

PENERAPAN METODE RANDOM FOREST PREDIKSI POPULARITAS LAGU DI SPOTIFY BERBASIS FITUR AUDIO



✖ ✖ LATAR BELAKANG ✖ ✖

Spotify sebagai **platform streaming musik terbesar** menawarkan jutaan lagu dan ribuan artis yang menarik perhatian pendengar. Popularitas suatu lagu tidak hanya ditentukan oleh penyanyinya, tetapi juga **faktor teknis** seperti *danceability*, *energy*, tempo, dan lainnya. Data popularitas tidak hanya membantu memahami tren dan selera musik pendengar di seluruh dunia, tetapi juga berperan penting dalam **memprediksi lagu-lagu yang berpotensi menjadi hits**, sehingga mendukung strategi promosi artis dan label rekaman.

✖ ✖ TUJUAN ✖ ✖

Menerapkan metode Random Forest dalam memprediksi popularitas lagu di Spotify berdasarkan fitur-fitur audio.

DATA DAN SUMBER DATA

Data yang digunakan bersumber dari **Kaggle** dengan judul **Spotify Tracks Dataset**, yang terdiri dari **11 peubah numerik** dan **83795 amatan**. Pada penelitian ini peubah popularity sebagai peubah respon dikelompokkan menjadi *popular* (≥ 50) dan *unpopular* (< 50).

Peubah yang digunakan

✖ Popularity

✖ Duration_ms

✖ Danceability

✖ Energy

✖ Loudness

✖ Speechiness

✖ Acousticness

✖ Liveness

✖ Valence

✖ Tempo

✖ Instrumentalness

PROSEDUR ANALISIS

Input data - *Praprocessing data* -
Eksplorasi data

Pembagian data (67.036 data latih
dan 16.759 data uji)

Penanganan *imbalance data*
dengan teknik SMOTE - Pemodelan
data latih

Evaluasi model dengan data uji

Peningkatan spesifisitas dengan
menyesuaikan nilai *threshold* (0.6)

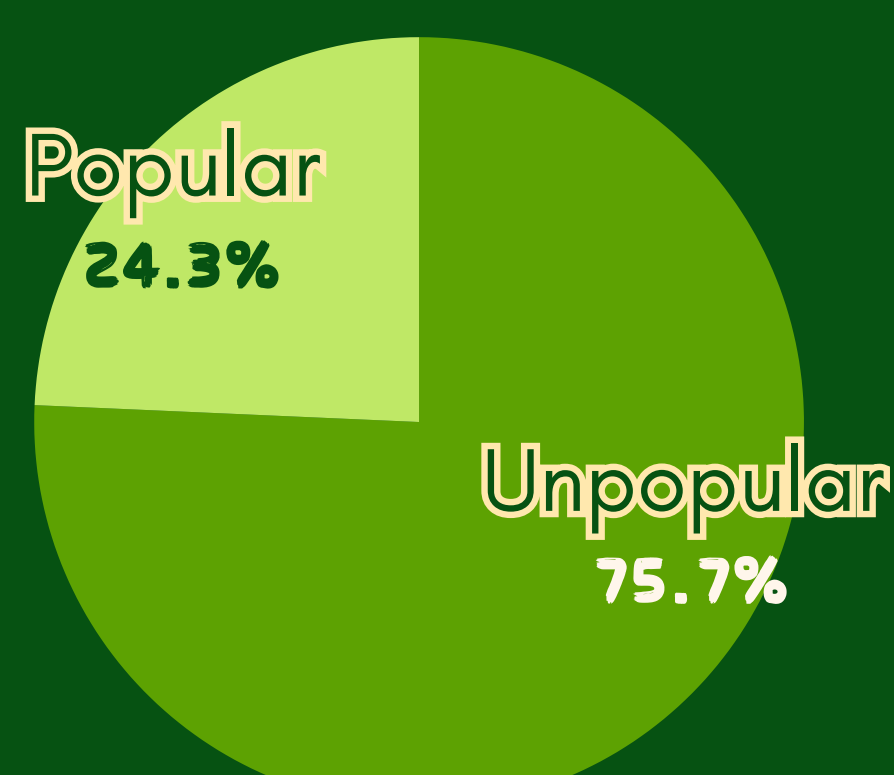
Evaluasi dan perbandingan model -
Interpretasi hasil

