

Music Classifier Machine Learning

1. Pendahuluan

Proyek Music Classifier Machine Learning bertujuan untuk mengembangkan model yang mampu mengklasifikasikan genre musik berdasarkan berbagai fitur dari dataset musik. Algoritma machine learning digunakan untuk mempelajari pola-pola dalam data musik dan memprediksi genre secara otomatis. Dengan pendekatan ini, proyek ini berusaha meningkatkan otomatisasi dalam pengkategorian musik yang dapat diterapkan oleh platform streaming, aplikasi musik, atau alat analisis musik lainnya.

2. Tujuan Proyek

Proyek ini memiliki beberapa tujuan utama:

1. Membangun model klasifikasi musik yang dapat memprediksi genre berdasarkan fitur numerik seperti durasi, popularitas, tempo, dan fitur lainnya.
2. Mengolah data musik untuk memastikan bahwa semua fitur diubah ke dalam format yang cocok untuk model machine learning.
3. Mengukur akurasi dan performa model menggunakan metrik seperti classification report agar dapat mengidentifikasi seberapa baik model dalam mengenali genre yang berbeda.

3. Metodologi

Proyek ini menggunakan pendekatan yang sistematis dalam membangun model klasifikasi, melalui beberapa tahapan berikut:

- a. Pengumpulan dan Pembersihan Data
 - Data yang digunakan dalam proyek ini diambil dari file CSV yang berisi informasi musik seperti durasi, genre, popularitas, dan berbagai fitur lainnya.
 - Data diproses untuk menghapus atau mengatasi masalah seperti nilai yang hilang, serta memastikan setiap fitur disiapkan dengan baik.
- b. Preprocessing Data
 - Pemrosesan data numerik dan kategorikal: Fitur numerik dinormalisasi menggunakan StandardScaler untuk memastikan data dalam skala yang konsisten. Fitur kategorikal diubah menjadi representasi numerik menggunakan LabelEncoder.
 - Labeling berdasarkan durasi: Lagu diklasifikasikan menjadi tiga kategori berdasarkan durasi mereka: Short (kurang dari 180.000 ms), Medium (kurang dari 240.000 ms), dan Long (lebih dari 240.000 ms).
- c. Pelatihan Model

- Model dilatih menggunakan dataset yang telah diproses dengan algoritma machine learning.
- Model yang dilatih kemudian disimpan dalam bentuk file model dengan menggunakan pustaka joblib, sehingga bisa digunakan untuk evaluasi lebih lanjut tanpa harus melatih ulang model.

d. Evaluasi Model

- Model dievaluasi menggunakan metrik classification report yang mencakup precision, recall, dan F1-score. Ini membantu mengukur seberapa akurat model dapat memprediksi genre musik yang berbeda.
- Hasil evaluasi memberikan wawasan tentang genre mana yang lebih mudah atau lebih sulit diklasifikasikan oleh model.

4. Hasil dan Diskusi

- Model yang dilatih pada proyek ini menghasilkan hasil yang memadai dalam memprediksi genre musik dengan akurasi sekitar [X%] (sesuai hasil evaluasi).
- Beberapa genre lebih sulit untuk diklasifikasikan dengan benar, menunjukkan bahwa fitur yang tersedia mungkin tidak sepenuhnya memadai untuk membedakan beberapa genre tertentu.
- Analisis tambahan pada fitur seperti tempo, popularitas, dan durasi menunjukkan bahwa ada pola yang dapat diprediksi dalam musik dengan genre yang berbeda.

5. Kesimpulan

Proyek Music Classifier Machine Learning berhasil menghasilkan model yang dapat mengklasifikasikan genre musik berdasarkan fitur numerik dan kategorikal. Meski akurasi model cukup baik, ada peluang untuk meningkatkan performa lebih lanjut dengan menggunakan lebih banyak data atau menambahkan fitur yang lebih relevan.

6. Rekomendasi

Beberapa hal yang dapat dilakukan untuk peningkatan di masa depan:

1. Menambah data: Dataset yang lebih besar dan lebih beragam dapat meningkatkan performa model.
2. Memperbaiki fitur: Menambahkan fitur-fitur baru seperti lirik, akord musik, atau sentimen dapat membantu dalam membedakan genre yang lebih kompleks.
3. Eksperimen dengan model lain: Menggunakan algoritma yang lebih kompleks seperti Gradient Boosting atau Neural Networks mungkin dapat memberikan hasil yang lebih baik dalam klasifikasi genre.

7. Kode Sumber

Seluruh kode yang digunakan dalam proyek ini dapat ditemukan dalam struktur direktori proyek, mencakup kode untuk pemrosesan data, pelatihan model, dan evaluasi:

<https://github.com/haekalsetiawan/Music-Classfier-Machine-Learning.git>

8. Alat dan Pustaka yang Digunakan

1. Pandas: Untuk memuat, memanipulasi, dan preprocessing data.
2. scikit-learn: Untuk pelatihan, preprocessing, dan evaluasi model machine learning.
3. matplotlib: Potensial untuk visualisasi (meskipun tidak secara eksplisit digunakan dalam kode yang diberikan).
4. joblib: Untuk menyimpan dan memuat model yang sudah dilatih.