Aplikasi Prediksi Popularitas Lagu Spotify

Menggunakan Linear Regression

Rafiah Rosada (4122022)

Ilmu Komputer

Machine Learning

Dosen :
Dikky Suryadi, S.Kom,M.Kom

LATAR BELAKANG

Spotify memiliki banyak lagu dengan karakteristik audio yang berbeda.

Kita ingin memprediksi seberapa populer sebuah lagu berdasarkan fitur-fiturnya, seperti energy, tempo, dan acousticness.

Tujuan Proyek

- Memprediksi skor popularitas lagu (0–100)
- Menggunakan model Linear Regression
- Membuat aplikasi interaktif dengan Streamlit

Dataset

Nama Dataset: SpotifyFeatures.csv
Jumlah Data: 232.725 lagu
Fitur Penting:

- acousticness (akustik)
- danceability (kemampuan menari)
- energy (energi)
- tempo
- loudness (kekerasan)
- valence (valensi)
- speechiness (kemampuan bicara)
- instrumentalness, dll

Dataset

Delimiter: ,

	genre	artist_name	track_name	track_id	popularity	acousticness	danceability	duration
1	Movie	Henri Salvador	est beau de faire un Show	jO6ga9RKCKjfDqeFgWV	0	0.611	0.389	9
2	Movie	Martin & les fées	avance (par Gad Elmaleh)	C1NfoEOOusryehmNudP	1	0.246	0.59	13
3	Movie	Joseph Williams	Let Me Be Lonely Tonight	SDzoNIKCRs124s9uTVy	3	0.952	0.663	17
4	Movie	Henri Salvador	Monsieur Gordon Cooper	c6TVm52BwZD07Ki6tIvf	0	0.703	0.24	15
5	Movie	Fabien Nataf	Ouverture	slXpMROHdEPvSl1fTQK	4	0.95	0.331	8
6	Movie	Henri Salvador	etit souper aux chandelles	Mf1jKa8eNAf1a4PwTbizj	0	0.749	0.578	16
7	Movie	Martin & les fées	que & Gérard Lenorman)	JiKYRd6jt1LKMYGkUdnZ	2	0.344	0.703	21
8	Movie	Laura Mayne	Let Me Let Go	bIF9YVD505GutwotpB5C	15	0.939	0.416	24
9	Movie	Chorus	Helka	T6uPfvaPpJLtQwhE6KfC	0	0.00104	0.734	22
10	Movie	Le Club des Juniors	es bisous des bisounours	qZ3KStsjcfERGdcWpFO	10	0.319	0.598	15
11	Movie	Leopold Stokowski	ro Energico E Passionato	XKgegoxLclihK3Klpfo3N	0	0.921	0.191	56
12	Movie	Randy Newman	ture Score) - Remastered	rxsuRM5vVCOfaM7l3gQ	2	0.0383	0.23	12
13	Movie	Idoles De La Musique	J'ai demandé à la lune	HUm18fg6QQCLHhfhC0	4	0.215	0.68	21
14	Movie	Chorus	Mangala Aarti	0oRAp1T4mezDyEhOad	3	0.958	0.758	30
15	Movie	Richard M. Sherman	Keys of Love	wl2CRP5awxHsF9eET3L	0	0.97	0.4	15
16	Movie	Michel Roux	sène Lupin: 813 - Partie 6	UjxM7oDPKpb3T2y3oZm	0	0.548	0.588	244
17	Movie	Jean Claude Corbel	Diane	S7Zid3q0dMyX8uMnlQ8s	0	0.7	0.625	52
18	Movie	Bernard Minet	Ultra Man 80	xSaoSfQkOYUnG1nbga0	3	0.488	0.744	17
19	Movie	Henri Salvador	Veunise	3pHPsGwyeklEjLhnB3ZT	1	0.381	0.451	19

Model Machine Learning

Model yang Digunakan:

- Linear Regression

Langkah-langkah:

1. Preprocessing (standarisasi data)
2. Split data: 80% train, 20% test
3. Latih model
4. Simpan model (linreg_model.pkl)

Evaluasi Model

- MAE: ~12.78
- RMSE: ~15.97
- R^2 Score: ~0.23

Artinya, Linear Regression bisa menjelaskan 23% variasi popularitas. Cukup untuk baseline model.

