

BAB I

PENDAHULUAN

1.1 Latar Belakang

Transformasi digital telah memberikan dampak besar pada berbagai sektor, termasuk industri musik. Kehadiran platform streaming seperti Spotify yang menawarkan akses tak terbatas ke jutaan lagu, memungkinkan pengguna untuk menikmati musik dari seluruh dunia secara instan. Menurut laporan terbaru, Spotify memiliki lebih dari 515 juta pengguna aktif pada tahun 2023, menjadikannya salah satu layanan yang paling banyak digunakan untuk mendengarkan musik secara online di seluruh dunia (Imanulloh et al., 2023). Pertumbuhan yang pesat ini tidak hanya meningkatkan jumlah pengguna, tetapi juga meningkatkan persaingan dalam industri musik, di mana popularitas sebuah lagu menjadi faktor utama yang memengaruhi keberhasilan komersial sebuah karya musik. Tingkat popularitas yang tinggi dapat mendorong promosi yang lebih agresif dan meningkatkan peluang pendapatan bagi artis dan label musik (Kamilah et al., 2023).

Untuk memahami dan memprediksi popularitas sebuah lagu, perlu dilakukan analisis terhadap berbagai fitur audio yang dianggap mampu memberikan indikasi daya tarik sebuah lagu. Fitur audio seperti danceability, energy, valence, loudness, tempo, dan speechiness mencerminkan karakteristik musikal sebuah lagu yang dapat memengaruhi persepsi dan preferensi pendengar. Studi yang dilakukan oleh (Marlia et al., 2024) menunjukkan bahwa fitur audio ini memiliki peran penting dalam menentukan genre musik dan daya tarik sebuah lagu di platform streaming. Fitur danceability, misalnya, menggambarkan sejauh mana sebuah lagu cocok untuk menari dan sering dikaitkan dengan lagu-lagu yang memiliki tempo dan irama yang menarik. Demikian pula, energi dan valensi, yang berhubungan dengan intensitas dan nada emosional sebuah lagu, diketahui memengaruhi popularitasnya di kalangan pendengar (Kamilah et al., 2023).

Dengan banyaknya fitur audio yang tersedia, analisis popularitas lagu menjadi semakin rumit dan memerlukan teknik analisis data yang canggih. Machine learning telah menjadi alat yang sangat populer dalam analisis big data, yang memungkinkan identifikasi pola dan tren dalam data yang kompleks. Di antara berbagai algoritma machine learning, XGBoost dan Random Forest dikenal memiliki keunggulan dalam menangani set data besar dan variabel yang kompleks. XGBoost, yang dikembangkan oleh Tianqi Chen pada tahun 2016, adalah algoritma berbasis pohon keputusan yang menggunakan teknik boosting untuk meningkatkan akurasi prediksi. Telah terbukti mampu menangani set data besar dengan efisiensi tinggi, menjadikannya salah satu pilihan utama dalam berbagai studi prediksi big data (Charchyan & Abrahamyan, 2024).

Penelitian yang dilakukan oleh (Dahiya & Sahni, 2019) menyatakan bahwa XGBoost memberikan hasil optimal pada data dengan variabel kompleks yang bersifat numerik dan kategoris. XGBoost menggunakan mekanisme pembobotan pada setiap iterasi pohon keputusan, sehingga berfokus pada kesalahan sebelumnya. Hal ini membantu algoritme mengurangi risiko overfitting dan memberikan hasil prediksi yang lebih stabil, terutama pada data besar yang sering ditemui dalam industri musik (Dahiya & Sahni, 2019). Sebaliknya, Random Forest merupakan algoritma pembelajaran ensemble yang menggabungkan beberapa pohon keputusan untuk membuat prediksi yang lebih akurat. Dalam sebuah studi yang dilakukan oleh (Kamilah et al., 2023), algoritma ini menunjukkan kinerja tinggi dalam prediksi data musik, di mana fitur-fitur seperti *danceability*, *energy*, *valence*, *loudness*, *tempo*, *speechiness*, *acousticness*, *instrumentalness*, dan *liveness* dapat mencerminkan prediksi tingkat popularitas yang lebih akurat.

Pada penelitian terdahulu (Saragih, 2023) yang berjudul *Predicting song popularity based on Spotify's audio features: insights from the Indonesian streaming users* melakukan analisis klasifikasi fitur audio untuk memprediksi populer pada spotify indonesia menggunakan beberapa model seperti Decision Tree, Logistic Regression, Random Forest dan beberapa model lainnya. Pada hasil

analisis tersebut yang memiliki evaluasi terbaik yaitu Rndom Forest dengan *accuracy* 0.70, *precision* 0.65 *recall* 0.69, *f1 score* 0.69.

Sedangkan pada penelitian (Tian et al., 2019) *A Music Recommendation System Based on logistic regression and eXtreme Gradient Boosting*, yang membandingkan performa model logistic regression dan xgboost untuk merekomendasikan musik mendapatkan hasil (*AUC score* 0.76, *error* 0.27) untuk xgboost dan (*AUC score* 0.72, *error* 0.30) untuk logistic regression.

