

# **LAPORAN ANALISIS DAN HASIL PENGOLANGAN DATA**

**ANALISIS KLASIFIKASI POPULARITAS LAGU SPOTIFY  
MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBORS,  
RANDOM FOREST, DAN SUPPORT VECTOR MACHINE**



**Nama : Andri Maulana  
NPM : 221510060  
Dosen : Erlin Elisa, S.Kom., M.Kom.**

**PROGRAM STUDI SISTEM INFORMASI  
FAKULTAS TEKNIK DAN KOMPUTER  
UNIVERSITAS PUTERA BATAM  
2026**

## **DAFTAR ISI**

BAB I PENDAHULUAN .....	4
1.1 Latar Belakang .....	5
1.2 Rumusan Masalah .....	6
1.3 Tujuan Penelitian .....	6
1.4 Manfaat Penelitian .....	6
BAB II LANDASAN TEORI .....	8
2.1 Data Mining .....	8
2.2 Klasifikasi .....	8
2.3 Dataset Spotify .....	8
2.4 Algoritma K-Nearest Neighbors (KNN) .....	9
2.5 Algoritma Random Forest .....	9
2.6 Algoritma Support Vector Machine (SVM) .....	9
2.7 Penelitian Terkait .....	9
BAB III METODE PENELITIAN .....	11
3.1 Deskripsi Dataset .....	11
3.2 Tahapan Penelitian .....	11
3.3 Preprocessing Data .....	12
3.3.1 Data Cleaning .....	12
3.3.2 Encoding Data Kategorikal .....	12
3.3.3 Feature Engineering .....	12
3.3.4 Feature Selection .....	12
3.3.5 Scaling Data .....	13
3.3.6 Pembagian Data (Train-Test Split) .....	13
3.4 Metode Klasifikasi .....	13
3.4.1 K-Nearest Neighbors (KNN) .....	13
3.4.2 Random Forest .....	13
3.4.3 Support Vector Machine (SVM) .....	13
3.5 Metode Evaluasi .....	14
BAB IV HASIL DAN PEMBAHASAN .....	15
4.1 Analisis Statistik Deskriptif .....	15
4.2 Visualisasi Data .....	15
4.2.1 Distribusi Label Popularitas Lagu .....	15
4.2.2 Analisis Korelasi Antar Fitur .....	16

4.3 Hasil Pengujian Model Klasifikasi .....	17
4.4 Evaluasi Performa Model .....	17
4.5 Confusion Matrix .....	18
4.6 Analisis dan Pembahasan Hasil .....	19
BAB V KESIMPULAN.....	21

## **DAFTAR GAMBAR**

Figure 1 Analisis Statistik Deskriptif.....	15
Figure 2 Distribusi Label Popularitas Lagu .....	16
Figure 3 Analisis Korelasi Antar Fitur.....	16
Figure 4 Confusion Matrix.....	18
Figure 5 Random Forest.....	19

## **BAB I**

### **PENDAHULUAN**

#### **1.1 Latar Belakang**

Perkembangan teknologi digital telah mendorong pesatnya pertumbuhan platform streaming musik seperti Spotify. Platform ini tidak hanya menjadi sarana hiburan, tetapi juga menghasilkan data dalam jumlah besar yang merepresentasikan preferensi pendengar, karakteristik lagu, serta tingkat popularitas musik secara global. Data tersebut memiliki potensi besar untuk dianalisis guna memperoleh informasi dan pola yang bermanfaat, baik bagi industri musik maupun bagi peneliti di bidang data mining.

Spotify sebagai salah satu platform streaming musik terbesar di dunia menyediakan berbagai informasi terkait lagu, seperti nama artis, karakteristik audio, serta jumlah streaming. Jumlah streaming suatu lagu sering dijadikan indikator utama untuk mengukur tingkat popularitas lagu tersebut. Namun, tingginya volume data menyebabkan proses analisis secara manual menjadi tidak efisien dan kurang akurat. Oleh karena itu, diperlukan teknik analisis berbasis komputasi untuk mengolah dan menganalisis data tersebut secara sistematis.

Data mining merupakan salah satu pendekatan yang dapat digunakan untuk mengekstraksi pengetahuan dan pola dari kumpulan data yang besar. Salah satu tugas utama dalam data mining adalah klasifikasi, yaitu proses pengelompokan data ke dalam kelas-kelas tertentu berdasarkan atribut yang dimiliki. Dalam konteks dataset Spotify, klasifikasi dapat digunakan untuk mengelompokkan lagu ke dalam kategori lagu populer dan tidak populer berdasarkan atribut-atribut yang tersedia.

