

BAB I

PENDAHULUAN

1.1.Latar Belakang

Musik merupakan karya seni yang menggambarkan ide, pikiran dan perasaan manusia melalui alunan irama dan nada-nada yang indah dan teratur[1]. Musik berfungsi sebagai cermin budaya dan sosial yang mencerminkan nilai-nilai, identitas, dan pengalaman kolektif[2]. Musik bermula sebagai bentuk primitif komunikasi, yang tumbuh dari alat untuk memperlancar pekerja, atau untuk pengiring upacara keagamaan [3]. Namun, seiring dengan kemajuan zaman, peran musik kini jauh lebih luas. Peran musik tidak hanya sebagai pengiring upacara, melainkan digunakan untuk media hiburan hingga media pembelajaran. Peran musik yang luas didukung oleh platform *streaming* yang memudahkan audiens mengakses musik dimana saja dan kapan saja. Hal tersebut didukung oleh digitalisasi yang memperluas pasar dan meraup lebih banyak keuntungan karena perusahaan dapat memotong anggaran untuk memproduksi barang-barang fisik seperti kaset dan beralih ke platform digital [4]. Platform digital tersebut berupa platform *streaming* seperti *Spotify*. *Spotify* adalah layanan musik digital, *podcast*, dan jutaan lagu dan konten lain dari kreator seluruh dunia [5]. Laporan lembaga riset statistik yang berbasis di Hamburg, Jerman, Statista GmbH mengungkapkan pendapatan (revenue) di segmen *streaming* musik di Indonesia diprediksi mencapai US\$148 Juta pada tahun 2020 dan proyeksi pendapatan musik *streaming* sebesar US\$148 Juta meningkat 15% dari tahun ke tahun, dan diperkirakan menjadi pertumbuhan tertinggi dalam rentang 2018 hingga 2024 [6]. Pertumbuhan pendapatan tersebut memberikan insentif bagi para musisi untuk terus berkarya dan berlomba-lomba untuk memproduksi lagu baru. Hal ini dapat mendorong persaingan ketat antara musisi dalam upaya menciptakan lagu yang berpeluang diminati audiens.

Persaingan ketat dalam industri musik mendorong para musisi untuk terus berinovasi dalam menciptakan karya musik yang berpeluang untuk populer. Dinamika budaya populer yang terus berubah menjadi salah satu tantangan prediksi terhadap keberhasilan sebuah karya musik. Lagu baru yang dirilis setiap bulan dan jutaan catatan pengguna platform musik, menyebabkan

timbulnya kebutuhan mendesak untuk menganalisis dan memanfaatkan data besar ini untuk memahami tren musik [7]. Kebutuhan tersebut mendorong musisi melakukan upaya-upaya untuk melakukan analisis mendalam terhadap berbagai faktor yang dapat mempengaruhi popularitas sebuah lagu.

Upaya menganalisis popularitas sebuah lagu dihadapkan pada tantangan, mengingat setiap karya musik memiliki karakteristik yang unik sehingga sulit untuk digeneralisasi. Oleh sebab itu, dibutuhkan pengembangan model prediksi untuk mengetahui peluang lagu akan populer berdasarkan fitur-fitur yang mempengaruhi popularitas lagu seperti fitur audio dan fitur metadata [8]. Memperkirakan skor popularitas itu tidak mudah karena banyaknya dan kompleksitas faktor yang mempengaruhi popularitas [9]. Tujuan prediksi popularitas lagu adalah untuk memprediksi keberhasilan lagu sebelum dirilis. Label rekaman dapat menggunakan model ini untuk sistem untuk kemudian menentukan lagu mana yang harus dipromosikan [10]. Model prediksi yang akurat memberikan keuntungan bagi berbagai pihak, keuntungan tersebut mengenai tren popularitas musik, dapat meningkatkan pengalaman pengguna dan meningkatkan pendapatan platform, selain itu juga meningkatkan popularitas penyanyi [11]. Keuntungan lain adalah membantu artis dan label rekaman memaksimalkan pengembalian komersial [12]. Penelitian yang dilakukan oleh Xiang Liu mengungkapkan bahwa saat ini, banyak metode atau model yang dikembangkan untuk prediksi popularitas lagu, tetapi banyak dari model dan metode tersebut tidak memiliki dasar ilmiah yang kuat [13]. Dalam beberapa tahun terakhir, beberapa penelitian telah dilakukan untuk memprediksi popularitas lagu melalui berbagai teknik analisis data, termasuk regresi. Salah satu metode statistik yang paling umum, regresi telah diterapkan untuk mengeksplorasi hubungan antara fitur audio dan popularitas lagu di platform *streaming* seperti *Spotify*.

