to create a better customer experience as well as more effective marketing efforts for TikTok app developers. In this study, cluster analysis was carried out on a dataset totaling 6746 tracks, which were the most popular on the TikTok application from 2004 to 2021. In this study, four processes were carried out in the research method: data collection, data pre-processing, application of the K-Means algorithm, and analysis. clustering results. The dataset from the data collection process has eleven attributes, namely danceability, key, energy, loudness, speechiness, acousticness, instrumentalness, liveness, valence, tempo, and duration. The cluster results show that there are two groups of data where the grouped data is song or music data with a popularity value of more than 50. The first cluster contains 1846 records, while the second contains 2876 records. From the cluster results, it can be seen that there are several attributes that make songs or music tracks trending on the TikTok application, namely the instrumentalness attribute with a high value, the long duration of playback (duration_minutes), danceability, loudness, speechiness, valence, and tempo, which also have a high value.*

***Keywords:*** *TikTok, K-Means, Data Clustering, Attribute, Song*

## 1. PENDAHULUAN

TikTok merupakan salah satu aplikasi paling populer di dunia dengan ratusan juta pengguna, yang kebanyakan dari mereka anak-anak dan remaja. TikTok digunakan untuk mengunggah, menonton,


*Received: 11 Oktober 2022*
*Accepted: 14 Desember 2022*

menelusuri dan menyinkronkan video dan meme. TikTok dikembangkan oleh ByteDance, sebuah perusahaan Cina dengan fitur untuk mengunggah video *lip-sync* durasi 60 detik dengan berbagai fitur kreatif dan interaktif. Sebagai aplikasi Tiktok tumbuh dengan cepat dan menempati peringkat ketujuh yang paling banyak diunduh dalam 10 tahun terakhir. Antara rentang waktu Januari 2018 hingga Agustus 2020 jumlah pengguna aktif di TikTok meningkat sebesar 800% yaitu dari 100 juta pengguna di Amerika Serikat menjadi 700 juta pengguna di seluruh dunia [1]. Di Amerika Serikat pengguna TikTok memiliki rentang umur 10 hingga 19 tahun yang merupakan 25% dari total akun yang ada [2]. Pada kuartal pertama tahun 2020 saja TikTok mengumpulkan lebih dari 315 juta pemasangan di App Store dan Google Play, dan terus meningkat menghasilkan lebih dari 2 miliar unduhan secara global sejak peluncurannya [3]. Di antara pengguna TikTok ini, 90% menyatakan bahwa menggunakannya setiap hari [4]. Selama masa pandemi covid-19 aplikasi TikTok lebih banyak lagi menarik penggunanya dan memiliki popularitas di kalangan pengguna berusia muda. Layanan streaming musik secara online melalui jaringan internet telah merevolusi manusia dalam mendengarkan musik. Tidak hanya efisiensi menurunkan biaya menikmati musik namun juga menjangkau pendengar musik dari berbagai genre musik secara lebih luas. Aplikasi TikTok selain menyediakan fitur untuk upload video juga sebagai platform untuk mengupload, mendownload dan mendengarkan track musik dengan genre apapun. Aplikasi TikTok sebagai layanan streaming musik yang populer menyediakan akses ke lebih dari 50 juta track hingga 200 juta pengguna yang mengakses track tersebut [5]. Dalam beberapa tahun terakhir, TikTok telah memungkinkan pengguna untuk menemukan track musik dan membuat daftar putar eksklusif berdasarkan preferensi, genre, artis favorit dan bahkan suasana hati dari penggunanya. Fitur-fitur musik pada TikTok tersebut membantu pengguna mencari warna musik kesukaan pada jutaan track pada database. Algoritma pemahaman di dalam aplikasi TikTok tentang karakteristik dan penggunaan daftar putar memberikan rekomendasi yang sesuai bagi penggunanya.

K-means merupakan metode clustering yang mencari pusat cluster yang mewakili wilayah tertentu dari data. Pengelompokan K-means memiliki banyak keuntungan, seperti tidak memerlukan matriks jarak pada *clustering* hirarkis dan memiliki waktu komputasi yang cepat [5]. Analisis algoritma K-Means sebelumnya digunakan untuk mengelompokkan mahasiswa baru menentukan pola pemilihan program studi. Hasil pengujian dengan menggunakan metode Silhouette Coefficient mendapatkan nilai terbaik 0,690754 dengan jumlah 3 cluster dan jumlah data 15 [6]. Metode K-Means dan K-Medoid digunakan untuk membandingkan hasil atribut *genre* pada data *Spotify*. Dengan menggunakan data global Top 50 lagu dan menghitung 3 *cluster* dihasilkan *cluster* 1 berisi 2.833 anggota, *cluster* 2 berisi 21 anggota. Hasil perhitungan K-Means berdasarkan rata-rata pada *cluster* 1 tertinggi adalah atribut tempo dengan nilai 118 dan *cluster* 2 tertinggi adalah atribut tempo dengan nilai 125, sedangkan *cluster* 3 atribut tertinggi juga tempo dengan nilai 123 [7]. Pengelompokan daerah dengan penyakit DBD tertinggi dan terendah di Kabupaten Karawang menggunakan K-Medoids dengan jarak *euclidean distance* dan proses data dengan melakukan seleksi data, pembersihan data, transformasi data, mining dan evaluasi. Sehingga hasil yang didapatkan dari dataset penyakit DBD di Kabupaten Karawang Tahun 2020 memiliki 2 *cluster* optimal. Adapun *cluster* 1 dengan 6 daerah di kategorikan tinggi, sedangkan *cluster* 2 dengan 22 daerah di kategorikan rendah [8]. Pendataan asuransi produk perusahaan nasional, menggunakan 3 atribut yaitu premi, jumlah pelanggan, dan tahun pelepasan untuk setiap produk. Hasil pengujian menggunakan K-Means dengan nilai terkecil K = 5 yaitu 0,118. Sedangkan pada K-Medoids nilai terkecil adalah K = 2 yaitu 0,027 [9]. Penggunaan K-Means juga dilakukan oleh Parlina dkk dalam penentuan pegawai yang layak mengikuti program SDP di PT. Bank Syariah. Hasil dari pengelompokan diperoleh tiga kelompok yaitu Lolos, Hampir Lolos dan Tidak Lolos, dimana pusat cluster 1= 8;66;13, cluster-2= 10;71;14 dan cluster-3=7;60;12 [10]. Majhi dan Biswal mengkombinasikan algoritma K-means dan *Ant Lion Optimization* (ALO). Hasil klaster kombinasi tersebut dibandingkan dengan hasil klaster K-Means, KMeans-PSO, KMeans-FA dan DBSCAN. Hasil yang diperoleh menunjukkan bahwa kombinasi K-Means dan ALO bekerja lebih baik daripada ketiga algoritma lainnya dalam hal jumlah jarak intracluster [11]. Penelitian lain membandingkan K-Means dan K-Medoids menggunakan dataset tanaman Iris dari website UCI repository, dari penelitian menunjukkan bahwa K-Medoids memiliki hasil lebih baik jika dilihat dari nilai akurasi klaster yang dihasilkan yaitu 92% untuk algoritma K-Medoids dan 88.7% untuk algoritma K-Means [12]. Pengembangan K-Means digunakan untuk mengelompokkan titik GPS taksi pada suatu peta tanpa memerlukan input jumlah klaster dengan nama algoritma *NoiseClust.* Didapatkan

