

## **BAB III**

### **METODE PENELITIAN**

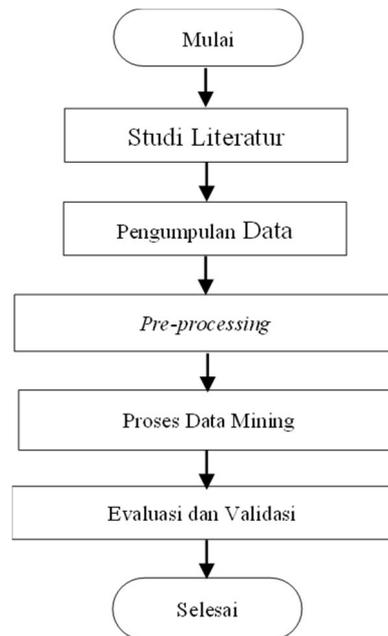
#### **3.1 Alat dan Bahan Penelitian**

Pada Penelitian ini peneliti menggunakan alat dan bahan sebagai berikut :

1. Perangkat Keras (*Hardware*)
  - a. Laptop Asus Vivobook Core I7
  - b. Hardisk 1 TB
  - c. Wifi Bizznet 100 Mbps
2. Perangkat Lunak (*Software*)
  - a. Website Kaggle.com
  - b. Google Colab
  - c. OS Windows 11
  - d. Ms. Word 2019
  - e. Ms. Excel 2019
  - f. Ms. Power Point 2019

#### **3.2 Tahapan Penelitian**

Pada bab ini akan membahas langkah-langkah dari proses penelitian yang akan dilaksanakan. Dalam melakukan analisa dan mencari pola data untuk dijadikan sebuah dataset dalam memudahkan penelitian dan dapat berjalan dengan sistematis dan memenuhi tujuan, maka dibuat alur dalam tahapan penelitian yang akan dilakukan sebagai berikut:



Gambar 3. 1 Tahapan Penelitian

### 3.2.1. Studi Literature

Pada tahap ini peneliti melakukan review jurnal yang berkaitan dengan judul penelitian yang akan diteliti oleh peneliti. Jurnal-jurnal yang menjadi rujukan tentu memiliki ruang lingkup dan metode yang sama dilakukan oleh peneliti

### 3.2.2. Pengumpulan Dataset

Data yang digunakan merupakan dataset publik berupa data yang diperoleh dari Kaggle dengan link <https://www.kaggle.com/datasets/yasserh/breast-cancer-dataset>. yang terdiri dari informasi demografis, kebiasaan, dan catatan medis historis. Beberapa pasien memutuskan untuk tidak menjawab beberapa pertanyaan karena masalah privasi (missing value).

### **3.2.3. Pre-processing dan Pengolahan Data (Cleaning Data).**

Dalam tahap preprocessing dan pengolahan data yang dilakukan di Google Colab, beberapa langkah penting diambil untuk mempersiapkan dataset agar siap digunakan dalam analisis lebih lanjut:

#### **a) Normalisasi Data**

Normalisasi dilakukan untuk menskalakan fitur-fitur dalam dataset agar berada dalam rentang yang seragam. Proses ini penting untuk memastikan bahwa model tidak bias terhadap fitur dengan skala yang berbeda. Dalam hal ini, normalisasi dilakukan dengan menggunakan StandardScaler dari pustaka sklearn, yang menerapkan standarisasi fitur dengan menghilangkan rata-rata dan menskalakan varians.

#### **b) Penghapusan Fitur yang Tidak Berguna**

Fitur yang dianggap tidak memberikan kontribusi signifikan terhadap model dihapus dari dataset. Penghapusan fitur yang tidak relevan ini membantu menyederhanakan model dan mengurangi risiko overfitting. Langkah ini dilakukan dengan mengidentifikasi dan menghapus fitur yang tidak memberikan informasi tambahan yang berarti untuk proses klasifikasi.

#### **c) Penanganan Data yang Hilang**

Data yang hilang dapat memengaruhi hasil analisis dan kualitas model. Untuk mengatasi hal ini, metode imputasi digunakan untuk menggantikan nilai yang hilang dengan nilai rata-rata fitur yang bersangkutan. Proses ini dilakukan menggunakan SimpleImputer dari pustaka sklearn, yang mengisi data yang hilang dengan nilai rata-rata.

