ANALISIS PREDIKSI PENGELOMPOKKAN MOOD MUSIK BTS *(BANGTAN BOYS)* MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING DAN RANDOM FOREST PADA APLIKASI RAPID MINER

SKRIPSI

Skripsi diajukan untuk memenuhi persyaratan
memperoleh gelar sarjana

Disusun oleh:

BIMA RIZKY RAMADHAN

200111401026

JURUSAN TEKNIK INFORMATIKA

UNIVERSITAS GLOBAL JAKARTA

2023

ABSTRAK

Spotify adalah platform streaming musik yang terus mengupdate fitur-fitur musik terbaru dengan beragam variasi. Penggemar musik dari grup Bangtan Boys, atau lebih dikenal dengan BTS, sangat populer dan sering mendengarkan lagu-lagu mereka. Namun, penelitian tentang klasifikasi lagu BTS berdasarkan suasana hati dengan menggunakan valensi jarang dilakukan. Setiap lagu memiliki energi emosional yang tercermin dari tingkat energinya dan berhubungan erat dengan psikologi manusia. Masalah yang dihadapi Spotify adalah kurangnya fitur untuk mendengarkan lagu berdasarkan suasana hati. Kategorisasi lagu pop berdasarkan suasana hati dapat membantu pengguna untuk memilih lagu BTS yang sesuai dengan perasaan mereka pada waktu tertentu. Penelitian ini bertujuan untuk mengelompokkan data musik pop berdasarkan 4 kategori suasana hati model Thayer's menggunakan algoritma k-means dan random forest. Metode penelitian yang digunakan adalah SEMMA, yang meliputi tahap sample, explore, modify, model, dan assess. Atribut yang digunakan meliputi judul, artis, tanggal rilis, tingkat kecocokan, energi, kunci, kebisingan, mode, tingkat bicara, keacakan, instrumental, kehidupan, valensi, tempo, ID, dan durasi. Dari atribut tersebut, dilakukan pengklasteran data menggunakan algoritma k-means dan perhitungan kinerja menggunakan indeks Davies-Bouldin pada aplikasi RapidMiner. Hasil penelitian menghasilkan empat suasana hati yaitu angry, sad, calm, dan happy. Suasana hati yang paling sering ditemui adalah suasana hati kecuali happy. Evaluasi dilakukan menggunakan confusion matrix yang menghasilkan tingkat akurasi sebesar 98,64%.

Kata kunci: Spotify; BTS; Suasana Hati; SEMMA; K-Mean Clustering; Random Forest; Rapidminer; Davies Bouldin Index; Confusion Matrix;

ABSTRACT

Spotify is a music streaming platform that continuously updates its features with various music variations. Fans of the Bangtan Boys group, also known as BTS, are very popular and often listen to their songs. However, research on classifying BTS songs based on mood using valence is rarely done. Each song has emotional energy reflected in its energy level and is closely related to human psychology. The problem faced by Spotify is the lack of features to listen to songs based on mood. Categorizing pop songs based on mood can help users choose BTS songs that match their feelings at certain times. This study aims to group pop music data based on 4 categories of mood from Thayer's model using the k-means and random forest algorithms. The research method used is SEMMA, which includes the stages of sample, explore, modify, model, and assess. The attributes used include title, artist, release date, compatibility level, energy, key, loudness, mode, speechiness, acousticness, instrumentalness, liveness, valence, tempo, ID, and duration. From these attributes, data clustering is performed using the k-means algorithm and performance calculations using the Davies-Bouldin index in the RapidMiner application. The research results in four moods: angry, sad, calm, and happy. The most common mood encountered is all moods except happy. Evaluation is done using a confusion matrix, resulting in an accuracy rate of 98.64%.