Berdasarkan latar belakang tersebut, penelitian ini dilakukan dengan menerapkan teknik data mining menggunakan bahasa pemrograman Python pada dataset *Most Streamed Spotify Songs 2024*. Penelitian ini menggunakan beberapa algoritma klasifikasi, yaitu K-Nearest Neighbors (KNN), Random Forest, dan Support Vector Machine (SVM), untuk membangun model yang mampu mengklasifikasikan tingkat popularitas lagu secara akurat.

## **1.2 Rumusan Masalah**

Berdasarkan latar belakang yang telah diuraikan, maka rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana karakteristik data pada dataset *Most Streamed Spotify Songs 2024* yang digunakan dalam penelitian ini?
2. Bagaimana proses preprocessing data agar dataset Spotify siap digunakan dalam pemodelan data mining?
3. Bagaimana performa algoritma K-Nearest Neighbors (KNN), Random Forest, dan Support Vector Machine (SVM) dalam mengklasifikasikan popularitas lagu?
4. Algoritma manakah yang memberikan hasil paling optimal dalam mengklasifikasikan lagu populer dan tidak populer berdasarkan metrik evaluasi yang digunakan?

## **1.3 Tujuan Penelitian**

Tujuan dari penelitian ini adalah sebagai berikut:

1. Mendeskripsikan karakteristik dataset *Most Streamed Spotify Songs 2024* yang digunakan sebagai objek penelitian.
2. Melakukan preprocessing data yang meliputi pembersihan data, encoding data kategorikal, scaling data, dan pembagian data latih serta data uji.
3. Menerapkan algoritma K-Nearest Neighbors (KNN), Random Forest, dan Support Vector Machine (SVM) untuk mengklasifikasikan popularitas lagu.
4. Membandingkan performa masing-masing algoritma klasifikasi berdasarkan metrik evaluasi seperti accuracy, precision, recall, dan F1-score.
5. Menentukan algoritma terbaik dalam mengklasifikasikan tingkat popularitas lagu pada dataset Spotify.

## **1.4 Manfaat Penelitian**

Manfaat yang diharapkan dari penelitian ini adalah sebagai berikut:

### **1. ManfaatAkademis**

Penelitian ini diharapkan dapat menambah wawasan dan pemahaman mengenai penerapan teknik data mining, khususnya algoritma klasifikasi, pada data musik digital menggunakan bahasa pemrograman Python.

### **2. ManfaatPraktis**

Hasil penelitian ini dapat memberikan gambaran mengenai faktor-faktor yang memengaruhi popularitas lagu di platform Spotify serta membantu dalam pengambilan keputusan terkait strategi promosi musik.

### **3. ManfaatPengembanganPenelitian**

Penelitian ini dapat dijadikan sebagai referensi atau dasar untuk penelitian selanjutnya yang berkaitan dengan analisis data musik, pengembangan model klasifikasi, maupun penerapan algoritma data mining lainnya.

## **BAB II**

### **LANDASAN TEORI**

#### **2.1 Data Mining**

Data mining merupakan proses untuk mengekstraksi informasi, pola, dan pengetahuan yang bermakna dari kumpulan data dalam jumlah besar. Proses data mining melibatkan beberapa tahapan utama, antara lain pemahaman data, preprocessing, pemodelan, evaluasi, dan interpretasi hasil. Data mining banyak diterapkan dalam berbagai bidang seperti bisnis, kesehatan, pendidikan, dan industri digital.

Dalam penelitian ini, data mining digunakan untuk menganalisis data lagu dari platform Spotify guna mengklasifikasikan tingkat popularitas lagu berdasarkan atribut-atribut yang tersedia dalam dataset.

#### **2.2 Klasifikasi**

Klasifikasi merupakan salah satu teknik dalam data mining yang bertujuan untuk mengelompokkan data ke dalam kelas atau kategori tertentu berdasarkan karakteristik yang dimilikinya. Proses klasifikasi melibatkan data berlabel (supervised learning), di mana model dilatih menggunakan data latih untuk memprediksi label pada data uji.