Penelitian yang dilakukan oleh Beinuo Guo dengan judul A Model for Predicting Pop Music Popularity and Its Different Characteristics Based on Multiple Linear Regression meneliti mengenai prediksi popularitas musik pop berdasarkan berbagai karakteristik lagu. Karakteristik tersebut meliputi waktu publikasi, jumlah pengikut, dan durasi lagu. Fitur-fitur tersebut digunakan dalam model regresi yaitu regresi linier berganda. Hasil dari penelitian tersebut

menunjukkan bahwa karakteristik tertentu dari lagu pop memiliki pengaruh signifikan terhadap popularitasnya [14]. Karya Harriman Samuel Saragih dengan judul Predicting Song Popularity Based On Spotify's Audi Features: Insights From The Indonesian *Streaming* Users meletakkan dasar untuk eksplorasi fitur audio seperti *loudness*, tempo, dan *danceability* dalam memprediksi popularitas lagu di platform seperti *Spotify*. Model *Random Forest* dalam penelitian ini mampu memprediksi sebesar 68%. Penelitian ini menunjukkan bahwa faktor teknis dari sebuah lagu dapat dikaitkan dengan popularitasnya, dengan algoritma seperti *Random Forest* memberikan akurasi tinggi dalam memprediksi lagu-lagu hit berdasarkan fitur audio [15]. Hasil penelitian oleh Votter & Mayerl (2021) yang berjudul Novel Datasets for evaluating song popularity prediction tasks menunjukkan bahwa model *Random Forest* bekerja dengan baik dalam memprediksi berbagai tingkat popularitas lagu, seperti posisi puncak di chart dan jumlah pendengar. Pada dataset HSP-S, model ini mendapai nilai 43% untuk prediksi [16]. Berdasarkan penelitian yang sudah dipaparkan sebelumnya, *Random Forest* memiliki performa yang cukup baik dan tidak banyak penelitian lain menggunakan *Random Forest* untuk algoritma untuk model regresi, oleh sebab itu *Random Forest* digunakan dalam metode penelitian ini.

Meskipun penelitian tentang prediksi popularitas lagu telah menghasilkan banyak wawasan, namun terdapat beberapa keterbatasan yang cukup signifikan, keterbatasan pertama adalah genre, penelitian sering berfokus pada genre yang paling populer, sehingga kurang mempertimbangkan karakteristik unik dari genre lain yang dapat mempengaruhi popularitas. Keterbatasan kedua adalah kurang analisis tentang bagaimana fitur-fitur tersebut berinteraksi satu sama lain dalam popularitas lagu. Penelitian yang dilakukan oleh Ahmet Çimen dan Enis Kayış berjudul “A Longitudinal Model for Song Popularity Prediction”, berfokus pada fitur akustik dan terkait artis untuk memprediksi popularitas lagu di platform *streaming* seperti *Spotify*. Penelitian ini menganalisis fitur yang berkontribusi pada kesuksesan lagu, tetapi tidak secara jelas mengeksplorasi interaksi antar fitur tersebut[17]. Penelitian lain yang dilakukan oleh Zehao Xing, mencatat pentingnya fitur-fitur tertentu yang mempengaruhi popularitas lagu, tetapi tidak ada analisis mendalam mengenai

