

BAB III

METODE PENELITIAN

3.1. Objek Penelitian

Objek pada penelitian ini menggunakan data penyakit diabetes dengan jumlah 768 dataset dan sembilan atribut (Khare, 2022). Pada dataset tersebut terdapat variabel yang dibagi menjadi kategori yaitu menyatakan diabetes (1) dan tidak diabetes (0). Dataset tersebut ditunjukkan pada Tabel 3.1.

Tabel 3. 1 Dataset Penyakit Diabetes

No	Pregnan	Glucose	BloPres	...	Diab Pedig Func	Age	Outcome
0	6	148	72	...	0,627	50	1
1	1	85	66	...	0,351	31	0
2	8	183	64	...	0,672	32	1
3	1	89	66	...	0,167	21	0
4	0	137	40	...	2,288	33	1
...
763	10	101	76	...	0,171	63	0
764	2	122	70	...	0,340	27	0
765	5	121	72	...	0,245	30	0
766	1	126	60	...	0,349	47	1
767	1	93	70	...	0,315	23	0

3.2 Lokasi dan Waktu

KARAWANG

Lokasi penelitian dilakukan di Universitas Buana Perjuangan Karawang sejak Februari 2023 hingga Juli 2023. Perincian penelitian ditunjukkan pada Tabel 3.2.

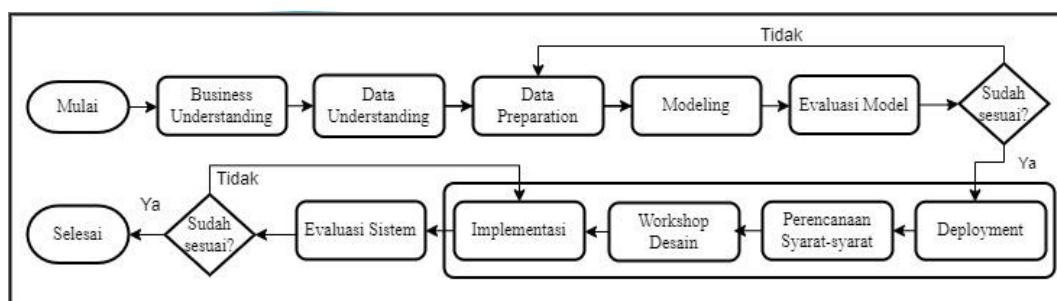
Tabel 3. 2 Rincian Waktu Penelitian

No	Kegiatan	2023				
		Februari	Maret	April	Mei	Juni
1	<i>Business Understanding</i>					
2	<i>Data Understanding</i>					
3	<i>Data Preparation</i>					
4	<i>Modeling</i>					
5	Evaluasi Model					
6	<i>Deployment</i>					
	a. Perencanaan Kebutuhan					
	b. Workshop Desain					

No	Kegiatan	2023				
		Februari	Maret	April	Mei	Juni
c. Implementasi						
7 Evaluasi Sistem						

3.3. Prosedur Penelitian

Prosedur dalam penelitian ini dimulai dari tahap *business understanding*, *data understanding*, *data preparation*, *modeling*, evaluasi model, *deployment*, dan evaluasi sistem. Alur prosedur penelitian ditunjukkan pada Gambar 3.1.



Gambar 3.1 Prosedur Penelitian

3.3.1. *Business Understanding*

Tahap *business understanding* dilakukan pada penelitian ini berdasarkan meningkatnya jumlah kematian dan penderita diabetes. Penelitian ini merancang sebuah model klasifikasi menggunakan Algoritma *Logistic regression* dengan teknik SMOTE. Kemudian model tersebut diimplementasikan kedalam sebuah sistem. Sistem tersebut diharapkan dapat membantu pihak medis dalam menguatkan diagnosa apakah seorang pasien terkena penyakit diabetes atau tidak.

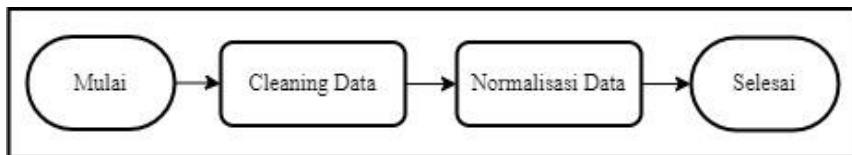
3.3.2. *Data Understanding*

Tahap data *understanding* yang dilakukan pada penelitian ini yaitu dengan mengumpulkan dataset penyakit diabetes. Dataset yang digunakan diperoleh dari *Kaggle* yang berjumlah 768 dengan sembilan atribut dan berformat csv. Dari sembilan atribut terdapat delapan variabel yang menyatakan hubungan antara kondisi seseorang dengan penyakit diabetes. Atribut-atribut tersebut terdiri dari *pregnancies*, *glucose*, *bloodpressure*, *skintickness*, *insulin*, *bmi*, *diabetespedigreefunction*, dan *age*. Selain itu, juga terdapat variabel *outcome* yang menyatakan apakah seseorang masuk dalam kategori memiliki penyakit diabetes

atau tidak. *Type* data dari dataset yang digunakan terdiri dari int sebanyak tujuh variabel, dan float sebanyak dua variabel.