Coding

Fitur

```python
# 2. Pilih fitur dan target
features = [
    'acousticness', 'danceability', 'duration_ms', 'energy',
    'instrumentalness', 'liveness', 'loudness', 'speechiness',
    'tempo', 'valence'
]
X = df[features]
y = df['popularity']
```

Model

```python
# 5. Latih model
model = LinearRegression()
model.fit(X_train, y_train)

# 6. Simpan model ke file
joblib.dump(model, "model/linreg_model.pkl")
```

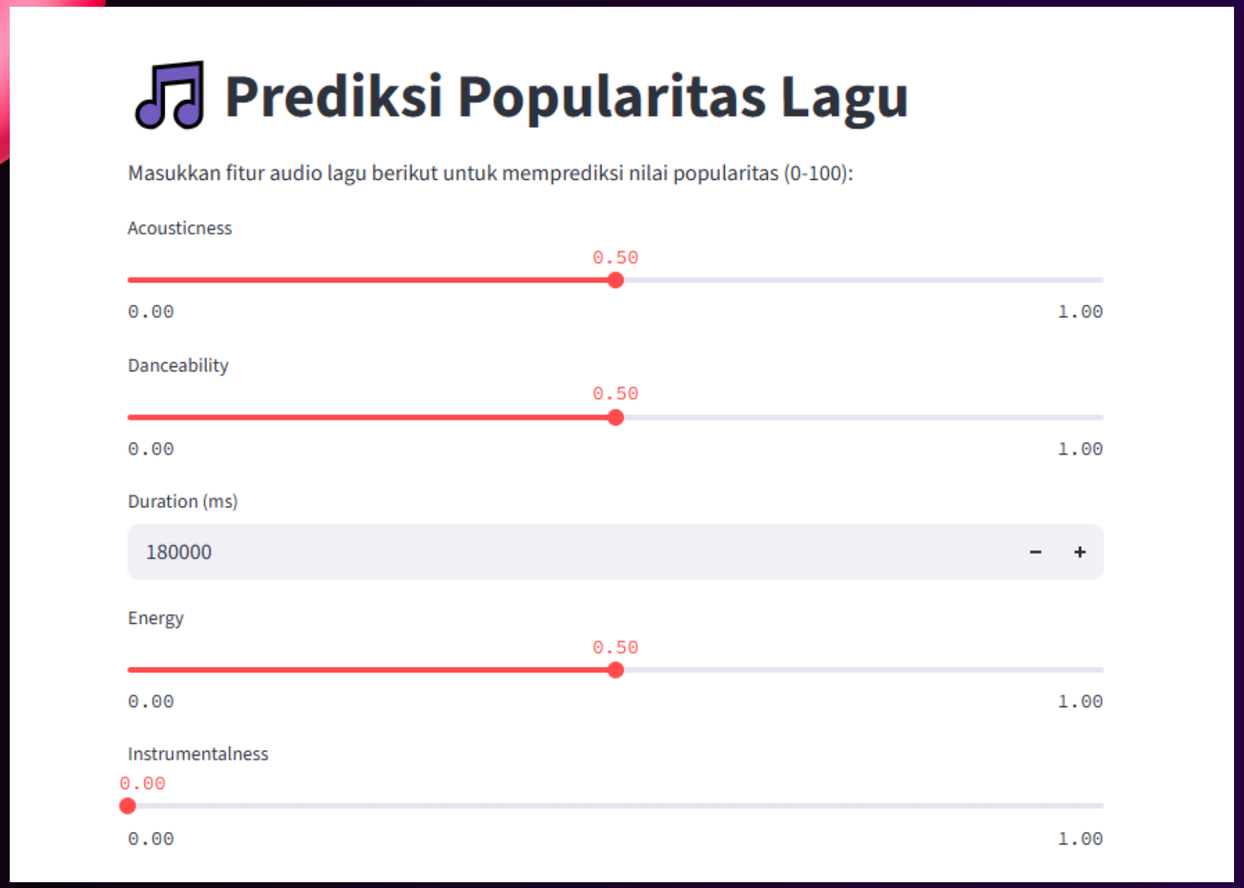
```
['model/linreg_model.pkl']
```

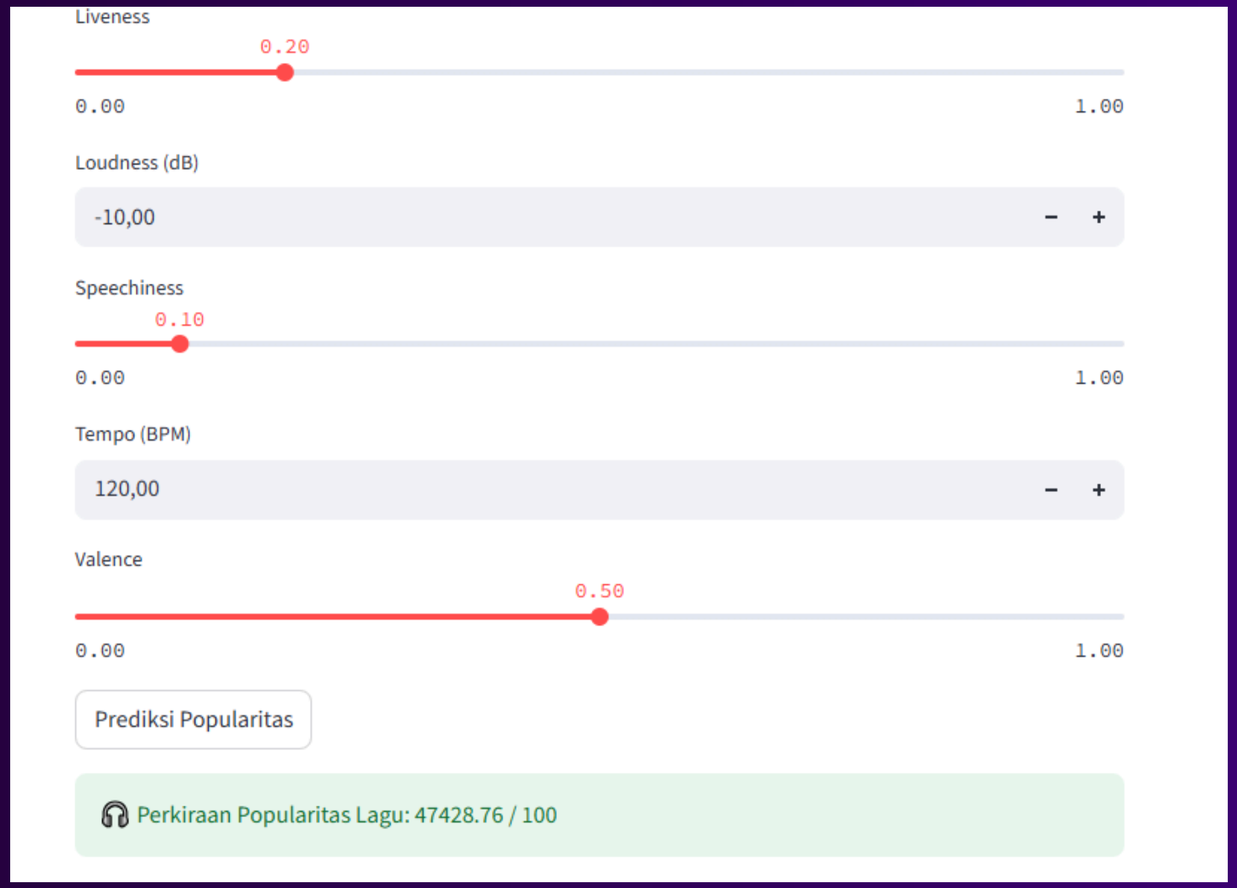
Standarisasi

```python
# 3. Standarisasi
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)

# 4. Split data
X_train, X_test, y_train, y_test = train_test_split(X_scaled, y, test_size=0.2, random_state=42)
```

Hasil Streamlit





Penjelasan

- **Acousticness :** sebuah fitur audio yang menunjukkan seberapa akustik sebuah lagu.
- **Danceability :** fitur audio yang menggambarkan seberapa cocok sebuah lagu untuk menari, berdasarkan kombinasi beberapa elemen musik.
- **Duration :** lama atau durasi sebuah lagu, biasanya dinyatakan dalam milidetik (ms).