hasil klaster yang lebih optimal namun memerlukan waktu komputasi lebih lama dibandingkan K-Means [13]. Heil dkk membandingkan K-Means dengan Fuzzy K-Means pada data pertanian. Pada hasil penelitian ditemukan bahwa Fuzzy K-Means dapat mengelompokkan data lebih baik dibandingkan dengan K-Means yang mana tidak dapat mendeteksi beberapa data untuk dikelompokkan [14].

Dari beberapa penelitian menggunakan K-Means tersebut belum ada penelitian yang menganalisis data *track* musik pada aplikasi TikTok untuk mengetahui apakah *track* musik yang *trending* atau populer terkait dengan atribut atau fitur *audio* seperti *key, valence, danceability, energy, loudness, speechiness* dan *acousticness*. Karena daftar putar musik disesuaikan berdasarkan algoritma rekomendasi aplikasi TikTok, lagu-lagu tertentu mulai berulang kali muncul di daftar lagu teratas yang mengakibatkan *trending* di platform TikTok. Setiap lagu di TikTok memiliki fitur *audio* seperti *danceability, energy, loudness, speechiness* dan *acousticness*. Penelitian ini bermaksud untuk mengetahui apakah keberhasilan lagu-lagu trending terkait dengan beberapa atribut *audio* tersebut.

## 2. TINJAUAN PUSTAKA

### 2.1 Aplikasi TikTok

TikTok adalah aplikasi sosial media populer yang memungkinkan pengguna membuat, menonton, dan berbagi video berdurasi 15 detik yang direkam dari perangkat seluler atau *webcam*. Dengan fitur *feed* yang dipersonalisasi dari video pendek unik yang dilengkapi musik dan efek suara, aplikasi TikTok terkenal karena videonya yang adiktif dan tingkat keterlibatan pengguna lain yang tinggi [7]. Pembuat konten TikTok (amatir atau profesional) dapat menambahkan efek seperti *filter*, musik latar, dan *sticker* dalam konten video mereka, dan dapat berkolaborasi pada konten dan membuat video duet layar terpisah meskipun berada di lokasi yang berbeda [8]. Format video atau *track* TikTok lebih ditujukan untuk hiburan dan komedi. Namun saat ini semakin banyak digunakan untuk *infotainment*. *Influencer* yang mendapatkan *audiens* tetap di aplikasi TikTok menawarkan potongan saran dan tips bersama dengan promosi diri. Topik kecantikan, mode, keuangan pribadi, dan memasak adalah topik-topik yang populer untuk tujuan video informasi. Selain itu video digunakan untuk mempromosikan dan menjual produk [9]. Dalam perkembangannya TikTok menyediakan fitur TikTok *sound* melalui *SoundOn* dimana sebagai *platform* promosi dan distribusi musik TikTok, yang memungkinkan artis menciptakan kreasi suara ataupun musik sehingga dapat berkembang dengan basis penggemar mereka dan karya mereka dapat didengar di seluruh dunia. Melalui *platform SoundOn* artis dapat mengunggah musik mereka di TikTok dan mendapatkan royalti saat musik itu digunakan. Musik atau kreasi suara dapat digunakan sebagai latar suara bagi pembuat konten video. Musik atau kreasi suara tertentu yang digunakan berkali-kali oleh pembuat konten maka *track sound/song* tersebut menjadi viral di TikTok [15].

### 2.2 K-Means

K-means clustering adalah metode *clustering* yang mencari pusat cluster yang mewakili wilayah tertentu dari data. Pengelompokan K-means memiliki banyak keuntungan, seperti tidak memerlukan matriks jarak pada *clustering* hirarkis dan memiliki waktu komputasi yang cepat [5]. K-Means diperkenalkan oleh J.B. MacQueen pada tahun 1976. Metode ini mempartisi data ke dalam *cluster-cluster* sehingga data dengan karakteristik yang sama dikelompokkan ke dalam *cluster* yang sama dan karakteristik yang berbeda dikelompokkan ke dalam grup lain [16]. Berikut langkah-langkah algoritma K-Means [16]:

Langkah 1: Tentukan jumlah K-cluster yang ingin Anda bentuk.

Langkah 2: Untuk pusat *cluster* awal (*centroid*), buat k nilai acak.

Langkah 3: Hitung jarak setiap titik data input ke setiap *centroid* menggunakan rumus jarak Eucledian (*Eucledian Distance*) untuk mencari jarak terdekat dari setiap titik data ke *centroid*. Berikut adalah persamaan Jarak Eucledian (Persamaan 2):

$$d(x_i, \mu_j) = \sqrt{(x_i - \mu_j)^2} \quad (2)$$

Langkah 4: Klasifikasikan setiap kumpulan data berdasarkan kedekatannya dengan centroid (jarak terkecil).

