

BAB III

METODE PENELITIAN

3.1 Objek Penelitian

Dalam penelitian ini, dataset yang digunakan terdiri dari 569 data dengan 32 fitur dalam format .csv (UCI, 2017). Adapun 32 fitur merupakan hasil dari Fine Needle Aspiration (FNA) atau Aspirasi Jarum Halus pada massa payudara. 32 fitur yang ada digunakan untuk menggambarkan karakteristik pada inti sel dari kanker payudara. Pada dataset terdapat juga fitur target yang memberikan informasi diagnosis dari jenis kanker payudara, yaitu Malignant (ganas) atau Benign (jinak). Adapun dataset kanker payudara ditunjukkan pada Tabel 3.1.

Tabel 3. 1 Data Kanker Payudara

No	id	diagnosis	radius_mean	texture_se	fractal_dimension_worst
0	842302	M	17.99	0.9053	0.1189
1	842517	M	20.57	0.7339	0.0890
2	84300903	M	19.69	0.7869	0.0875
3	84348301	M	11.42	1.156	0.173
4	84358402	M	20.29	0.7813	0.0767
...					
564	926424	M	21.56	1.256	0.0711
565	926682	M	20.13	2.463	0.0663
566	926954	M	16.6	1.075	0.0782
567	927241	M	20.6	1.595	0.124
568	92751	B	7.76	1.428	0.0703

Sumber : (UCI, 2017)

3.1.1 Waktu dan Tempat Penelitian

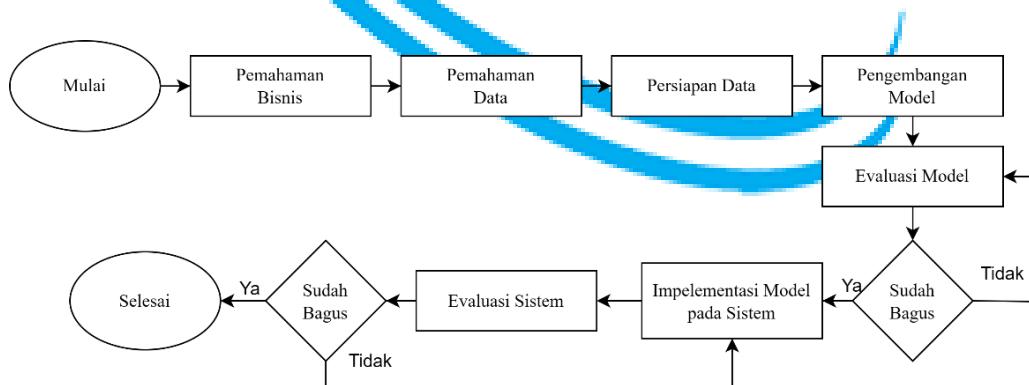
Penelitian ini dilakukan di Laboratorium Riset Universitas Buana Perjuangan Karawang untuk membuat laporan. serta RS Kanker Dharmais sebagai tempat validasi pakar. Adapun penelitian dilakukan dari bulan Februari 2023 sampai Desember 2023 yang ditunjukkan pada Tabel 3.2.

Tabel 3. 2 Waktu Penelitian

No	Kegiatan	2023									
		Feb	Mar	Apr	Mei	Jun	Jul	Okt	Nov	Des	
1.	Studi Literatur										
2.	Pemahaman Data										
3.	Persiapan Data										
4.	Model										
5.	Evaluasi										
6.	Deployment										

3.2 Prosedur Penelitian

Penelitian ini, metode data mining yang digunakan adalah tahapan *Cross Industry Standard Process for Data Mining* (CRIPS-DM). Tahapan dimulai dari pemahaman bisnis dan data, yang dilanjutkan dengan mempersiapkan data menggunakan beberapa teknik. Adapun proses yang dilakukan yaitu pembersihan data, normalisasi data, transformasi data, dan seleksi fitur. Setelah persiapan data selesai, model dikembangkan menggunakan algoritma Logistic Regression, yang kemudian dievaluasi menggunakan metode *Confusion Matrix*. Pada tahapan terakhir, model akan diimplementasikan ke dalam sistem. Pengembangan sistem ini menggunakan pendekatan pengembangan RAD dan diimplementasikan menggunakan *framework* Flask, sistem juga diujikan menggunakan *blackbox testing* untuk memastikan kinerjanya. Adapun alur dari prosedur penelitian ditunjukkan pada Gambar 3.1.