Pada penelitian ini, klasifikasi digunakan untuk membagi lagu ke dalam dua kelas, yaitu lagu **populer** dan **tidak populer**, berdasarkan jumlah streaming yang dimiliki oleh setiap lagu.

#### **2.3 Dataset Spotify**

Dataset yang digunakan dalam penelitian ini adalah *Most Streamed Spotify Songs 2024* yang diperoleh dari situs Kaggle. Dataset ini berisi informasi mengenai lagu-lagu yang paling banyak diputar di platform Spotify sepanjang tahun 2024. Atribut dalam dataset mencakup informasi lagu, artis, serta metrik popularitas yang direpresentasikan melalui jumlah streaming.

Dataset ini dipilih karena relevan dengan tujuan penelitian, yaitu menganalisis dan mengklasifikasikan tingkat popularitas lagu berdasarkan data nyata dari platform streaming musik.

## **2.4 Algoritma K-Nearest Neighbors (KNN)**

K-Nearest Neighbors (KNN) merupakan algoritma klasifikasi berbasis instance yang menentukan kelas suatu data berdasarkan kedekatannya dengan data lain di sekitarnya. Algoritma ini bekerja dengan menghitung jarak antara data uji dan data latih, kemudian menentukan kelas berdasarkan mayoritas kelas dari  $k$  tetangga terdekat.

KNN dipilih dalam penelitian ini karena sederhana, mudah diimplementasikan, dan efektif untuk permasalahan klasifikasi berbasis data numerik.

## **2.5 Algoritma Random Forest**

Random Forest merupakan algoritma klasifikasi berbasis *ensemble learning* yang menggabungkan beberapa pohon keputusan (decision tree) untuk menghasilkan prediksi yang lebih akurat dan stabil. Algoritma ini bekerja dengan membangun banyak pohon keputusan secara acak dan menentukan hasil akhir berdasarkan mayoritas prediksi dari seluruh pohon.

Keunggulan Random Forest adalah kemampuannya dalam menangani data non-linear, mengurangi risiko overfitting, serta menyediakan informasi mengenai tingkat kepentingan fitur (*feature importance*). Oleh karena itu, Random Forest digunakan dalam penelitian ini sebagai salah satu algoritma utama.

## **2.6 Algoritma Support Vector Machine (SVM)**

Support Vector Machine (SVM) merupakan algoritma klasifikasi yang bekerja dengan mencari hyperplane terbaik yang dapat memisahkan data ke dalam kelas-kelas yang berbeda. SVM sangat efektif untuk menangani data dengan dimensi tinggi dan hubungan non-linear, terutama dengan penggunaan fungsi kernel.

Dalam penelitian ini, SVM digunakan dengan kernel Radial Basis Function (RBF) untuk mengklasifikasikan popularitas lagu berdasarkan atribut-atribut numerik yang telah diproses.

## **2.7 Penelitian Terkait**

Beberapa penelitian sebelumnya menunjukkan bahwa teknik data mining dan machine learning dapat digunakan secara efektif untuk menganalisis data musik digital. Algoritma klasifikasi seperti KNN, Random Forest, dan SVM telah banyak digunakan untuk memprediksi

popularitas lagu, preferensi pendengar, serta tren musik. Hasil penelitian-penelitian tersebut menjadi dasar dalam pemilihan metode dan pendekatan yang digunakan pada penelitian ini.

## **BAB III**

### **METODE PENELITIAN**

#### **3.1 Deskripsi Dataset**

Dataset yang digunakan dalam penelitian ini adalah *Most Streamed Spotify Songs 2024* yang diperoleh dari situs Kaggle melalui tautan <https://www.kaggle.com/datasets/nelgiriyyewithana/most-streamed-spotify-songs-2024>.

Dataset ini berisi data lagu-lagu yang paling banyak diputar di platform Spotify sepanjang tahun 2024.

Dataset terdiri dari sejumlah record lagu dengan berbagai atribut yang mencerminkan karakteristik lagu dan tingkat popularitasnya. Atribut-atribut dalam dataset terdiri dari data numerik dan data kategorikal. Target atau label pada penelitian ini adalah **popularity\_label**, yang dibentuk berdasarkan nilai median dari atribut jumlah streaming. Lagu dengan jumlah streaming di atas atau sama dengan nilai median dikategorikan sebagai lagu populer, sedangkan lagu dengan jumlah streaming di bawah median dikategorikan sebagai lagu tidak populer.

