

## BAB 3 METODOLOGI PENELITIAN

Adapun langkah-langkah untuk melakukan penelitian pada penerapan implementasi algoritma *naive bayes* dan *information gain* untuk rekomendasi jenis musik dalam mendukung kesehatan mental, terdapat metodologi penelitian yang akan dilakukan.

### 3.1 Studi Literatur

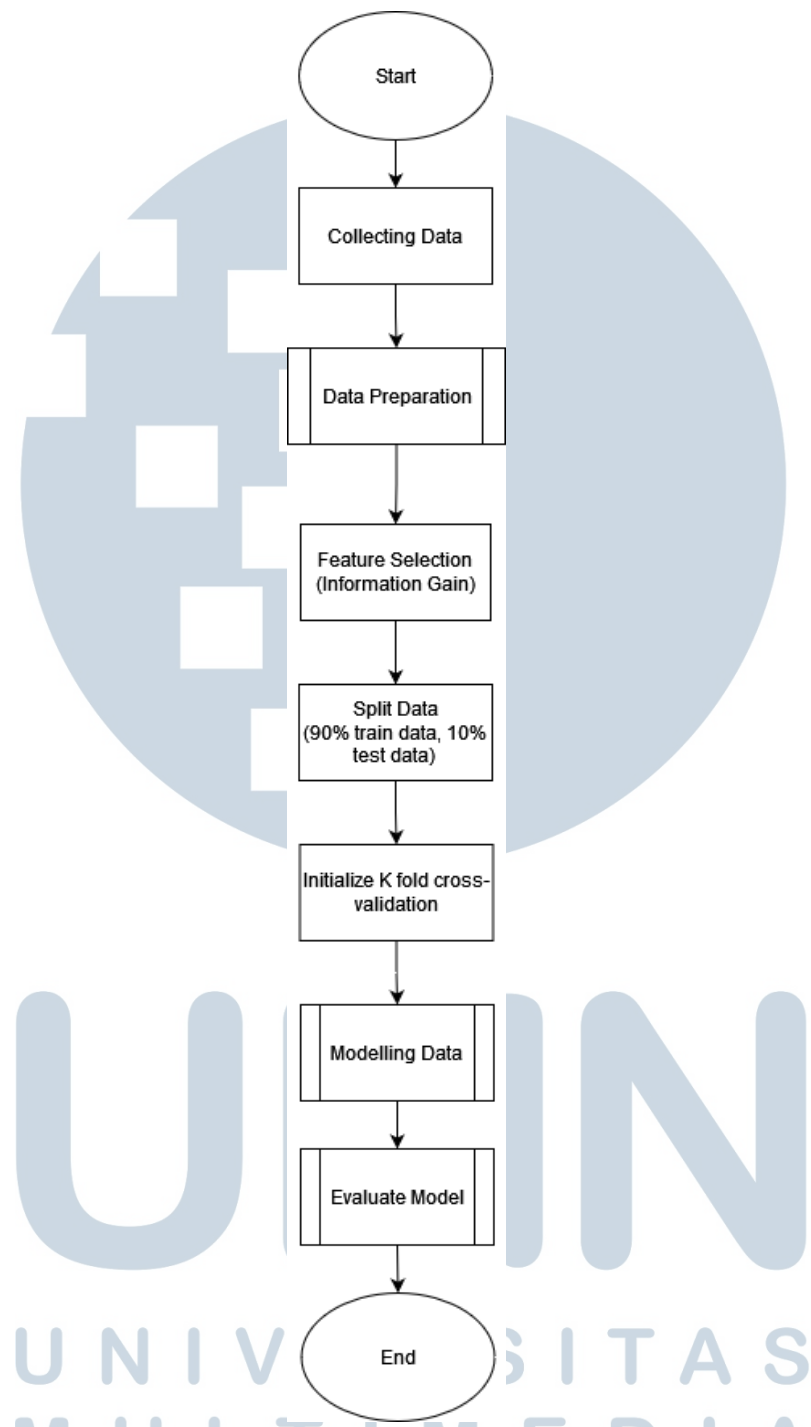
Tahap pertama dari metodologi penelitian ini yaitu studi literatur. Pada bagian ini bertujuan untuk mempelajari beberapa teori-teori yang dibutuhkan dalam proses penelitian. Adapun studi literatur yang dilakukan pada penelitian ini adalah Kesehatan Mental, Jenis Musik, *Naïve Bayes Classifier*, *Feature Selection*, *Information Gain*, dan *Confusion Matrix*.

### 3.2 Pengumpulan data

Tahap kedua yaitu pengumpulan data. Pengumpulan data dilakukan dengan memperoleh data dari *kaggle* yaitu *music & mental health survey results* pada tahun 2022. Data dari *kaggle* tersebut akan dikumpulkan dan variabel yang diambil akan disesuaikan dengan data *preprocessing*.