fitur-fitur tersebut berinteraksi satu sama lain [18]. Hit song prediction : leveraging low-and high-level audio features oleh Eva Zengerle mengungkapkan keterbatasan penelitiannya yaitu tidak memanfaatkan kombinasi fitur low-level dan high-level secara bersamaan, dan terlalu fokus pada satu jenis fitur akustik low-level atau high-level, tanpa mengeksplorasi potensi sinergi keduanya [19]. Ketiga penelitian tersebut menunjukkan meskipun fitur penting telah diidentifikasi, ada kebutuhan untuk analisis lebih lanjut mengenai interaksi antar fitur untuk mendapatkan pemahaman yang lebih komprehensif tentang faktor yang mempengaruhi popularitas lagu. Keterbatasan ketiga terletak pada matrik evaluasi regresi. Penelitian Hariman Samuel Saragih yang berjudul Predicting Song Popularity Based On Spotify's Audio Features: Insights From The Indonesian Streaming Users hanya menggunakan R-Score dan RMSE (Root Mean Squared Error) untuk matrik evaluasi regresi[15]. Kedua matrik tersebut bisa memiliki keterbatasan dalam memberikan wawasan yang lebih mendalam, maka di penelitian ini menambahkan matrik evaluasi lain seperti MAE (Mean Absolute Error) dan MSE (Mean Squared Error). Hal ini diharapkan dapat meningkatkan validitas hasil prediksi dan memberikan wawasan yang luas bagi pengembangan model selanjutnya.

1.2.Rumusan Masalah

1. Bagaimana penerapan algoritma *Random Forest* dalam prediksi popularitas lagu ?
2. Bagaimana interaksi antarfitur dapat mempengaruhi popularitas sebuah lagu ?

1.3.Tujuan Penelitian

1. Menerapkan algoritma *Random Forest* dalam prediksi popularitas lagu.
2. Menganalisis interaksi antara fitur audio dan metadata, dapat mempengaruhi popularitas sebuah lagu.

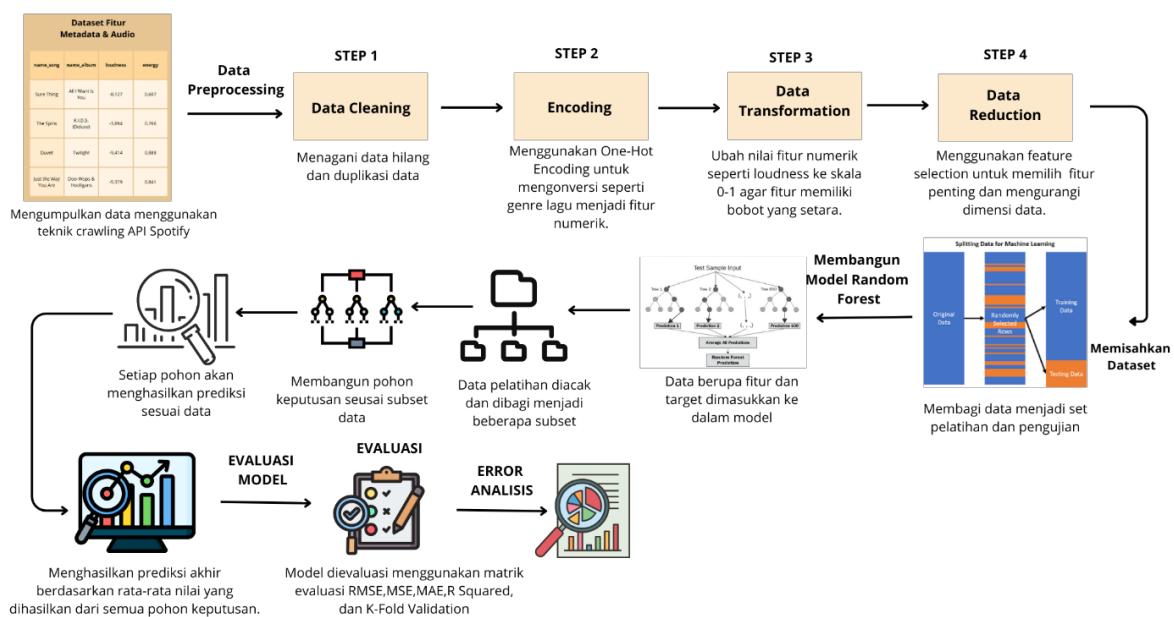
1.4.Batasan Masalah

1. Model prediksi menggunakan *Random Forest*.
2. Penelitian ini terbatas pada dataset yang diambil dari API *Spotify*.
3. Penelitian ini terbatas pada periode waktu lagu yang dirilis 5-10 tahun terakhir.

