

**BIG DATA STREAMING MENGGUNAKAN
API SPOTIFY UNTUK REKOMENDASI
MUSIK BERDASARKAN MODEL
CLUSTERING**

PROPOSAL
PROJECT
IF5270 SISTEM
BIG DATA



TEAM

1

Iwan Kurniawan (23521058)

2

Hamzah Misbachul Adlan
(23521070)

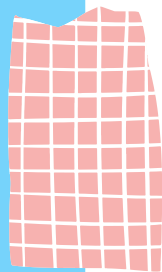
3

Brilliant Fahmi Fauzi Perdana
(23522024)



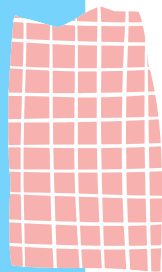
**BIG DATA STREAMING MENGGUNAKAN
API SPOTIFY UNTUK REKOMENDASI
MUSIK BERDASARKAN MODEL
CLUSTERING**

LATAR BELAKANG



Dalam beberapa tahun terakhir, permintaan untuk rekomendasi musik yang personal telah semakin meningkat. Hal ini disebabkan oleh peningkatan jumlah pengguna layanan streaming musik seperti Spotify. Dalam hal ini, Big Data Streaming dapat dimanfaatkan untuk memberikan rekomendasi musik yang lebih akurat dan personal.

LATAR BELAKANG



Dengan menggunakan data Spotify API yang tersedia, Big Data Streaming dapat menganalisis preferensi musik pengguna secara real-time dan memberikan rekomendasi musik yang relevan. Data yang diambil melalui Spotify API dapat dianalisis menggunakan berbagai teknik analisis data seperti machine learning, clustering, dan association rule mining.



Data Spotify API adalah data yang dapat diakses secara langsung dari Spotify, salah satu layanan streaming musik terbesar di dunia. Data Spotify API berisi informasi mengenai lagu, album, artis, playlist, dan sebagainya. Beberapa contoh data yang tersedia di Spotify API antara lain:

- Informasi mengenai lagu, seperti judul, artis, durasi, album, dan genre.
- Informasi mengenai artis, seperti nama, popularitas, dan genre.
- Informasi mengenai playlist, seperti judul, jumlah lagu, dan deskripsi.



PERSOALAN

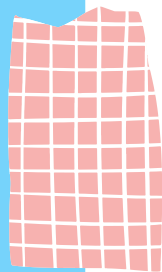
Are You Ready?

Berdasarkan latar belakang yang telah disebutkan, waktu yang dihabiskan untuk listing data dari spotify membutuhkan waktu. Dalam tugas besar ini, teknologi big data seperti Hadoop, Spark, dan Apache Kafka dapat digunakan untuk memproses data Spotify API. Dengan menggunakan teknologi ini, data dari API Spotify dapat diolah dengan lebih efisien dan cepat.

Hadoop dapat digunakan untuk melakukan proses ETL (Extract, Transform, Load) pada data Spotify API, sehingga data tersebut dapat diolah lebih mudah dan cepat. Apache Spark dapat digunakan untuk melakukan proses analisis data dan machine learning dengan skala yang besar dan dalam waktu yang lebih cepat.



KEPENTINGAN PERSOALAN



Pentingnya persoalan ini diselesaikan adalah karena kebutuhan akan rekomendasi musik yang tepat sangat penting dalam dunia industri musik dan hiburan. Dalam beberapa industri, penggunaan teknologi big data dan machine learning dapat membantu meningkatkan pengalaman pengguna dengan memberikan rekomendasi musik yang sesuai dengan preferensi dan minat mereka.

TANTANGAN

1

BEBERAPA
TANTANGAN YANG
AKAN DIHADAPI

AUTENTIKASI

Untuk mengakses data Spotify, Kita harus memiliki akses token atau otentikasi. Token ini dapat berubah setiap saat, sehingga kita harus memperbarui token secara teratur agar tetap dapat mengakses data.

RATE LIMIT

Spotify memiliki batasan tarif yang membatasi jumlah permintaan API yang dapat dilakukan dalam satu waktu. Jika kita melakukan permintaan terlalu sering atau terlalu banyak dalam waktu singkat, kita bisa dibatasi dan tidak dapat mengakses data selama beberapa waktu.

PENGOLAHAN DATA YANG BESAR

Spotify memiliki miliaran lagu dan artis yang tersedia, dan memproses semua data tersebut dalam waktu yang singkat bisa sangat menantang. Oleh karena itu, memilih teknik pengolahan data yang tepat dan mengoptimalkan kinerja kode kita menjadi sangat penting.