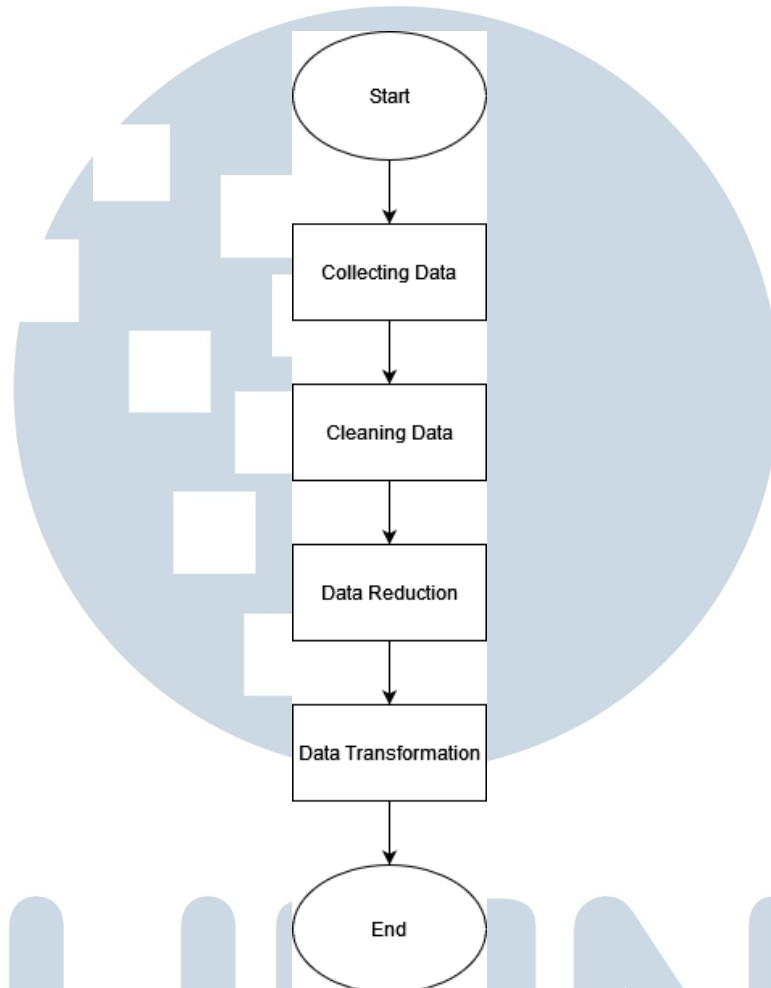
### 3.3 Perancangan Sistem

Tahap ketiga yaitu perancangan sistem. Perancangan sistem dibuat dengan membuat sebuah *flowchart* (diagram alir) yang digunakan sebagai gambaran untuk mempermudah alur sistem cara kerja penelitian ini dan juga membantu peneliti untuk membuat sistem yang menjadi lebih terstruktur sehingga dapat dikembangkan yang lebih baik pada penelitian yang akan datang. Tahapan dalam alur kerja sistem tersebut diuraikan pada gambar 3.1 ini.



Gambar 3.1. Flowchart perancangan sistem

### 3.3.1 Data Preparation



Gambar 3.2. *Flowchart data preparation*

Data *preparation* merupakan tahapan setelah melakukan pengumpulan data dari *dataset* yang sudah ada. Berikut urutan untuk data *preparation* sesuai gambar

3.2. *Flowchart data preparation*:

1. Pembersihan data: melakukan penghapusan nilai yang tidak relevan dan mengisi nilai-nilai yang hilang atau tidak valid.
2. Reduksi data: melakukan reduksi pada atribut data dan memilih serta mengambil atribut yang mendukung dalam pemodelan.
3. Transformasi data: melakukan tahap representasi dengan melakukan teknik *one hot encoding* yaitu merubah data kategorikal menjadi data numerik agar data lebih mudah untuk dibaca pada proses pemodelan *naive bayes*.

### 3.3.2 Feature Selection

Setelah melakukan data *preparation* terhadap *dataset* lama, selanjutnya yaitu melakukan proses fitur seleksi dengan menggunakan *information gain*. Fungsi dari fitur seleksi ini adalah dapat memberikan atribut atau fitur yang relevan untuk *dataset* yang baru agar dapat dilakukan proses pemodelan *naive bayes* serta melakukan perbandingan *dataset* lama dan *dataset* yang baru.