BAB II

METODE PENELITIAN

Pada Bab metode penelitian berikut membahas alur penelitian menggunakan metode regresi dengan algoritma *Random Forest*. Metode regresi digunakan untuk memprediksi popularitas lagu menggunakan dataset yang telah didapat menggunakan teknik crawling data dari API *Spotify*. Berikut alur penelitian dijelaskan pada **Gambar 1**.



Gambar 1 Alur Penelitian Prediksi Menggunakan Random Forest

Gambar 1 menunjukkan langkah penelitian menggunakan *Random Forest*. Langkah pertama adalah mengumpulkan dataset, teknik pengumpulan dataset yang digunakan dalam penelitian ini adalah menggunakan teknik crawling melalui API *Spotify*. Langkah kedua adalah melakukan Data Preprocessing yang meliputi pembersihan data, transformasi data dan reduksi data. Langkah ketiga adalah melakukan splitting data atau pembagian data. Pembagian data yang dilakukan dengan proporsi yang bervariatif. Langkah keempat yaitu penerapan algoritma *Random Forest* menggunakan regresi. Langkah ke lima atau terakhir yaitu tahap evaluasi menggunakan matrik evaluasi seperti MAE, MSE, RMSE dan R-Score.

2.1 Dataset

Pengumpulan dataset menggunakan teknik crawling data dari API *Spotify*. Dataset memiliki 24 fitur yang terbagi menjadi 2 kategori fitur yaitu fitur metadata dan fitur audio. Fitur metadata meliputi *name_song*, *name_album*, *name_artist*, *release_date*, dan *release_method*. Fitur audio meliputi *length*, *length_format*, *popularity*, *loudness*, *energy*, *danceability*, *acousticness*, *tempo*, *valence*, *time_signature*, *speechiness*, *instrumentalness*, *key*, *mode*, *liveness*, *track_type*, *noise_level*, *brightness*, *harmonics*. Dataset diambil dari tahun 2010 hingga 2024 dengan jumlah 2.416 *instance*. Dataset akan digunakan sebagai data latih dan uji dalam penelitian. Berikut ini adalah contoh table dataset :

Fitur Metadata				
name_song	name_album	name_artist	release_date	release_method
Sure Thing	All I Want Is You	Miguel	2010-11-26	album
The Spins	K.I.D.S. (Deluxe)	Mac Miller	2010-08-13	album
Duvet	Twilight	bôa	2010-04-20	album
Just the Way You Are	Doo-Wops & Hooligans	Bruno Mars	2010-05-11	album
Girl With The Tattoo Enter.lewd	All I Want Is You	Miguel	2010-11-26	album
Hey, Soul Sister	Save Me, San Francisco (Golden Gate Edition)	Train	2010-12-01	album
TiK ToK	Animal (Expanded Edition)	Kesha	2010-01-01	album
Fitur Audio				
length	length_formatted	popularity	loudness	energy
195373	03:15	85	-8.127	0.607
195873	03:15	81	-3.894	0.766
203600	03:23	83	-9.414	0.888
220734	03:40	84	-5.379	0.841
102813	01:42	80	-11.314	0.239

216773	03:36	83	-4.44	0.886
199693	03:19	83	-2.718	0.837

Tabel 1 Contoh Dataset dari API Spotify

2.2 Pra Pemrosesan Data

Tahapan penelitian selanjutnya adalah pemrosesan data, pemrosesan data adalah tahapan untuk menghilangkan permasalahan yang dapat mengganggu saat pemrosesan data seperti data kosong, data yang terduplikat, data tidak konsisten . Tahapan pra pemrosesan data adalah sebagai berikut :

a. Data Cleaning

Data Cleaning atau pembersihan data yaitu data mentah kemudian diseleksi kembali. Data track yang diambil dari *Spotify* memungkinkan data memiliki *missing value* atau nilai kosong, sehingga data harus dibersihkan untuk menangani data yang tidak lengkap, tidak relevan, dan tidak akurat dengan menggunakan median (nilai tengah), mean (rata-rata), atau modus (nilai yang paling sering muncul).