Keyword: *Spotify; BTS; Mood; SEMMA; K-Mean Clustering; Random Forest; Rapidminer; Davies Bouldin Index; Confusion Matrix;*

DAFTAR ISI

PERNYATAAN ORISINALITAS SKRIPSI ..... **ii**
HALAMAN PENGESAHAN PEMBIMBING ..... **iii**
HALAMAN PENGESAHAN DEWAN PENGUJI ..... **iv**
KATA PENGANTAR/UCAPAN TERIMA KASIH ..... **v**
ABSTRAK ..... **vii**
ABSTRACT ..... **viii**
DAFTAR ISI ..... **ix**
DAFTAR GAMBAR ..... **xii**
DAFTAR TABEL ..... **xiii**
BAB I **1**
PENDAHULUAN **1**
1.1 Latar Belakang 1
1.2 Rumusan Masalah 3
1.3 Tujuan Penelitian 3
1.4 Manfaat Penelitian 4
1.5 Batasan Masalah 4
BAB II **6**
STUDI PUSTAKA **6**
2.1 Landasan Teori 6
2.1.1 Lagu 6
2.1.2 Spotify 6
2.1.3 Pola Pendengaran 7
2.1.4 Data Mining 7
2.1.5 Clustering 9
2.1.6 Klasifikasi 10
2.1.7 Algoritma K-Means 10
2.1.8 Algoritma *Random Forest* 11

2.1.9 Rapid Miner 12
2.2 Penelitian Terkait 12
BAB III 18
METODE PENELITIAN 18
3.1 Metode Penelitian yang Diusulkan 18
3.1.1 *Sample* 18
3.1.2 *Explore* 19
3.1.3 *Modify* 22
3.1.4 *Model* 23
3.1.5 *Assess* 23
3.2 Lokasi dan Obyek Penelitian 24
3.2.1 Lokasi penelitian 24
3.2.2 Obyek Penelitian 24
3.2.3 *Timeline* Penelitian 24
3.3 Metode Pengumpulan Data 25
3.4 Diagram Alir Penelitian 26
BAB IV 29
HASIL DAN PEMBAHASAN 29
4.1 Data Sample 29
4.2 Implementasi *K-Mean Clustering* 30
4.2.1 Model Algoritma *K-Mean Clustering* 30
4.3 Implementasi *Random Forest* 33
4.3.1 Model Algoritma *Random Forest* 33
4.4 Pengukuran Akurasi 36
4.4.1 Hasil Pengukuran Akurasi Algoritma *K-Mean Clustering* 37
4.4.2 Hasil Pengukuran Akurasi Algoritma Random Forest 38
BAB V 39
KESIMPULAN 39
5.1 Kesimpulan 39
5.2 Saran 39

DAFTAR PUSTAKA **41**
LAMPIRAN **43**

DAFTAR GAMBAR

Gambar 3. 1 Alur Metode Penelitian Model SEMMA 18
Gambar 3. 2 Diagram Alir Penelitian 26
Gambar 4. 1 *Data Sample* 30
Gambar 4. 2 Model Algoritma *K-Mean Clustering* pada aplikasi *Rapidminer* 30
Gambar 4. 3 Hasil proses pada tahap retrieve 31
Gambar 4. 4 Pengaturan parameter K-Means dan Hasil pemetaan cluster 32
Gambar 4. 5 Hasil data *export* pemetaan klaster 32
Gambar 4. 6 Model Algoritma *Random Forest* 33
Gambar 4. 7 Data batasan nilai *cluster* dan grafik model Thayer 34
Gambar 4. 8 Hasil data *export* pemetaan mood by valence 35
Gambar 4. 9 Hasil pohon keputusan secara keseluruhan 36
Gambar 4. 10 Hasil Akurasi Algoritma K-Mean Clustering 37
Gambar 4. 11 Hasil Evaluasi *Confusion Matrix* 38

DAFTAR TABEL

Table 2. 1 Peneliti Terkait 14
Table 3. 1 *Audio Features Attributes* 19
Table 3. 2 Mood by Valence 23
Table 3. 3 *Timeline* Penelitian 24

BAB I
PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi zaman sekarang semakin pesat, dalam mengikuti laju perkembangan teknologi informasi kita harus dapat menggunakan serta memanfaatkan layanan teknologi informasi secara optimal (Soraya & Hendry, 2023). Penerapan teknologi diberbagai bidang maupun aspek dalam kehidupan masyarakat tentunya untuk memudahkan berbagai pekerjaan, menyajikan informasi dengan fokus dan spesifik, serta meningkatkan akurasi data dan informasi dengan pengawasan yang lebih mudah dan penggunaan teknologi yang maksimal.