PERUBAHAN DALAM API

Spotify secara teratur memperbarui API mereka, sehingga kode yang telah ditulis sebelumnya mungkin tidak lagi berfungsi dengan benar. Kita harus memperbarui kode kita secara berkala untuk memastikan bahwa kita selalu dapat mengakses data terbaru.

TANTANGAN

1

BEBERAPA
TANTANGAN YANG
AKAN DIHADAPI

PENGUNAAN MODEL MACHINE LEARNING

karena rekomendasi
dibuat berdasarkan
clustering
diharapkan dapat
dibuat model
clustering yang
cepat dan mampu
mengelompokkan
secara baik
berdasarkan musik-
musik yang telah
didengarkan

PERLINDUNGAN HAK CIPTA

Dalam mengambil
data dari Spotify, kita
perlu memastikan
bahwa kita tidak
melanggar hak cipta
mereka. kita tidak
dapat mengambil
data tanpa izin atau
memperbanyak data
yang dilindungi hak
cipta tanpa izin dari
pemilik hak cipta.





KONSEP SISTEM


Gambaran umum tentang konsep
tool big data dengan data Spotify:

1. Pengambilan Data: Data Spotify dapat diambil dari API Spotify atau dengan cara web scraping menggunakan library seperti BeautifulSoup. Data dapat diambil dalam format JSON atau CSV.
2. Penyimpanan Data: Data Spotify kemudian disimpan dalam sistem penyimpanan big data, seperti Hadoop Distributed File System (HDFS), Apache Cassandra, MongoDB atau Amazon S3. Data yang disimpan dapat dianalisis secara cepat dan efisien menggunakan alat analisis data big data.
3. Proses Data: Data Spotify kemudian diproses menggunakan alat analisis big data, seperti Apache Spark atau Apache Flink. Alat-alat ini memungkinkan pengolahan data dalam skala besar dan real-time, yang memungkinkan pengolahan data Spotify secara efisien dan cepat.
4. Model Clustering: Model clustering dibuat untuk mengelompokkan lagu dan artis ke dalam kelompok-kelompok berdasarkan kesamaan karakteristik, seperti genre, tempo, dan lain-lain. Model clustering menggunakan algoritma machine learning, seperti K-Means atau Hierarchical Clustering, untuk menghasilkan kelompok-kelompok yang optimal.
5. Visualisasi Data: Setelah model clustering terbentuk, sistem akan memberikan rekomendasi musik berdasarkan kelompok-kelompok yang terbentuk. Rekomendasi musik ini dapat diberikan kepada pengguna melalui antarmuka aplikasi web atau aplikasi mobile yang terhubung dengan sistem.