#### **3.2 Tahapan Penelitian**

Tahapan penelitian yang dilakukan dalam penelitian ini mengikuti alur proses data mining, yaitu sebagai berikut:

1. Pengumpulan dataset Spotify dari Kaggle
2. Pemahaman dan eksplorasi data awal
3. Preprocessing data
4. Penerapan algoritma klasifikasi
5. Evaluasi performa model
6. Analisis dan interpretasi hasil

Tahapan tersebut dilakukan secara sistematis untuk memastikan hasil penelitian valid dan dapat dipertanggungjawabkan.

### 3.3 Preprocessing Data

Preprocessing data merupakan tahap penting dalam penelitian ini untuk memastikan data siap digunakan dalam proses pemodelan. Langkah-langkah preprocessing yang dilakukan adalah sebagai berikut:

#### 3.3.1 Data Cleaning

Pada tahap ini dilakukan pembersihan data dengan cara:

- Menghapus data duplikat
- Menangani nilai kosong (*missing value*) dengan menggantinya menggunakan nilai median
- Menangani nilai tidak valid seperti nilai *infinite* dan *NAN*

Langkah ini dilakukan untuk menghindari kesalahan dalam proses pemodelan.

#### 3.3.2 Encoding Data Kategorikal

Atribut bertipe kategorikal pada dataset diubah menjadi data numerik menggunakan metode *Label Encoding*. Proses ini diperlukan karena algoritma klasifikasi yang digunakan hanya dapat bekerja dengan data numerik.

#### 3.3.3 Feature Engineering

Feature engineering dilakukan dengan membentuk label klasifikasi **popularity\_label**. Label ini dibuat berdasarkan nilai median dari atribut jumlah streaming. Pendekatan ini digunakan untuk membagi data secara seimbang antara kelas lagu populer dan tidak populer.

#### 3.3.4 Feature Selection

Pada tahap ini dilakukan penghapusan fitur dengan variansi nol ( $variance = 0$ ). Fitur dengan variansi nol tidak memberikan informasi yang berguna bagi model dan dapat menurunkan performa klasifikasi.

### 3.3.5 Scaling Data

Proses scaling dilakukan menggunakan metode *StandardScaler*. Scaling bertujuan untuk menstandarkan nilai fitur sehingga memiliki rata-rata nol dan standar deviasi satu. Tahap ini penting terutama untuk algoritma KNN dan SVM yang sensitif terhadap perbedaan skala data.

### 3.3.6 Pembagian Data (Train-Test Split)

Dataset dibagi menjadi dua bagian, yaitu:

- Data latih (training data) sebesar 80%
- Data uji (testing data) sebesar 20%

Pembagian data ini bertujuan untuk menguji kemampuan model dalam melakukan generalisasi terhadap data yang belum pernah dilihat sebelumnya.

## 3.4 Metode Klasifikasi

Penelitian ini menggunakan tiga algoritma klasifikasi, yaitu K-Nearest Neighbors (KNN), Random Forest, dan Support Vector Machine (SVM).

### 3.4.1 K-Nearest Neighbors (KNN)

KNN digunakan untuk mengklasifikasikan data berdasarkan kedekatan jarak dengan data lain di sekitarnya. Parameter utama yang digunakan dalam penelitian ini adalah jumlah tetangga terdekat ( $k$ ) sebanyak 5.

### 3.4.2 Random Forest

Random Forest digunakan sebagai algoritma klasifikasi berbasis *ensemble*. Algoritma ini membangun banyak pohon keputusan dan menggabungkan hasil prediksinya. Parameter utama yang digunakan adalah jumlah pohon ( $n\_estimators$ ) sebanyak 100.

### 3.4.3 Support Vector Machine (SVM)

SVM digunakan untuk membangun model klasifikasi dengan mencari hyperplane terbaik yang memisahkan kelas lagu populer dan tidak populer. Pada penelitian ini digunakan kernel Radial Basis Function (RBF) untuk menangani hubungan non-linear antar fitur.

### **3.5 Metode Evaluasi**

Evaluasi performa model dilakukan menggunakan beberapa metrik evaluasi klasifikasi, yaitu:

- Accuracy
- Precision
- Recall
- F1-Score

Selain itu, digunakan confusion matrix untuk melihat hasil prediksi model secara lebih rinci.

