

## **BAB 3**

### **METODE PENELITIAN**

#### **3.1 ALAT DAN BAHAN**

Penelitian ini membahas mengenai pengujian parameter pada penelitian yang berjudul “IMPLEMENTASI METODE *FORWARD SELECTION* PADA ALGORITMA *NAÏVE BAYES* PADA DATA TES PSIKOLOGI MENGENAI KLASIFIKASI GAYA BELAJAR MAHASISWA”.

##### **3.1.1 Dataset**

Penelitian ini dilakukan untuk mengklasifikasikan gaya belajar mahasiswa IT Telkom Purwokerto. Sumber data pada penelitian ini diambil dari dataset bimbingan dan konseling yang berisi data tes psikologi mahasiswa IT Telkom Purwokerto. Data tersebut berupa sampel data bimbingan dan konseling mahasiswa IT Telkom Purwokerto tahun 2020. Data ini memiliki atribut berupa nama mahasiswa, program studi, 36 aspek pertanyaan terbagi ke dalam tiga tipe pertanyaan mengenai tes psikologi gaya belajar dimana didalamnya sudah berisi hasil inputan dari masing-masing mahasiswa, serta hasil yang merupakan jenis gaya belajar mahasiswa.

##### **3.1.2 Perangkat Keras (*Hardware*)**

Perangkat keras yang digunakan dalam penelitian ini yaitu sebuah laptop dengan spesifikasi sebagai berikut:

1. Intel® Core™ i5-8250U CPU @ 1.60GHz
2. Windows 11 (64-bit)
3. RAM 7,88 GB

##### **3.1.3 Perangkat Lunak (*Software*)**

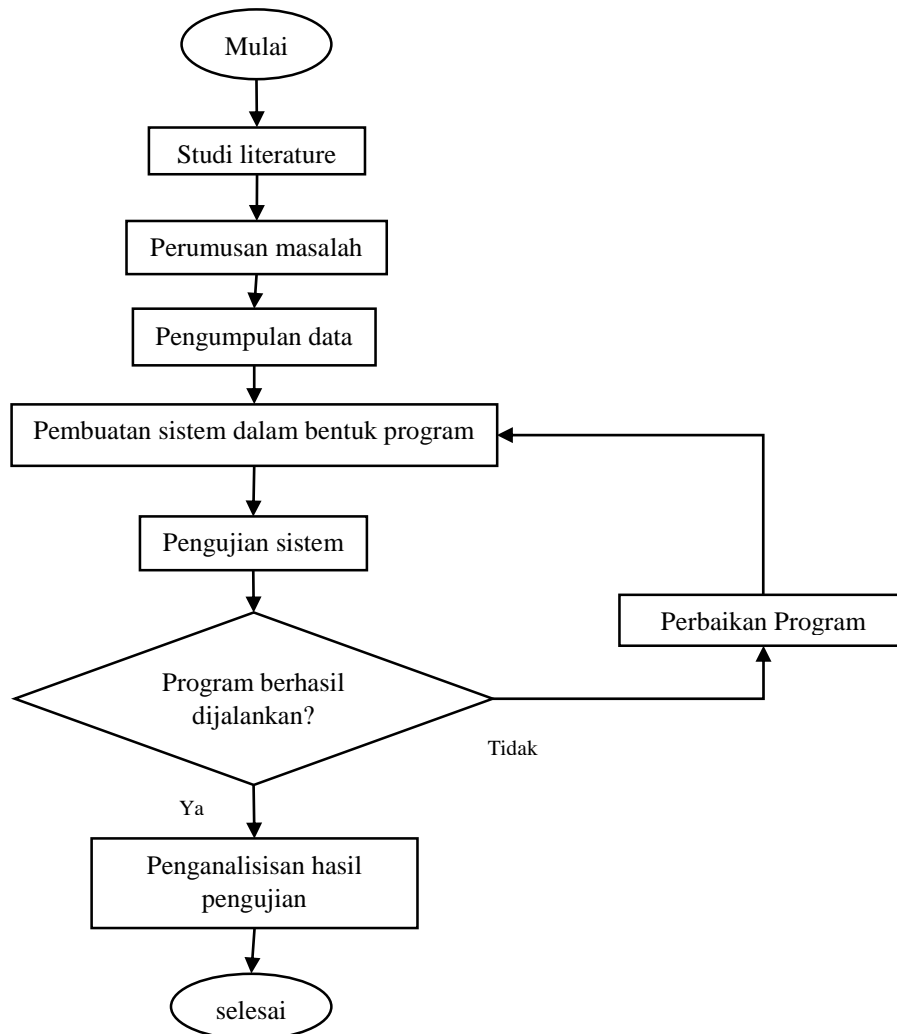
Perangkat lunak yang digunakan dalam pembuatan sistem pengklasifikasian gaya belajar mahasiswa adalah sebagai berikut:

1. *Google Colaboratory*

## 2. Python 3.10.0

### 3.2 ALUR PENELITIAN

Penelitian ini dilakukan dalam beberapa tahap, yaitu studi literatur, perumusan masalah, pengumpulan data, perancangan sistem dalam bentuk program, pengujian program, analisis hasil pengujian. Alur penelitian ini dapat dilihat pada gambar berikut.