### **3.2.3.1 Seleksi Fitur RFE Recursive Feature Elimination (RFE)**

Seleksi fitur menggunakan Recursive Feature Elimination (RFE) dilakukan untuk meningkatkan performa model dengan memilih fitur-fitur yang paling relevan. RFE bekerja dengan cara menghapus fitur yang kurang penting secara bertahap dan memilih subset fitur terbaik untuk model.

#### **a) Inisialisasi dan Pelatihan RFE**

RFE diinisialisasi dengan menggunakan model RandomForestClassifier sebagai estimator. Selanjutnya, RFE diterapkan pada dataset untuk memilih jumlah fitur yang diinginkan. Proses ini melibatkan pelatihan model pada subset data dan menghilangkan fitur yang memberikan kontribusi terkecil terhadap performa model.

#### **b) Evaluasi Hasil Seleksi Fitur**

Setelah proses seleksi fitur selesai, fitur yang dipilih dievaluasi untuk menentukan fitur-fitur yang paling berpengaruh dalam model. Fitur yang dipilih merupakan fitur-fitur dengan kontribusi terbesar terhadap kinerja model, dan hasil evaluasi ini digunakan untuk melanjutkan tahap analisis berikutnya.

### **3.2.4 Proses Data Mining**

Pada tahap ini, Anda akan menerapkan algoritma-algoritma klasifikasi yang telah dipilih, yaitu ANN, dan KNN, pada data kanker payudara yang telah diproses

sebelumnya. Setiap algoritma akan digunakan untuk membangun model klasifikasi berdasarkan data yang ada.

### 3.2.5 Evaluasi dan Validasi

Setelah model klasifikasi dibangun, langkah selanjutnya adalah mengevaluasi kinerja masing-masing model. Evaluasi dilakukan dengan menggunakan metrik-metrik seperti akurasi, presisi, recall, dan F1-score. Hal ini membantu peneliti memahami seberapa baik setiap model dalam mengklasifikasikan kanker payudara.

Tahap terakhir adalah validasi, di mana peneliti menguji keandalan dan generalisasi dari model-model yang telah dibangun. Validasi dapat dilakukan dengan menggunakan metode cross-validation atau pembagian dataset menjadi subset training dan testing. Tujuannya adalah untuk memastikan bahwa model-model yang dibangun tidak hanya berperforma baik pada data training, tetapi juga pada data baru yang belum pernah dilihat sebelumnya.

#### 1. Akurasi (Accuracy)

Akurasi mengukur seberapa sering model membuat prediksi yang benar secara keseluruhan.

$$\text{Akurasi} = \frac{\text{Jumlah Prediksi Benar}}{\text{Total Jumlah Data}} = \frac{TP + TN}{TP + TN + FP + FN}$$

## 2. Presisi (Precision)

Presisi mengukur seberapa banyak prediksi positif yang benar dari semua prediksi positif yang dihasilkan oleh model.

$$\text{Presisi} = \frac{\text{Jumlah Prediksi Positif Benar}}{\text{Total Prediksi Positif}} = \frac{TP}{TP + FP}$$

## 3. Recall (Sensitivitas atau True Positive Rate)

Recall mengukur seberapa baik model mengidentifikasi semua sampel positif sebenarnya dari keseluruhan data positif.

$$\text{Recall} = \frac{\text{Jumlah Prediksi Positif Benar}}{\text{Total Kasus Positif}} = \frac{TP}{TP + FN}$$

### Keterangan:

- **TP (True Positive):** Prediksi positif yang benar (model memprediksi positif dan hasilnya memang positif).
- **TN (True Negative):** Prediksi negatif yang benar (model memprediksi negatif dan hasilnya memang negatif).
- **FP (False Positive):** Prediksi positif yang salah (model memprediksi positif tetapi hasilnya negatif).
- **FN (False Negative):** Prediksi negatif yang salah (model memprediksi negatif tetapi hasilnya positif).