Langkah 5: Hitung ulang nilai centroid. Nilai centroid baru diperoleh dari rata-rata cluster yang bersangkutan dengan menggunakan rumus yang ditunjukkan pada Persamaan 2:

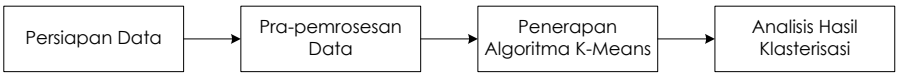
$$\mu_j(t+1) = \frac{1}{N_{S_j}} \sum j = S_j\, X_j \quad (3)$$

Dimana:
$\mu_j(t+1)$ adalah *centroid* baru pada iterasi ke (t+1)
$N_{S_j}$ adalah data di *cluster* $S_j$

Langkah 6: Ulangi langkah 2-5 sampai anggota setiap cluster tetap konstan.
Langkah 7: Jika langkah 6 telah terpenuhi, maka nilai rata-rata pusat cluster (j) pada iterasi terakhir akan digunakan sebagai parameter untuk fungsi radial basis pada lapisan tersembunyi.

## 3. METODE PENELITIAN

Dalam penelitian ini beberapa proses dilakukan untuk mengetahui keterkaitan antara atribut pengklasifikasi pada dataset lagu aplikasi TikTok, diantaranya adalah persiapan data kemudian dilanjutkan dengan pemilihan atribut pengklasifikasi, klasterisasi dengan K-Means Clustering dan analisis hasil klasterisasi. Gambaran proses-proses dalam metode penelitian ditunjukkan pada Gambar 1.




Gambar 1. Diagram alir proses penelitian

### 3.1 Persiapan Data

Proses awal yang dilakukan dalam persiapan data adalah pengumpulan dataset lagu TikTok yang bersumber dari website Kaggle [17]. Dataset lagu TikTok yang didapatkan memiliki komponen fitur audio, data fitur-fitur audio ditampilkan pada Tabel 1. Dalam beberapa tahun TikTok mengkompilasi berbagai daftar yang menampilkan artis, lagu dan album top untuk kemudian dikategorikan menjadi beberapa daftar berdasarkan wilayah dan genre musik. Untuk menganalisa *trend* musik populer dan untuk bisa lebih memahami atribut apa saja yang mempengaruhi, pada dataset penelitian ini digunakan 6746 baris data lagu yang paling populer dengan tahun rilis dari tahun 2004 hingga tahun 2021.

Tabel 1. Fitur audio TikTok [17]

| Atribut | Tipe Data | Deskripsi |
|---|---|---|
| *Key* | Integer | Nilai kunci keseluruhan lagu, nilai bertipe data integer ini memetakan ke pitch dengan standar pitch notasi kelas, misalnya 0=C, 1=C/D, 2=D dan seterusnya. |
| *Danceability* | Float | Menjelaskan seberapa cocok lagu digunakan untuk menari berdasarkan kombinasi elemen musik seperti tempo, stabilitas ritme, kekuatan ketukan dan keteraturan secara keseluruhan. Jika nilainya 0.0 maka yang paling tidak cocok dan jika 1.0 maka yang paling cocok. |
| *Energy* | Float | Rentang nilai 0.0 hingga 1.0 dan mewakili ukuran persepsi intensitas dan aktifitas. Jika lagu memiliki energi yang terasa cepat, keras dan berisik seperti contohnya lagu metal maka memiliki nilai energi yang tinggi. Sedangkan lagu jazz memiliki skala nilai yang rendah. Fitur-fitue yang berkontribusi terhadap nilai atribut ini adalah rentang dinamis, kenyaringan yang dirasakan, timbre, tingkat onset dan entropi. |
| *Loudness* | Float | Atribut sensasi pendengaran suara yang dapat diurutkan pada skala tertentu baik dari lirih ke keras. |
| *Mode* | Integer | Tangga nada mayor atau minor dari sebuah lagu. Mayor diwakili oleh 1 dan minor adalah 0. |
| *Speechiness* | Float | Mendeteksi adanya kata dalam pengucapan pada lagu. Jika lagu berisi banyak kata-kata maka bernilai antara 0.33 hingga 0.66 seperti lagu rap. Jika lagu tidak mengandung kata-kata apapun maka nilai *speechiness*nya kurang dari 0.33. |

| Atribut | Tipe Data | Deskripsi |
|---|---|---|
| *Acousticness* | Float | Nilai *confidence* dengan rentang nilai 0.0 hingga 1.0 apakah lagu akustik atau tidak. Jika bernilai 1.0 maka lagu merupakan lagu akustik. |
| *Instrumentalness* | Float | Mewakili jumlah vokal dalam lagu, jika nilai lebih dari 1.0 maka lagu tidak mengandung konten vokal. |
| *Liveness* | Float | Atribut yang menjelaskan apakah lagu direkam dengan penonton langsung atau secara (*live)* saat di atas panggung. Jika bernilai lebih dari 0.8 maka lagu tersebut direkam dengan kondisi *live.* |
| *Valence* | Float | Atribut yang menggambarkan nuansa musik positif yang disampaikan sebuah lagu, dengan rentang nilai 0.0 hingga 1.0. Jika lagu memiliki nilai *valence* yang tinggi maka lagu terdengar positif (nuansa bahagia, ceria dan gembira) sementara lagu dengan nilai *valence* yang rendah terdengar lebih negatif (nuansa sedih, tertekan dan marah). |
| *Tempo* | Float | Atribut yang menjelaskan kecepatan musik pada lagu dimainkan. |
| *Duration_ms* | Integer | Durasi lagu dalam satuan *millisecond* atau milli detik. |