b. Data Transformation

Transformasi data yaitu proses mengubah data dari satu format ke format lain untuk analisis. Transformasi data biasa dilakukan seperti normalisasi atau menyesuaikan skala data numerik agar data berada dalam rentang yang sama, misalnya mengubah skor tes dari skala 0-100 ke skala 0-1. Data yang telah dilakukan pembersihan data, kemudian di encoding atau mengubah data kategorik menjadi numerik. Encoding menggunakan one-hot encoding karena dapat mengindikasikan keberadaan fitur tertentu dalam data. encoding penting karena algoritma *Random Forest* memerlukan input numerik untuk melakukan perhitungan [20]. Data yang telah di encoding menjadikan data memiliki rentang yang berbeda, oleh sebab itu diperlukan transformasi data berupa normalisasi untuk mengubah ke rentang yang sama yaitu 0-1. Transformasi data penting untuk menyelaraskan format data dari *Spotify* dari berbagai sumber data

untuk memastikan keseragaman, sehingga memudahkan integrasi sekaligus analisis.

c. Reduksi Data

Reduksi data adalah proses pengurangan volume data dan merepresentasikannya dalam volume yang jauh lebih kecil tetapi data asli tetap terjaga. Fitur yang banyak dalam dataset dapat meningkatkan kompleksitas komputasi sehingga diperlukan reduksi data untuk mengurangi waktu pemrosesan dan membuat data lebih mudah dianalisis tanpa kehilangan informasi penting.

2.3 Pembagian Data

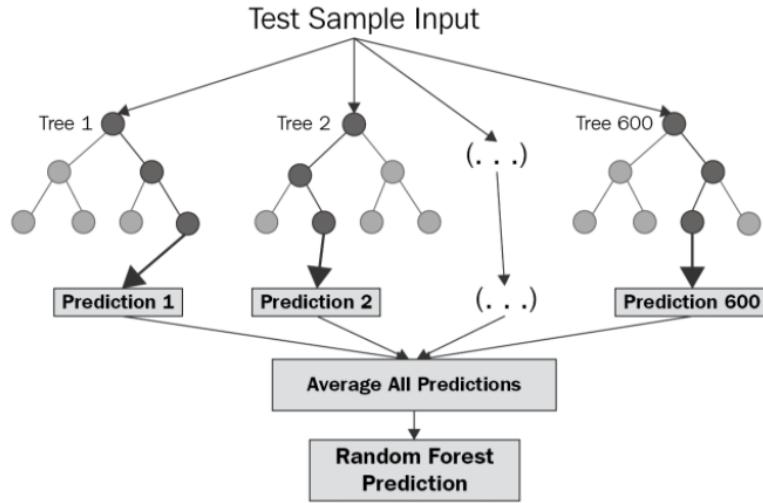
Dataset dibagi menjadi pelatihan dan pengujian, data pelatihan digunakan untuk melatih model, sedangkan data uji digunakan untuk menilai kinerja model secara keseluruhan. Pembagian data track *Spotify* dibagi sebagai berikut :

- a. 70% pelatihan dan 30% pengujian
- b. 80% pelatihan dan 20% pengujian
- c. 50% pelatihan dan 50% pengujian
- d. 60% pelatihan dan 40% pengujian

Dengan membagi data dengan proporsi secara bervariatif tersebut, diharapkan dapat mengidentifikasi masalah *overfitting* (model terlalu cocok dengan data pelatihan) dan *underfitting* (model tidak cukup kompleks untuk menangkap pola dalam data). Pembagian data yang dilakukan secara bervariatif, juga membantu untuk menentukan pembagian data yang sesuai dengan akurasi tinggi.

2.4 Metode *Random Forest*

Random Forest merupakan metode pembelajaran ensemble yang menggabungkan prediksi dari beberapa pohon keputusan untuk menghasilkan prediksi yang akurat. Penerapan *Random Forest* untuk regresi merujuk pada penggunaan model untuk memprediksi nilai kontinu berdasarkan fitur yang tersedia. Metode *Random Forest* akan dievaluasi untuk kemampuannya dalam memilih fitur yang paling relevan untuk regresi, serta untuk mengoptimalkan jumlah fitur yang digunakan dalam model, sehingga meningkatkan efisiensi dan akurasi prediksi [14].



Gambar 2 Arsitektur Random Forest Regression

Hasil dari metode *Random Forest* untuk prediksi popularitas lagu adalah nilai prediksi dalam bentuk skor kontinu.