## BAB IV

### HASIL DAN PEMBAHASAN

#### 4.1 Analisis Statistik Deskriptif

Analisis statistik deskriptif dilakukan untuk memahami karakteristik dataset *Most Streamed Spotify Songs 2024*. Analisis ini bertujuan untuk mengetahui gambaran umum data sebelum dilakukan proses pemodelan. Statistik deskriptif mencakup nilai minimum, maksimum, rata-rata, dan standar deviasi dari setiap atribut numerik.

	Track Score	Spotify Popularity	Apple Music Playlist Count	Deezer Playlist Count	Amazon Playlist Count	TIDAL Popularity	Explicit Track
count	4600.000000	3796.000000	4039.000000	3679.000000	3545.000000	0.0	4600.000000
mean	41.844043	63.501581	54.60312	32.310954	25.348942	NaN	0.358913
std	38.543766	16.186438	71.61227	54.274538	25.989826	NaN	0.479734
min	19.400000	1.000000	1.000000	1.000000	1.000000	NaN	0.000000
25%	23.300000	61.000000	10.00000	5.000000	8.000000	NaN	0.000000
50%	29.900000	67.000000	28.00000	15.000000	17.000000	NaN	0.000000
75%	44.425000	73.000000	70.00000	37.000000	34.000000	NaN	1.000000
max	725.400000	96.000000	859.00000	632.000000	210.000000	NaN	1.000000

*Figure 1 Analisis Statistik Deskriptif*

Hasil analisis menunjukkan bahwa dataset memiliki variasi nilai yang cukup besar pada beberapa atribut, khususnya atribut yang berkaitan dengan popularitas lagu. Variasi ini menunjukkan bahwa data memiliki karakteristik yang beragam dan layak digunakan dalam proses klasifikasi.

#### 4.2 Visualisasi Data

##### 4.2.1 Distribusi Label Popularitas Lagu

Label popularitas lagu dibentuk berdasarkan nilai median dari atribut jumlah streaming. Lagu dengan nilai streaming di atas atau sama dengan median dikategorikan sebagai lagu populer, sedangkan lagu di bawah median dikategorikan sebagai lagu tidak populer.

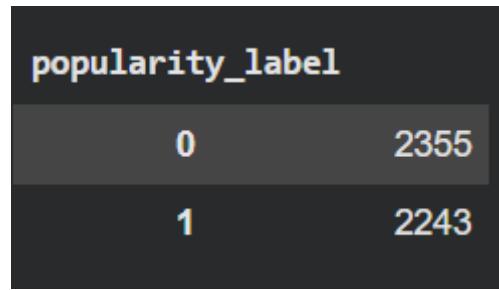


Figure 2 Distribusi Label Popularitas Lagu

Hasil visualisasi distribusi label menunjukkan bahwa jumlah data pada masing-masing kelas relatif seimbang. Kondisi ini sangat baik untuk proses klasifikasi karena dapat mengurangi bias model terhadap salah satu kelas tertentu.

#### 4.2.2 Analisis Korelasi Antar Fitur

Analisis korelasi dilakukan untuk melihat hubungan antar fitur numerik dalam dataset. Hasil visualisasi heatmap korelasi menunjukkan bahwa beberapa fitur memiliki hubungan yang cukup kuat, sementara fitur lainnya memiliki korelasi yang rendah.