- **Energy :** fitur audio yang mengukur intensitas dan aktivitas suatu lagu, biasanya dalam skala 0.0 sampai 1.0.
- **Instrumentalness :** fitur audio yang mengukur seberapa besar kemungkinan sebuah lagu adalah instrumental (tanpa vokal atau lirik manusia).
- **Liveness:** fitur audio yang mengukur kemungkinan sebuah lagu direkam dalam suasana live (langsung di depan penonton).
- **Loudnes :** fitur audio yang mengukur seberapa keras (volume keseluruhan) dari sebuah lagu, biasanya dinyatakan dalam satuan desibel (dB).
- **Speechiness :** fitur audio yang mengukur seberapa banyak unsur ucapan manusia (seperti berbicara, bukan bernyanyi) dalam sebuah lagu.
- **Tempo :** fitur audio yang menunjukkan kecepatan musik, biasanya diukur dalam satuan BPM (beats per minute) atau ketukan per menit.
- **Valence :** fitur audio yang menunjukkan keberanian emosional positif dari suatu lagu — dalam kata lain, seberapa "bahagia" atau "ceria" lagu itu terdengar.

Kesimpulan

- Model dapat memprediksi popularitas lagu dari fitur audio
- Streamlit memudahkan pembuatan aplikasi ML
- Model bisa dikembangkan lagi (Random Forest, XGBoost)

Thank You