

BAB III

METODE PENELITIAN

3.1 Subjek dan Objek Penelitian

2.1.1 Subjek Penelitian

Subjek pada penelitian ini adalah metode seleksi fitur menggunakan *Recursive Feature Elimination* (RFE) dalam diagnosis kanker payudara.

2.1.2 Objek Penelitian

Objek pada penelitian ini adalah dataset kanker payudara WDBC (*Wisconsin Diagnostic Breast Cancer*) yang digunakan untuk proses klasifikasi antara kelas kanker yang bersifat jinak (*benign*) dan yang bersifat ganas (*malignant*).

3.2 Alat dan Bahan

3.2.1 Alat

Alat yang digunakan dalam penelitian ini dapat dikelompokkan menjadi dua jenis, yaitu perangkat keras (*hardware*) dan perangkat lunak (*software*) yang dapat dilihat pada Tabel 3.1 dan Tabel 3.2.

Tabel 3. 1 Spesifikasi Perangkat Keras

| Perangkat Keras | Tipe/Ukuran |
|-----------------|---|
| Device | Hp Pavilion 15 |
| Processor | AMD A10-8700P Radeon R6, 10 Compute Cores 4C+6G 1.80 GHz. |
| Memory | 8.00 GB |

Tabel 3. 2 Spesifikasi Perangkat Lunak

| Perangkat Lunak | Kegunaan |
|------------------------|---|
| Windows 8.1 Pro 64-bit | Sistem Operasi |
| DeepL | Menerjemahkan referensi berbahasa Inggris |

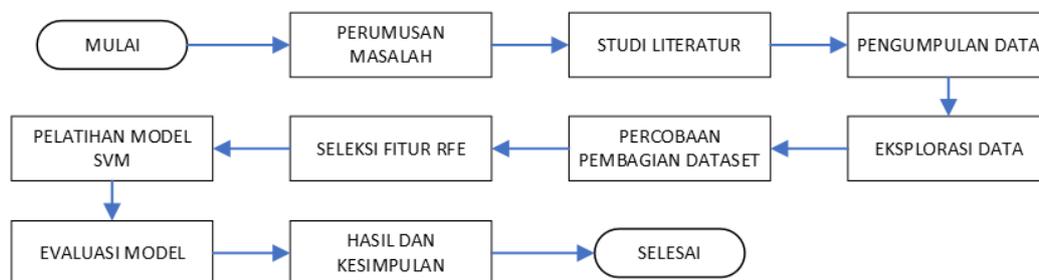
| Perangkat Lunak | Kegunaan |
|----------------------|---------------------------------|
| MathType | Menulis persamaan matematis |
| Mendeley | Membuat sitasi otomatis |
| Microsoft Visio | Membuat diagram alir |
| Microsoft Word | Menulis laporan penelitian |
| Google Scholar | Mencari sumber referensi ilmiah |
| Google Collaboratory | Melakukan pemrograman |
| Google Docs | Memeriksa kesalahan penulisan |

3.2.2 Bahan

Data yang digunakan dalam penelitian ini adalah seluruh data dari dataset WDBC selain kolom ID. Totalnya mencakup 569 baris data dengan 30 kolom fitur dan 1 kolom target yang berisi label kelas jinak dan ganas.

3.3 Diagram Alir Penelitian

Penelitian ini melibatkan beberapa tahapan. Berikut adalah gambaran alur kerja penelitian:



Gambar 3. 1 Diagram Alir Penelitian

3.3.1 Perumusan Masalah

Tahap perumusan masalah adalah awal dari proses penelitian dimana penulis mengidentifikasi masalah yang akan dijadikan fokus penelitian. Tahap ini mencakup pertanyaan penelitian, menentukan batasan-batasan yang relevan, tujuan penelitian, dan mempertimbangkan metode yang tepat untuk diimplementasikan dalam penelitian.

3.3.2 Studi Literatur

Pada tahap studi literatur ini, penulis melakukan pencarian dan analisis berbagai sumber referensi yang relevan dengan topik penelitian mengenai diagnosis kanker payudara dan penggunaan metode *Recursive Feature Elimination* (RFE). Referensi yang digunakan mencakup jurnal ilmiah, buku, artikel, serta literatur terkait lainnya yang mendukung sebagai dasar penelitian. Tujuan utamanya adalah mengidentifikasi kontribusi penelitian ini terhadap pengembangan ilmu pengetahuan dan sosial, baik dalam hal teori maupun metode penelitian yang digunakan sebelumnya.

3.3.3 Pengumpulan Data

Tahap pengumpulan data dalam penelitian ini melibatkan penggunaan dataset kanker payudara WDBC (*Wisconsin Diagnostic Breast Cancer*) yang diperoleh dari UCI Machine Learning Repository. Dataset ini digunakan untuk melakukan klasifikasi antara kanker payudara yang bersifat jinak (*benign*) dan yang bersifat ganas (*malignant*).

3.3.4 Eksplorasi Data

Dataset yang telah berhasil dikumpulkan kemudian memasuki tahap eksplorasi. Pada tahap ini, dataset akan diperiksa untuk memastikan kondisinya, seperti keberadaan data yang hilang, adanya duplikasi data, dan bagaimana tipe datanya. Tujuan dari tahap ini adalah untuk memahami kondisi keseluruhan dataset yang akan digunakan serta merencanakan langkah-langkah yang perlu diambil sebelum menggunakannya dalam proses klasifikasi.