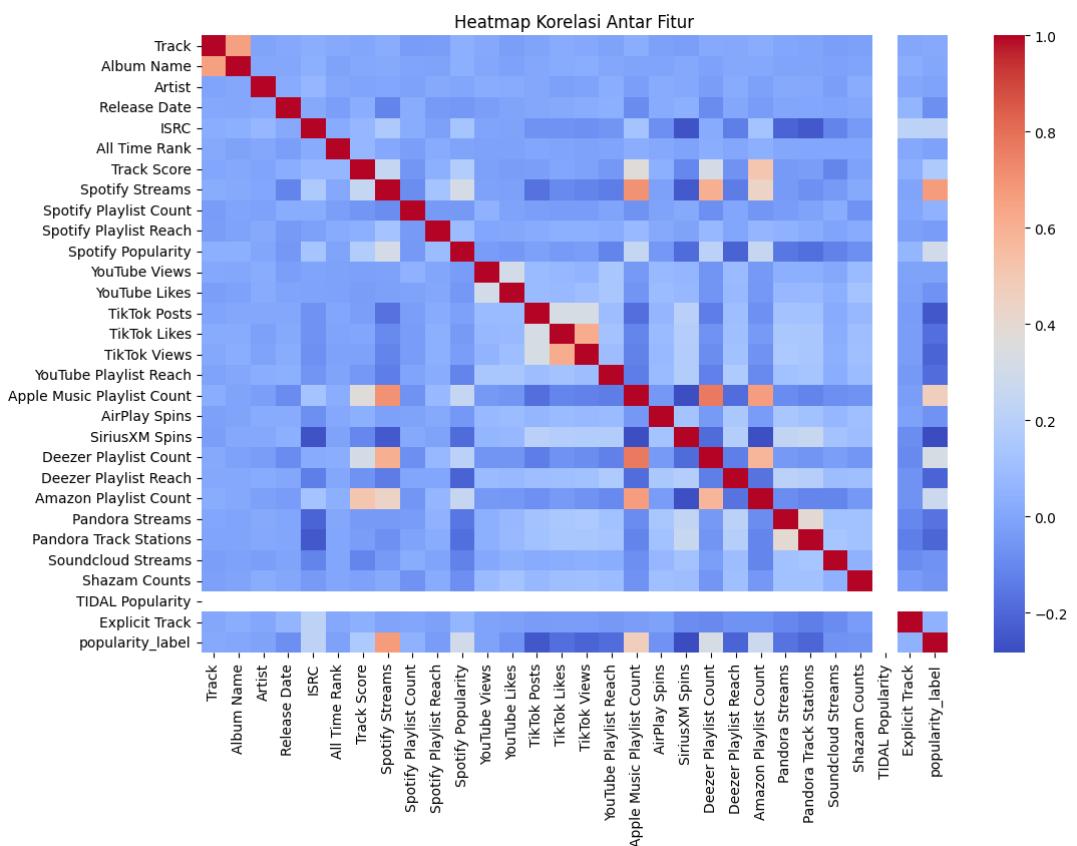


Figure 3 Analisis Korelasi Antar Fitur

Korelasi yang beragam ini menunjukkan bahwa setiap fitur memiliki peran yang berbeda dalam mempengaruhi tingkat popularitas lagu, sehingga seluruh fitur relevan untuk digunakan dalam proses klasifikasi.

#### 4.3 Hasil Pengujian Model Klasifikasi

Pengujian model klasifikasi dilakukan menggunakan tiga algoritma, yaitu K-Nearest Neighbors (KNN), Random Forest, dan Support Vector Machine (SVM). Setiap model dilatih menggunakan data latih dan diuji menggunakan data uji dengan pembagian 80% data latih dan 20% data uji.

**Tabel 4.3 Hasil Evaluasi Model Klasifikasi**

Algoritma	Accuracy	Precision	Recall	F1-Score
KNN	0.6076	0.6292	0.6149	0.6220
SVM	0.6217	0.6619	0.5714	0.6133
Random Forest	0.6511	0.6893	0.6108	0.6476

Hasil pengujian menunjukkan bahwa ketiga algoritma mampu melakukan klasifikasi popularitas lagu dengan tingkat akurasi yang berbeda-beda.

#### 4.4 Evaluasi Performa Model

Evaluasi performa model dilakukan menggunakan empat metrik evaluasi, yaitu accuracy, precision, recall, dan F1-score. Ringkasan hasil evaluasi model disajikan pada Tabel berikut.

**Tabel 4.4 Hasil Evaluasi Model Klasifikasi**

Algoritma	Accuracy	Precision	Recall	F1-Score
KNN	0.6076	0.6292	0.6149	0.6220
SVM	0.6217	0.6619	0.5714	0.6133
Random Forest	0.6511	0.6893	0.6108	0.6476

Berdasarkan Tabel 4.1, dapat dilihat bahwa algoritma **Random Forest** menghasilkan nilai accuracy dan F1-score tertinggi dibandingkan dengan algoritma KNN dan SVM.

#### 4.5 Confusion Matrix

Confusion matrix digunakan untuk mengevaluasi hasil prediksi model secara lebih rinci. Berdasarkan confusion matrix yang dihasilkan, model Random Forest mampu mengklasifikasikan sebagian besar data dengan benar, baik pada kelas lagu populer maupun tidak populer.