### 3.2 Pra-Pemrosesan Data

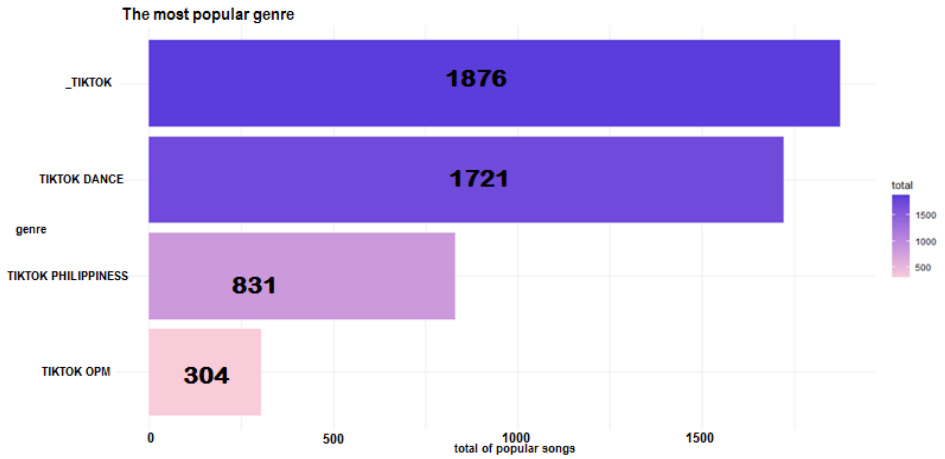
Pada tahap pra pemrosesan data akan didefinisikan berdasarkan atribut popularitas (*popularity*) sebagai atribut biner dan untuk memilih lagu atau *track* tertentu yang memiliki nilai popularitas lebih atau sama dengan 50 (≥ 50). Nilai 50 ditentukan dalam penelitian ini karena skala nilai pada atribut popularitas adalah 0 s/d 100. *Track* atau lagu yang memiliki nilai popularitas ≥ 50 pada penelitian ini akan didefinisikan sebagai populer dan diganti dengan kode 1 sedangkan *track* atau lagu yang nilai popularitasnya dibawah 50 (< 50) diganti dengan nilai 0 dan dianggap pada penelitian ini sebagai data yang tidak populer. Selanjutnya data yang dipilih adalah *track* atau lagu yang memiliki nilai popularitas 1 seperti yang diperlihatkan pada Gambar 2. Ada sejumlah 4722 baris data yang memiliki nilai popularitas 1 dan 2024 data yang nilai popularitasnya 0.

| | duration_mins | genre | popularity.conv |
|---|---|---|---|
| 1 | 3.98 | TIKTOK DANCE | 1 |
| 2 | 3.98 | TIKTOK DANCE | 1 |
| 3 | 2.69473333333333 | TIKTOK DANCE | 1 |
| 4 | 3.63825 | TIKTOK DANCE | 1 |
| 5 | 2.0462 | TIKTOK DANCE | 1 |
| 6 | 2.0462 | TIKTOK DANCE | 1 |
| 7 | 3.85931666666667 | TIKTOK DANCE | 1 |
| 8 | 3.4243 | TIKTOK DANCE | 1 |
| 9 | 2.09298333333333 | TIKTOK DANCE | 1 |
| 10 | 2.09298333333333 | TIKTOK DANCE | 1 |
| 11 | 2.74068333333333 | TIKTOK DANCE | 1 |
| 12 | 2.74068333333333 | TIKTOK DANCE | 1 |
| 13 | 3.66155 | TIKTOK DANCE | 1 |
| 14 | 4.63475 | TIKTOK DANCE | 1 |
| 15 | 4.63475 | TIKTOK DANCE | 1 |
| 16 | 2.84776666666667 | TIKTOK DANCE | 1 |
| 17 | 2.84776666666667 | TIKTOK DANCE | 1 |
| 18 | 2.46193333333333 | TIKTOK DANCE | 1 |
| 19 | 4.19981666666667 | TIKTOK DANCE | 1 |
| 20 | 3.85666666666667 | TIKTOK DANCE | 1 |
| 21 | 3.4811 | TIKTOK DANCE | 1 |
| 22 | 3.4811 | TIKTOK DANCE | 1 |
| 23 | 3.4811 | TIKTOK DANCE | 1 |
| 24 | 3.4811 | TIKTOK DANCE | 1 |

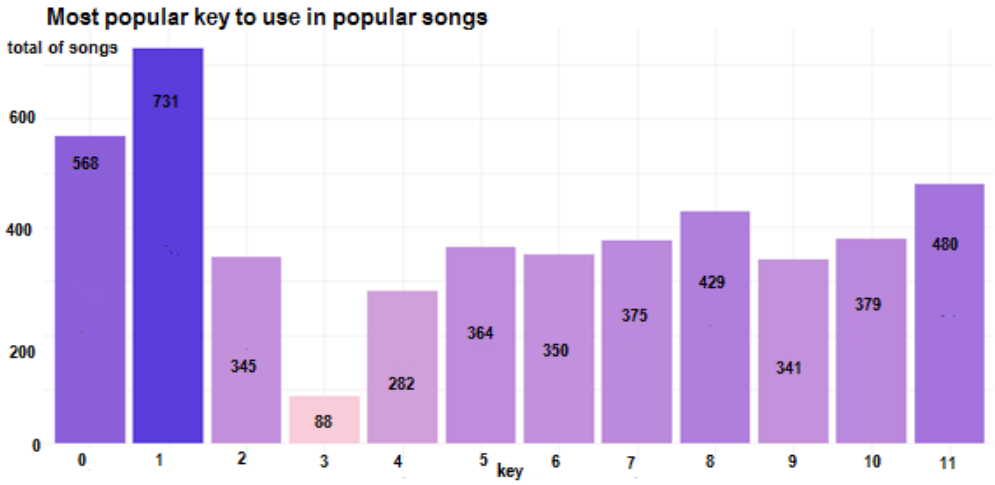
Gambar 2. Data dengan nilai atribut popularitas 1

Setelah memilih data lagu atau musik dengan nilai popularitas 1, dilakukan agregasi data yaitu menggabungkan genre paling populer berdasarkan seberapa sering genre tersebut muncul pada daftar lagu populer. Sehingga didapatkan grafik seperti yang ditunjukkan pada Gambar 3 terlihat bahwa jumlah data yang paling banyak adalah lagu dengan genre TIKTOK DANCE diikuti dengan genre _TIKTOK di posisi kedua, genre TIKTOK PHILIPPINES di posisi ketiga dan genre TIKTOK OPM yang jumlah lagunyanya paling sedikit. Namun jika data lagu dieksplorasi lebih lanjut berdasarkan nilai popularitasnya seperti yang ditunjukkan pada Gambar 3, genre _TIKTOK memiliki nilai popularitas tertinggi, yang populer kedua genre TIKTOK PHILIPPINES, populer ketiga genre TIKTOK DANCE dan yang paling rendah nilai popularitasnya genre TIKTOK OPM. Selain memvisualisasikan genre tertentu, digunakan juga parameter *key* untuk melihat kunci musik apa yang digunakan pada daftar lagu populer seperti yang ditunjukkan pada Gambar 4.

