

Sistem Rekomendasi Lagu Teratas – Global Berdasarkan Properti Trek menggunakan Metode Content Based Filtering

Rolly Maulana Awangga
Jurusan Teknik Informatika
Politeknik Pos Indonesia
Bandung, Indonesia
awangga@poltekpos.ac.id

Naomi C.H Tampubolon'
Jurusan Teknik Informatika
Politeknik Pos Indonesia
Bandung, Indonesia
naomicindi.hermina@gmail.com

Spotify saat ini merupakan platform musik, podcast dan video yang banyak digemari banyak penggunanya. Spotify juga memberikan kemudahan akses musik. Platform musik spotify sangat banyak digemari penggunanya, terutama kalangan young adults yaitu berusia 18 hingga 25 tahun. (Wahyu Dewatara & Monik Agustin, 2019) Spotify menyediakan banyak fitur seperti yaitu sekarang memutar, cari, koleksi, dengarkan offline, podcast dan acara serta banyak lagi. Properti trek di Spotify termasuk genre, nama artis, dan banyak lainnya. (Fitriani, n.d.) Berdasarkan properti trek, bagaimana cara merekomendasikan lagu berdasarkan properti trek, dan bagaimana cara memastikan bahwa lagu berdasarkan properti trek disukai. Masalah ini akan menjadi sumber dalam penelitian. *Dataset* yang digunakan dalam penelitian ini adalah 50 Lagu Teratas – Global (*50 Top Songs - Global*) yang diambil dari Spotify playlist chart yang kemudian diconvert ke dalam *file* dengan *format cvs* dan kemudian nantinya akan dikelola atau diproses. Penelitian ini bertujuan untuk merekomendasikan lagu berdasarkan trek properti yang ada dan merekomendasikan berdasarkan trek properti disukai. Metode Content-Based Filtering digunakan dalam penelitian ini. *Content-Based Filtering* berfungsi merekomendasikan berdasarkan kemiripan konten seperti *genres* yang merupakan kemiripan jarak *cosinusnya* dan terdapat 5 lagu teratas yaitu *STAYING ALIVE (feat. Drake & Lil Baby)*, *Bam Bam (feat. Ed Sheeran)*, *About Damn Time*, *Bad Decisions (with BTS & Snoop Dogg)*, dan *Ferrari* yang direkomendasikan menggunakan metode *Content-Based Filtering*.