RANDOM FOREST

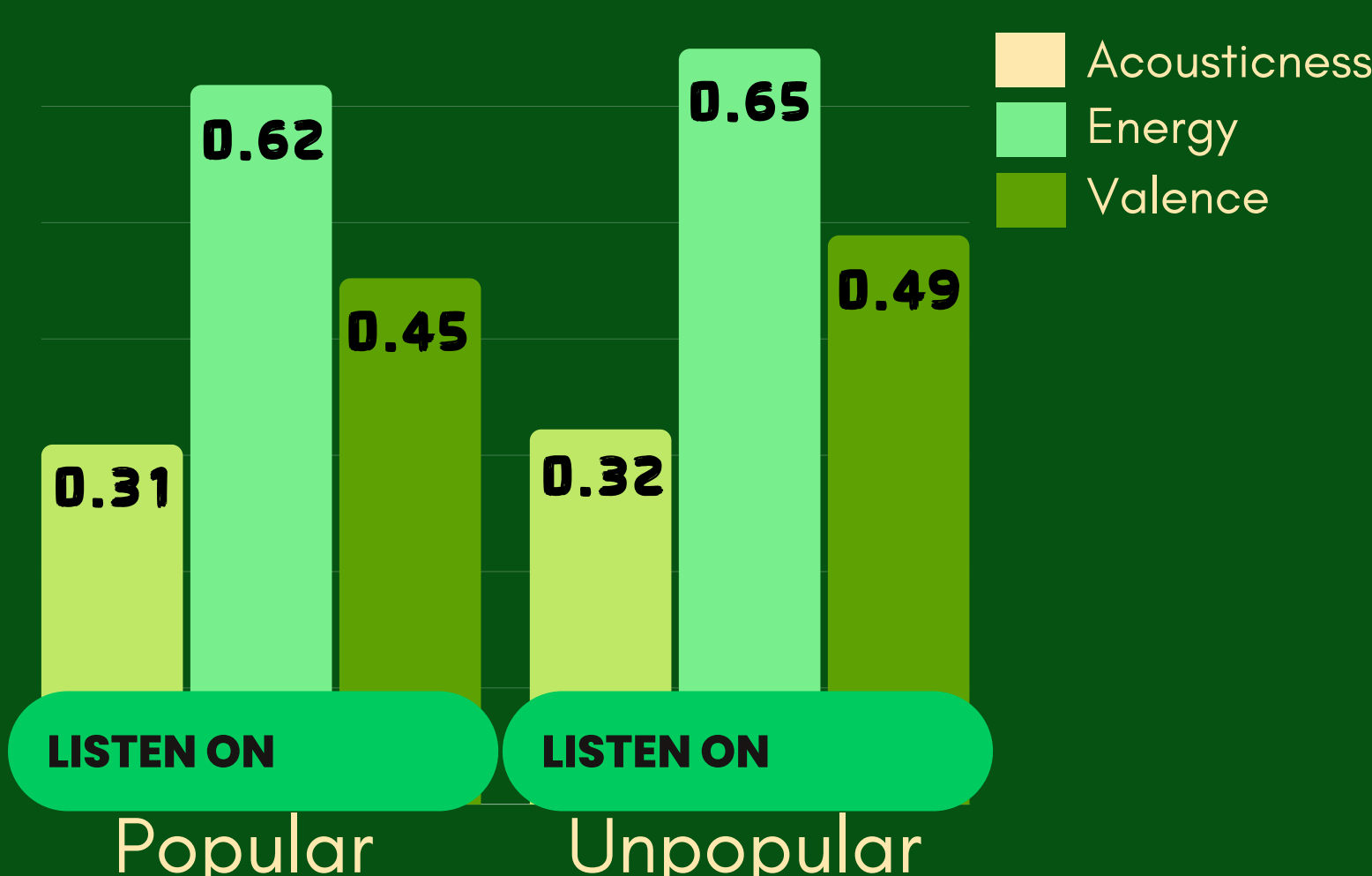
Random Forest merupakan metode yang dapat meningkatkan hasil akurasi, efisien dalam penyimpanan sebuah data, serta mampu mengambil fitur terbaik yang dapat meningkatkan performa model (Supriyadi *et al.* 2020).