*p-ISSN: 2460-173X*
*e-ISSN: 2598-5841*

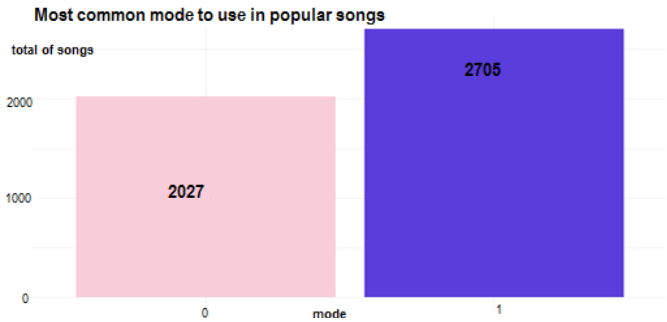



Gambar 3. Grafik jumlah lagu berdasarkan genre




Gambar 4. Grafik sebaran kunci musik (*key*) yang digunakan pada daftar lagu terpopuler

Pada Gambar 4 grafik menunjukkan ada 12 kunci tangga nada kromatik pada sumbu x dari 0 hingga 11 ini mewakili nada A hingga G#. Dari grafik pada Gambar 4 terlihat bahwa kunci tangga nada yang paling banyak dipakai pada daftar lagu populer adalah kunci nada 1 atau A# yang berjumlah 731 data. Untuk data mode atau penggunaan tangga nada minor (mode bernilai 0) dan mayor (mode bernilai 1) yang paling banyak dipakai pada daftar lagu populer adalah tangga nada mayor (mode bernilai 1) dimana datanya berjumlah 2705 data seperti yang ditunjukkan pada grafik Gambar 5.

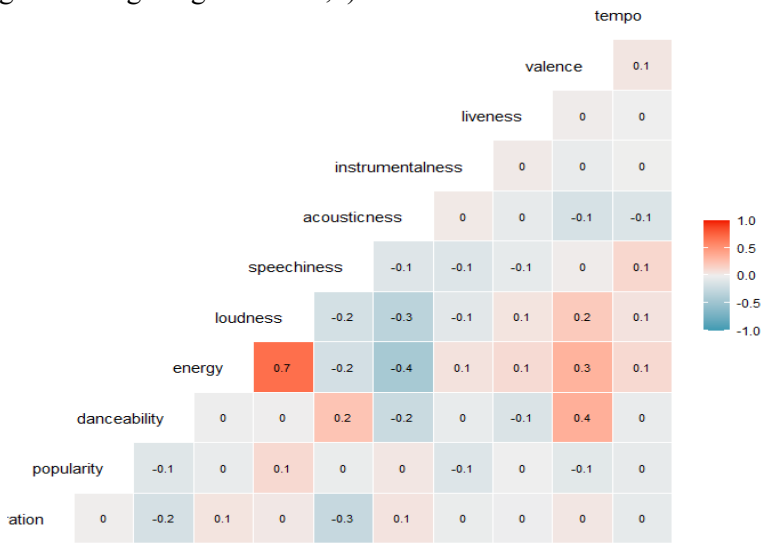



Gambar 5. Grafik penggunaan tangga nada mayor atau minor (*mode*) pada daftar lagu terpopuler

Selanjutnya dilakukan pengujian hubungan atau korelasi antara setiap atribut numerik seperti yang ditunjukkan pada Gambar 6. Memang ada beberapa atribut yang memiliki korelasi sedang

*p-ISSN: 2460-173X*
*e-ISSN: 2598-5841*

dan kuat satu sama lain. Atribut yang terkait secara positif adalah *loudness* & *energy* (berkorelasi positif kuat dengan nilai 0,7), *valence* dan *danceability* (berkorelasi positif sedang dengan nilai 0,4), atribut *valence* dan *energy* (berkorelasi positif sedang dengan nilai 0,3) dan atribut *valence* dan *loudness* (berkorelasi positif sedang dengan nilai 0,2). Variabel terkait negatif adalah atribut *energy* dan *acousticness* (berkorelasi negatif kuat dengan nilai -0,4), atribut *speechiness* dan *duration* (berkorelasi negatif sedang dengan nilai -0,3 kuat) dan atribut *acouticness* dan *loudness* (berkorelasi negatif sedang dengan nilai -0,3).




