

BAB III METODE PENELITIAN

3.1. Objek Penelitian

Objek pada penelitian ini menggunakan data penyakit kanker payudara. Adapun data tersebut memiliki jumlah 569 data dan 32 atribut (Buddini w, 2016). Kemudian, data tersebut dibagi menjadi dua kelas yaitu kanker jinak (*benign*) dan kanker ganas (*malignant*). Dataset tersebut ditunjukkan pada Tabel 3.1.

Tabel 3. 1 Dataset Kanker Payudara

id	diagnosis	radius _mean	...	concave_points_ worst	symmetry_ worst	fractal_ dimension_worst
853201	M	17.57	...	1.456	2.756	7.919
853401	M	18.63	...	1.848	3.444	9.782
853612	M	11.84	...	1.546	4.761	1.402
...
854268	M	14.25	...	1.447	3.591	1.014
854941	B	13.03	...	5.013	1.987	6.169
855133	M	14.99	...	2.899	1.565	5.504

3.2. Lokasi dan Waktu Penelitian

Lokasi penelitian ini dilakukan di Universitas Buana Perjuangan Karawang yang dimulai dari Februari 2023. Adapun rincian waktu penelitian ditunjukkan pada Tabel 3.2.

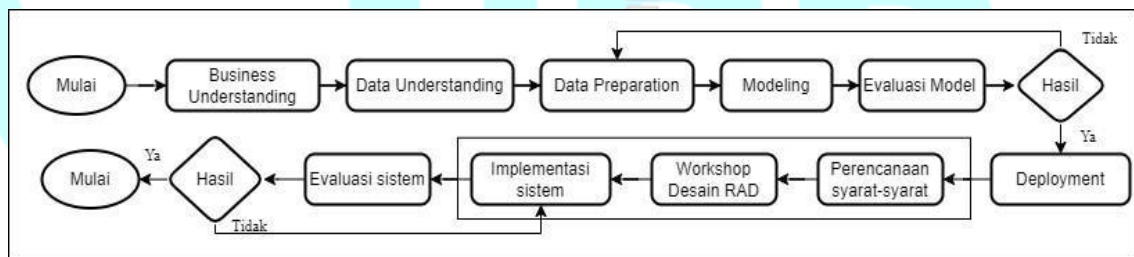
Tabel 3. 2 Rincian waktu penelitian

No	Kegiatan	2023				
		Feb	Mar	Apr	Mei	Jun
1.	<i>Bussiness Understanding</i>					
2.	<i>Data Understanding</i>					
3.	<i>Data Preparation</i>					
4.	<i>Modeling</i>					

No	Kegiatan	2023				
		Feb	Mar	Apr	Mei	Jun
5.	Evaluasi					
6.	Deployment					
	a. Perencanaan Syarat-syarat					
	b. Workshop Desain RAD					
	c. Implementasi					
7.	Evaluasi Sistem					

3.3. Prosedur penelitian

Prosedur penelitian ini dimulai dengan tahap *business understanding*, *data understanding*, *data preparation*, *modeling*, evaluasi model, *deployment* dan Evaluasi sistem. Adapun prosedur penelitian ini ditunjukkan pada Gambar 3.1.



Gambar 3. 1 Prosedur penelitian

3.3.1. Business Understanding

Berdasarkan prevalensi penderita kanker payudara dinegara berkembang yang meningkat dengan cepat, maka penelitian ini dilakukan untuk mengimplementasikan model klasifikasi jenis kanker payudara pada sistem berbasis web menggunakan Algoritma SVM dengan teknik SMOTE. Hal ini diharapkan dapat membantu para tenaga medis dalam menangani penderita penyakit kanker payudara dengan cepat dan tepat.

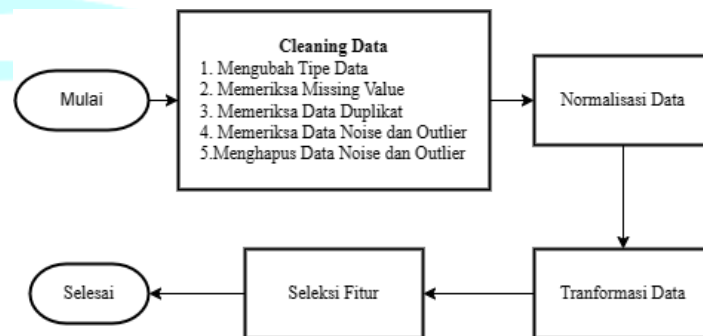
3.3.2. Data Understanding

Pada tahap ini, pengumpulan dataset dilakukan pada penelitian ini. Adapun dataset yang digunakan diperoleh dari *Kaggle* dengan jumlah 569 dan 32 atribut. Selanjutnya, berdasarkan Tabel 3.1 dataset tersebut dibagi menjadi tiga fitur yaitu

Mean (rata-rata), Standar deviasi (*se*), dan terburuk (*worst*). Kemudian, fitur tersebut diuraikan pada setiap kolom dengan berbagai ukuran yang terdiri dari *radius*, *teksture*, *perimeter*, *area*, *smoothness*, *compactness*, *concavity*, *concave points*, *symmetry*, dan *fractal dimension*.

