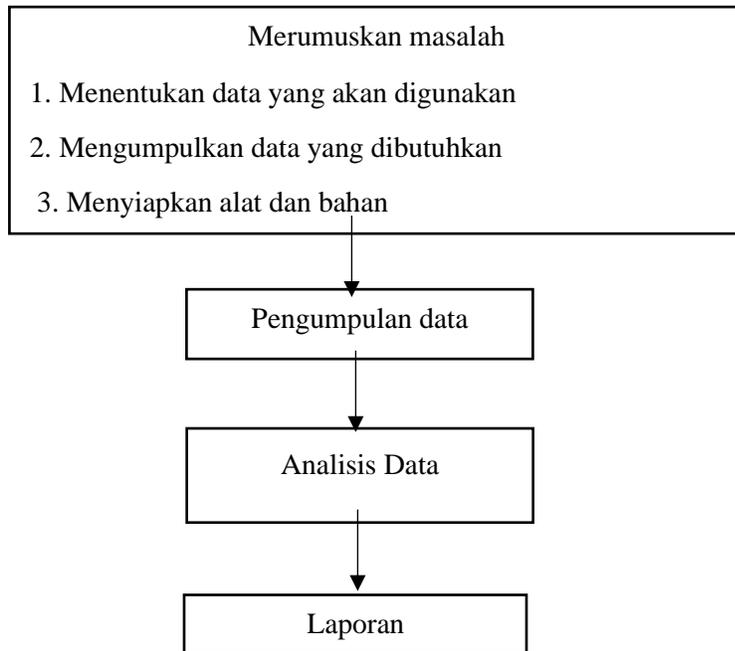


BAB III

METODOLOGI PENELITIAN

3.1 Metode 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 untuk memudahkan penelitian dan dapat berjalan dengan sistematis sehingga memenuhi tujuan yang diinginkan maka dibuat alur tahapan penelitian secara detail pada gambar 3.1 :



Gambar 3.1 alur tahapan penelitian

3.2 Teknik Pengumpulan Data

Berikut adalah beberapa teknik pengumpulan data yang dilakukan pada penelitian ini:

1. Observasi

Melakukan observasi digital terhadap permasalahan bertujuan untuk mengumpulkan informasi yang akurat melalui analisis konten daring, data daring, dan jejak digital terkait. Dengan melakukan observasi ini, peneliti dapat mengidentifikasi masalah

secara jelas dan terperinci. Langkah ini diperlukan agar tujuan penelitian dapat dicapai dengan lebih efektif, memastikan bahwa analisis dilakukan berdasarkan data yang relevan.

2. Mengumpulkan Literature

peneliti ini juga melakukan pengumpulan literatur melalui paper, buku, jurnal, dan penelitian lain yang relevan dengan **Algoritma C4.5 dan K-Nearest Neighbors (KNN) Dalam Diagnosa *Postpartum Depression***. Tujuan dari pengumpulan literatur ini adalah untuk mendapatkan pemahaman yang mendalam tentang konsep, dan perkembangan terbaru dari algoritma-algoritma tersebut serta pemahaman tentang hubungan antara algoritma dan ***Postpartum Depression*** Metode ini digunakan untuk mencari sumber – sumber data yang diperlukan dalam penelitian yang biasa diperoleh dari membaca dan mengutip referensi serta jurnal penelitian yang pernah dilakukan sebelumnya yang mendukung dan berkaitan dengan penelitian. Adapun data yang di peroleh adalah data pasien penderita postpartum depression.

3. Mengumpulkan Data

Dalam penelitian ini, peneliti memanfaatkan dataset publik yang tersedia di platform Kaggle sebagai sumber data utama. Sebelum menggunakan dataset dari Kaggle, yaitu data peneliti melakukan pemeriksaan terhadap validitas serta keandalannya. Langkah ini menjadi penting guna memastikan bahwa data yang digunakan sesuai dengan tujuan penelitian dan dapat diandalkan. Hasil pemeriksaan tersebut membantu memastikan integritas data sebelum dilibatkan dalam analisis statistik atau eksplorasi lebih lanjut.

3.3 Penerapan Data Mining

3.3.1 Penerapan Algoritma *K-Nearest Neighbor* (K-NN)

Penerapan algoritma *K-Nearest Neighbor* (KNN) dalam klasifikasi kanker payudara melibatkan beberapa langkah berikut:

1. Persiapan Data:

Kumpulkan dataset yang berisi informasi tentang pasien, termasuk fitur-fitur yang relevan untuk deteksi kanker payudara seperti usia, ukuran tumor, bentuk tumor, kepadatan, dan sebagainya. Lakukan preprocessing data seperti membersihkan data dari missing values atau outlier, dan transformasi data jika diperlukan.

2. Pembagian Data:

Bagi dataset menjadi dua subset: data training dan data testing.

Data training digunakan untuk melatih model KNN, sedangkan data testing digunakan untuk menguji kinerja model yang telah dilatih.

3. Normalisasi Fitur:

Jika fitur-fitur pada dataset memiliki skala yang berbeda, normalisasikan fitur-fitur tersebut ke dalam rentang yang seragam. Hal ini penting untuk memastikan bahwa semua fitur memiliki pengaruh yang seimbang pada perhitungan jarak dalam algoritma KNN.

4. Pemilihan Parameter:

Pilih nilai K yang optimal, yaitu jumlah tetangga terdekat yang akan digunakan dalam pengklasifikasian.