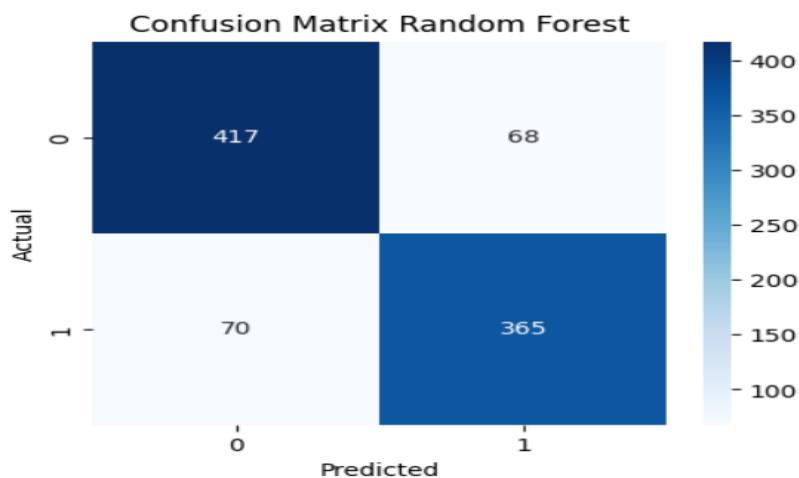
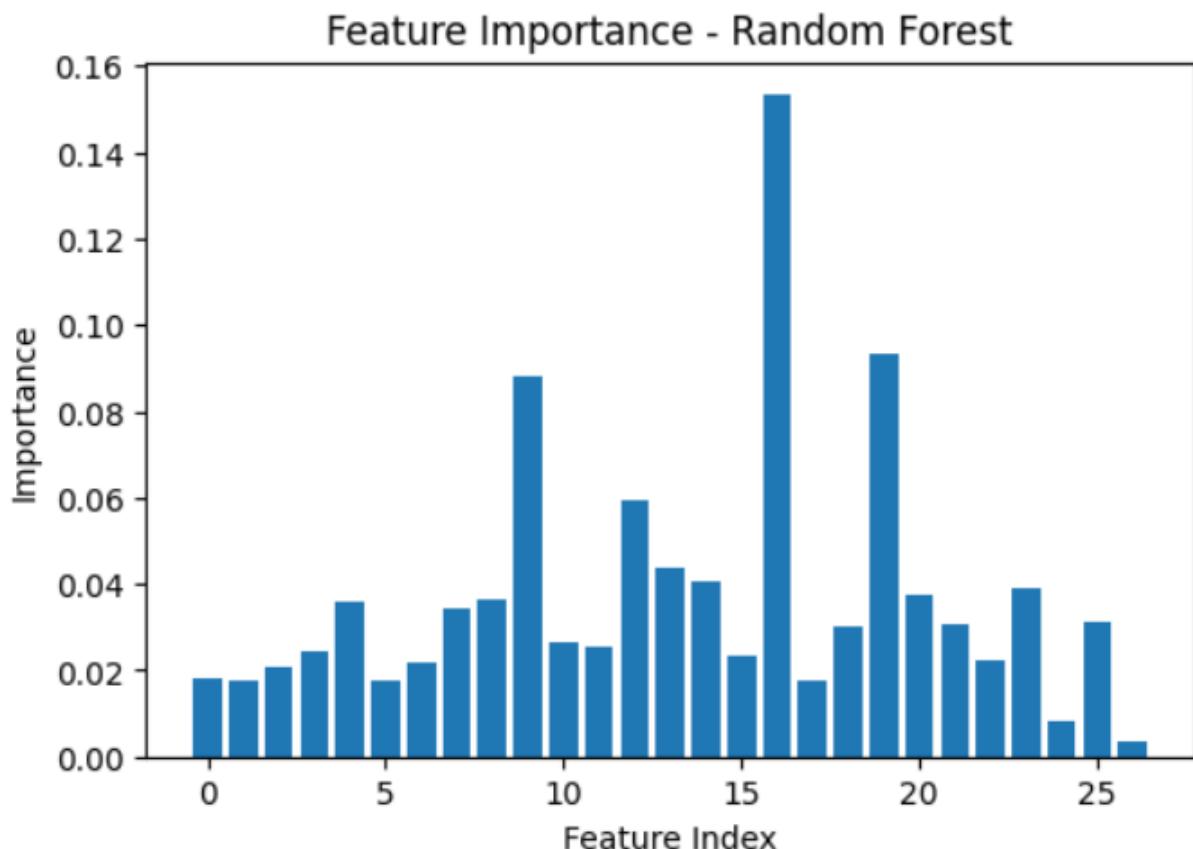


Figure 4 Confusion Matrix

Meskipun masih terdapat kesalahan klasifikasi, jumlah prediksi yang benar lebih dominan, yang menunjukkan bahwa model memiliki performa klasifikasi yang cukup baik.