INFRASTRUKTUR

1. Apache Hadoop
2. Apache Kafka
3. Apache Spark
4. MongoDB
5. Apache Hive
6. Spark ML
7. Laravel





GAMBARAN OUTPUT



https://



Music Recommendation

Song Listened

Add

Mahalini - Melawan Restu

Tulus - Sewindu

Sheila On 7 - Dan

Kahitna - Soulmate

Glenn Fredly - Januari

Isyana Sarasvati - Tetap Dalam Jiwa

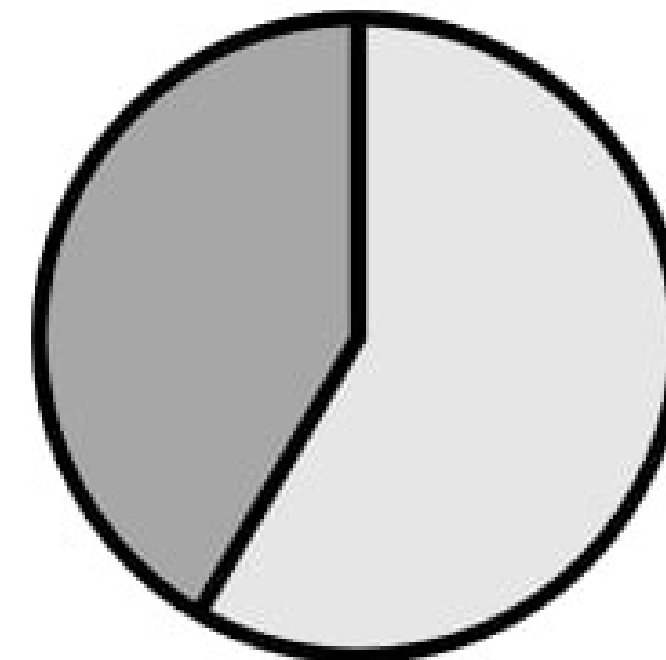
Naff - Terendap Laraku

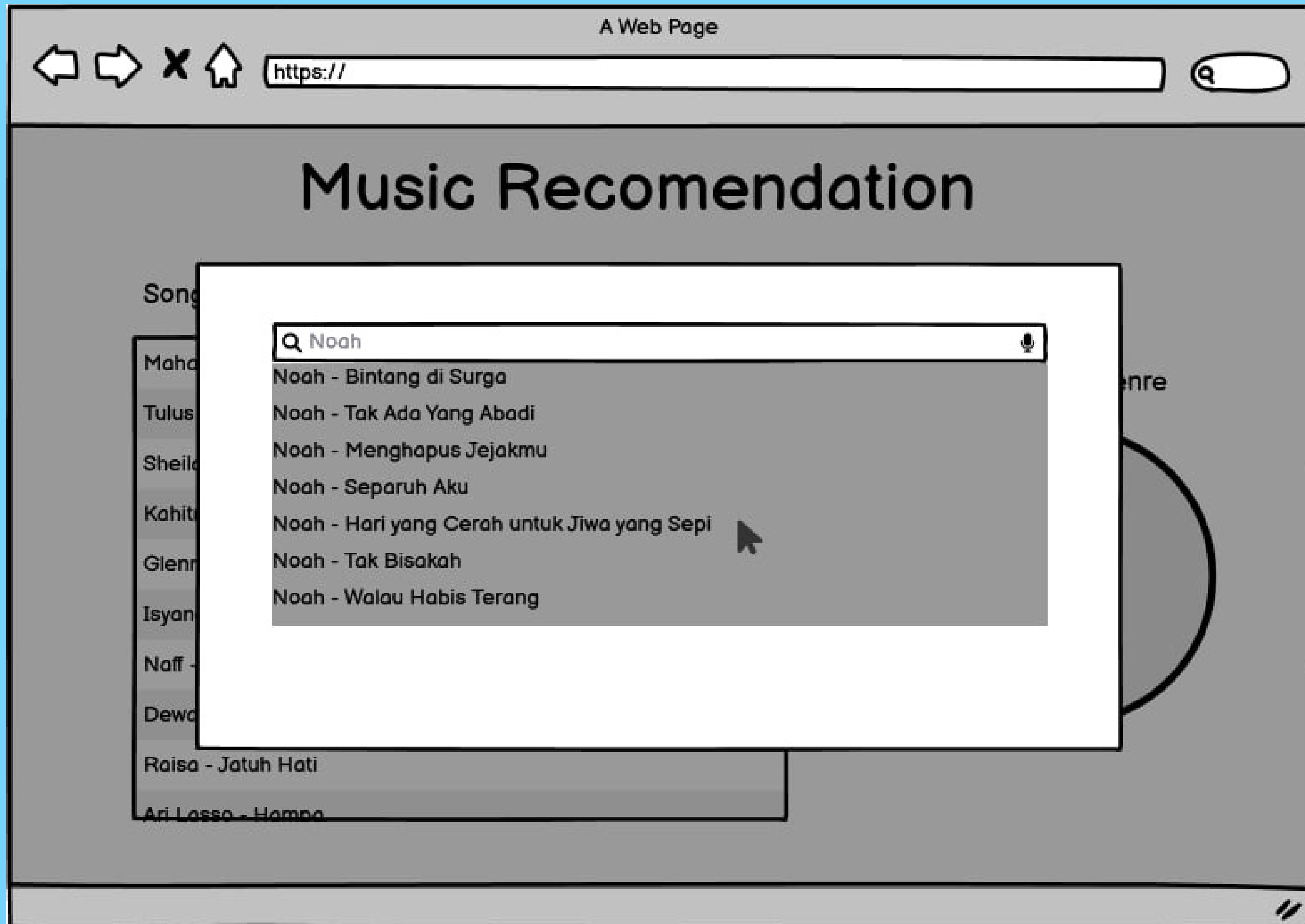
Dewa 19 - Kangen

Raisa - Jatuh Hati

Ari Lasso - Hampa

Your Top Genre







https://



Music Recommendation

New Release Recommendation

Usai – Tiara Andini

Motion (feat. Money Man)