✖ ✖ EKSPLORASI DATA ✖ ✖

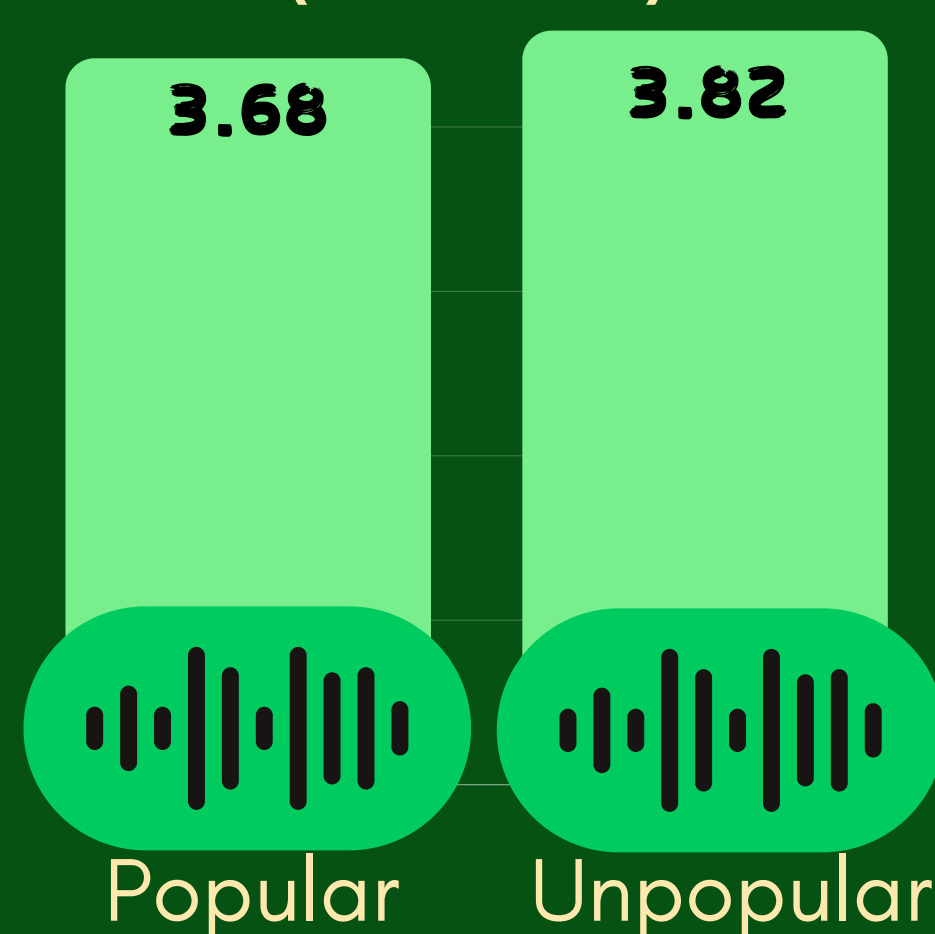
Persentase Tingkat Popularitas