Kata Kunci: *Sistem rekomendasi, musik, spotify, content-based filtering.*

I. PENDAHULUAN

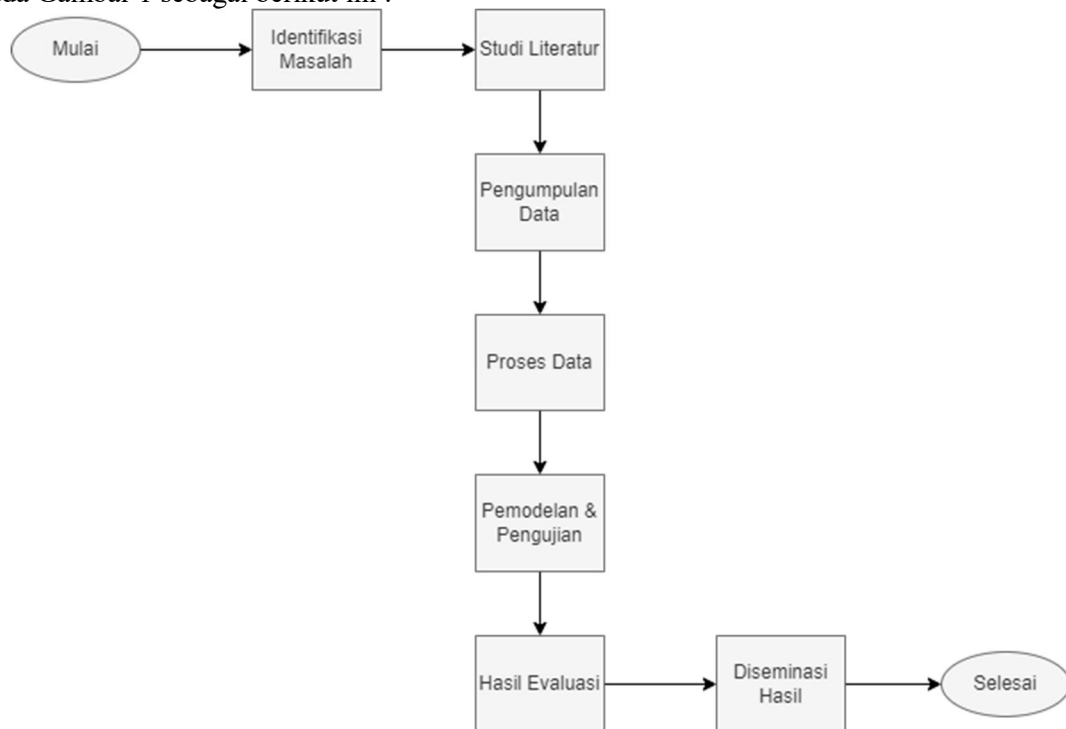
Musik terdiri dari beberapa unsur, seperti melodi, harmoni, ritme, dan timbre. Musik adalah fenomena *intuitif* yang dapat dikembangkan, ditingkatkan, dan disebarluaskan dengan berbagai cara. Musik adalah fenomena unik yang dihasilkan oleh berbagai alat musik. Musik memang menjadi salah satu hobi yang paling digemari, hal tersebut terbukti dengan manfaat yang bisa didapatkan dengan mendengarkan musik. Musik memiliki kemampuan untuk mengurangi stres. Manfaat mendengarkan musik adalah kita bisa mendengarkannya sambil melakukan hal lain. Musik adalah komponen yang tak terpisahkan dari kehidupan dan teknologi. Perkembangan yang semakin berkembang pesat juga mempengaruhi dunia industri. (Muslimin M, n.d.) Dimana pada jaman era digital saat ini tidak perlu repot dalam mendengarkan musik dimanapun dan kapanpun, dikarenakan kecanggihan teknologi tersebut. *Spotify* adalah layanan *streaming* musik terpopuler di Indonesia, telah mencapai 200 juta pengguna perbulan. (Noviani et al., 2020) Data yang digunakan pada penelitian ini adalah data lagu yang diambil dari *playlist chart spotify 50 Lagu Teratas – Global (50 Top Songs - Global)*. (Fitriani, n.d.) *Spotify* memiliki properti lagu yang berfungsi untuk mengetahui lebih mendalam mengenai lagu atau sebuah musik. Ada beberapa fitur atau properti lagu pada *spotify* yaitu, sebagai berikut: *track name, album name, artist name, duration, popularity, genres, danceability, energi, key, loudness, speechiness, acousticness, instrumentaness, liveness, tempo* dan *valence*. Berdasarkan properti trek tersebut kita dibutuhkan sistem yang dapat merekomendasikan berdasarkan properti trek yang ada. Maka dari itu bagaimana caranya untuk merekomendasikan sebuah lagu berdasarkan properti trek yang ada. Dari permasalahan tersebut dibutuhkan sebuah

sistem yang dapat merekomendasikan lagu berdasarkan properti trek yang ada. Properti trek ‘*genres*’ merupakan properti trek yang digunakan dalam rekomendasi sistem.

Dalam penelitian ini sistem rekomendasi akan dianalisa dan dibangun dengan menggunakan metode *content-based filtering* karena dengan algoritma ini berdasarkan *similarity* konten sehingga rekomendasi yang dihasilkan mempunyai kualitas yang baik. Visualisasi hasil rekomendasi tersebut akan diimplementasikan dengan *framework Jupyter Notebook*. Target luaran yang ingin dicapai pada penelitian ini adalah jurnal nasional terakreditasi.

II. METODOLOGI PENELITIAN

Metode penelitian yang digunakan dalam penelitian ini akan diwakilkan oleh flowchart pada Gambar 1 sebagai berikut ini :



Gambar 1. Diagram Alur Metodologi Penelitian

A. Identifikasi Masalah

Identifikasi masalah merupakan langkah pertama dalam tahapan metode penelitian yang berfungsi untuk mengidentifikasi masalah. Pada tahap ini peneliti melakukan penelitian dengan cara memnukan masalah (*problem*) yang ada, kemudian melakukan indentifikasi sumber permasalahan (*root cause*) sehingga menciptakan permasalahan (*problem statement*) yang menjelaskan permasalahan yang sudah diidentifikasi. Adapun Permasalahan yang terdapat pada penelitian ini adalah bagaimana cara merekomendasikan lagu berdasarkan properti trek.

