

BAB 3

METODE PENELITIAN

Penelitian ini merupakan upaya untuk memperbaiki dan mengoptimalkan performa algoritma Random Forest dengan menerapkan pendekatan *Feature Selection* dan *Hyperparameter Tuning* dalam konteks klasifikasi genre musik. Langkah awal penelitian ini melibatkan pemahaman mendalam terhadap latar belakang dan tantangan yang dihadapi dalam klasifikasi genre musik, analisis terhadap proses-proses yang terlibat dalam pengklasifikasian tersebut, pengidentifikasian faktor-faktor krusial yang memengaruhi hasil klasifikasi, dan akhirnya merancang serta mengembangkan sistem yang mampu meningkatkan akurasi klasifikasi genre musik dengan memanfaatkan algoritma Random Forest yang telah dioptimalkan. Proses pengembangan sistem ini melibatkan serangkaian eksperimen yang mencakup pemilihan fitur-fitur yang relevan dengan genre musik, penyetelan *hyperparameter* dan evaluasi menyeluruh terhadap kinerja sistem yang dikembangkan.

3.1 BAHAN DAN ALAT PENELITIAN

Penelitian ini menggunakan jenis data kuantitatif yang bersumber dari data sekunder. Data sekunder merujuk pada data yang telah dikumpulkan oleh pihak lain dan tidak terkait langsung dengan objek penelitian. Data tersebut diperoleh dari sumber kedua, yang biasanya tersedia di berbagai situs web yang menyediakan akses terhadap kumpulan data global. Informasi yang dikumpulkan dan dipublikasikan dalam situs data publik tersebut dapat digunakan oleh peneliti lain untuk melakukan eksperimen atau penelitian. Sumber data pada penelitian ini merupakan dataset Spotify Tracks Dataset dari website Kaggle yang dapat diakses melalui tautan <https://www.kaggle.com/datasets/maharshipandya/-spotify-tracks-dataset>.

Peralatan yang diperlukan dalam penelitian ini adalah sebuah komputer dengan spesifikasi yang memadai untuk menjalankan sistem operasi dan perangkat lunak pengembangan, serta memiliki koneksi internet yang stabil.