Putus Tapi Cinta – Andmesh

Jika – Buha Bahrin Naufal Azrin

Tak Ingin Usai – Keisya Levronka

Sempurnakan Cinta – Hari Putra & Putri

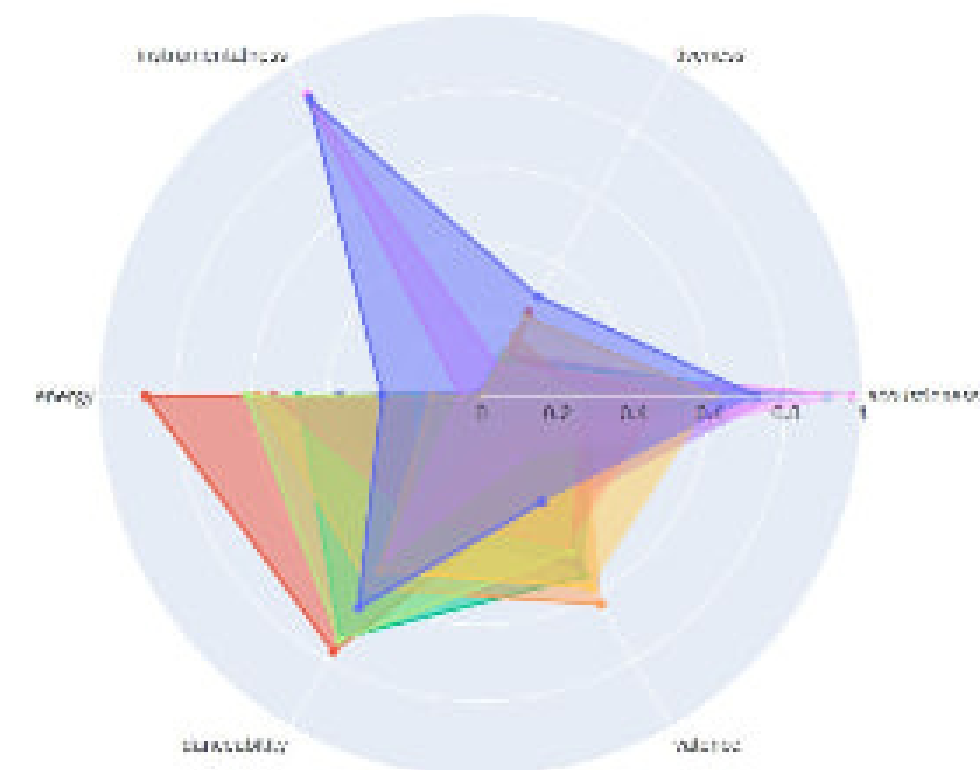
Oh Wanita – Khai Bahar

Too Many Pieces – Circle the Earth

Kota Mati – Noah

Pilihan yang Terbaik – Ziva Magnolia

Your Characteristic



PERAN

1, Analisis kebutuhan dan desain sistem:

- Menganalisis kebutuhan dan spesifikasi sistem.
- Mendesain arsitektur sistem dan skema database.

2. Pengambilan data dari API Spotify:

- Menerapkan autentikasi dengan menggunakan OAuth2.
- Mengambil data dari API Spotify dengan menggunakan Python dan Kafka.

PERAN

3.Data preprocessing dan pipeline:

- Membersihkan dan mengubah format data Spotify menjadi format yang sesuai untuk pemrosesan di Spark.
- Menyimpan data hasil preprocessing di MongoDB.
- Menyediakan pipeline data

4.Pemodelan dan clustering:

- Membuat model machine learning untuk clustering menggunakan Spark ML.
- Menentukan jumlah cluster yang optimal.
- Mengelompokkan data berdasarkan cluster yang telah ditentukan.

5.Pengembangan aplikasi web:

- Membangun aplikasi web dengan framework Laravel.
- Menghubungkan aplikasi web dengan MongoDB untuk menampilkan hasil clustering.
- Menyediakan fitur untuk merekomendasikan lagu berdasarkan lagu lagu yang dipilih.

PERAN

6. Uji Coba Dan Evaluasi

- Melakukan uji coba sistem dan mengevaluasi performa serta akurasi sistem.
- Mengoptimalkan performa sistem dengan mempertimbangkan faktor-faktor seperti latency dan throughput.

7. Dokumentasi

- Mendokumentasikan seluruh proses pengembangan sistem, mulai dari tahap analisis kebutuhan hingga tahap evaluasi dan optimasi.
- Menyediakan dokumentasi penggunaan sistem untuk memudahkan pengguna dalam menggunakan sistem.



PERAN TIM

1

IWAN KURNIAWAN

Team Leader, Documenter

2

HAMZAH M ADLAN

ML ENGINEER

3

Brilliant Fahmi Fauzi Perdana

Data Engineer, Web Developer

Nomor	Tugas	PIC
1	Analisis kebutuhan dan desain sistem	Iwan Kurniawan
2	Pengambilan data dari API Spotify	<u>Brillian</u> Fahmi Fauzi Perdana
3	Data preprocessing dan pipeline	Hamzah Misbachul Adlan
4	Pemodelan dan clustering	Hamzah Misbachul Adlan
5	Membangun aplikasi web dengan framework Laravel	<u>Brillian</u> Fahmi Fauzi Perdana
6	Uji coba dan evaluasi	Iwan Kurniawan
7	Dokumentasi	Iwan Kurniawan



**THANK
YOU!**

Have a
great day
ahead.