Gambar 3. 1 Alur Prosedur Penelitian

3.2.1 Pemahaman Bisnis

Berdasarkan risiko kematian yang terjadi pada penderita kanker payudara, maka penelitian ini dilakukan untuk mengembangkan sistem berbasis *web*. Sistem yang dibuat dapat mengklasifikasikan jenis kanker payudara berdasarkan data yang dimasukan. Hal ini diharapkan dapat membantu para petugas kesehatan dalam memberikan penangan yang lebih cepat kepada pasien penderita kanker payudara.

3.2.2 Pemahaman Data

Pemahaman data dilakukan untuk menjelaskan berbagai aspek data, termasuk atribut, kelas, dan tipe data. Dalam penelitian ini, dataset yang digunakan memiliki 32 fitur. Fitur dalam data ini dikelompokkan lebih lanjut menjadi *radius*, *tekstur*, *perimeter*, *area*, *smoothness*, *compactness*, *concavity*, *concave points*, *symmetry*, dan *fractal dimension*. Pada setiap fitur dibagi menjadi Mean (rata-rata), Standar Deviasi (se), dan kondisi terburuk (worst). Adapun penjelasan dari fitur-fitur ditunjukkan pada Tabel 3.3.

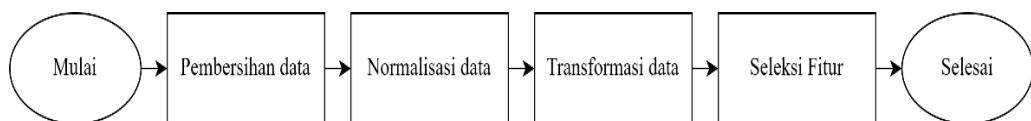
Tabel 3. 3 Deskripsi Atribut Kanker Payudara

Fitur	Deskripsi
<i>id</i>	Nomor pengenal untuk setiap sampel
<i>diagnosis</i>	Kategori dari jenis kanker payudara, Malignan = ganas dan Benign = jinak
<i>radius</i>	Rata-rata jarak dari pusat titik di sekitar tepi inti sel
<i>texture</i>	Standar deviasi dari nilai abu-abu dari inti sel
<i>perimeter</i>	Keliling dari inti sel
<i>area</i>	Luas dari inti sel
<i>smoothness</i>	Variasi dalam Panjang radius inti sel
<i>compactness</i>	Proporsi area inti sel terhadap area lingkaran dengan radius yang sama
<i>concavity</i>	Tingkat kedalaman bagian cekung dari kontur inti sel
<i>concave points</i>	Jumlah bagian cekung dari kontur inti sel
<i>symmetry</i>	Simetri dari inti sel
<i>fractal</i>	Dimensi fractal dari inti sel
<i>dimension</i>	

Sumber : (UCI, 2017)

3.2.3 Persiapan Data

Persiapan data dilakukan untuk mendapatkan kualitas data yang baik. Penelitian ini menggunakan beberapa teknik dengan tahapan yang ditunjukkan pada Gambar 3.2.



Gambar 3. 2 Alur persiapan data

1. Pembersihan Data

Pada tahap ini, dilakukan pengecekan dan pembersihan data dari *missing values*, duplikasi data, *outlier*, dan *noise*. Kemudian dilakukan juga perubahan tipe data dari ‘object’ menjadi ‘category’ pada kolom fitur diagnosis. Hal ini dilakukan untuk memastikan kualitas data sebelum dilanjutkan pada tahap selanjutnya.

