



**LAPORAN KECERDASAN KOMPUTASIONAL – IF184503**

## **FINAL PROJECT KECERDASAN KOMPUTASIONAL**

**Kelompok 6 :**

**Alie Husaini R.  
05111840000097**

**Ammar Alifian  
0511184000007**

**Nodas Uziel Putra Serpara  
05111840007007**

## **ABSTRAK**

Dataset yang digunakan yaitu dataset lagu-lagu di Spotify yang dirilis selama abad terakhir yang diunduh dari situs Kaggle. Dataset berisi data umum dari lagu, seperti artist, judul lagu, dan tanggal rilis, serta data karakteristik lagu yang ditentukan sendiri oleh Spotify. Terhadap dataset dilakukan eksplorasi data, seleksi fitur, normalisasi data, ekstraksi fitur, clustering, serta klasifikasi. Klasifikasi bertujuan untuk memprediksi era setiap lagu dalam satuan dekade. Digunakan beberapa model yang berbeda dalam uji coba klasifikasi. Hasil dari setiap uji coba dibandingkan untuk menentukan model yang paling efektif.

**Kata Kunci:** Spotify, Eksplorasi Data, Clustering, Klasifikasi

## DAFTAR ISI

ABSTRAK	ii
DAFTAR ISI	iii
Bab 1.	1
PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Perumusan Masalah	1
Bab 2.	2
DESAIN DAN IMPLEMENTASI	2
2.1. Persiapan Data	2
2.2. Skenario Uji Coba	2
Bab 3.	3
HASIL UJI COBA DAN DISKUSI	3
3.1. Eksplorasi Data	3
3.2. Hasil Uji Coba Skenario	3
3.3. Pembahasan dan Kesimpulan	3
DAFTAR PUSTAKA	4

## **Bab 1.**

### **PENDAHULUAN**

#### **1.1. Latar Belakang**

Laporan ini dibuat untuk memenuhi tugas akhir mata kuliah Kecerdasan Komputasional. Dataset yang digunakan berasal dari website Kaggle tentang data lagu di Spotify dari tahun 1921-2020, Dengan target popularitas suatu lagu. Selain itu, eksplorasi data akan dilakukan untuk membantu menganalisis data.

#### **1.2. Perumusan Masalah**

Masalah yang dirumuskan adalah sebagai berikut :

1. Artis terpopuler tiap dekade?
2. Fitur apa yang sangat mempengaruhi popularitas suatu lagu?
3. Secara garis besar, lagu di Spotify dapat digolongkan menjadi berapa kelompok?
4. Model klasifikasi apa yang paling bagus untuk mengklasifikasi lagu berdasarkan era masing-masing?

## **Bab 2.**

### **DESAIN DAN IMPLEMENTASI**

#### **2.1. Persiapan Data**

Untuk Final Project ini, kami menggunakan dataset dari Kaggle untuk menunjukkan data mendetail dari lagu-lagu yang ada di Spotify. Data-data yang ada dalam dataset tersebut adalah :

- Data numerical, terdiri dari :
  - acousticness (Representasi dari seberapa akustik lagu tersebut, bernilai dari 0 hingga 1)
  - danceability (Kelayakan lagu tersebut untuk digunakan sebagai pengiring tarian, bernilai dari 0 hingga 1)
  - energy (Energi yang terpancar dari lagu tersebut, bernilai dari 0 hingga 1)
  - duration\_ms (Durasi lagu dalam milisecond)
  - instrumentalness (Instrumentalitas dalam lagu tersebut, bernilai dari 0 hingga 1)
  - valence (Positivitas yang terpancar dari lagu tersebut, bernilai dari 0 hingga 1)
  - popularity (Seberapa populer lagu tersebut, bernilai dari 0 hingga 100)
  - tempo (Tempo lagu dalam BPM)
  - liveness (Keberadaan penonton, bernilai dari 0 hingga 1)
  - loudness (Kerasnya lagu rata-rata dalam dB)
  - speechiness (Frekuensi lirik dalam lagu, bernilai dari 0 hingga 1)
  - year (Tahun rilis lagu, berkisar dari 1921 hingga 2020)
- Dummy, terdiri dari :
  - mode (Tipe kunci lagu, 0 = Minor, 1 = Major)
  - explicit (Sifat suatu lagu, apakah eksplisit (mengandung umpatan, unsur seksual, narkoba, dsb) atau tidak. 0 = Tidak eksplisit, 1 = Eksplisit)
- Categorical, terdiri dari :

- key (Kunci lagu yang dikodekan dengan integer, berkisar dari 0 - 11, kunci C = 0, C# = 1, dan seterusnya)
- artists (Daftar artis dari lagu tersebut)
- release\_date (Tanggal rilis lagu, dalam format yyyy-mm-dd)
- name (Judul lagu)

## **2.2. Skenario Uji Coba**

Untuk penyelesaian masalah nomor 1, digunakan pencarian dengan cara mencari lagu terpopuler dari masing-masing dekade, lalu kami mengambil nama artist dari lagu tersebut.