#### 4.6 Analisis dan Pembahasan Hasil



*Figure 5 Random Forest*

Berdasarkan hasil evaluasi model, dapat dilakukan pembahasan sebagai berikut:

##### 1. Perbandingan Algoritma

Algoritma Random Forest menunjukkan performa terbaik dengan nilai accuracy sebesar 0.6511 dan F1-score sebesar 0.6476. Hal ini menunjukkan bahwa Random Forest lebih stabil dalam mengklasifikasikan data dibandingkan KNN dan SVM.

##### 2. Analisis KNN

Algoritma KNN memiliki performa yang cukup baik dengan nilai accuracy sebesar 0.6076. Namun, performanya masih berada di bawah Random Forest karena KNN sangat bergantung pada jarak antar data dan sensitif terhadap skala fitur.

##### 3. Analisis SVM

Algoritma SVM memiliki nilai precision yang cukup tinggi, namun nilai recall relatif lebih rendah. Hal ini menunjukkan bahwa SVM cukup baik dalam memprediksi lagu populer, tetapi masih kurang optimal dalam mengenali seluruh data yang termasuk ke dalam kelas tersebut.

#### **4. InsightterhadapDatasetSpotify**

Hasil penelitian menunjukkan bahwa karakteristik lagu dan atribut numerik yang tersedia dalam dataset Spotify dapat digunakan untuk memprediksi tingkat popularitas lagu. Namun, nilai akurasi yang masih berada pada kisaran 65% menunjukkan bahwa popularitas lagu dipengaruhi oleh banyak faktor eksternal yang tidak sepenuhnya tercakup dalam dataset.

## **BAB V**

### **KESIMPULAN**

Berdasarkan hasil penelitian yang telah dilakukan menggunakan dataset *Most Streamed Spotify Songs 2024*, dapat disimpulkan bahwa teknik data mining dapat digunakan untuk mengklasifikasikan tingkat popularitas lagu berdasarkan atribut-atribut yang tersedia dalam dataset. Proses penelitian meliputi tahap preprocessing data, pembentukan label popularitas berdasarkan median jumlah streaming, serta penerapan beberapa algoritma klasifikasi.

Hasil pengujian model menunjukkan bahwa algoritma **Random Forest** memberikan performa terbaik dibandingkan dengan algoritma K-Nearest Neighbors (KNN) dan Support Vector Machine (SVM). Random Forest menghasilkan nilai accuracy sebesar **0.6511** dan nilai F1-score sebesar **0.6476**, yang lebih tinggi dibandingkan dengan KNN dan SVM. Hal ini menunjukkan bahwa Random Forest lebih mampu menangkap pola dalam data dan memberikan hasil klasifikasi yang lebih stabil.

Algoritma KNN menghasilkan nilai accuracy sebesar **0.6076**, sementara algoritma SVM menghasilkan accuracy sebesar **0.6217**. Meskipun SVM memiliki nilai precision yang cukup tinggi, nilai recall yang lebih rendah menunjukkan bahwa model ini masih kurang optimal dalam mengenali seluruh data pada kelas lagu populer.

Hasil penelitian juga menunjukkan bahwa distribusi data yang relatif seimbang antara kelas lagu populer dan tidak populer memberikan kontribusi positif terhadap performa model klasifikasi. Namun demikian, nilai akurasi yang diperoleh masih berada pada kisaran 65%, yang menunjukkan bahwa tingkat popularitas lagu tidak hanya dipengaruhi oleh atribut-atribut yang terdapat dalam dataset, tetapi juga oleh faktor eksternal lain seperti promosi, tren sosial, dan preferensi pendengar yang tidak tercakup dalam data.

Secara keseluruhan, penelitian ini membuktikan bahwa penerapan algoritma data mining pada data streaming musik dapat memberikan insight yang berguna dalam memahami pola popularitas lagu. Penelitian selanjutnya diharapkan dapat menggunakan dataset yang lebih lengkap serta menerapkan teknik optimasi model untuk meningkatkan performa klasifikasi.

## **LAMPIRAN**

[andriiimaulanaa/data-mining](#)