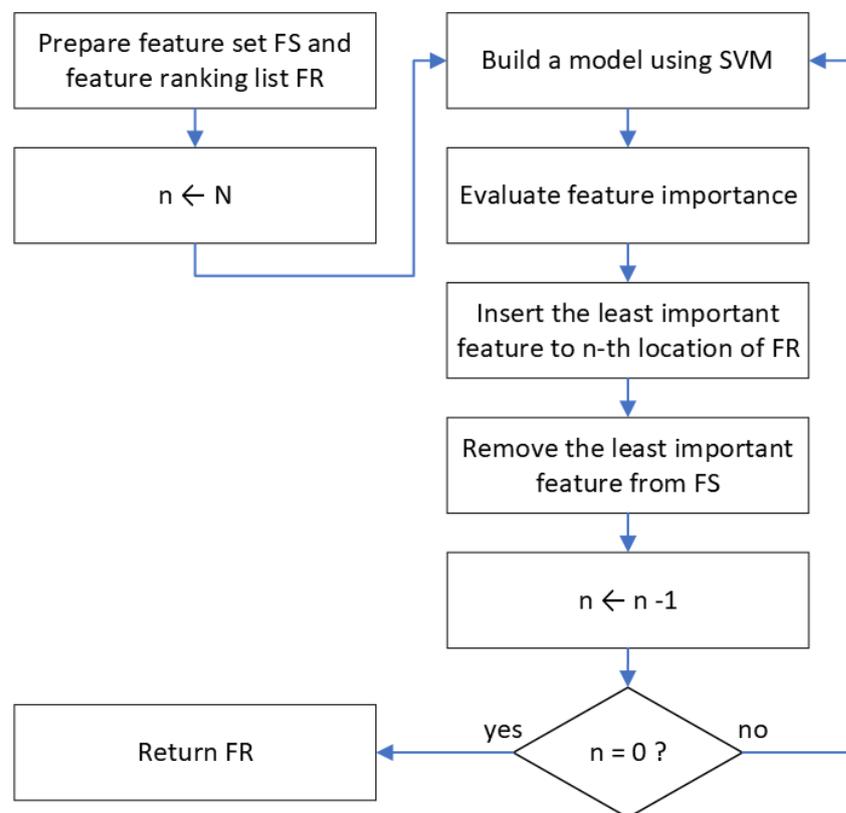
3.3.5 Percobaan Pembagian Dataset

Pada tahap ini, dataset akan dibagi untuk persiapan pelatihan model. Dalam penelitian ini, dilakukan pembagian dataset menggunakan empat skenario berbeda: skenario pertama dengan perbandingan 60% data untuk pelatihan dan 40% untuk pengujian, skenario kedua dengan perbandingan 70% data untuk pelatihan dan 30% untuk pengujian, skenario ketiga dengan

perbandingan 80% data untuk pelatihan dan 20% untuk pengujian, serta skenario terakhir dengan perbandingan 90% data untuk pelatihan dan 10% untuk pengujian [5]. Tujuan dari eksperimen ini adalah untuk memilih skenario terbaik yang akan digunakan dalam proses klasifikasi.

3.3.6 Seleksi Fitur RFE

Setelah mendapatkan skenario terbaik untuk pembagian dataset, selanjutnya adalah proses seleksi fitur terhadap dataset awal melalui metode RFE, seperti yang ditunjukkan dalam Gambar 3.2.



Gambar 3. 2 Proses RFE [35]

Gambar 3.2 menunjukkan bagaimana RFE (*Recursive Feature Elimination*) melatih seluruh fitur dalam dataset yang tersimpan dalam *Feature Set* (FS) menggunakan basis model klasifikasi SVM untuk menghitung bobot fitur yang mencerminkan tingkat kepentingan masing-masing fitur. Perhitungan bobot mengikuti persamaan (2.1), dan kemudian

fitur diurutkan berdasarkan bobot menggunakan persamaan (2.2). Fitur dengan nilai bobot terendah dipindahkan ke dalam *Feature Ranking* (FR) dan dihapus dari FS. Proses ini berlanjut hingga jumlah fitur (n) mencapai nilai 0, yang menunjukkan bahwa tidak ada lagi fitur yang tersisa dalam FS untuk dilatih. Hasilnya adalah semua fitur diurutkan dalam FR berdasarkan tingkat kepentingan fitur.

3.3.7 Pelatihan Model SVM

Pada tahap ini, dilakukan pelatihan model klasifikasi menggunakan *Support Vector Machine* (SVM). Setelah mendapatkan ranking fitur dari *Recursive Feature Elimination*, langkah selanjutnya adalah eksperimen pengurangan fitur secara bertahap. Dimulai dengan 30 fitur, penelitian ini bertujuan untuk mengamati dampak pengurangan fitur dengan peringkat rendah terhadap akurasi model. Pengurangan dilakukan secara bertahap dengan kelipatan 5, yaitu dari 30 fitur menjadi 25, 20, 15, 10, hingga 5 fitur.

3.3.8 Evaluasi Model

Pada tahap ini dilakukan evaluasi untuk menilai seberapa baik model klasifikasi yang dibangun dengan algoritma SVM dalam mengklasifikasikan data. Evaluasi dilakukan untuk membandingkan kinerja model ketika menggunakan seluruh fitur dengan menggunakan sebagian fitur yang dianggap relevan oleh RFE. Penilaian kinerja model akan berfokus pada metrik akurasi, presisi, *recall*, dan *F1-score*.

3.3.9 Hasil dan Kesimpulan

Tahap terakhir penelitian ini melibatkan analisis hasil klasifikasi menggunakan algoritma SVM. Evaluasi model mempertimbangkan akurasi, presisi, *recall*, dan *F1-score*. Hasil analisis ini memberikan kesimpulan mengenai pengaruh penggunaan RFE terhadap akurasi dalam diagnosis kanker payudara, serta perbandingan kinerja model dengan dan tanpa seleksi fitur RFE. Tahap ini menyimpulkan fokus penelitian yang diajukan.