Untuk penyelesaian masalah nomor 2, digunakan *library* SKLearn dan menggunakan fungsi SelectKBest pada SKLearn untuk menentukan keberpengaruhannya suatu aspek pada popularitas lagu.

Untuk penyelesaian masalah nomor 3, digunakan Principal Component Analysis (PCA) dan K-means untuk membagi tiap Playlist (Cluster) berdasarkan keunggulan fitur audio masing-masing Playlist.

Untuk penyelesaian masalah nomor 4, digunakan model SVM, KNN, MLP, dan Decision Tree dalam mengklasifikasi. dan dibandingkan hasilnya untuk mencari klasifikasi yang terbaik untuk dataset Spotify.

## Bab 3.

# HASIL UJI COBA DAN DISKUSI

### 3.1. Eksplorasi Data

Eksplorasi data dilakukan untuk menganalisis data. Analisis dibantu dengan pertanyaan yang telah dirumuskan pada sub-bab Rumusan Masalah. Sub-bab ini akan mengeksplorasi data berdasarkan dan digunakan untuk menjawab pertanyaan yang telah dirumuskan sebelumnya.

#### 1. Artis terpopuler tiap dekade?

Dalam menjawab masalah ini, dilakukan pencarian artis dengan nilai popularitas tertinggi pada tiap 1 dekade dari 1921-2020. Dan hasil yang didapat adalah :

Nama Artis	Popularitas	Tahun
Louis Armstrong	52	1921-1930
Billie Holiday, Eddie Heywood	64	1930-1939
Bing Crosby, Ken Darby Singers, John Scott Trotter & His Orchestra	76	1940-1949
Dean Martin	81	1950-1959
Brenda Lee	85	1960-1969

Fleetwood Mac	89	1970-1979
AC/DC	84	1980-1989
The Police	84	1980-1989
a-ha	84	1980-1989
Mariah Carey	88	1990-1999
Coldplay	84	2000-2009
Linkin Park	84	2000-2009
Eminem, Nate Dogg	84	2000-2009
Harry Styles	94	2010-2019

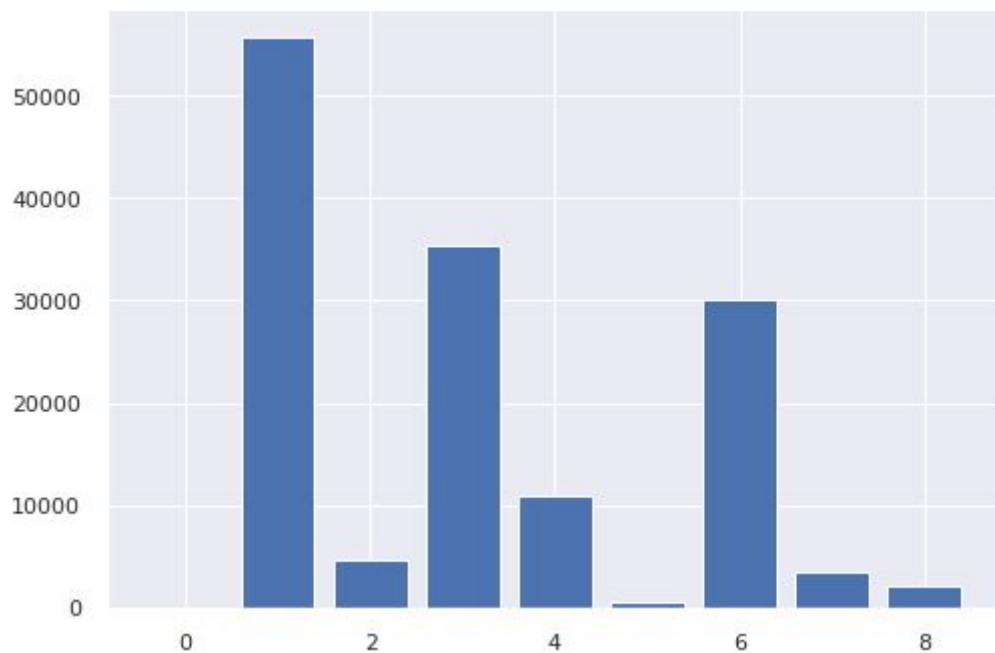
Tabel 1. Artis terpopuler tiap dekade

Dapat disimpulkan bahwa terdapat beberapa nama artis yang memiliki popularitas tinggi berdasarkan lagu yang dinyanyikan bersama atau lagu yang diciptakan berdasarkan kolaborasi di antara 2 atau lebih artis. seperti Billie Holiday dan Eddie Heywood dengan judul lagu *Wherever you are (Live)*.