Ada banyak informasi musik yang mudah diakses saat ini. Misalnya informasi tentang musik dan karya seni yang diperoleh dari berbagai media seperti label rekaman, website, lirik dari database yang sudah diterbitkan, dan komentar lain tentang musik dapat dilihat di blog atau majalah musik online ternama. Di setiap tahap kehidupan, musik dikenal memiliki kemampuan untuk menangkap dan mengekspresikan berbagai emosi, seperti kebahagiaan, kesedihan, gairah, dan pendidikan. Selain itu, musik memiliki nilai yang sangat bermakna bagi segala upaya intelektual, serta fungsi kemasyarakatan yang dapat mengangkat dan memperkaya ritual tertentu. Musik adalah bahasa universal yang dimiliki oleh semua anggota masyarakat manusia dengan cara yang berbeda-beda dalam mengapresiasi berbagai jenis musik. Setiap orang memiliki selera musik yang berbeda-beda, bahkan terkadang dalam konteks geografis dan budaya yang sama meskipun berbeda di seluruh dunia (Ruddin et al., 2022).

Data Mining adalah proses penggunaan kumpulan data besar dengan gagasan mengekstraksi pengetahuan baru dari proses pengumpulan informasi menggunakan teknik tertentu. *Clustering* dan klasifikasi adalah dua teknik yang digunakan dalam penambangan data. Algoritma K-Means dan Random Forest saat ini banyak digunakan dalam algoritma *clustering* dan klasifikasi. Algoritma K-Means *Clustering* adalah algoritma yang dirancang untuk mengelompokkan data ke dalam *cluster* berdasarkan kesamaan karakteristik yang dimiliki. Di sisi lain, Random Forest adalah

algoritma *unsamble* yang membangun beberapa pohon keputusan (atau pohon) dan menggabungkan hasil prediksi dari setiap pohon untuk memberikan prediksi yang lebih akurat dan stabil (Duman et al., 2022).

Dalam penelitian yang dilakukan oleh I Putu Bayu Wira Brata dan I Dewa Made Bayu Atmaja Darmawan dengan judul "Klasifikasi Mood Lagu Bali dengan Metode Pengelompokan K-Means Berdasarkan Fitur Konten Audio", hasil penelitian menunjukkan bahwa mood lagu dapat diklasifikasikan berdasarkan fitur audio seperti energi dan valensi yang diperoleh dari dataset Spotify API. Data yang digunakan terdiri dari empat kelas mood pada setiap lagu, dan tidak ada lagu dari 50 lagu yang termasuk dalam kategori mood marah. Klasifikasi menggunakan tiga klaster menghasilkan akurasi sebesar 40%, sementara menggunakan empat klaster menghasilkan akurasi 32%. Penyebabnya adalah karena data hanya mencakup mood senang, santai, dan sedih. (Brata & Darmawan, 2021).

Penggunaan *random forest* pernah dibandingkan oleh Prashant dkk dengan judul penelitiannya "*Predicting Music Popularity Using Machine Learning Algorithm and Music Metrics Available in Spotify*" yang menghasilkan kesimpulan terbaik yang didapat dari penelitian tersebut dalam penggunaan *random forest* sebesar 89% tingkat akurasinya, dan hasil ini menjadikan *random forest* dengan tingkat akurasi dibandingkan 2 algoritma lainnya yaitu *K-Nearest Neighbour* dan *Linear Support Vector Classifier* (Sharma et al., 2022).

Berdasarkan kedua penelitian tersebut, akurasi yang dihasilkan metode Random Forest cukup baik, menghasilkan akurasi yang cukup tinggi, dan metode K-Mean dapat dengan cepat mengelompokkan data dalam jumlah besar. Faktor lain yang mendorong peneliti untuk menggunakan algoritma K-Mean adalah K-Means merupakan salah satu algoritma clustering yang paling cepat dan mudah sehingga efektif jika diterapkan pada dataset yang besar (Nurhalimah et al., 2022). Adapun dilain hal, peneliti memilih algoritma Random Forest karena secara konsisten

memberikan kinerja tinggi dalam berbagai tugas klasifikasi dan menghasilkan hasil klasifikasi yang relatif tinggi.