Gambar 6. Grafik korelasi antar atribut yang bernilai numerik

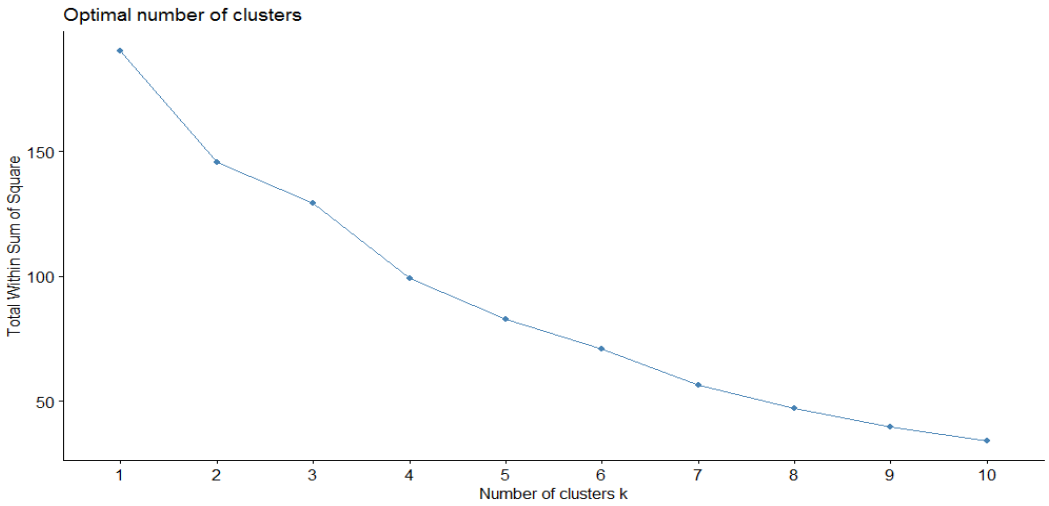
Untuk mempersiapkan dataset yang berisi nilai numerik yang digunakan dalam algoritma K-Means perlu dilakukan *feature scalling.* Hal ini dibutuhkan karena akan diterapkannya algoritma K-Means yang menggunakan perhitungan *ecludian distance*. Nilai *ecludian distance* antara atribut duration dengan atribut lainnya memiliki jarak yang sangat jauh maka akan menimbulkan masalah pada pemodelan K-Means, karena itu dilakukan *feature scalling,* yaitu teknik membuat data numerik pada dataset memiliki rentang nilai (*scale*) yang sama, sehingga tidak ada atribut data yang mendominasi atribut data yang lain istilah sederhananya adalah menormalisasi nilai dari tiap atribut pada dataset. Setelah melakukan *feature scalling*, dilakukan pengurangan jumlah pengamatan dari dataset dengan memilih secara acak lagu yang akan disimpan untuk analisis lebih lanjut. Tujuan utama dari melakukan hal ini adalah untuk mengurangi waktu komputasi pada tahap selanjutnya, terutama pada proses mencari jumlah K yang optimal untuk pengelompokan.

Pada dataset yang digunakan *track_id, track_name, artist_id, artist_name, album_id, release_date, popularity, playlist_id* dan *playlist_name* yang tidak sesuai dalam analisis klaster dihapus. Selanjutnya untuk memvisualisasikan dataset, diputuskan juga untuk menghapus kolom *time_signature* yang memiliki perbedaan nilai yang rendah yang hanya berisi signature_time 3 dan 4. Kemudian semua baris dengan nilai nol dihapus.

## 4. PEMBAHASAN

Setelah data dibersihkan, dilakukan normalisasi semua nilai non-kategoris untuk memastikan semua atribut memiliki kepentingan yang sama ketika jarak dihitung. Langkah terakhir pada tahap pra-pemrosesan data adalah membuat *dummy* atribut yaitu menambah kolom baru untuk kategori *genre*, karena kategori *genre* tidak disediakan oleh TikTok sehingga perlu dikumpulkan secara manual dan dimasukkan dalam kumpulan data. *Genre* dikeluarkan dari *cluster* analisis dan disimpan untuk dibandingkan dengan *cluster* yang dihasilkan. Kemudian, dilanjutkan untuk menentukan jumlah optimal *cluster* untuk algoritma k-means menggunakan bahasa pemrograman R. Untuk membentuk kelompok data, dilakukan pemilihan nilai K yang optimum untuk jumlah cluster seperti yang ditunjukkan Gambar 7. Untuk penentuan nilai K dilihat dari garis tertinggi pertama pada grafik Gambar 7 dimana nilai K adalah 2.

*p-ISSN: 2460-173X*
*e-ISSN: 2598-5841*




Gambar 7. Grafik untuk penentuan nilai K yang optimum

Setelah menentukan nilai K, maka dataset dibagi menjadi 2 klaster. Dari hasil klastering yang ditunjukkan pada Gambar 8 sedangkan grafik visualisasi pengelompokkan datanya diperlihatkan pada Gambar 9. Pada Gambar 8 memperlihatkan tiga bagian, dimana bagian **pertama** menunjukkan ukuran atau jumlah data pada tiap klaster. Pada klaster pertama ada 1846 data sedangkan pada klaster kedua ada 2876 data. Untuk bagian **kedua** memperlihatkan nilai rata-rata (*centroid*) dari tiap klaster. Di bagian **ketiga** yaitu berupa *clustering vector* dimana menunjukkan *vector* berisi angka 1 sampai dengan 2 sesuai dengan jumlah klaster yang ditentukan di awal. Hasil *vector* menunjukkan dimulai dari angka 1 ini artinya data pertama dari dataset dikelompokkan pada klaster pertama. Dari gambar juga terlihat isi vector pada baris kedua atau data ke-4783 bernilai 2 yang artinya data ke-4783 dari dataset masuk pada klaster kedua, dan seterusnya hingga data terakhir.

Jika dianalisis lebih lanjut hasil pengelompokkan data (*data clustering*) pada Gambar 8 bagian kedua menunjukkan karakteristik atribut *audio* pada kedua klaster. Pada klaster pertama berisi data lagu atau *track* musik yang memiliki nilai rendah (bernilai negatif) pada atribut *danceability, loudness, speechiness, valence* dan *tempo.* Data pada klaster pertama memiliki nilai *instrumentalness* yang tinggi (bernilai positif) dan dengan durasi pemutaran (*duration_minutes*) yang lama (bernilai positif). Untuk klaster kedua nilai rendah hanya ada pada atribut *instrumentalness.* Untuk atribut *danceability, loudness, speechiness, valence* dan *tempo* pada klaster kedua memiliki nilai yang tinggi. Sedangkan durasi pemutaran data lagu atau *track* musik pada klaster kedua sangat cepat (dikarenakan nilai *duration_minutes* bernilai negatif).