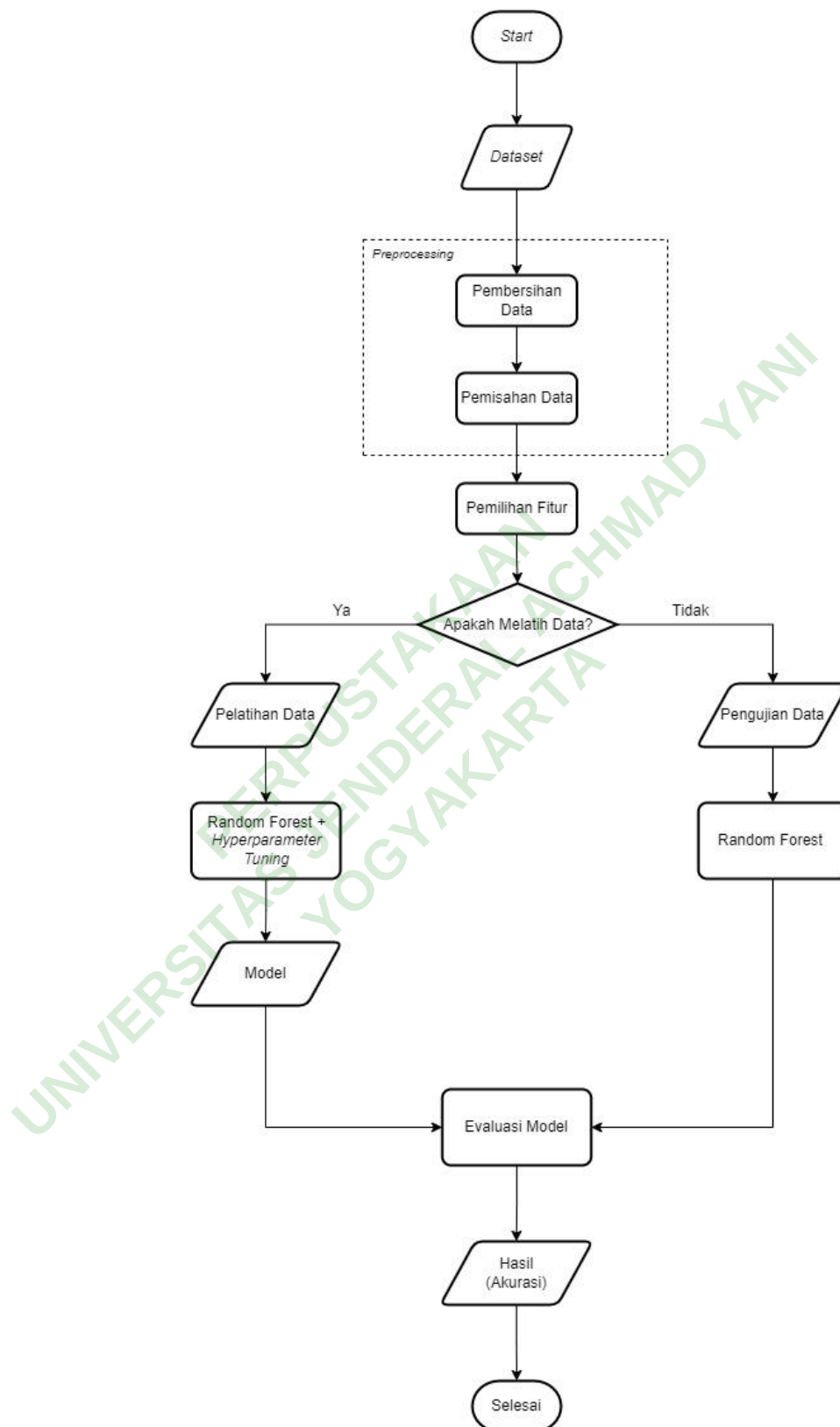
Sistem Operasi dan program-program aplikasi yang dipergunakan dalam pengembangan aplikasi ini adalah:

1. Sistem Operasi: Windows 10 atau lebih baru.
2. Bahasa pemrograman: Python dengan *library* yang digunakan diantaranya Numpy, Pandas, Matplotlib, Seaborn, Scikit-Learn, dan Flask.
3. Code editor : Visual Studio Code atau yang lebih baik.
4. Web browser : Google Chrome atau yang lebih baik.
5. Lingkungan pengembangan interaktif : Google Colab.

3.2 JALAN PENELITIAN

Penelitian ini dilakukan menggunakan bahasa pemrograman Python dengan memerlukan beberapa tahapan agar memudahkan penelitian untuk membangun mekanisme atau langkah-langkah yang harus dilakukan dalam penelitian ini. Untuk mengetahui lebih jelasnya maka akan gambarkan dalam bentuk *flowchart* pada Gambar 3.1.

Secara umum, penelitian ini dimulai dengan pengumpulan data yang akan digunakan untuk proses klasifikasi. Setelah itu data yang telah dikumpulkan dilakukan pembersihan data untuk meningkatkan kualitas data agar lebih akurat. Kemudian atribut-atribut penting yang dapat merepresentasikan kelas dipilih menggunakan *Feature Selection*. Data yang telah dibersihkan kemudian dibagi menjadi data *train*, *test* dan *validation*. Selanjutnya data tersebut dilakukan klasifikasi dengan menggunakan metode Random Forest serta menerapkan *Hyperparameter Tuning*. Evaluasi performa model dengan menggunakan metrik evaluasi ROC AUC.