Pada penelitian (Essa et al., 2022) yang berjudul *Predicting the Song Popularity Using Machine Learning Algorithm* telah menguji Fitur *danceability, energy, valence, loudness, tempo, speechiness, acouticness, instrumentalness, liveness* beserta *track artist* dengan hasil Gradient Boosting *accuracy* 0.89 *precision* 0.86 *recall* 0.70, *f1 score* 0.77 dan hasil dari Random Forest mendapatkan *accuracy* 0.89 *precision* 0.86 *recall* 0.69, *f1 score* 0.76

Algoritma Random Forest juga dikenal karena kemampuannya menangani data yang tidak seimbang, dengan membangun sejumlah besar pohon keputusan yang dibor pada sampel data yang berbeda. Random Forest menggunakan mekanisme agregasi bootstrap atau bagging untuk mengurangi varians prediksi, membuatnya lebih andal dan tidak mudah mengalami kesalahan dibandingkan dengan model pohon Tunggal (Britanthia et al., 2020). Dalam penelitian ini, XGBoost dan Random Forest akan digunakan untuk menganalisis dan memprediksi popularitas lagu berdasarkan kesamaan fitur audio di Spotify, guna mengetahui hasil evaluasi model mana yang lebih efektif dan akurat.

Melalui penelitian ini, diharapkan dapat ditemukan model prediksi yang lebih baik untuk mendukung strategi pemasaran dan promosi musik. Dengan kemampuan memprediksi popularitas lagu berdasarkan fitur audio, industri musik dapat mengidentifikasi lagu-lagu yang berpotensi populer lebih awal, sehingga dapat menyesuaikan strategi promosi yang lebih tepat sasaran dan tepat waktu. Penelitian ini juga akan memberikan kontribusi bagi pengembangan algoritma

machine learning di ranah hiburan digital, khususnya dalam analisis musik berbasis data.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, dapat diidentifikasi beberapa permasalahan utama yang menjadi fokus penelitian ini :

1. Bagaimana cara fitur audio seperti (*danceability, energy, valence, loudness, tempo, speechiness, acouticness, instrumentalness, dan liveness*) memprediksi popularitas lagu di Spotify?
2. Manakah algoritma machine learning yang lebih efektif dan akurat dalam memprediksi popularitas lagu?
3. Apakah fitur-fitur audio tertentu memiliki pengaruh terhadap popularitas lagu?

1.3 Tujuan Penelitian

Berdasarkan pada rumusan masalah di atas, maka tujuan dari penelitian ini adalah sebagai berikut :

1. Membandingkan performa prediktif algoritma XGBoost dan Random Forest dalam memprediksi popularitas lagu.
2. Sejauh mana fitur audio seperti (*danceability, energy, valence, loudness, tempo, speechiness, acouticness, instrumentalness, dan liveness*) dalam prediksi popularitas lagu pada Spotify.

1.4 Manfaat Penelitian

1. Membantu pengelola industri musik dalam merencanakan langkah-langkah promosi yang lebih tepat.
2. Mendorong pengembangan model prediksi musik menggunakan machine learning yang akan digunakan.
3. Memberikan wawasan untuk memilih lagu dengan potensi tinggi agar promosi lebih efisien.

1.5 Batasan Masalah

Untuk menjaga fokus dan keterbatasan penelitian, beberapa batasan yang diberlakukan dalam penelitian ini adalah :

1. Algoritma yang dibandingkan hanya XGBoost dan Random Forest, mengingat keterbatasan waktu dan ruang lingkup penelitian serta untuk memastikan perbandingan yang efektif.
2. Popularitas lagu diukur berdasarkan data di Spotify tanpa mempertimbangkan preferensi pada platform streaming lainnya.
3. Penelitian ini hanya akan menggunakan fitur audio yang diekstrak dari dataset Spotify, seperti *danceability*, *energy*, *valence*, *loudness*, *tempo*, *speechiness*, *acousticness*, *instrumentalness*, dan *liveness*. Fitur lain seperti metadata artis atau genre akan diabaikan untuk fokus pada analisis karakteristik musik itu sendiri.

1.6 Kontribusi

Penelitian ini diharapkan memberikan kontribusi sebagai berikut :

1. Memberikan wawasan baru tentang faktor-faktor audio yang berkontribusi pada popularitas lagu di Spotify, yang dapat berguna untuk memahami dinamika pasar musik.
2. Menyediakan perbandingan empiris antara XGBoost dan Random Forest dalam prediksi popularitas lagu, yang dapat menjadi referensi untuk penelitian lebih lanjut.
3. Memberikan kontribusi bagi pengembangan strategi pemasaran dalam industri musik, terutama dalam pemanfaatan fitur audio untuk meningkatkan daya tarik lagu di Spotify, serta memberikan dasar bagi perusahaan rekaman dalam merencanakan promosi yang lebih efisien.