2.5 Evaluasi Model

Tahap evaluasi model adalah untuk mengetahui seberapa baik model melakukan prediksi pada data yang belum pernah dilihat sebelumnya. Model dilatih menggunakan data pelatihan dan dievaluasi menggunakan data pengujian. Model kemudian dievaluasi kinerjanya menggunakan matrik evaluasi. Matrik evaluasi yang digunakan adalah *Mean Absolute Error (MAE)*, *Mean Squared Error (MSE)*, *Root Mean Square Error (RMSE)*, dan *R-Squared (R²)*. Evaluasi yang dilakukan selain menggunakan matrik yaitu divisualisasikan hasilnya dengan plot prediksi vs aktual untuk membandingkan nilai prediksi dengan nilai aktual untuk melihat seberapa baik model bekerja. Teknik validasi seperti K-Fold Cross Validation digunakan dalam eror analisis dengan tujuan mendapatkan estimasi yang lebih baik tentang kinerja model dengan membagi data menjadi beberapa subset dan melatih serta menguji model beberapa kali.

2.6 Error Analysis

Error analysis adalah menggunakan residual plot terhadap nilai prediksi atau variabel independen. Penerapan error analysis adalah misalkan model memprediksi akurasi R-Score sebesar 30% kemudian error 70% error tidak dapat dijelaskan oleh model. Residual plot membantu menemukan pola dalam

model seperti apabila di dalam plot terbentuk pola yang spesifik maka model tidak sepenuhnya menangkap hubungan antara fitur dan target. Sebaliknya, apabila dalam residual plot tidak membentuk pola, maka model menangkap hubungan antara fitur audio atau metadata dan popularitas lagu dengan baik.

BAB III

KESIMPULAN

Penelitian ini berfokus pada penerapan algoritma *Random Forest* untuk analisis dan prediksi popularitas lagu menggunakan data dari *Spotify*, dengan tujuan untuk memahami interaksi antarfitur yang dapat mempengaruhi popularitas lagu. Berdasarkan penelitian sebelumnya, genre yang digunakan kurang bervariatif sehingga kurang mempertimbangkan karakteristik unik dari genre lain yang dapat mempengaruhi popularitas. Oleh karena itu, penelitian ini menambahkan genre yang lebih bervariatif. Penelitian ini menambahkan evaluasi matrik untuk memberi wawasan tambahan, evaluasi matriks tersebut meliputi MAE, MSE, RMSE, dan R^2 . Dataset yang digunakan dalam penelitian ini diambil melalui teknik Crawling dari API *Spotify*, mencakup berbagai fitur audio dan metadata dari lagu-lagu yang dirilis dalam beberapa tahun terakhir.

Daftar Pustaka

- [1] D. W. Suci, “Manfaat Seni Musik Dalam Perkembangan Belajar Siswa Sekolah Dasar,” *Edukatif: Jurnal Ilmu Pendidikan*, vol. 1, no. 3, pp. 177–184, 2019, doi: 10.31004/edukatif.v1i3.45.
- [2] S. Lepa, J. Steffens, M. Herzog, and H. Egermann, “Popular music as entertainment communication: How perceived semantic expression explains liking of previously unknown music,” *Media Commun*, vol. 8, no. 3, pp. 191–201, 2020, doi: 10.17645/mac.v8i3.3153.
- [3] Prenika Yuniar, J. K. Sitoena, D. M. Matius, and G. B. Obed, “Sejarah Musik sebagai Dasar Pengetahuan dalam Pembelajaran Teori Musik,” *Clef: Jurnal Musik dan Pendidikan Musik*, vol. 3, no. 2, pp. 141–150, 2022, doi: 10.51667/cjmpm.v3i2.1098.
- [4] I. Ruddin, H. Santoso, and R. E. Indrajit, “Digitalisasi Musik Industri : Bagaimana Teknologi Informasi Mempengaruhi Industri Musik di Indonesia