3.3.3. Data Preparation

Pada tahap ini dilakukan *preprocessing* data dengan membagi beberapa tahapan. Adapun tahapan tersebut ditunjukkan pada Gambar 3.2.



Gambar 3. 2 Tahap *preprocessing*

Berdasarkan Gambar 3.2 tahapan *preprocessing* dimulai *cleaning* data dengan mengubah tipe data. Perubahan tipe data ini dilakukan untuk memudahkan dalam pembacaan data. Selanjutnya, dilakukan pemeriksaan *missing value*, data duplikat, *noise* dan *outlier*. Kemudian, dilakukan pembersihan data *noise* dan *outlier* dengan menghapusnya. Adapun *Cleaning* data ini digunakan agar data terhindar dari data dari *noisy*, *outlier*, *missing value* dan duplikat data, sehingga siap digunakan pada tahap selanjutnya (Nikmatun and Waspada, 2019). Kemudian, tahap *preprocessing* selanjutnya adalah normalisasi data. Normalisasi data digunakan untuk mengatasi perbedaan skala antar atribut pada dataset. Pada tahap ini teknik normalisasi yang digunakan adalah *Min-Max Scalling*. Selanjutnya, tahap selanjutnya adalah transformasi data dengan teknik *label encoding*. Tahap ini digunakan untuk mengubah dataset kategorik menjadi numerik. Kemudian, dilakukannya seleksi fitur dengan menggunakan teknik *correlation based fitur*. Teknik *correlation based fitur* ini dapat menemukan subset fitur asli dengan menggunakan pendekatan berbeda berdasarkan informasi

(Dikshit *et al.*, 2022). Adapun seleksi fitur ini digunakan untuk mengetahui hubungan antar atribut pada dataset (Tiana and Wahyuni, 2020).

3.3.4. Modeling

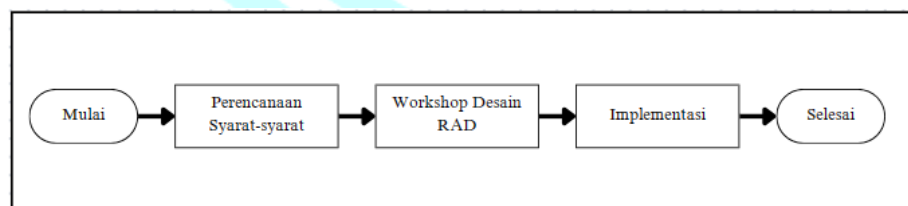
Pada tahap modeling ini dilakukan penerapan algoritma SVM dengan *kernel Polynomial*. Namun, berdasarkan dataset yang diperoleh pada tahap sebelumnya menunjukkan adanya ketidakseimbangan data, maka dataset tersebut perlu diterapkan teknik *oversampling*. Adapun teknik *oversampling* yang digunakan yaitu teknik SMOTE. Kemudian setelah diterapkan teknik SMOTE, selanjutnya dataset tersebut dipartisi terlebih dahulu menjadi data latih dan data uji menggunakan *K-Fold Cross Validation*. Adapun data latih ini digunakan untuk melatih model. Sedangkan data uji digunakan untuk tahap selanjutnya.

3.3.5. Evaluasi model

Pada tahap ini dilakukan evaluasi model menggunakan *Confusion matrix*. Adapun data yang digunakan adalah data uji yang telah dipartisi sebelumnya menggunakan *K-Fold Cross Validation*. Kemudian, data diuji sebanyak k dan menghasilkan nilai akurasi, presisi dan *recall*. Selanjutnya, model yang memiliki nilai akurasi, presisi dan *recall* tertinggi akan diimplementasikan kedalam sistem pada tahap selanjutnya. Sedangkan, jika hasil evaluasi model yang telah dilakukan belum sesuai yang diharapkan, maka akan kembali ketahap data *preparation*.

3.3.6. Deployment

Berdasarkan hasil implementasi model yang telah dilakukan pada tahap sebelumnya, maka model tersebut akan diterapkan kedalam sistem berbasis web dengan menggunakan *flask*. Adapun tahapannya dapat dilihat pada gambar 3.3.



Gambar 3. 3 Tahapan *Deployment*

a. Perencanaan Syarat-Syarat

Pada tahap ini terdiri dari identifikasi pengguna, analisis kebutuhan fungsional dan non fungsional

1. Identifikasi pengguna

Identifikasi pengguna ini dilakukan untuk mengidentifikasi kebutuhan pengguna pada sistem, sehingga sistem ini dapat membantu pengguna dalam menangani penyakit kanker payudara. Adapun identifikasi pengguna dapat dilihat pada Tabel 3.3.

Tabel 3. 1 Identifikasi Pengguna

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

Berdasarkan Tabel 3.3 menunjukkan bahwa pengguna sistem pada penelitian ini adalah tenaga medis. Tenaga medis ini dapat melakukan klasifikasi kanker payudara.

2. Analisis kebutuhan fungsional

Adapun kebutuhan fungsional pada dapat dilihat pada Tabel 3.4.