3.3.3. Data Preparation

Tahap data *preparation* yang dilakukan pada penelitian ini yaitu dengan melakukan *preprocessing* yang terbagi dalam beberapa proses. Tahapan *preprocessing* ditunjukkan pada Gambar 3.2.



Gambar 3. 2 Tahap *Preprocessing*

Berdasarkan Gambar 3.2. tahap *preprocessing* yang dilakukan dimulai dengan tahap *cleaning* data. Hal yang dilakukan pada tahap ini yaitu mengecek *missing value* dan duplikasi data. Kemudian, mengubah nilai dari variabel yang memiliki nilai angka 0, hal ini bertujuan agar dataset yang digunakan tidak banyak yang dibuang pada saat proses penghapusan *outlier* dan *noise*. Pada proses ini nilai angka 0 diubah menjadi *NaN* (*Not a Number*). Nilai *NaN* tersebut kemudian diisi dengan nilai yang sesuai berdasarkan nilai tengah (*median*) dari variabel *outcome*. Kemudian, mengecek dan menangani *outlier* menggunakan teknik *Interquartile Range* (IQR). Selanjutnya, mengecek dan menghilangkan *noise* pada data menggunakan metode *Z-score*. Tahapan *cleaning* data tersebut dilakukan agar data yang digunakan menjadi lebih optimal dan menghasilkan model terbaik. Setelah tahap *cleaning* data, hal yang dilakukan selanjutnya yaitu proses normalisasi data menggunakan metode *Min-Max Scaling*. Proses ini dilakukan untuk mengatasi perbedaan skala antar variabel dari dataset yang digunakan.

3.3.4. Modeling

Tahap *modeling* yang dilakukan pada penelitian ini menggunakan Algoritma *Logistic Regression* dengan teknik SMOTE. Teknik SMOTE dilakukan karena adanya ketidakseimbangan pada data yang digunakan sebelum penerapan algoritma. Penerapan teknik SMOTE ini juga berguna untuk meningkatkan performa model. Selanjutnya, setelah penerapan teknik SMOTE dilakukan proses pembagian data menjadi data latih dan uji menggunakan *k-fold cross validation*.

3.3.5. Evaluasi Model

Tahap evaluasi model yang dilakukan pada penelitian ini yaitu dengan menggunakan *confusion matrix*. Data yang digunakan berasal dari pembagian data yang dilakukan pada tahap *modeling* menggunakan *k-fold cross validation*. Teknik *confusion matrix* digunakan untuk mengetahui performa dari model yang dibangun dengan menampilkan nilai akurasi, presisi dan *recall*.

3.3.6. Deployment

Tahap *deployment* pada penelitian ini dilakukan setelah proses membangun model. Model tersebut diterapkan ke dalam sebuah sistem berbasis *website* yang digunakan untuk mengklasifikasikan penyakit diabetes. Sistem tersebut dibangun menggunakan *framework streamlit*. Tahap *deployment* yang diterapkan berdasarkan proses *Rapid Application Development* (RAD) yang merupakan pengembangan perangkat lunak. Tahap tersebut ditunjukkan pada Gambar 3.3.



Gambar 3. 3 Tahap *Deployment*

a. Perencanaan Syarat-Syarat

Pada tahap ini dilakukan proses identifikasi pengguna, kebutuhan fungsional dan non fungsional.

1. Identifikasi pengguna

Identifikasi pengguna bertujuan untuk mengetahui kebutuhan pengguna pada sistem yang dibangun, sehingga dapat membantu pengguna dalam proses penanganan penyakit diabetes. Identifikasi pengguna dapat dilihat pada Tabel 3.3.

Tabel 3. 3 Identifikasi Pengguna

No	Pengguna	Keterangan
1	Tenaga Medis	Melakukan klasifikasi penyakit diabetes

Berdasarkan Tabel 3.3 pada penelitian ini identifikasi pengguna untuk sistem yang dibangun ditujukan kepada tenaga medis. Tenaga medis dapat menggunakan sistem untuk melakukan klasifikasi penyakit diabetes.

2. Kebutuhan Fungsional

Kebutuhan fungsional dari sistem dapat dilihat pada Tabel 3.4.

Tabel 3. 4 Kebutuhan Fungsional

No	Deskripsi	Aksi
1	Sistem menampilkan <i>form</i> untuk melakukan klasifikasi	Melakukan klasifikasi
2	Sistem menampilkan hasil klasifikasi	Melihat hasil klasifikasi

Berdasarkan Tabel 3.4. kebutuhan fungsional dari sistem yang dibangun memiliki dua fungsi yang berguna untuk melakukan klasifikasi penyakit diabetes serta menampilkan hasil klasifikasi. Hasil yang ditampilkan pada sistem menunjukkan apakah seseorang menderita penyakit diabetes atau tidak.