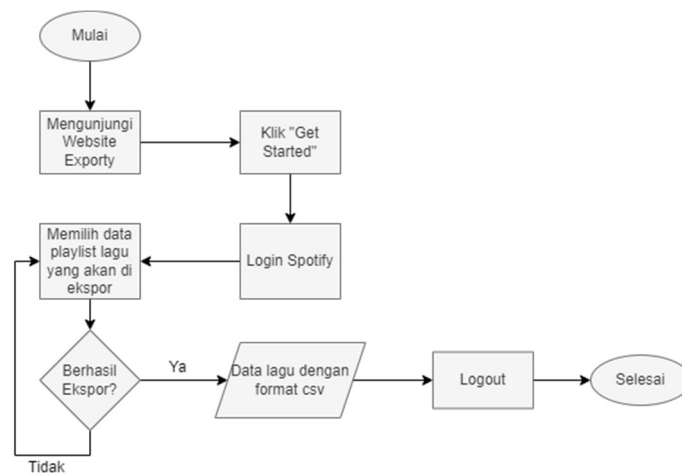
B. Studi Literatur

Setelah dilakukan tahap indentifikasi masalah langkah selanjutnya dalam tahapan metode penelitian adalah studi literatur. Pada tahap ini peneliti melakukan pencarian informasi yang relevan dengan masalah yang akan diteliti dan mengkaji beberapa teori dasar yang relevan dengan permasalahan yang akan diteliti kemudian memperdalam pengetahuan peneliti tentang permasalahan yang akan diteliti dan mengkaji hasil-hasil penelitian terdahulu yang berkaitan dengan penelitian yang akan diteliti dan mendapatkan

informasi tentang aspek-aspek mana dari suatu permasalahan yang sudah pernah diteliti untuk menghindari agar tidak terjadi *plagiarisme* penelitian.

C. Pengumpulan Data

Pengumpulan data adalah tahap setelah studi literatur. Tujuan pengumpulan data adalah untuk memperoleh data yang akan digunakan dalam penelitian. Peneliti melakukan pengumpulan studi literatur yang relevansi dan sesuai dengan apa yang dibutuhkan untuk menunjang penelitian untuk dapat melakukan teknik pengumpulan data studi literatur. Maka dari itu data yang digunakan pada penelitian ini adalah data yang di *convert* dari website <https://exportify.net/#playlists> kedalam bentuk *format csv*. Website tersebut berfungsi untuk mengekspor daftar putar *spotify* menggunakan *Web API*. Alur dari pengumpulan data dapat dilihat dari *flowchart* berikut:



Penjelasan *flowchart*, dimulai dengan cara mengunjungi website *Exportify* kemudian akan muncul tampilan awal dari website. Klik “*Get Started*”, kemudian akan dialihkan ke halaman *login* dan melakukan proses *login*. Setelah masuk *login* selanjutnya akan di tampilkan daftar *spotify* yang sudah ditambahkan kedalam koleksi *playlist* pada *spotify*. Kemudian klik *export* lalu *playlist* yang berisi lagu yang ingin diambil datanya akan terekspor menjadi data yang berbetuk *file csv*. kemudian *logout* setelah selesai melakukan *export* data.

D. Proses Data

Setelah dilakukan pengumpulan data langkah selanjutnya dalam tahap metodologi penelitian adalah proses data. Pada tahap proses data atau pengolahan data dilakukan data *cleaning*. *Cleaning* data adalah prosedur untuk memastikan keakuratan, konsistensi, dan kegunaan data dalam kumpulan data. Prosedurnya adalah mendeteksi kesalahan atau kerusakan pada data dan kemudian memperbaiki atau menghapusnya jika perlu.

E. Pemodelan dan Pengujian

Tahapan selanjutnya yaitu Pemodelan dan Pengujian. Tahap ini ialah tahap menggunakan metode rekomendasi dalam pemodelan dan pengujian data. *Content-Based Filtering* merupakan metode, model atau algoritma yang akan digunakan pada penelitian ini. Pemodelan ada tahap yang dilakukan sebelum pengujian. Dalam pemodelan dipertimbangkan dulu model yang akan dipakai dalam mengelola data. Kemudian langkah selanjutnya setelah pemodelan adalah langkah Pengujian menggunakan model *Content-Based Filtering*.

F. Hasil Evaluasi

Tahap hasil evaluasi adalah tahap performansi dari pemodelan dan pengujian. Hasil pengujian data akan diimplementasi pada tahap hasil evaluasi.

G. Diseminasi Hasil

Tahap deseminasi hasil adalah tahap untuk artikel yang diterbitkan dalam laporan penelitian dan *draft* jurnal.

III. PEMBAHASAN DAN HASIL

A. PEMBAHASAN

Penelitian ini dilakukan dengan mengambil data selama 1 minggu terakhir dari *playlist* 50 Lagu Teratas - Global (*50 Top Songs - Global*) sebanyak 50 data lagu. Untuk hasil data mentah atau data yang belum dilakukan proses data yaitu data yang ada pada *spotify*, dapat dilihat sebagai berikut:

#	JUDUL	PEMUTARAN	ALBUM	
1	Quevedo: Bzrp Music Sessions, Vol. 52 Bizarap, Quevedo	49.200.772	Quevedo: Bzrp Music Sessions, Vo...	3:18
2	As It Was Harry Styles	39.156.789	Harry's House	2:47
3	Me Porto Bonito Bad Bunny, Jhay Cortez	36.028.246	Un Verano Sin Ti	2:58
4	Titi Me Preguntó Bad Bunny	32.101.875	Un Verano Sin Ti	4:03
5	Running Up That Hill (A Deal With God) Kate Bush	30.032.752	Hounds Of Love	4:58
6	DESPECHÁ ROSALÍA	28.107.868	DESPECHÁ	2:37
7	Efecto Bad Bunny	26.359.490	Un Verano Sin Ti	3:33
8	Glimpse of Us Joji	25.333.978	Glimpse of Us	3:53
9	Ojitos Lindos Bad Bunny, Jhay Cortez	23.486.675	Un Verano Sin Ti	4:18
10	Bad Habit Steve Lacy	21.288.974	Gemini Rights	3:52
11	Moscow Mule Bad Bunny	20.845.139	Un Verano Sin Ti	4:05
12	Heat Waves Glass Animals	20.321.676	Dreamland (+ Bonus Levels)	3:58
13	La Bachata Manuel Turizo	20.180.692	La Bachata	2:42
14	PROVENZA KAROL G	19.745.401	PROVENZA	3:30
15	Bad Decisions (with BTS & Snoop Dogg) benny blanco, BTS, Snoop Dogg	19.500.644	Bad Decisions (with BTS & Snoop ...	2:52
16	I Ain't Worried OneRepublic	19.361.468	I Ain't Worried (Music From The M...	2:28

17		Left and Right (Feat. Jung Kook of BTS) Charlie Puth, Jung Kook, BTS	18.378.990	Left and Right (Feat. Jung Kook of ...	♥	2:34
18		Doja Central Cee	17.262.833	Doja	♥	1:37
19		BREAK MY SOUL Beyoncé	16.262.031	RENAISSANCE		4:38
20		About Damn Time Lizzo	15.991.437	Special		3:11
21		STAYING ALIVE (feat. Drake & Lil Baby) DJ Khaled, Drake, Lil Baby	15.991.390	STAYING ALIVE (feat. Drake & Lil ...		2:58
22		STAY (with Justin Bieber) The Kid LAROI, Justin Bieber	15.847.791	F*CK LOVE 3: OVER YOU	♥	2:21
23		Tarot Bad Bunny, Jhay Cortez	15.746.632	Un Verano Sin Ti		3:57
24		Cold Heart - PNAU Remix Elton John, Dua Lipa, PNAU	15.351.197	The Lockdown Sessions		3:22
25		Te Felicito Shakira, Rauw Alejandro	15.230.719	Te Felicito		2:52
26		SNAP Rosa Linn	15.219.493	SNAP PACK		2:59
27		Late Night Talking Harry Styles	14.395.673	Harry's House		2:57
28		Party Bad Bunny, Rauw Alejandro	14.346.290	Un Verano Sin Ti		3:47
29		I Like You (A Happier Song) (with Doja Cat) Post Malone, Doja Cat	14.133.871	Twelve Carat Toothache	♥	3:12
30		Bam Bam (feat. Ed Sheeran) Camila Cabello, Ed Sheeran	14.016.418	Familia	♥	3:26
31		Jimmy Cooks (feat. 21 Savage) Drake, 21 Savage	13.188.576	Honestly, Nevermind		3:38
32		Under The Influence Chris Brown	13.145.783	Indigo (Extended)		3:04
33		Sunroof Nicky Youre, dazy	13.141.161	Sunroof	♥	2:43
34		Sweater Weather The Neighbourhood	12.617.763	I Love You.		4:00
35		Blinding Lights The Weeknd	12.478.897	After Hours		3:20
36		Ferrari James Hype, Miggy Dela Rosa	12.308.447	Ferrari		3:06
37		Ghost Justin Bieber	12.221.064	Justice	♥	2:33
38		La Corriente Bad Bunny, Tony Dize	12.129.312	Un Verano Sin Ti		3:18
39		One Kiss (with Dua Lipa) Calvin Harris, Dua Lipa	11.931.800	One Kiss (with Dua Lipa)	♥	3:34
40		Vegas (From the Original Motion Picture Soundtrack ...) Doja Cat	11.912.237	Vegas (From the Original Motion P...		3:02
41		Die For You The Weeknd	11.891.349	Starboy		4:20
42		Until I Found You Stephen Sanchez	11.831.698	Until I Found You	♥	2:57
43		MIDDLE OF THE NIGHT Elley Duhé	11.830.078	MIDDLE OF THE NIGHT	♥	3:04
44		First Class Jack Harlow	11.457.282	Come Home The Kids Miss You		2:53
45		Another Love Tom Odell	11.450.164	Long Way Down (Deluxe)		4:04
46		Una Noche en Medellín Cris MJ	11.162.073	Una Noche en Medellín		2:33
47		Kesariya (From "Brahmastra") Pritam, Arijit Singh, Amitabh Bhattacharya	10.934.301	Kesariya (From "Brahmastra")		4:28
48		Después de la Playa Bad Bunny	10.923.650	Un Verano Sin Ti		3:50
49		Dandelions Ruth B.	10.619.440	Safe Haven	♥	3:53
50		Shivers Ed Sheeran	10.503.440	=	♥	3:27

Kemudian data *diconvert* dalam *format file csv*.