## 2. Fitur apa yang sangat mempengaruhi popularitas suatu lagu?

Dalam menjawab masalah ini, dilakukan menggunakan Sklearn dan SelectKBest, membuat Data Train dan Data Test, lalu membuat function yang mengecek seberapa pengaruh tiap fitur lagu yang ada. Jika divisualisasikan maka akan menjadi seperti Gambar 2.





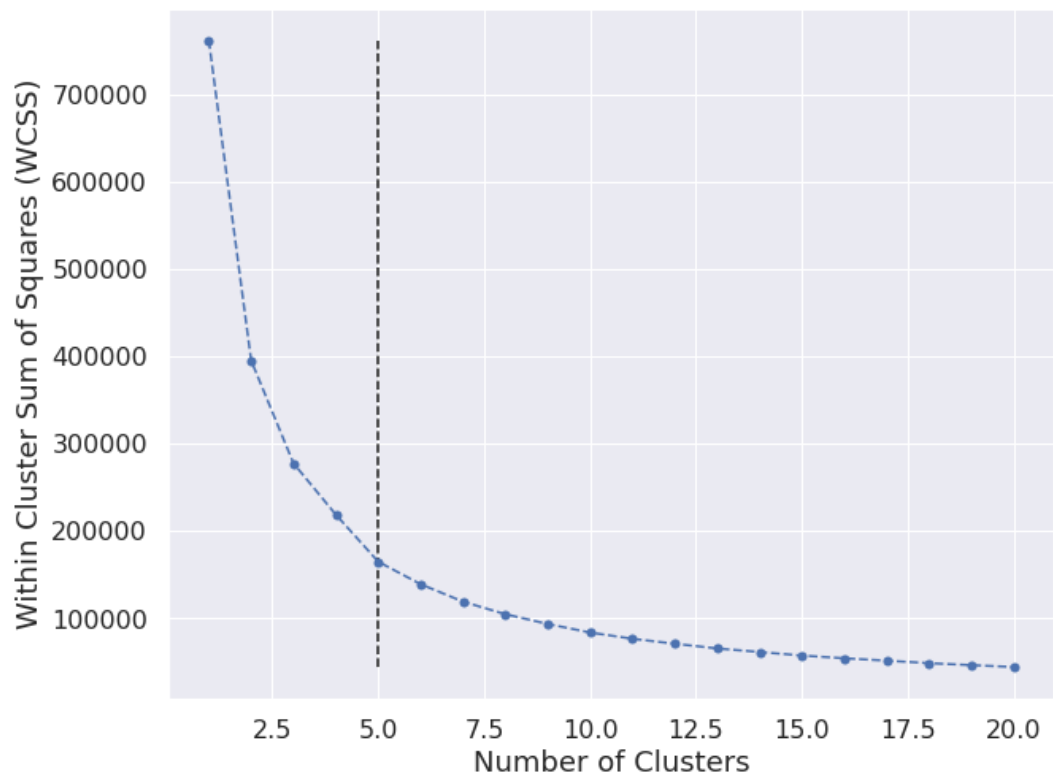
Gambar 2. Histogram Fitur Audio yang Mempengaruhi Nilai Popularitas Suatu Lagu

Dalam Gambar 2, sumbu Y sebagai Nilai Korelasi dan sumbu X sebagai Fitur audio. Fitur audio 0-8 adalah *valence* (0), *acousticness* (1), *danceability* (2), *energy* (3), *instrumentalness* (4), *liveness* (5), *loudness* (6), *speechiness* (8), dan *tempo*(9). Sehingga dapat disimpulkan bahwa fitur audio yang paling mempengaruhi popularitas suatu lagu adalah *acousticness*, dan dua fitur lainnya adalah *energy* dan *loudness*.