2. Normalisasi data

Tahap normalisasi ini dilakukan untuk menyesuaikan skala nilai atribut pada data, sehingga data dapat terletak pada suatu rentang tertentu. Proses ini diperlukan ketika nilai data memiliki variasi yang signifikan. Pada penelitian ini, normalisasi diterapkan menggunakan teknik *Min-Max scaling*, yang melibatkan transformasi linear pada data.

3. Transformasi data

Proses transformasi dilakukan untuk mengubah atribut diagnosis menjadi ‘numerik’ menggunakan teknik *label encoding*. Label yang diberikan yaitu 0 sebagai Benign (kankae jinak) dan 1 sebagai Malignan (kanker ganas).

4. Seleksi Fitur

Dalam proses seleksi fitur, dilakukan pemilihan fitur yang sesuai dari sekumpulan fitur, dengan mengurangi fitur yang tidak relevan. Tujuan dari pemilihan fitur adalah agar algoritma yang digunakan dapat memproses data lebih efisien, sehingga hasil klasifikasi dapat dicapai lebih cepat.

3.2.4 Pengembangan Model

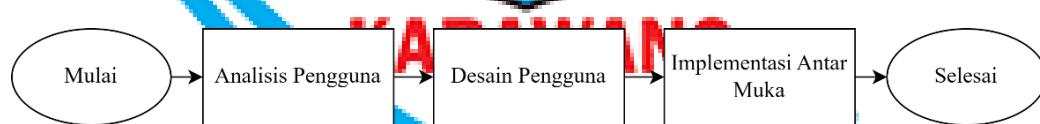
Pengembangan model dilakukan setelah data dianggap layak digunakan. Pada tahap ini, teknik SMOTE diterapkan untuk mentasi permasalahan ketidak seimbangan jumlah data, sehingga distribusi jumlah data menjadi lebih merata. Selanjutnya, Algoritma Logistic Regression akan diimplementasikan pada model klasifikasi.

3.2.5 Evaluasi Model

Model yang telah dibuat akan dievaluasi guna mengukur kinerjanya. Pada penelitian ini, evaluasi model akan menggunakan metode *Confusion Matrix* dengan menerapkan *K-fold Cross Validation*. *K-fold Cross Validation* digunakan untuk membagi data menjadi set pengujian dan set peatihan, hal ini dilakukan untuk mencari hasil terbaik dalam hal akurasi, presisi, dan recall.

3.2.6 Implementasi Model pada Sistem

Setelah model selesai dibuat dan dievaluasi, selanjutnya akan diterapkan pada sistem berbasis *web*. Sistem yang dibuat akan dikembangkan menggunakan pendekatan RAD untuk implementasi yang cepat dan efisien. Tahapan pengembangan sistem ditunjukkan pada Gambar 3.3.



Gambar 3. 3 Tahapan Pengembangan Sistem

1. Analisis Pengguna

Pada analisis pengguna dilakukan identifikasi pengguna, kebutuhan fungsional dan non fungsional, serta *usecase* pengguna. Adapun identifikasi pengguna ditunjukkan pada Tabel 3.4, serta kebutuhan fungsional ditunjukkan pada Tabel 3.5.

Tabel 3. 4 Identifikasi Pengguna

No	Pengguna	Keterangan
1.	Tenaga Medis	Tenaga Medis merupakan pengguna yang dapat melakukan klasifikasi

Berdasarkan Tabel 3.4 identifikasi pengguna adalah Tenaga Medis yang dapat menggunakan sistem dan melakukan klasifikasi.