Spotify ID,Artist IDs,Track Name,Album Name,Artist Name(s),Release Date,Duration,Popularity,Added By

2tTmW7RDtMQtBk7m2rYeSw,"716NhGYqD1j2w11Qkgq36,52iwsT98xCoGgiC4Dvkj6JhhA12EX05fT7y2e,"6KlMcVD70vtIoJWnq6nGn3","As It Was","Harry's 6Sq7ltF9Qa7SNFBsV5Cogx,"4q3ewBCX7sLwd24euuV69X,37230BxxYs9ksS70l1IHWI5LamUGEuP4ozKQ5XZ,"4q3ewBCX7sLwd24euuV69X","Tití Me Preguntó 75FEaRjZTKLhTrFGsfMUXR,"1aSxMhuvixZ8h9dK9jIDwL","Running Up That Hill 5ildQOEKmJuWGI2vRkFdYc,"7ltDVBr6mKbRvohxheJ9h1","DESPECHÁ","DESPE 5Eax0qFko2dh7RI2IYs3bx,"4q3ewBCX7sLwd24euuV69X","Efecto","Un Verano 6xGruZOHLs39ZbVccQTuPZ,"3MZsBdqDrRTJihTHQrO6Dq","Glimpse of Us","G 3k3NWokhRRkEPHCzPmV8TW,"4q3ewBCX7sLwd24euuV69X,5n9bMYfz9qss2'

Proses selanjutnya dalam penelitian ini adalah proses pemodelan dan pengujian data. Penelitian ini menggunakan *content-based filtering* dalam pemodelannya. Langkah-langkah dalam pemodelan menggunakan *content-based filtering*.

a) Import Library

```
In [1]: import pandas as pd
import numpy as np
import sklearn
import nltk
from sklearn.metrics.pairwise import cosine_distances
from sklearn.feature_extraction.text import CountVectorizer
from nltk.tokenize import word_tokenize
```

b) Menampilkan 5 dataset teratas.

```
In [2]: df = pd.read_csv("lagu_teratas_-_global.csv")
df.head()
```

Out[2]:

	Spotify ID	Artist IDs	Track Name	Album Name	Artist Name(s)	Release Date	Duration (ms)	Popularity	Added By
0	2tTmW7RDtMQtBk7m2rYeSw	716NhGYqD1j2w11Qkgq36,52iwsT98xCoGgiC4Dvkj6JhhA12EX05fT7y2e	Quevedo: Bzrp Music Sessions, Vol. 52	Quevedo: Bzrp Music Sessions, Vol. 52	Bizarrap, Quevedo	2022-07-06	198937	100	spotify
1	4Dvkj6JhhA12EX05fT7y2e	6KlMcVD70vtIoJWnq6nGn3	As It Was	Harry's House	Harry Styles	2022-05-20	167303	93	spotify
2	6Sq7ltF9Qa7SNFBsV5Cogx	4q3ewBCX7sLwd24euuV69X,37230BxxYs9ksS70l1IHWI5LamUGEuP4ozKQ5XZ	Me Porto Bonito	Un Verano Sin Ti	Bad Bunny, Jhay Cortez	2022-05-06	178567	99	spotify
3	1IHWI5LamUGEuP4ozKQ5XZ	4q3ewBCX7sLwd24euuV69X	Tití Me Preguntó	Un Verano Sin Ti	Bad Bunny	2022-05-06	243716	98	spotify
4	75FEaRjZTKLhTrFGsfMUXR	1aSxMhuvixZ8h9dK9jIDwL	Running Up That Hill (A Deal With God)	Hounds Of Love	Kate Bush	1985	298933	95	spotify

5 rows x 23 columns

c) Cleaning Data

Cleaning data adalah prosedur untuk memastikan keakuratan, konsistensi, dan kegunaan data dalam kumpulan data. Prosedurnya adalah mendeteksi kesalahan atau kerusakan pada data dan kemudian memperbaiki atau menghapusnya jika perlu. Berikut adalah proses data yang telah dilakukan *cleaning* data:

- *Drop* tabel

Drop tabel yang berfungsi untuk menghilangkan tabel yang tidak digunakan atau diperlukan dalam proses penelitian. Terdapat beberapa data yang tidak diperlukan dalam penelitian ini, dan di lakukan *drop* tabel. Dapat dilihat sebagai berikut:

```
In [4]: df=df.drop(['Spotify ID', 'Artist IDs', 'Added At', 'Added By'], axis=1)
```

```
In [5]: df
```

Out[5]:

	Track Name	Album Name	Artist Name(s)	release	duration	Popularity	Genres	Danceability	Energy	Key	Loudness	Mode	Spe
0	Quevedo: Bzrp Music Sessions, Vol. 52	Quevedo: Bzrp Music Sessions, Vol. 52	Bizarrap, Quevedo	2022-07-06	198937	100	argentine hip hop, pop, venezolano, trap, argentin...	0.621	0.782	2	-5.548	1	
1	As It Was	Harry's House	Harry Styles	2022-05-20	167303	93	pop	0.520	0.731	6	-5.338	0	
2	Me Porto Bonito	Un Verano Sin Ti	Bad Bunny, Jhay Cortez	2022-05-06	178567	99	latin, reggaeton, trap, latino	0.911	0.712	1	-5.105	0	
3	Tití Me Preguntó	Un Verano Sin Ti	Bad Bunny	2022-05-06	243716	98	latin, reggaeton, trap, latino	0.650	0.715	5	-5.198	0	
4	Running Up That Hill (A Deal With God)	Hounds Of Love	Kate Bush	1985	298933	95	art pop, art rock, baroque pop, new wave, pop, perm...	0.629	0.547	10	-13.123	0	

- Cek Data

Mengecek apakah data tersebut memiliki *missing value* atau tidak. Setelah dilakukan pengecekan data terdapat 2 data yang memiliki *missing value*: Drop tabel yang tidak digunakan.

```
In [6]: df.isnull().sum()
Out[6]:
```

Track Name	0
Album Name	0
Artist Name(s)	0
release	0
duration	0
Popularity	0
Genres	2
Danceability	0
Energy	0
Key	0
Loudness	0
Mode	0
Speechiness	0
Acousticness	0
Instrumentalness	0
Liveness	0
Valence	0
Tempo	0
Time Signature	0
dtype: int64	

Lalu dilakukan proses mengisi nilai *NA/NaN (Not a Number)* menjadi 0.

```
In [7]: df.fillna(0).head()
```

```
Out[7]:
```

	Track Name	Album Name	Artist Name(s)	release	duration	Popularity	Genres	Danceability	Energy	Key	Loudness	Mode	Speechiness	Acousticness
0	Quevedo: Bizarro Music Sessions, Vol. 52	Quevedo: Bizarro Music Sessions, Vol. 52	Bizarrap, Quevedo	2022-07-08	196937	100	argentine hip hop pop venezolano trap argentine...	0.621	0.782	2	-5.548	1	0.0440	0.01
1	As It Was	Harry's House	Harry Styles	2022-05-20	187303	93	pop	0.520	0.731	6	-5.338	0	0.0557	0.34
2	Me Porto Bonito	Un Verano Sin Ti	Bad Bunny, Jhay Cortez	2022-05-09	178567	96	latin reggaeton trap latino	0.911	0.712	1	-5.105	0	0.0817	0.01
3	Titi Me Preguntó	Un Verano Sin Ti	Bad Bunny	2022-05-09	243716	96	latin reggaeton trap latino	0.650	0.715	5	-5.198	0	0.2530	0.01
4	Running Up That Hill (That Dead Zynthy)	Hounds Of Love	Kate Bush	1985	296933	95	art pop art rock baroque pop new wave pop perm...	0.629	0.547	10	-13.123	0	0.0550	0.72

Kemudian menghapus atau menghilangkan semua entri dengan nilai *NaN* atau 0 seperti berikut:

```
In [8]: df = df.dropna(axis = 0)
df.shape
```

```
Out[8]: (48, 19)
```

- d) Pemodelan sistem rekomendasi berdasarkan konten.

Pada pemodelan menggunakan *content-based filtering* memiliki beberapa tahap dalam menampilkan rekomendasi berdasarkan konten. Dapat dilihat sebagai berikut:

- *Encode Data*

Pada tahap ini *genres* di *encode* dengan cara mengkonversi *teks* menjadi *vector* menggunakan *CountVectorizer* dengan konten yang digunakan dalam merekomendasikan yaitu kemiripan konten index 0 pada properti *trek genres*. Dengan langkah sebagai berikut:

content based filtering

encode data

[illegible]

- Mencari *similarity*.