Berdasarkan uraian di atas, penulis mempunyai gagasan untuk melakukan penelitian dengan judul "Analisis Prediksi Pengelompokkan Mood Lagu BTS (*Bangtan Boys*) Menggunakan Algoritma *K-Mean Clustering* dan *Random Forest* Pada Aplikasi Rapid Miner". Dataset yang digunakan dalam penelitian ini adalah 147 lagu BTS pada *Spotify* beserta dataset pola pendengarannya. Hasil dari penelitian ini dapat digunakan sebagai bahan pertimbangan bagi pengguna *spotify* dalam memilih lagu BTS berdasarkan suasana hati yang dirasakan. Namun *mood* tersebut dapat diambil berdasarkan analisis prediksi nilai akurasi kedua algoritma tersebut. Selain itu juga penelitian ini dapat dijadikan referensi bagi peneliti selanjutnya.

1.2 Rumusan Masalah

Dari penjelasan diatas dapat diambil rumusan masalah yang akan menjadi pembahasan penelitian yaitu:

1. Bagaimana hasil analisis pengelompokkan klasifikasi dan *clustering mood* musik BTS (*Bangtan Boys*) menggunakan algoritma *k-mean* dan *random forest*?
2. Bagaimana prediksi akurasi yang dihasilkan oleh metode algoritma *k-mean* dan *random forest* dalam melakukan klasifikasi mood music BTS (Bangtan Boys)?

1.3 Tujuan Penelitian

Berdasarkan masalah yang telah dirumuskan sebelumnya maka tujuan dari penelitian ini adalah:

1. Menghasilkan analisis klasifikasi dan *clustering mood* musik BTS (*Bangtan Boys*) menggunakan algoritma *k-mean* dan *random forest.*
2. Menghasilkan nilai akurasi yang dihasilkan oleh metode algoritma *k-mean* dan *random forest* dalam melakukan klasifikasi *mood* musik *BTS (Bangtan Boys).*

1.4 Manfaat Penelitian

Berdasarkan penelitian ini manfaat yang diperoleh dari penelitian ini untuk penulis, pembaca, dan akademik adalah:

1. Bagi Penulis

   Dengan menerapkan pengetahuan yang diperoleh selama kuliah, penulis dapat meningkatkan pemahaman dalam bidang data mining dan data science. Dalam penelitian ini, penulis memperoleh pengetahuan tentang proses klasifikasi dan pengelompokan (clustering) dengan menggunakan fitur audio dari musik BTS. Penelitian ini melibatkan penggunaan algoritma pengelompokan k-means untuk mengelompokkan data dan algoritma random forest untuk melakukan klasifikasi.
2. Bagi Pembaca

   Memberikan wawasan dan referensi tentang ilmu data mining atau data science dan mendapatkan gambaran langkah proses clustering menggunakan algoritma k-mean dan klasifikasi menggunakan algoritma random forest.
3. Bagi Akademik

   Sebagai bahan referensi bagi penulis lain untuk dijadikan perbandingan dalam menyusun proposal dan skripsi pada penelitian selanjutnya dan juga sebagai evaluasi sejauh mana kemampuan mahasiswa dalam menerapkan ilmu pengetahuan yang diberikan.

1.5 Batasan Masalah

Klasifikasi dan clustering ini memiliki cakupan yang luas, untuk itu, agar penelitian lebih focus, Batasan masalah yang diidentifikasi dalam penelitian ini, ialah:

a. Menggunakan algoritma k-mean untuk clustering audio features music BTS (Bangtan Boys).
b. Menggunakan algoritma random forest untuk klasifikasi audio features music BTS (Bangtan Boys).
c. Data yang digunakan dalam penelitian diambil dari dataset public (kaggle) berupa data audio feature BTS (Bangtan Boys) berjumlah 147 lagu.

d. Variabel yang digunakan sebagai parameter adalah nilai valence antara 0 dan 1 yang menghasilkan 4 mood yaitu *happy*, *sad*, *relaxed*, dan *angry* berdasarkan thayer's model.
e. Tools atau aplikasi yang digunakan dalam penelitian ini yaitu rapid miner.