### 3.2. Hasil Uji Coba Skenario

3. Secara garis besar, lagu di Spotify dapat digolongkan menjadi berapa kelompok?

Dalam menjawab masalah ini, digunakan Standard Scaler dan Principal Component Analysis (PCA), mencari optimal cluster yang terbaik, lalu divisualisasikan menjadi :

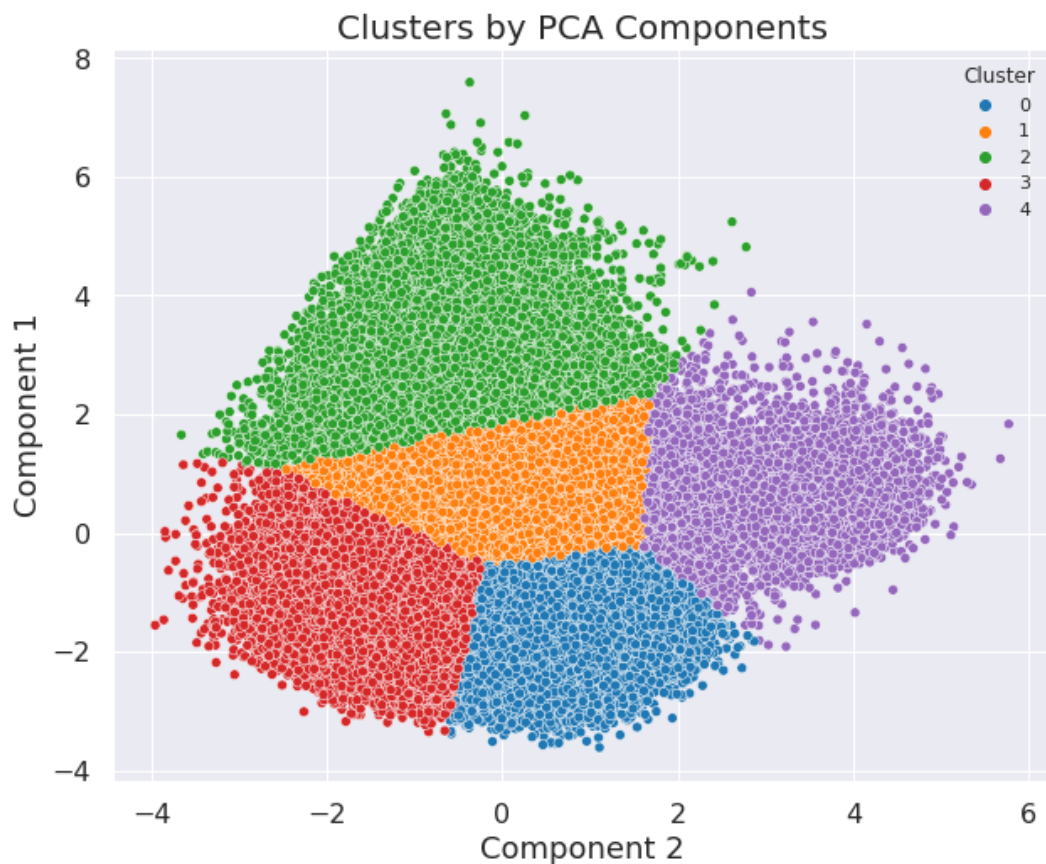


Gambar 3. Visualisasi Cluster yang Optimal

Setelah itu memakai K-means untuk mengcluster lagu-lagu yang ada berdasarkan fitur audionya dan jika dijabarkan ke dalam tabel dan scatterplot menjadi :

	valence	acousticness	danceability	energy	instrumentalness	liveness	loudness	speechiness	tempo	Component 1	Component 2	Cluster
0	0.0594	0.982	0.279	0.211	0.878000	0.665	-20.096	0.0366	80.954	3.638700	-1.207302	2
1	0.9630	0.732	0.819	0.341	0.000000	0.160	-12.441	0.4150	60.936	-0.250806	3.368148	4
2	0.0394	0.961	0.328	0.166	0.913000	0.101	-14.850	0.0339	110.339	3.188101	-1.213979	2
3	0.1650	0.967	0.275	0.309	0.000028	0.381	-9.316	0.0354	100.109	1.489517	-1.190830	1
4	0.2530	0.957	0.418	0.193	0.000002	0.229	-10.096	0.0380	101.665	1.429597	-0.384092	1