### 3.3.3 Inisialisasi Kfold Cross-Validation

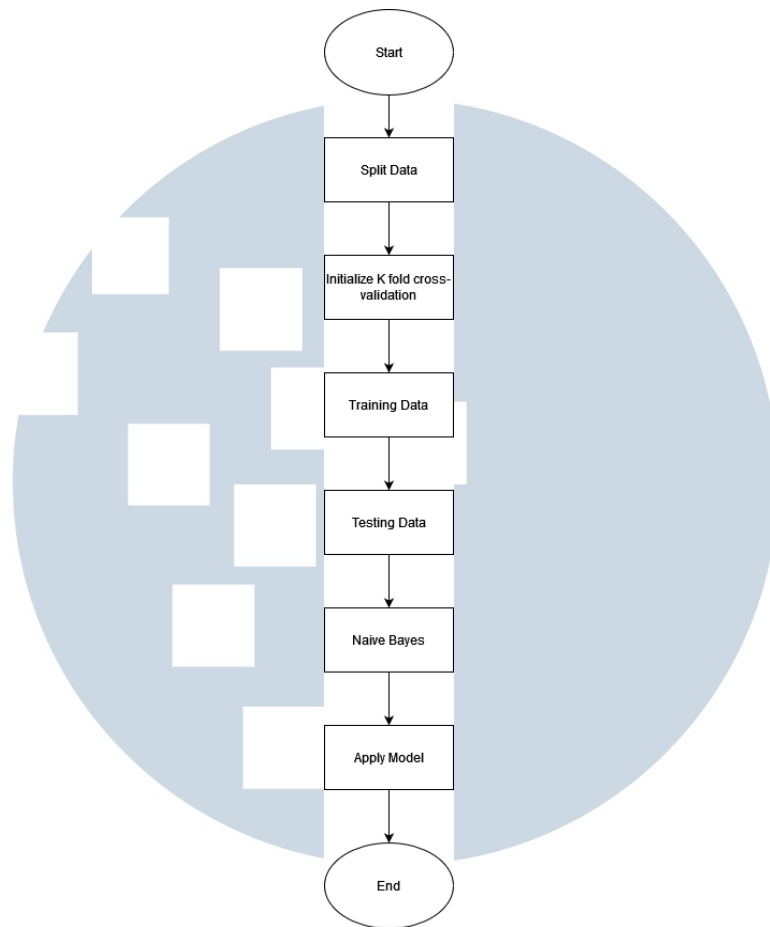
Setelah mendapatkan *dataset* yang baru, selanjutnya yaitu melakukan tahap inisialisasi *kfold cross-validation* dengan menggunakan *information gain*. Fungsi dari inisialisasi tersebut ini yaitu untuk membantu proses pemodelan data yang berbeda dan memberikan data yang seimbang agar lebih stabil dan lebih realistis. Hal ini untuk mempengaruhi hasil pengujian akurasi dan performa pada model.

### 3.3.4 Split Data

Setelah melewati *feature selection* terhadap pembuatan *dataset* baru, selanjutnya melakukan tahapan pembagian data antara *dataset* lama dan *dataset* baru yaitu sebesar 90% *train* data dan 10% *test* data. Fungsi dari pembagian data ini agar mendapatkan hasil akurasi yang baik dari *dataset*.

### 3.3.5 Modelling Data

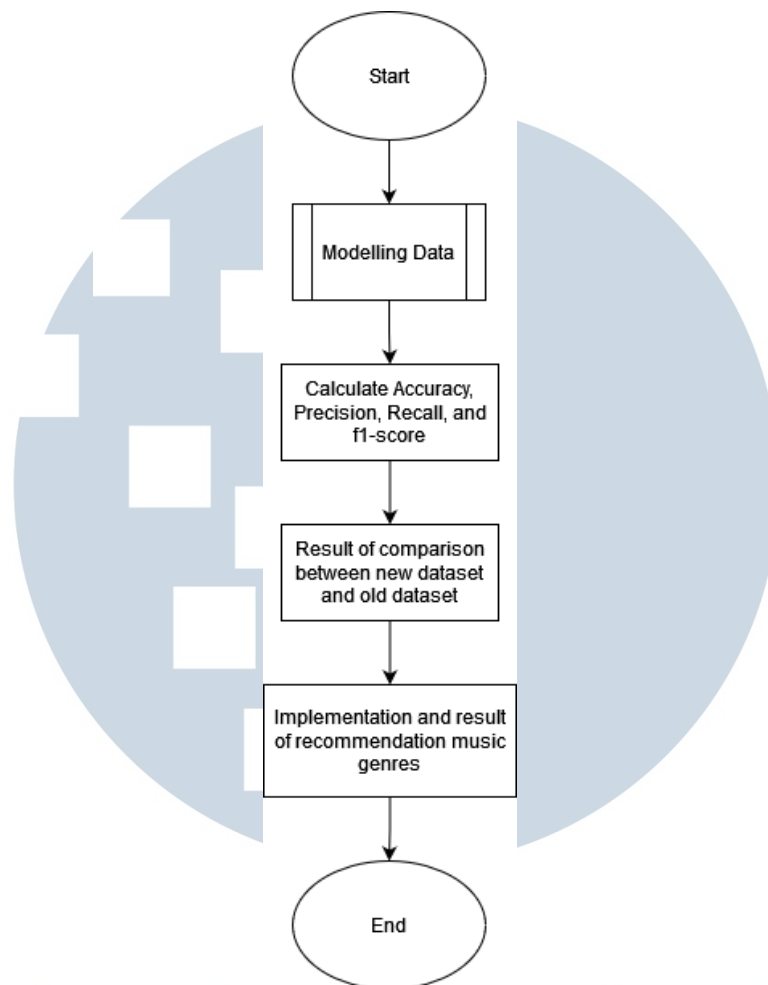
Setelah melakukan tahap *split* data, selanjutnya melakukan proses antara *training* data dan *testing* data. Pada tahap ini data akan dilakukan pemodelan pada *naive bayes* untuk *training* data, sedangkan *testing* data akan langsung diimplementasikan dan dilakukan tahapan evaluasi model. Tahapan dalam alur kerja *modelling* data tersebut diuraikan pada gambar 3.3 ini.