- Jurnal Pendidikan Sains dan Komputer," *Jurnal Pendidikan Sains dan Komputer*, vol. 2, no. 1, pp. 124–136, 2022.
- [5] U. K-means, S. Marlia, K. Setiawan, and C. Juliane, "Analisis Fitur Musik dan Tren Popularitas Lagu di *Spotify* menggunakan K - Means dan CRISP - DM Analysis of Music Features and Song Popularity Trends on *Spotify*," vol. 13, pp. 595–607, 2024.
 - [6] L. H. TM, "Peringkat 18 Dunia, Nilai Pasar Musik *Streaming* di Indonesia Tembus Rp2,1 Triliun." [Online]. Available: <https://teknologi.bisnis.com/read/20200708/84/1262980/peringkat-18-dunia-nilai-pasar-musik-streaming-di-indonesia-tembus-rp21-triliun>
 - [7] X. Liu, "Music Trend Prediction Based on Improved LSTM and *Random Forest* Algorithm," *J Sens*, vol. 2022, 2022, doi: 10.1155/2022/6450469.
 - [8] A. I. Journal, "1 , 2 , 3 , 4," vol. 5119, no. 11, pp. 10–19, 2022.
 - [9] H. H. Lin, J. Da Lin, J. J. M. Ople, J. C. Chen, and K. L. Hua, "Social Media Popularity Prediction Based on Multi-Modal Self-Attention Mechanisms," *IEEE Access*, vol. 10, pp. 4448–4455, 2022, doi: 10.1109/ACCESS.2021.3136552.
 - [10] M. Vötter, M. Mayerl, E. Zangerle, and G. Specht, "Song Popularity Prediction using Ordinal Classification," *Proceedings of the Sound and Music Computing Conferences*, vol. 2023-June, pp. 346–353, 2023.
 - [11] Y. Xu, M. Wang, H. Chen, and F. Hu, "Prediction Model of Music Popular Trend Based on NNS and DM Technology," *Journal of Function Spaces*, vol. 2022, 2022, doi: 10.1155/2022/6104056.
 - [12] C. V. S. Araujo, M. A. P. Cristo, and R. Giusti, "A model for predicting music popularity on *streaming* platforms," *Revista de Informatica Teorica e Aplicada*, vol. 27, no. 4, pp. 108–117, 2020, doi: 10.22456/2175-2745.107021.
 - [13] X. Liu, "Music Trend Prediction Based on Improved LSTM and *Random Forest* Algorithm," *J Sens*, vol. 2022, 2022, doi: 10.1155/2022/6450469.
 - [14] E. Izquierdo-Verdiguier and R. Zurita-Milla, "An evaluation of Guided Regularized *Random Forest* for classification and regression tasks in remote sensing," *International Journal of Applied Earth Observation and Geoinformation*, vol. 88, no. June 2019, 2020, doi: 10.1016/j.jag.2020.102051.
 - [15] H. S. Saragih, "Predicting song popularity based on *Spotify*'s audio features: insights from the Indonesian *streaming* users," *Journal of Management Analytics*, vol. 10, no. 4, pp. 693–709, doi: 10.1080/23270012.2023.2239824.
 - [16] M. Votter, M. Mayerl, G. Specht, and E. Zangerle, "Novel Datasets for Evaluating Song Popularity Prediction Tasks," *Proceedings - 23rd IEEE International Symposium on Multimedia, ISM 2021*, pp. 166–173, 2021, doi: 10.1109/ISM52913.2021.00034.

- [17] A. Çimen and E. Kayış, “A longitudinal model for song popularity prediction,” *Proceedings of the 10th International Conference on Data Science, Technology and Applications, DATA 2021*, no. Data, pp. 96–104, 2021, doi: 10.5220/0010607700960104.
- [18] Z. Xing, “Popularity Prediction of Music by Machine Learning Models,” *Highlights in Science, Engineering and Technology*, vol. 47, pp. 37–45, 2023, doi: 10.54097/hset.v47i.8162.
- [19] E. Zangerle, R. Huber, M. Vötter, and Y. H. Yang, “Hit song prediction: Leveraging low- and high-level audio features,” *Proceedings of the 20th International Society for Music Information Retrieval Conference, ISMIR 2019*, pp. 319–326, 2019.
- [20] A. A. Liu *et al.*, “A review of feature fusion-based media popularity prediction methods,” *Visual Informatics*, vol. 6, no. 4, pp. 78–89, 2022, doi: 10.1016/j.visinf.2022.07.003.