Tabel 2. Pembagian Lagu berdasarkan Fitur Audio



Gambar 4. Scatter Plot Kelompok Lagu(Cluster)

	valence	acousticness	danceability	energy	instrumentalness	liveness	loudness	speechiness	tempo	Component 1	Component 2
Cluster											
0	0.747830	0.274110	0.699702	0.653766	0.030027	0.184995	-8.191141	0.100244	118.629215	-1.549457	0.517091
1	0.519472	0.731202	0.530495	0.340465	0.159562	0.201118	-12.555180	0.058100	113.669836	0.664586	0.149136
2	0.232189	0.911841	0.334032	0.158629	0.525554	0.179245	-19.117041	0.047170	100.058107	2.909646	-0.505705
3	0.469783	0.136713	0.459646	0.757666	0.098398	0.248745	-6.996387	0.063477	133.960292	-1.287291	-1.180112
4	0.624120	0.659276	0.703554	0.241589	0.067013	0.247415	-17.113946	0.635136	106.960114	0.575027	3.103868

Tabel 3. Fitur Audio pada masing-masing kelompok(Cluster)

Dapat disimpulkan bahwa, pembagian kelompok lagu berdasarkan fitur audio dapat dilakukan tetapi perbedaan tiap kelompok dalam fitur audio masing-masing tidak beda jauh. Dan tiap kelompok memiliki satu nilai fitur audio yang paling tinggi dari antar lain (Table 3).

#### 4. Model klasifikasi apa yang paling bagus untuk mengklasifikasi lagu berdasarkan era masing-masing?

Dalam menjawab masalah ini, digunakan klasifikasi model SVM, MLP, Decision Tree, dan KNN. tetapi dalam percobaan SVM terdapat kendala waktu yang sangat lama dalam mencari akurasi, walaupun sudah menggunakan 5% dari

dataset Spotify, proses pencarian yang dilakukan berlangsung sampai 1 setengah jam lebih, dan tidak membuahkan hasil apapun, sehingga SVM bukan model yang cocok untuk mengklasifikasi data ini. Dan dari percobaan model Decision tree, MLP, dan KNN, klasifikasi terbaik dalam dataset Spotify adalah Model Decision Tree dan MLP. Berikut gambar uji coba yang dilakukan :

Classifier	Test Size	Criterion	Splitter	Max Depth	Accuracy
Decision Tree	30%	Gini	Best	3	0.303
				5	0.306
				7	0.327
				9	0.348
		Random	7	0.3	
			9	0.317	
	Entropy	Best	7	0.328	
			9	0.347	
	50%	Gini		Best	0.344
		Entropy			0.339
	70%	Gini			0.338
		Entropy			0.336

Tabel 4. Hasil percobaan dengan menggunakan Decision Tree

Classifier	Test Size	Weights	N Neighbors	Accuracy
K-Nearest Neighbor	30%	Uniform	1	0.232
			11	0.248
			21	0.249
			31	0.248
		Distance	21	0.265
			31	0.265
	50%	Distance	21	0.283
			31	0.283
	70%	Distance	21	0.283
			31	0.283

Tabel 5. Hasil percobaan dengan menggunakan KNN

Classifier	Test Size	Solver	Activation	Alpha	Accuracy
Multilayer Perceptron	30%	Adam	ReLU	0.0001	0.345
				0.001	0.352
				0.01	0.333
				0.1	0.328
				1	0.302
			Identity	0.0001	0.313
				0.001	0.31
			Logistic	0.0001	0.348
				0.001	0.341
			Tanh	0.0001	0.34
				0.001	0.338
	50%	SGD	ReLU	0.0001	0.256
			Logistic		0.268
	70%	Adam	ReLU		0.338
			Logistic		0.329
			ReLU		0.335
			Logistic		0.328

Tabel 6. Hasil percobaan dengan menggunakan MLP

Dari hasil uji coba skenario 4, rata-rata akurasinya 35% ke bawah, bisa jadi dikarenakan rentang era yang kami gunakan adalah per dekade sehingga ada cukup banyak keberagaman musik untuk setiap era.

### 3.3. Kesimpulan

Dari ujicoba diatas, kami menyimpulkan bahwa :

1. Artis terpopuler untuk tiap dekade adalah seperti yang tercantum dalam Tabel 1.
2. Fitur audio yang paling menentukan popularitas suatu lagu adalah *acousticness*, *energy*, dan *loudness*.
3. Pembagian kelompok lagu berdasarkan fitur audio dapat dilakukan tetapi perbedaan tiap kelompok dalam fitur audio masing-masing tidak beda jauh.
4. Metode klasifikasi terbaik per era adalah dengan metode Decision Tree dan MLP.

## DAFTAR PUSTAKA

Ay, Yamaç Eren (2020). Spotify Dataset 1921-2020, 160k+ Tracks.

<https://www.kaggle.com/yamaerenay/spotify-dataset-19212020-160k-tracks?select=data.csv>

Suprianto (2020). LAPORAN TUGAS AKHIR.

<https://core.ac.uk/download/pdf/35380679.pdf>