**Gambar 3. 1** *Flowchart* Alur Penelitian

Alur penelitian yang akan dilakukan dalam tugas akhir ini dimulai dari studi literatur. Studi literature dilakukan untuk mencari sumber yang berkaitan dengan penelitian *naïve bayes* menggunakan metode *forward selection*. Penelitian ini menggunakan data berupa data bimbingan dan konseling IT Telkom Purwokerto

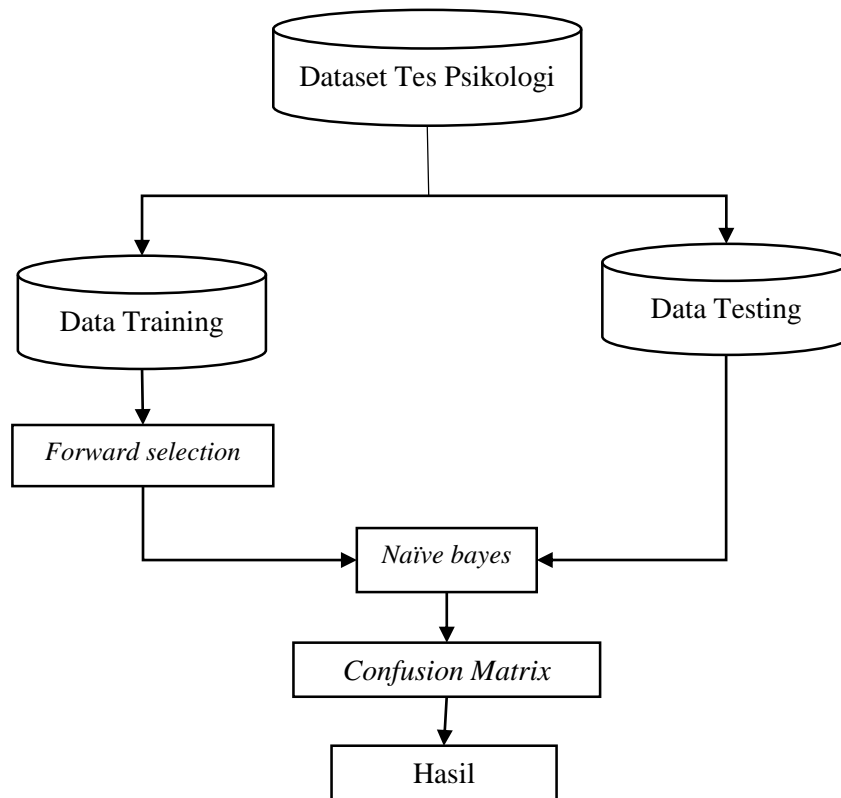
tahun 2020 yang terdiri dari 600 mahasiswa (responden). Atribut yang digunakan dalam klasifikasi berupa 36 aspek pertanyaan yang terbagi ke dalam tiga tipe pertanyaan mengenai tes psikologi gaya belajar dimana didalamnya sudah berisi inputan dari masing-masing mahasiswa, serta hasil yang merupakan jenis gaya belajar. Penelitian ini menggunakan skala pengukuran berupa skala likert. Skala likert digunakan untuk mengukur sikap, pendapat, dan persepsi seseorang mengenai fenomena sosial atau dalam penelitian disebut sebagai variabel penelitian. Dalam penelitian ini menggunakan skala yaitu 1 untuk jarang, 2 untuk kadang-kadang, dan 3 untuk sering.

Dalam penelitian data dibagi menjadi dua bagian yaitu berupa data training dan data testing dengan perbandingan 80% data latih yaitu 480 data dan 20% data uji yaitu 120 data. Penyeleksian atribut dilakukan dengan menggunakan fitur seleksi *forward selection*. Sistem yang akan dibuat merupakan sistem yang berbentuk kode program ditulis menggunakan bahasa pemrograman *Python 3.10.0*. Setelah rancangan sistem selesai dibuat, langkah selanjutnya rancangan sistem akan diuji untuk memastikan sistem berjalan sesuai dengan apa yang diinginkan atau tidak. Sistem berhasil dijalankan apabila sudah mendapatkan akurasi terbaik. Jika program tidak berhasil dijalankan maka akan dilakukan perbaikan program dan diujikan kembali. Pengujian ini bertujuan untuk mengetahui kerja dari algoritma *naïve bayes* dalam mengklasifikasikan data ke dalam kelas yang telah ditentukan. Apabila program berhasil dijalankan maka dilanjutkan ke tahap analisis pengujian.

Penganalisisan hasil pengujian dilakukan dengan menggunakan pengukuran evaluasi berupa *confusion matrix*. *Confusion matrix* merupakan sebuah metode berbentuk diagram yang berisikan informasi aktual mengenai perkiraan klasifikasi sebuah sistem dengan menggunakan data berbentuk matriks sebagai hasilnya. *Confusion matrix* digunakan untuk mempermudah menganalisis performa algoritma selain itu juga dapat digunakan untuk menampilkan informasi dalam bentuk angka sehingga bisa dilakukan perhitungan rasio keberhasilan klasifikasi.