Pada tahap ini kita menggunakan *cosine distance* dalam mencari kesamaan atau kemiripan konten dengan cara menghitung jarak *cosinus* yaitu ukuran

kesamaan antara dua vektor berdasarkan sudut kosinus di antara konten. (Usino et al., 2019) Kemudian mengembalikan nilai index bilangan bulat yang akan mengurutkan indeks. Dari hasil mengurutkan tersebut maka didapat hasil rekomendasi berdasarkan *index* 0 yaitu 5 lagu teratas yang mirip dengan konten yang sama dengan *index* 0 dan dapat dilihat sebagai berikut:

```
search
In [15]: dist = cosine_distances(code, bank)
          dist
Out[15]: array([[3.33066907e-16, 8.14304662e-01, 2.77814619e-01, 2.77814619e-01,
3.63258049e-01, 1.00000000e+00, 2.77814619e-01, 8.68693567e-01,
2.57218647e-01, 1.00000000e+00, 2.77814619e-01, 4.40107489e-01,
2.74981512e-01, 5.30223824e-01, 2.38088099e-01, 4.31426407e-01,
3.85817229e-01, 1.00000000e+00, 4.63943733e-01, 2.27951353e-01,
1.64173453e-01, 3.56732479e-01, 2.77814619e-01, 3.24668997e-01,
2.94784572e-01, 8.14304662e-01, 2.57218647e-01, 3.10641227e-01,
4.38510075e-01, 1.99986938e-01, 3.80477526e-01, 4.57917835e-01,
5.35761655e-01, 4.12112863e-01, 5.45141174e-01, 2.55930214e-01,
3.48108443e-01, 5.45141174e-01, 5.35761655e-01, 1.00000000e+00,
5.35761655e-01, 3.96490151e-01, 5.45141174e-01, 7.19255037e-01,
3.96490151e-01, 2.77814619e-01, 4.57917835e-01, 5.45141174e-01]])

In [24]: rec_idx = dist.argsort()[0, 0:6]
          rec_idx
Out[24]: array([ 0, 20, 29, 19, 14, 35], dtype=int64)

In [25]: df.loc[rec_idx]
Out[25]:
```

	Track Name	Album Name	Artist Name(s)	release	duration	Popularity	Genres	Danceability	Energy	Key	Loudness	Mode	Speechiness	Acous
0	Quevedo: R3rp Music Sessions, Vol. 52	Quevedo: R3rp Music Sessions, Vol. 52	Bizarrap, Quevedo	2022-07-06	198937	100	argentine hip hop, pop, venezolano, trap, argentin...	0.621	0.782	2	-5.548	1	0.0440	
20	STAYING ALIVE (feat. Drake & Lil Baby)	STAYING ALIVE (feat. Drake & Lil Baby)	DJ Khaled, Drake, Lil Baby	2022-06-05	178176	85	dance pop, hip hop, miami hip hop, pop, rap, ra...	0.718	0.461	4	-7.676	0	0.0797	
29	Bam Bam (feat. Ed Sheeran)	Familia	Camila Cabello, Ed Sheeran	2022-04-07	206070	84	dance pop, pop, post-teen pop, uk pop	0.756	0.667	8	-6.377	1	0.0401	
19	About Damn Time	Special	Lizzo	2022-07-15	191937	80	dance pop, escape room, minnesota hip hop, pop, tr...	0.836	0.743	10	-6.305	0	0.0953	
14	Bad Decisions (with BTS & Snoop Dogg)	Bad Decisions (with BTS & Snoop Dogg)	benry, blanco.BTS, Snoop Dogg	2022-08-05	172714	88	electropop, pop, rap, k-pop, x-pop, boy group, g...	0.762	0.861	0	-3.296	1	0.0955	
35	Ferrari	Ferrari	James Hype, Miggy Dela Rosa	2022-04-01	186661	90	deep groove house, edm, house, pop, dance, tropical...	0.847	0.690	1	-7.877	0	0.0493	

- Pemodelan *Content-Based Filtering*.

Tahap pemodelan *content-based filtering* adalah pemodel sistem rekomendasi berdasarkan *content-based filtering*. Pada pemodelan ini kita dapat menentukan konten atau item yang ingin kita rekomendasikan berdasarkan *index* yang ingin digunakan. standar rekomendasi yang digunakan ialah *index* 0, kemudian akan menampilkan rekomendasi lagu berdasarkan *genre* yang mirip dengan *index* 0. Dapat dilihat sebagai berikut:

```
In [25]: class RecommenderSystem:
          def __init__(self, data, content_col):
              self.df = pd.read_csv(data)
              self.content_col = content_col
              self.encoder = None
              self.bank = None
              self.df.isnull().sum()
              self.df.fillna(0).head()
              self.df = df.dropna(axis = 0)

          def fit(self):
              self.encoder = CountVecorizer(stop_words="english", tokenizer=word_tokenize)
              self.bank = self.encoder.fit_transform(self.df[self.content_col])

          def recommend(self, idx, topk=5):
              content = df.loc[idx, self.content_col]
              code = self.encoder.transform([content])
              dist = cosine_distances(code, self.bank)
              recommendation_idx = dist.argsort()[0, 0:(topk+1)]
              return self.df.loc[recommendation_idx]

In [26]: recsys = RecommenderSystem("lagu_teratas_-_global.csv", content_col="Genres")
          recsys.fit()
```



```
In [28]: recsys.recommend(e)
```

```
Out[28]:
```