Rataan Fitur Audio

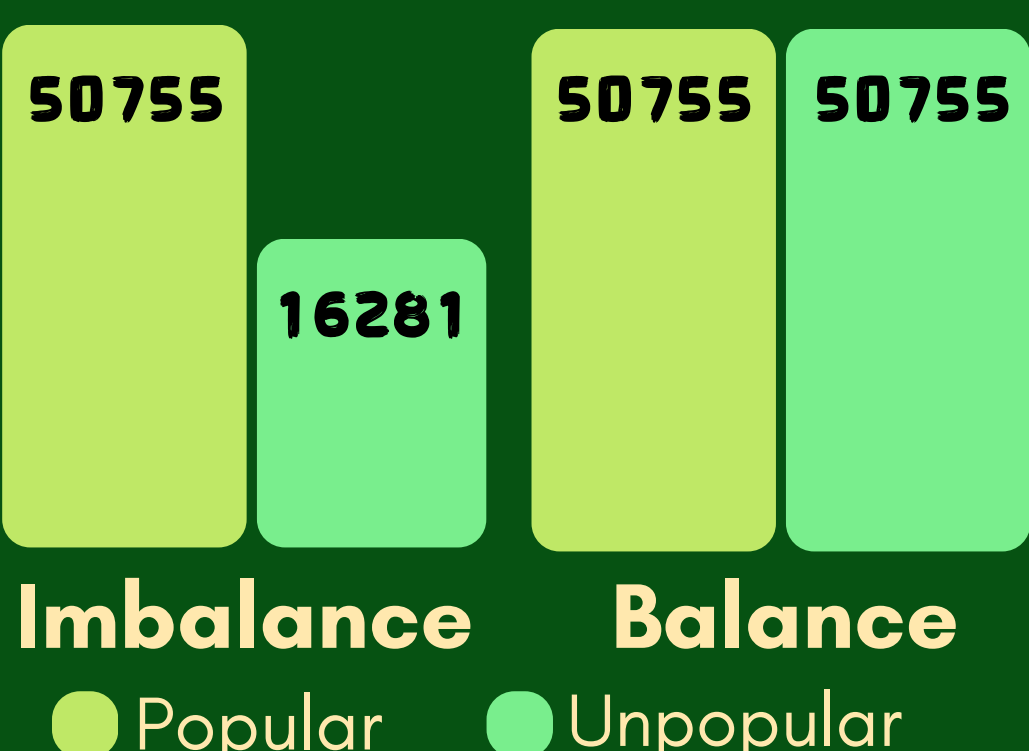


Rataan Durasi Lagu (minutes)