Gambar 3.1 Flowchart Jalan Penelitian

3.2.1 Dataset

Data yang digunakan pada penelitian ini berupa kumpulan fitur audio dari sebuah lagu yang diperoleh dari situs Kaggle melalui tautan <https://www.kaggle.com/datasets/maharshipandya/-spotify-tracks-dataset>. Data tersebut berisikan 114.000 lagu yang memiliki atribut track_id, artists, album_name, track_name, popularity, duration_ms, explicit, danceability, energy, key, loudness, mode, speechiness, acousticness, instrumentalness, liveness, valance, tempo, time_signature dan track_genre.

3.2.2 Preprocessing

Preprocessing merupakan serangkaian proses untuk membersihkan, mengubah, dan menyesuaikan data mentah sebelum dimasukkan ke dalam model klasifikasi dengan tujuan untuk meningkatkan akurasi dan kualitas model klasifikasi. Adapun langkah yang dilakukan pada tahap ini meliputi :

1. *Cleaning data*, pembersihan data dilakukan untuk memastikan data siap untuk diproses oleh mesin. Berikut beberapa langkah *cleaning* data yang dilakukan pada dataset berbentuk numerik :
 - a. *Check missing value*, dilakukan untuk melihat data apakah memiliki nilai yang hilang atau kosong. Jika telah ditemukan data yang kosong maka akan dilakukan tindakan untuk mengatasinya seperti menghapus atribut yang memiliki nilai yang kosong.
 - b. *Class-aware data cleaning*, proses ini bertujuan untuk memilih nilai kelas yang akan digunakan pada proses klasifikasi.
2. *Splitting data*, merupakan proses memisahkan data menjadi tiga kategori yaitu data *training*, *validation*, dan *testing*.

3.2.3 Tahap pemilihan fitur (*Feature Selection*)

Feature Selection merupakan proses pemilihan subset fitur yang paling relevan dan informatif dari dataset. Tujuannya adalah untuk meningkatkan kinerja model dengan mengurangi dimensi data, kompleksitas, serta meningkatkan interpretasi hasil. Dengan memilih fitur-fitur yang paling signifikan, proses ini dapat mempercepat waktu komputasi dan meningkatkan akurasi prediksi model.

3.2.4 Tahap perancangan sistem

Implementasi algoritma dilakukan dalam bentuk program modular yang telah dirancang menggunakan bahasa pemrograman Python dengan menggunakan beberapa library seperti Pandas, Numpy, Matplotlib dan Sci-kit Learn. Aplikasi ini dijalankan pada platform *cloud computing* gratis yang diberikan oleh Google, yang dikenal sebagai Google Colab. Platform ini memungkinkan pengguna untuk mengakses lingkungan Jupyter Notebook secara *online* dan menjalankan program Python tanpa memerlukan instalasi Python atau *library* yang diperlukan pada komputer lokal. Selain itu, Google Colab menyediakan sumber daya komputasi yang cukup besar, seperti CPU, GPU, dan TPU yang dapat digunakan untuk melatih model dengan efisien.

3.2.5 Evaluasi Model

Tahap evaluasi dilaksanakan pada akhir proses klasifikasi. Tahap ini bertujuan untuk menguji model yang telah dilakukan pengujian. Metrik evaluasi yang digunakan pada penelitian ini yaitu ROC AUC. Langkah – langkah dalam proses evaluasi yaitu sebagai berikut.

1. Gunakan model yang telah dilatih untuk memprediksi probabilitas kelas positif pada data pengujian.
2. Hitung nilai True Positive Rate (TPR) dan False Positive Rate (FPR) pada berbagai threshold.
3. Plot TPR versus FPR untuk berbagai threshold untuk melihat kurva ROC.
4. Hitung Area Under Curve (AUC) dari ROC Curve untuk mendapatkan skor ROC AUC.
5. Interpretasi nilai ROC AUC membantu memahami seberapa baik model dapat membedakan antara kelas positif dan negatif.