	Track Name	Album Name	Artist Name(s)	release	duration	Popularity	Genres	Danceability	Energy	Key	Loudness	Mode	Speechiness	Acous
0	Quevedo: Bzrp Music Sessions, Vol. 52	Quevedo: Bzrp Music Sessions, Vol. 52	Bizarrap, Quevedo	2022-07-05	198937	100	argentine hip hop, pop, venezolano, trap, argentin...	0.621	0.782	2	-5.548	1	0.0440	
20	STAYING ALIVE (feat. Drake & Lil Baby)	STAYING ALIVE (feat. Drake & Lil Baby)	DJ Khaled, Drake, Lil Baby	2022-08-05	178176	85	dance pop, hip hop, miami hip hop, pop, rap, ra...	0.718	0.461	4	-7.676	0	0.0797	
29	Bam Bam (feat. Ed Sheeran)	Familia	Camila Cabello, Ed Sheeran	2022-04-07	206070	84	dance pop, pop, post-teen pop, uk pop	0.756	0.697	8	-6.377	1	0.0401	
19	About Damn Time	Special	Lizzo	2022-07-15	191937	80	dance pop, escape room, minnesota hip hop, pop, tr...	0.836	0.743	10	-6.305	0	0.0653	
14	Bad Decisions (with BTS & Snoop Dogg)	Bad Decisions (with BTS & Snoop Dogg)	blanco, BTS, Snoop Dogg	2022-08-05	172714	88	electropop, pop, pop rap, k-pop, k-pop boy group, g...	0.762	0.861	0	-3.296	1	0.0955	
35	Ferrari	Ferrari	James Hype, Miggy Dela Rosa	2022-04-01	189961	90	deep groove house, edm, house, pop dance, tropical...	0.847	0.690	1	-7.877	0	0.0493	

B. HASIL

Berikut tabel lagu yang di rekomendasikan berdasarkan properti trek menggunakan *content-based filtering*.

<i>Content-Based Filtering</i>
<i>Quevedo: Bzrp Music Sessions, Vol. 52</i>
<i>STAYING ALIVE (feat. Drake & Lil Baby)</i>
<i>Bam Bam (feat. Ed Sheeran)</i>
<i>About Damn Time</i>
<i>Bad Decisions (with BTS & Snoop Dogg)</i>
<i>Ferrari</i>

IV. KESIMPULAN DAN SARAN

A. KESIMPULAN

Berdasarkan hasil proses penelitian diatas, maka dapat diambil kesimpulan pada penelitian ini sebagai berikut:

- 1) 50 lagu tertas – global (*50 top songs - global*) dianalisa dengan cara melakukan *cleaning* data seperti: penyuntingan (*editing*), perumusan, pengujian data. Sistem rekomendasi digunakan dalam metode. *Content-Based Filtering* digunakan sebagai pemodelan dalam merekomendasikan lagu berdasarkan parameter dari properti trek serta menggunakan *Jupyter Notebook* untuk menampilkan hasil rekomendasi lagu berdasarkan properti trek.
- 2) Berdasarkan properti trek ‘*genre*’ terdapat 5 lagu teratas yang direkomendasikan menggunakan metode *Content-Based Filtering* yaitu *STAYING ALIVE (feat. Drake & Lil Baby)*, *Bam Bam (feat. Ed Sheeran)*, *About Damn Time*, *Bad Decisions (with BTS & Snoop Dogg)*, dan *Ferrari*.

B. SARAN

Adapun saran untuk penelitian ini adalah hasil dari penelitian ini dapat menjadi bahan untuk penelitian selanjutnya yang lebih baik lagi adanya.

REFERENCES

- Fitriani, S. A. (n.d.). *Analisis Klaster Atribut Musik pada Global Top 50 dengan Data Spotify dengan Menggunakan Algoritma K-Means*.
- Muslimin M. (n.d.). *Perkembangan Teknologi dalam Industri Media*.

- Noviani, D., Pratiwi, R., Silvianadewi, S., Alexandri, M. B., & Hakim, M. A. (2020). Pengaruh Streaming Musik Terhadap Industri Musik di Indonesia. In *Jurnal Bisnis STRATEGI* • (Vol. 29, Issue 1).
- Usino, W., Satria Prabuwno, A., Hamed Allehaibi, K. S., Bramantoro, A., & Amaldi, W. (2019). Document Similarity Detection using K-Means and Cosine Distance. In *IJACSA) International Journal of Advanced Computer Science and Applications* (Vol. 10, Issue 2). www.ijacsa.thesai.org
- Wahyu Dewatara, G., & Monik Agustin, S. (2019). Pemasaran Musik pada Era Digital Digitalisasi Industri Musik dalam Industri 4.0 di Indonesia. *WACANA: Jurnal Ilmiah Ilmu Komunikasi*, 18, 1–10.