Gambar 3.3. Flowchart modelling data

### 3.3.6 Evaluate Model

Setelah melakukan proses model, selanjutnya yaitu melakukan tahapan perhitungan pada *confusion matrix* yaitu akurasi, presisi, *recall*, dan *f1-score*. Setelah mendapatkan hasil perhitungan, selanjutnya yaitu menampilkan hasil perbandingan antara *dataset* baru yang dilakukan proses seleksi fitur menggunakan *information gain* dan *dataset* yang lama serta proses tersebut menampilkan hasil visualisasi dan nilai persentase. Setelah menampilkan hasil perbandingan dari *dataset* tersebut, tahap terakhir yaitu mengimplementasikan hasil evaluasi model *naive bayes* dari penerapan kedua *dataset*. Tahapan dalam alur kerja *evaluate data* tersebut diuraikan pada gambar 3.4 ini.



Gambar 3.4. *flowchart evaluate model*

### 3.4 Pemograman Sistem

Tahap keempat yaitu pemrograman sistem. Pemrograman sistem dilakukan sesuai alur dari *flowchart* yang telah dibuat sebelumnya atau proses dari perancangan sistem. Kegiatan yang akan dilakukan yaitu adalah melakukan metode data *prepration* yang sudah melewati beberapa tahapan sebelumnya. Selanjutnya melakukan tahap implementasinya dengan *modelling* data dan *evaluate* model dengan tahap pemograman pada *python*. Setelah semua rangkaian perancangan dan pemograman sistem telah selesai, selanjutnya melakukan tahap pengujian dan evaluasi.

### 3.5 Pengujian dan Evaluasi

Tahap kelima yaitu pengujian dan evaluasi. Pada tahap ini dilakukan uji serta evaluasi pada proses implementasi model *Naïve Bayes* sebelumnya, setelah itu dilakukannya analisa sehingga dapat menghasilkan informasi penelitian yang menjadi tujuan dari penelitian ini. Pengujian yang dilakukan dengan melakukan evaluasi untuk mendapatkan kalkulasi hasil performa dan akurasi dari metode *confusion matrix* yaitu akurasi, presisi, *recall*, dan *F1-score*. Pengujian dan evaluasi tersebut melakukan serangkaian dua model yang berbeda agar proses implementasi, akurasi, serta performa mendapatkan hasil yang terbaik. Pada evaluasi yang dilakukan adalah proses klasifikasi untuk menerapkan dan mengimplementasikan hasil algoritma dengan melakukan klasifikasi rekomendasi jenis musik sesuai kondisi kesehatan mental. Model *naive bayes* dilatih menggunakan data training untuk memprediksi kondisi kesehatan mental berdasarkan fitur *input*. Proses klasifikasi ini menghasilkan prediksi kondisi kesehatan mental untuk setiap *instance* data. Berdasarkan hasil klasifikasi, sistem memberikan rekomendasi jenis musik yang sesuai. Proses transformasi dari hasil klasifikasi ke rekomendasi dilakukan melalui sistem mapping yang telah ditentukan sebelumnya, menghubungkan kondisi kesehatan mental dengan jenis musik yang direkomendasikan.

### 3.6 Penulisan Laporan

Tahap akhir dari metodologi penelitian yaitu penulisan hasil penelitian ke dalam laporan penelitian. Dalam membuat laporan penelitian, semua aspek pengerjaan yang dilakukan dari awal perencanaan hingga akhir penelitian akan dicantumkan dan disusun selama penelitian berlangsung.

U N I V E R S I T A S  
M U L T I M E D I A  
N U S A N T A R A