Pemilihan nilai K dapat dilakukan melalui eksperimen dan validasi silang (cross-validation) untuk menentukan performa terbaik.

5. Perhitungan Jarak:

Hitung jarak antara data testing dengan setiap data training menggunakan metrik jarak seperti Euclidean, Manhattan, atau metrik jarak lainnya.

Pilih K tetangga terdekat berdasarkan jarak yang terkecil.

6. Voting:

Berdasarkan K tetangga terdekat, tentukan kelas mayoritas dari tetangga tersebut.

Gunakan metode voting (misalnya, majority voting) untuk menentukan kelas prediksi untuk data testing.

7. Evaluasi Model:

Evaluasi kinerja model KNN menggunakan metrik evaluasi seperti akurasi, presisi, recall, atau F1-score.

Gunakan data testing yang memiliki label kelas yang diketahui untuk membandingkan hasil prediksi dengan nilai sebenarnya dan mengukur kinerja model.

3.3.2 Penerapan Algoritma Decision Tree (C4.5)

Pada bagian ini, akan dilakukan penerapan algoritma Decision Tree dengan menggunakan metode C4.5 untuk mengklasifikasi kanker payudara. Algoritma C4.5 merupakan salah satu algoritma yang sering digunakan dalam data mining untuk melakukan klasifikasi atau pengelompokan berdasarkan aturan-aturan yang dihasilkan dari pohon keputusan.

Penerapan algoritma Decision Tree (C4.5) pada penelitian ini memberikan keuntungan berupa visualisasi dari pemecahan masalah yang diolah menggunakan teknik data mining. Pohon keputusan yang dihasilkan dari algoritma ini memungkinkan pengguna untuk memahami langkah-langkah yang diambil dalam proses klasifikasi dan prediksi. Hal ini membuat protokol dari prediksi dapat diamati secara visual, sehingga memudahkan interpretasi hasil dan pengambilan keputusan.

Pohon keputusan sendiri telah banyak digunakan dalam berbagai bidang ilmu pengetahuan, termasuk dalam bidang kesehatan untuk diagnosa penyakit pasien, ilmu komputer dalam analisis struktur data, psikologi dalam teori pengambilan keputusan, dan banyak bidang lainnya. Dalam konteks klasifikasi kanker payudara, pohon keputusan berdasarkan algoritma C4.5 dapat membantu dokter atau peneliti dalam melakukan klasifikasi pasien berdasarkan atribut-atribut yang relevan dan menghasilkan prediksi mengenai kemungkinan keberadaan kanker payudara.

Pada pohon keputusan, cabang-cabang yang terbentuk merupakan pertanyaan klasifikasi yang digunakan untuk memisahkan data menjadi kelompok-kelompok yang lebih homogen. Setiap cabang mewakili aturan-aturan yang digunakan untuk mengelompokkan data. Sementara itu, daun-daun pohon keputusan merupakan kelas-kelas atau segmen-segmen yang ditemukan setelah proses klasifikasi selesai.

Rumus yang digunakan dalam algoritma C4.5 untuk menghitung entropy merupakan salah satu kunci utama dalam pembentukan pohon keputusan. Entropy digunakan untuk mengukur tingkat ketidakpastian atau

ketidaksempurnaan di dalam data. Dengan menggunakan rumus entropy, algoritma C4.5 akan memilih atribut yang paling informatif untuk dibagi dan membentuk cabang-cabang pada pohon keputusan.

Dengan melakukan penerapan algoritma Decision Tree (C4.5) pada penelitian ini, diharapkan dapat diperoleh model yang efektif untuk mengklasifikasi kanker payudara berdasarkan atribut-atribut yang relevan. Pohon keputusan yang dihasilkan akan memberikan gambaran yang jelas dan dapat dimengerti tentang aturan-aturan yang digunakan untuk klasifikasi, sehingga dapat membantu dalam pengambilan keputusan medis yang lebih akurat dan cepat.

3.4 Jadwal Pelaksanaan Penelitian

Jadwal pelaksanaan penelitian ini dapat dilihat pada Tabel 3.1 berikut:

Tabel 3.1 Penjadwalan

No.	Kegiatan Penelitian	Tahun 2023/2024				
		Okt	Nov	Des	Jan	Feb
1	Penyusunan dan Pengajuan Judul	■	■			
2	Pengajuan Proposal	■	■			
3	Seminar Proposal			■		
4	Pengumpulan Data			■		
5	Analisis Data			■	■	
6	Penyusunan Laporan Tesis			■	■	■
7	Sidang Tesis					■

Keterangan:

■ : Telah dilakukan.

■ : Akan dilakukan.

3.5 Alat dan Bahan

A. Alat

Penelitian ini menggunakan perangkat keras Laptop Acer Aspire 314 36p C7NO Intel Inside Ghz, RAM 08 Gigabyte, VGA Nvidia 2Gb dengan Ruang penyimpanan berupa Solid State Disk kapasitas 265 Gb, sedangkan perangkat lunak yang terlibat adalah system operasi Windows 10, Microsoft Excel dan Rapidminer Studio sebagai aplikasi yang digunakan untuk pengolahan data.

B. Bahan

Sebagai bahan penelitian : adalah data set yang berasal dari dataset public berisi data sejumlah 1514 data record, berupa data pasien penderita *postpartum depression* dari berbagai latar belakang yang beraneka ragam, yang terbagi menjadi 11 attribut.