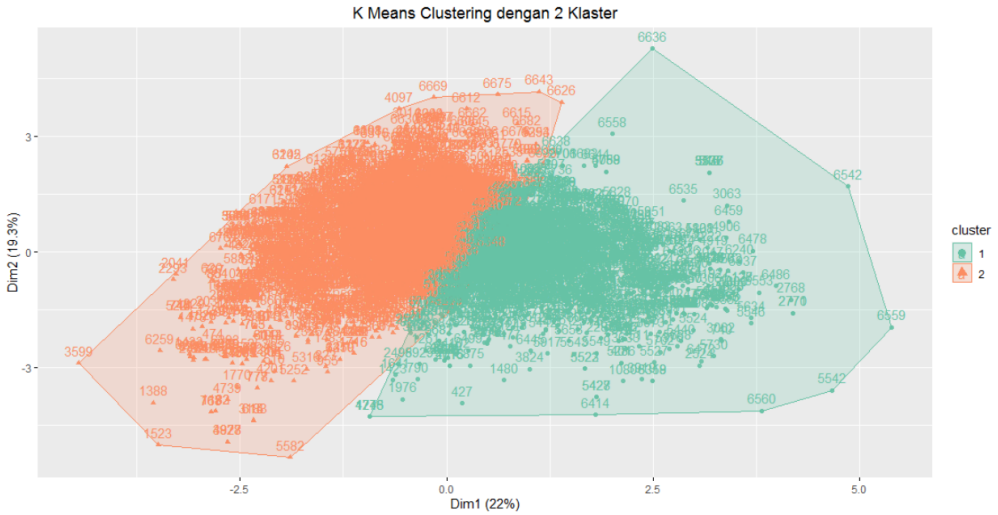
```
K-means clustering with 2 clusters of sizes 1846, 2876  (1)

Cluster means:  (2)
  danceability   loudness speechiness instrumentalness    valence      tempo duration_mins
1   -0.7606634 -0.2461131  -0.3909522       0.11601146 -0.7663664 -0.1565204     0.3096607
2    0.4882422  0.1579711   0.2509380      -0.07446354  0.4919028  0.1004647    -0.1987599

Clustering vector:  (3)
 381 3161 3263 5479 2497 3685 1148 4212 5945 1889 2685 5136 4506 1378 2407 2420 4646 3606
   1    1    1    1    2    2    2    2    2    1    1    2    1    2    1    2    1    2
4783 3622 5038 2826 1153 5179 5929 3691 1867 3281 6238 2343 6409 4669 5972 1212 4225 6641
   2    1    2    2    2    1    1    1    1    1    1    2    1    1    2    2    2    2
 875 2219 5804 5216 5548 4046 3294 5231 5927 1392 2058 2215 1331 1579 1841 3959 1697  827
   2    2    2    2    1    2    2    2    2    2    1    1    1    1    2    1    2    1
1539 3999 1415 3102 4328 6424 4523 2976 2392 3046 2977 1638 4639 2754 2189 3824 6456 4418
   2    2    2    1    1    2    1    2    2    1    1    1    2    2    2    1    1    2
4170 5717 5170 5564  611 3065 3997 6132 6552  252 3852 4887 1658 2004 4885 6040 1397 2385
   2    2    2    2    1    2    1    2    1    2    2    2    2    1    2    1    2    1
2984 6033 2592 3443  834  201 5133 2177 2590  273 2402 3795 4551 6452 4663   77 3556 5554
   1    2    2    1    2    2    2    1    2    2    2    2    2    2    2    1    2    1
5356  538 1586 6410  250 6079 4817 1332 5571 2630 2603 3132 3868 2335  189 6591 6341 3653
   1    2    2    2    2    2    1    2    2    2    2    1    2    2    1    2    1    2
 675 1575 5693 4886 3291 3837  108 3123  281 3062 4165 4450  576  949 6002  810 4814 6276
   2    1    1    1    2    2    2    2    1    1    1    1    2    2    2    2    1    2
 287  130 1314 3334 6113  914 1117 4001 5380 5561 5196  126 4601 5371 3745 3163 1063  588
   2    2    2    2    2    1    1    1    2    2    2    2    1    1    2    2    1    2
1072  176 4671 5009 5643 2877 2744 6664 5425 5215 2145 6297 4292 3014 3982 1897 4715 6064
   2    1    1    2    1    1    1    2    2    2    2    2    2    2    2    2    1    2
4432 1221 2285  811  710 1955 5499 6512 2867 1329 6320 4329 1954  782 3930  788 5165 2421
   2    2    2    2    1    1    2    1    1    2    2    2    2    1    2    2    2    1
```

Gambar 8. Hasil pengelompokan data (*data clustering*) menggunakan algoritma K-Means

*p-ISSN: 2460-173X*
*e-ISSN: 2598-5841*




Gambar 9. Grafik hasil pengelompokan data (*data clustering*) menggunakan algoritma K-Means

## 5. KESIMPULAN

Dari 6746 data TikTok [17] menunjukkan bahwa ada 4722 data yang memiliki nilai popularitas lebih dari 50 dan ada 2024 data yang nilai popularitasnya kurang dari 50. Data lagu terpopuler ini selanjutnya dikelompokan menjadi dua kelompok data. Klaster pertama berisi 1846 data sedangkan pada klaster kedua ada 2876 data. Dari hasil analisis pada klaster pertama dan kedua menunjukkan karakteristik nilai untuk masing-masing atribut audio. Pada klaster pertama berisi data lagu atau track musik yang memiliki nilai rendah (bernilai negatif) pada atribut danceability, loudness, speechiness, valence dan tempo. Data pada klaster pertama memiliki nilai instrumentalness yang tinggi (bernilai positif) dan dengan durasi pemutaran (duration_minutes) yang lama (bernilai positif). Untuk klaster kedua nilai rendah hanya ada pada atribut instrumentalness. Untuk atribut danceability, loudness, speechiness, valence dan tempo pada klaster kedua memiliki nilai yang tinggi. Sedangkan durasi pemutaran data lagu atau track musik pada klaster kedua sangat cepat (dikarenakan nilai duration_minutes bernilai negatif). Dari hasil klaster dapat diketahui bahwa terdapat beberapa atribut yang membuat lagu atau *track* musik pada aplikasi TikTok menjadi *trending* yaitu diantaranya atribut *instrumentalness* dengan nilai yang tinggi, durasi pemutaran yang lama, *danceability, loudness, speechiness, valence* dan *tempo* yang juga memiliki nilai yang tinggi.