Tabel 3. 5 Kebutuhan Fungsional

No	Deskripsi Kebutuhan Fungsional	Aksi
1.	Sistem dapat menampilkan hasil klasifikasi	Klasifikasi

Berdasarkan Tabel 3.5 mengenai kebutuhan fungsional, sistem beroperasi dengan melakukan proses klasifikasi setelah data dimasukkan. Hasil yang ditampilkan berupa diagnosis jenis kanker payudara, yang dibagi menjadi dua kategori yaitu Malignan (ganas) dan Benign (jinak). Selanjutnya, terdapat kebutuhan non fungsional yaitu sistem berbasis web yang dapat digunakan di berbagai perangkat lunak *web browser*, serta sistem dapat digunakan dengan mudah.

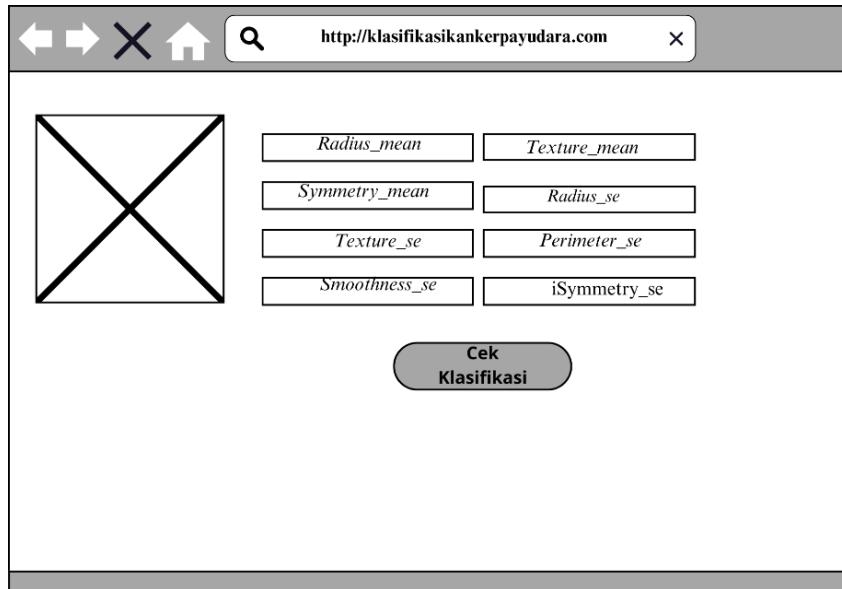
2. Desain Pengguna

Pada desain pengguna terdapat *usecase diagram* yang menggambarkan pengguna dari sistem, serta *mockup* dari tampilan sistem yang akan dibuat.



Gambar 3. 4 Usecase Diagram Klasifikasi Kanker Payudara

Berdasarkan Gambar 3.4 Tenaga medis dapat melakukan klasifikasi kanker payudara dengan menginput data pasien kedalam sistem.



Gambar 3.5 Mockup Tampilan Sistem

Berdasarkan Gambar 3.5 menunjukkan tampilan desain dari sistem klasifikasi kanker payudara, dengan fungsi utama yaitu cek klasifikasi. Sistem akan beroperasi ketika petugas medis memasukkan data pasien berdasarkan kolom yang tersedia. Setelah data dimasukkan, sistem akan mengolahnya dan menampilkan hasil klasifikasi.

3. Implementasi Antar Muka

Pada tahap ini, model yang telah dibuat akan diimplementasikan pada sistem berbasis *web*. Dalam proses implementasi, sistem dibuat dengan menggunakan *framework* Flask.

3.2.7 Evaluasi Sistem

Proses terakhir dilakukan evaluasi sistem untuk mengatahui kinerja sistem. Pada penelitian ini, sistem akan dievaluasi menggunakan pengujian *blackbox* dan uji validasi pakar. Adapun proses pengujian *blackbox* tersebut mencakup pengujian setiap fungsi yang tersedia pada sistem. Selanjutnya, uji validasi pakar dilakukan oleh pakar atau dokter Onkologi menggunakan data uji. Kemudian, hasil dari validasi pakar akan dilakukan penerapan perhitungan akurasi manual. Hal tersebut dilakukan untuk mengetahui keakuratan sistem tersebut dalam mengklasifikasikan kanker payudara (Ganas atau Jinak).