### **3.3 DESAIN PENELITIAN**

Berikut ini merupakan desain penelitian dalam bentuk *flowchart*, yaitu:



**Gambar 3. 2 Flowchart Desain Penelitian**

Pengumpulan data dilakukan dengan cara mengambil dataset data bimbingan dan konseling yang berupa data tes psikologi mahasiswa IT Telkom Purwokerto tahun 2020 sebanyak 600 data. Data penelitian ini berupa data numerik. Proses pemisahan dataset terbagi kedalam dua jenis data yaitu data *training* dan data *testing*. Data akan dibagi menjadi dua dengan perbandingan 80% data *training* dan 20% data *testing*.

Proses klasifikasi menggunakan algoritma *naïve bayes*, hasil proses klasifikasi tersebut kemudian dievaluasi menggunakan *confusion matrix*. *Confusion matrix* adalah tabel dengan empat kombinasi berbeda dari nilai prediksi dan nilai aktual. Nilai prediksi merupakan keluaran dari program dimana nilainya *Positive* dan *Negative*, sedangkan nilai aktual merupakan nilai sebenarnya dimana nilainya *True* dan *False*. Terdapat empat istilah yang merupakan representasi dari hasil proses klasifikasi *confusion matrix*, keempat istilah tersebut adalah *True Positive (TP)*, *True Negative (TN)*, *False Positive (FP)*, dan *False Negative (FN)*.

### 3.4 PENGUJIAN PENELITIAN

Pengujian penelitian mempunyai tujuan untuk menjelaskan teknik pengujian yang digunakan dalam penelitian. Teknik pengujian yang digunakan berupa tahap modeling klasifikasi gaya belajar mahasiswa menggunakan metode *forward selection* dan algoritma *naïve bayes*. Sedangkan penganalisisan hasil pengujian dilakukan dengan menggunakan pengukuran evaluasi berupa *confusion matrix*.

#### 3.4.1 *Forward Selection*

Metode *forward selection* dilakukan dengan cara memasukkan predictor secara bertahap. Dalam metode ini, variabel berprediktor yang dimasukkan dalam model tidak dapat dikeluarkan lagi. Proses tersebut akan dihentikan ketika prediktor baru tidak bisa lagi meningkatkan secara signifikan terhadap variable respon. Prosedur *forward selection* menjadi salah satu prosedur pemilihan terbaik dalam regresi dengan eliminasi variable bebas yang membangun model secara bertahap. Dalam metode ini, terdapat beberapa langkah yang dapat digunakan dalam pengujian. Langkah-langkah tersebut adalah sebagai berikut:

1. Menentukan model awal
2. Uji F terhadap peubah pertama yang dipilih

Jika ***Fhitung*** < ***Ftabel*** maka peubah terpilih dibuang dan proses dihentikan; Apabila ***Fhitung*** > ***Ftabel*** maka peubah terpilih memiliki pengaruh nyata terhadap peubah terkait y; sehingga layak untuk diperhitungkan di dalam model.

3. Masukkan peubah bebas terpilih ke dalam model
4. Uji F, jika ***Fhitung*** < ***Ftabel*** maka proses dihentikan dan model terbaik adalah model sebelumnya, namun jika ***Fhitung*** ≥ ***Ftabel*** variabel peubah bebas layak untuk dimasukan ke dalam model dan kembali ke langkah 3. Proses akan berakhir jika tidak ada lagi peubah yang bisa dimasukan ke dalam model.

#### 3.4.2 *Naïve Bayes*

Beberapa tahapan dari proses algoritma *naïve bayes*:

1. Menghitung jumlah kelas atau label,

2. Menghitung jumlah kasus per kelas,
3. Kalikan semua *variable* kelas,
4. Bandingkan hasil per kelas.

### 3.4.3 *Confusion Matrix*

Pada dasarnya *confusion matrix* memberikan informasi perbandingan dari hasil klasifikasi yang dilakukan oleh sistem (model) dengan hasil klasifikasi sebenarnya. *Confusion matrix* digunakan untuk mengukur nilai akurasi, presisi, dan *recall*.

Akurasi menggambarkan seberapa kuat model dapat mengklasifikasikan dengan benar. Berikut ini merupakan formulasi menghitung nilai akurasi:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} = \frac{TP \text{ Total}}{Dataset \text{ Total}}$$

Presisi menggambarkan tingkat akurasi antara data yang diminta dengan hasil prediksi yang diberikan oleh model. Presisi menunjukkan tingkat ketepatan atau ketelitian dalam pengklasifikasian. Berikut ini merupakan formulasi menghitung presisi:

$$Precision = \frac{TP}{TP + FP} = \frac{TP}{Prediction \text{ Total}}$$

*Recall* menggambarkan keberhasilan model dalam menemukan kembali sebuah informasi. *Recall* mempunyai fungsi untuk mengukur proporsi positif aktual yang benar diidentifikasi. Berikut ini merupakan formulasi menghitung *recall*:

$$Recall = \frac{TP}{TP + FN} = \frac{TP}{Actual \text{ Total}}$$