Tabel 3. 2 Kebutuhan fungsional

No	Deskripsi Kebutuhan	Aksi
1.	Sistem menyediakan tampilan untuk melakukan klasifikasi	Melakukan klasifikasi
2.	Sistem menyediakan tampilan hasil klasifikasi	Melihat hasil klasifikasi

Berdasarkan Tabel 3.4 menunjukan bahwa aksi dari sistem terdiri dari satu aksi saja yaitu klasifikasi. Sistem ini akan melakukan proses klasifikasi berdasarkan data yang diinputkan terlebih dahulu. Adapun output dari hasil klasifikasi akan menampilkan jenis diagnosis kanker payudara yang diderita yaitu jinak (*benign*) atau ganas (*malignan*).

3. Analisis kebutuhan non fungsional

1) Sistem ini berbasis web

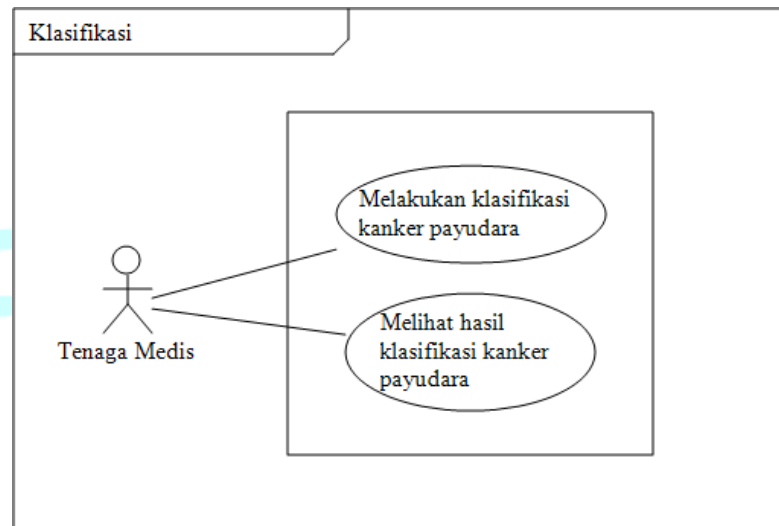
2) Sistem dibangun dengan menggunakan bahasa pemrograman python

b. *Workshop Desain RAD*

Adapun *workshop desain RAD* ini terdiri dari *usecase* diagram sistem dan *mockup* desain sistem.

1. *Usecase Diagram*

Usecase diagram sistem klasifikasi kanker payudara ini dapat dilihat pada Gambar 3.4.



Gambar 3. 4 *Usecase Diagram* sistem

Berdasarkan Gambar 3.4 menunjukkan bahwa tenaga medis dapat melakukan klasifikasi kanker payudara dengan menginputkan data pasien kesistem. Selain itu, tenaga medis dapat melihat hasil klasifikasi.

2. *Mockup desain sistem*

Adapun *mockup* desain sistem pada penelitian ini dapat dilihat pada Gambar 3.5.

The mockup shows a web browser window titled "A Web Page" with the URL "https://sistem klasifikasi kanker payudara". The page content is titled "Sistem Klasifikasi Kanker Payudara". On the left, there is a placeholder for an image, represented by a square with an 'X' inside. On the right, there are eight input fields for patient data: "Radius_mean", "Texture_mean", "Symmetry_mean", "radius_se", "Texture_se", "Perimeter_se", "Smoothness_se", and "Symmetry_se". At the bottom right, there is a button labeled "Cek Klasifikasi".

Gambar 3. 5 *Mockup Desain Sistem*

Berdasarkan pada Gambar 3.5 disajikan *mockup* desain sistem klasifikasi kanker payudara. Adapun fungsinya adalah untuk melakukan klasifikasi kanker payudara. Sistem ini akan berjalan ketika petugas medis memasukkan data pasien berdasarkan kolom yang ditampilkan. Setelah data pasien dimasukkan, sistem akan mengolahnya dan menampilkan hasil klasifikasi

c. Implementasi

Pada tahap ini dilakukan implementasi kedalam sistem berbasis web berdasarkan hasil dari tahap sebelumnya. Adapun implementasi sistem menggunakan bahasa pemrograman *python*. Selain itu, *database* yang digunakan pada sistem adalah *MySQL*. Kemudian, sistem ini dilakukan pengujian pada tahap selanjutnya.

3.3.7. Evaluasi Sistem

Adapun pada tahap ini dilakukan evaluasi sistem dengan menggunakan *blackbox testing* dan uji validasi. Pengujian dengan *blackbox testing* ini digunakan untuk mengetahui fungsi sistem berjalan sesuai yang diharapkan atau tidak (Wahyuni and Akbar, 2022), sedangkan uji validasi digunakan untuk mengetahui akurasi dari hasil klasifikasi penyakit kanker payudara (Nugraha, Shidiq and Rahayu, 2019). Selanjutnya, jika hasil evaluasi tidak sesuai yang diharapkan, maka akan kembali pada tahap implementasi sistem. Sedangkan, jika hasil evaluasi tersebut sesuai yang diharapkan, maka tahap penelitian dianggap selesai.