3. Kebutuhan Non Fungsional

Kebutuhan non fungsional dari sistem klasifikasi penyakit diabetes terdiri dari:

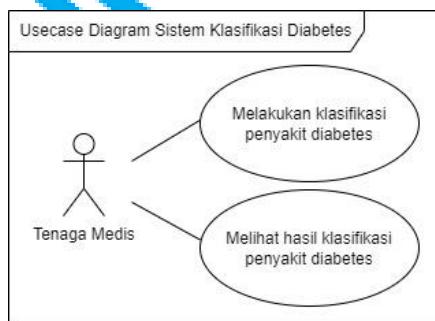
- a) Sistem mudah digunakan 
- b) Sistem dapat digunakan di beberapa *software web browser*

b. Workshop Desain

Tahap *workshop* desain yang dilakukan pada penelitian ini terdiri dari *activity* diagram dan *mockup* desain sistem yang dibangun.

1. Usecase Diagram

Use-case diagram dari sistem klasifikasi diabetes yang dibangun diperlukan dapat dilihat pada Gambar 3.4

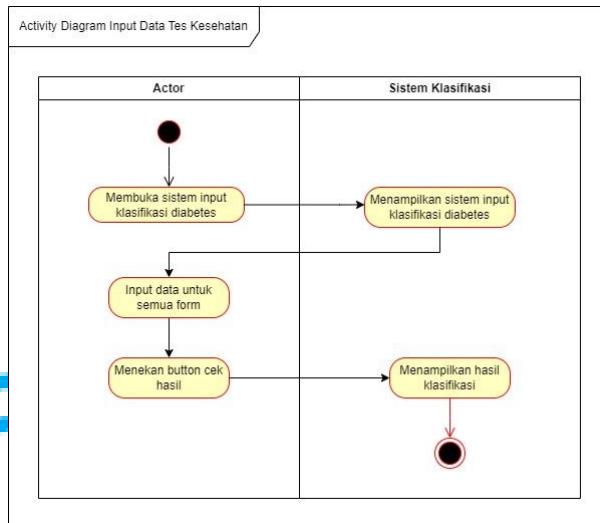


Gambar 3. 4 Use-case Diagram

Berdasarkan Gambar 3.4 menunjukkan bahwa tenaga medis dapat menginputkan nilai data kesehatan pasien pada form yang tersedia untuk melakukan klasifikasi dan melihat hasilnya.

2. Activity Diagram

Activity diagram dari sistem klasifikasi diabetes yang dibangun dapat dilihat pada Gambar 3.5.



Gambar 3. 5 Activity diagram sistem klasifikasi diabetes

Berdasarkan Gambar 3.5. menunjukkan tahapan aktivitas yang dilakukan tenaga medis dalam menggunakan sistem klasifikasi diabetes. Hal pertama yang dilakukan yaitu tenaga medis membuka sistem klasifikasi diabetes. Kemudian, sistem akan menampilkan halaman untuk menginputkan data kesehatan pasien. Selanjutnya, tenaga medis menginputkan data pada *form* yang tersedia dan menekan tombol cek hasil. Sistem akan memproses data yang diinputkan dan menampilkan hasil klasifikasi.

3. Mockup Desain

Mockup desain dari sistem klasifikasi diabetes yang dibangun dapat dilihat pada Gambar 3.6.

The mockup shows a web browser window titled "A Web Page" with the URL "https://klasifikasi-diabetes". The page itself is titled "Klasifikasi Diabetes" and contains a form with the following fields:

- Nilai Pregnancies
- Nilai Glucose
- Nilai Blood Pressure
- Nilai Skin Thickness
- Nilai Insulin
- Nilai BMI
- Nilai Diabetes Pedigree Function
- Nilai Age
-

Gambar 3. 6 Mockup desain sistem klasifikasi diabetes

Berdasarkan Gambar 3.6. menunjukkan mockup desain dari sistem klasifikasi diabetes. Sistem tersebut berfungsi untuk melakukan klasifikasi penyakit diabetes. Sistem akan berjalan jika tenaga medis telah menginputkan data pasien ke dalam *form* yang tersedia. Kemudian, setelah pengisian *form* tersebut sistem akan memproses data dan menampilkan hasil klasifikasi.

c. Implementasi

Pada tahap ini merupakan implementasi dari tahap-tahap yang telah dilakukan sebelumnya. Implementasi berupa sistem berbasis web yang dibangun menggunakan *framework streamlit* dengan bahasa pemrograman *python* dengan *database* yang digunakan yaitu MySQL. Setelah proses implementasi selanjutnya melakukan proses pengujian sistem.

3.3.7. Evaluasi Sistem

Tahap evaluasi sistem pada penelitian ini menggunakan *blackbox testing* serta uji validasi sistem dan pakar. Pengujian dengan *blackbox testing* dilakukan untuk mengetahui performa sistem yang dibuat apakah sudah sesuai dengan yang diinginkan atau tidak. Sedangkan, uji validasi sistem dan pakar dilakukan dengan membandingkan hasil sistem dan pakar, kemudian dilakukan perhitungan secara manual dengan rumus *confusion matrix* untuk mengetahui akurasi dari hasil klasifikasi penyakit diabetes.