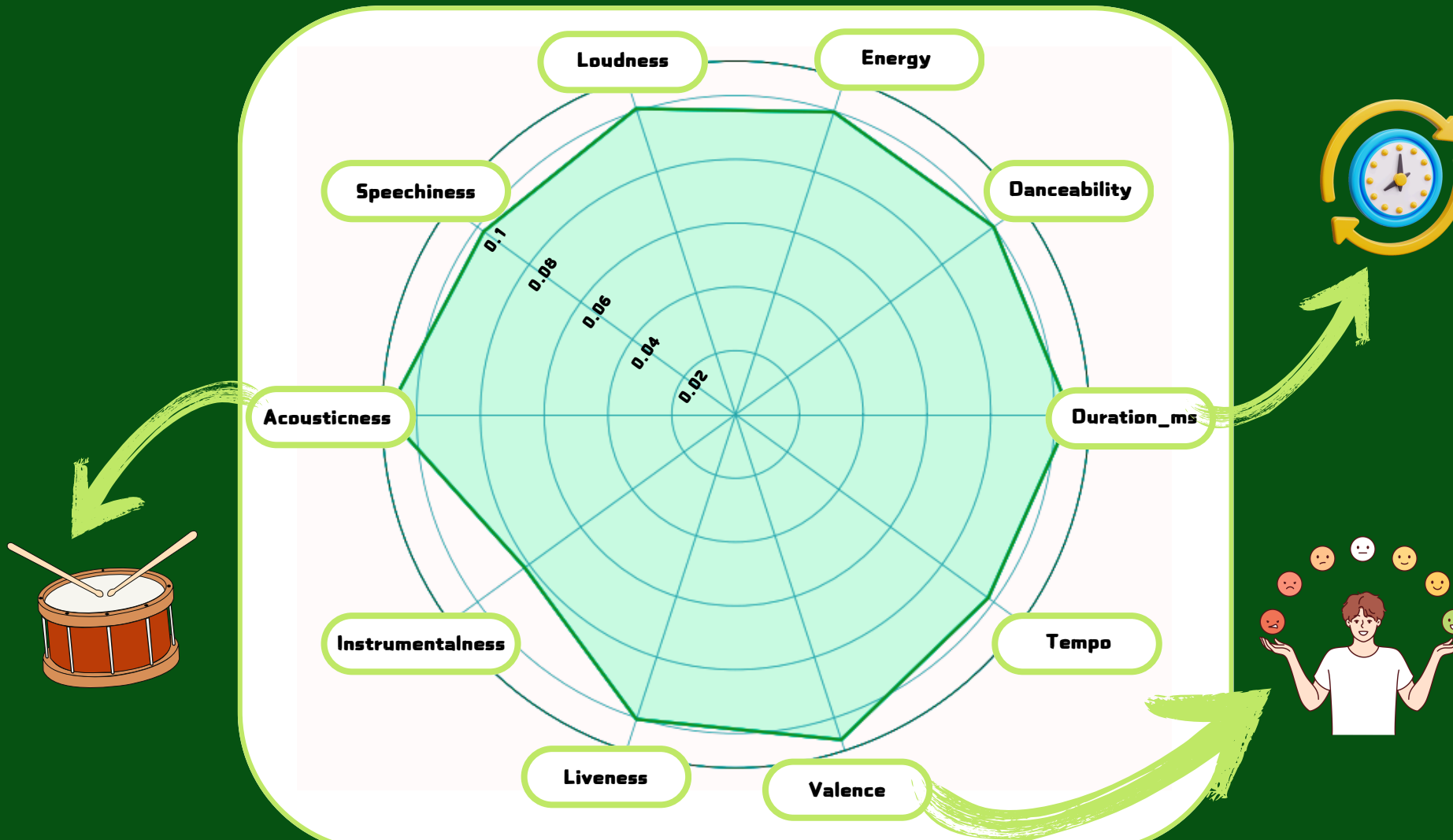
✖ ✖ HASIL DAN PEMBAHASAN ✖ ✖

BALANCING DATA (SMOTE)



SMOTE (*Synthetic Minority Over-sampling Technique*) merupakan teknik *oversampling* dengan membangkitkan data untuk kelas minoritas (Wijayanti *et al.* 2021)

VARIABEL PENTING



UJI PERFORMA MODEL

Prediksi				AKURASI	
Aktual	Popularity			0.81	
	Popularity	Unpopular		SENSITIVITAS	
	Popularity	Unpopular		0.87	
	Unpopular	Unpopular		SPESIFISITAS	
				0.64	
				Threshold	
				0.5	

Prediksi				AKURASI	
Aktual	Popularity			0.76	
	Popularity	Unpopular		SENSITIVITAS	
	Popularity	Unpopular		0.77	
	Unpopular	Unpopular		SPESIFISITAS	
				0.64	
				Threshold	
				0.6	

✖ ✖ KESIMPULAN ✖ ✖

Model dengan *threshold* 0.5 (hasil SMOTE) dipilih karena menghasilkan **nilai sensitivitas yang lebih tinggi** dan **nilai AUC yang lebih tinggi** dalam membedakan kelas negatif dan positif.

Model Random Forest dengan teknik SMOTE memiliki **performa yang baik** dalam prediksi popularitas lagu dengan akurasi 81%.

Acousticness, *duration_ms*, dan *valence* (berkaitan dengan emosional) cukup berpengaruh terhadap popularitas lagu, sehingga perlu perhatian terhadap fitur tersebut jika ingin membuat lagu populer dengan memerhatikan **pasar lagu** dan **musim**.