## DAFTAR PUSTAKA

[1] A. Sherman, "TikTok reveals detailed user numbers for the first time," *CNBC*, 2020. https://www.cnbc.com/2020/08/24/tiktok-reveals-us-global-user-growth-numbers-for-first-time.html (accessed Apr. 30, 2021).

[2] L. Ceci, "Distribution of TikTok users in the United States," *Statista*, 2021. https://www.statista.com/statistics/1095186/tiktok-us-users-age/ (accessed Apr. 30, 2021).

[3] C. Chapple, "TikTok Crosses 2 Billion Downloads After Best Quarter For Any App Ever," *SensorTower*, 2020. https://sensortower.com/blog/tiktok-downloads-2-billion (accessed May 23, 2020).

[4] C. Beer, "Is TikTok Setting the Scene for Music on Social Media?," *GWI*, 2019. https://blog.gwi.com/trends/tiktok-music-social-media/ (accessed May 23, 2020).

[5] M. Kundu, P. K. Kundu, and S. K. Damarla, *Chemometric Monitoring: Product Quality Assessment, Process Fault Detection, and Applications 1st Edition*. Boca Raton: CRC Press, 2017.

[6] N. P. E. Merliana, Ernawati, and A. J. Santoso, "Analisa Penentuan Jumlah Cluster Terbaik pada Metode K-Means Clustering," 2015.

[7] A. R. Zaidah, C. I. Septiarani, M. S. Nisa, A. Yusuf, and N. Wahyudi, "Komparasi Algoritma

K-Means, K-Medoid, Agglomeartive Clustering Terhadap Genre Spotify," *J. Ilm. Ilmu Komput. Fak. Ilmu Komput. Univ. Al Asyariah Mandar*, vol. 7, no. 1, pp. 49–54, 2021.

[8] D. R. Agustian and B. A. Darmawan, "Analisis Clustering Demam Berdarah Dengue Dengan Algoritma K-Medoids (Studi Kasus Kabupaten Karawang)," *JIKO (Jurnal Inform. dan Komputer)*, vol. 6, no. 1, pp. 18–26, 2022, doi: 10.26798/jiko.v6i1.504.

[9] F. Farahdinna, I. Nurdiansyah, A. Suryani, and A. Wibowo, "Perbandingan Algoritma K-Means Dan K-Medoids Dalam Klasterisasi Produk Asuransi Perusahaan Nasional," *J. Ilm. FIFO*, vol. 11, no. 2, pp. 208–214, 2019, doi: 10.22441/fifo.2019.v11i2.010.

[10] I. Parlina, A. P. Windarto, A. Wanto, and M. R. Lubis, "Memanfaatkan Algoritma K-Means Dalam Menentukan Pegawai Yang Layak Mengikuti Asessment Center. Memanfaatkan Algoritma K-Means Dalam Menentukan Pegawai Yang Layak Mengikuti Asessment Center Untuk Clustering Program SDP," *CESS (Journal Comput. Eng. Syst. Sci.*, vol. 3, no. 1, pp. 87–93, 2018, doi: 10.24114/cess.v3i1.8192.

[11] S. K. Majhi and S. Biswal, "Optimal cluster analysis using hybrid K-Means and Ant Lion Optimizer," *Karbala Int. J. Mod. Sci.*, vol. 4, no. 4, pp. 347–360, 2018, doi: 10.1016/j.kijoms.2018.09.001.

[12] K. G. Soni and A. Patel, "Comparative Analysis of K-means and K-medoids Algorithm on IRIS Data," *Int. J. Comput. Intell. Res.*, vol. 13, no. 5, pp. 899–906, 2017.

[13] X. Zhou *et al.*, "An automatic k-means clustering algorithm of gps data combining a novel niche genetic algorithm with noise and density," *ISPRS Int. J. Geo-Information*, vol. 6, no. 12, p. 392, 2017, doi: 10.3390/ijgi6120392.

[14] J. Heil, V. Häring, B. Marschner, and B. Stumpe, "Advantages of fuzzy k-means over k-means clustering in the classification of diffuse reflectance soil spectra: A case study with west African soils," *Geoderma*, vol. 337, pp. 11–21, 2019, doi: 10.1016/j.geoderma.2018.09.004.

[15] S. Hissong, "TikTok Is Giving a Niche Indie Band's 2008 Music Millions of New Streams," *RollingStone*, 2020. https://www.rollingstone.com/pro/news/mother-mother-viral-tiktok-charts-1079256/.

[16] J. J. Verbeek, N. A. Vlassis, and B. J. A. Kröse, "Segments algorithm to and principal curves," *Pattern Recognit. Lett.*, vol. 23, no. 8, pp. 1009–1017, 2002, doi: 10.1016/S0167-8655(02)00032-6.

[17] Y. Peleg, "TikTok Trending Tracks," *kaggle*, 2021. https://www.kaggle.com/datasets/yamqwe/tiktok-trending-tracks.

**Biodata Penulis**

***Karina Auliasari***, Merupakan dosen dan kepala Laboratorium Multimedia dan Pengolahan Citra Digital pada program studi Teknik Informatika S1 Institut Teknologi Nasional Malang. Minat bidang keilmuan dan keahlian penulis diantaranya *image processing, ui/ux design* dan *recommender system.* Selain mempublikasikan artikel ilmiah penulis juga menulis buku dengan judul Desain User Interface Menggunakan FIGMA.

***Mariza Kertaningtyas***, Merupakan dosen pada program studi Teknik Industri D3 Institut Teknologi Nasional Malang. Minat bidang keilmuan dan keahlian penulis diantaranya *decision making* dan *manajemen sains*. Selain aktif mengajar dan meneliti penulis juga aktif sebagai praktisi kesehatan yoga. Penulis banyak berperan dalam kegiatan pengabdian masyarakat dan kegiatan sosial untuk membiasakan pola hidup sehat.