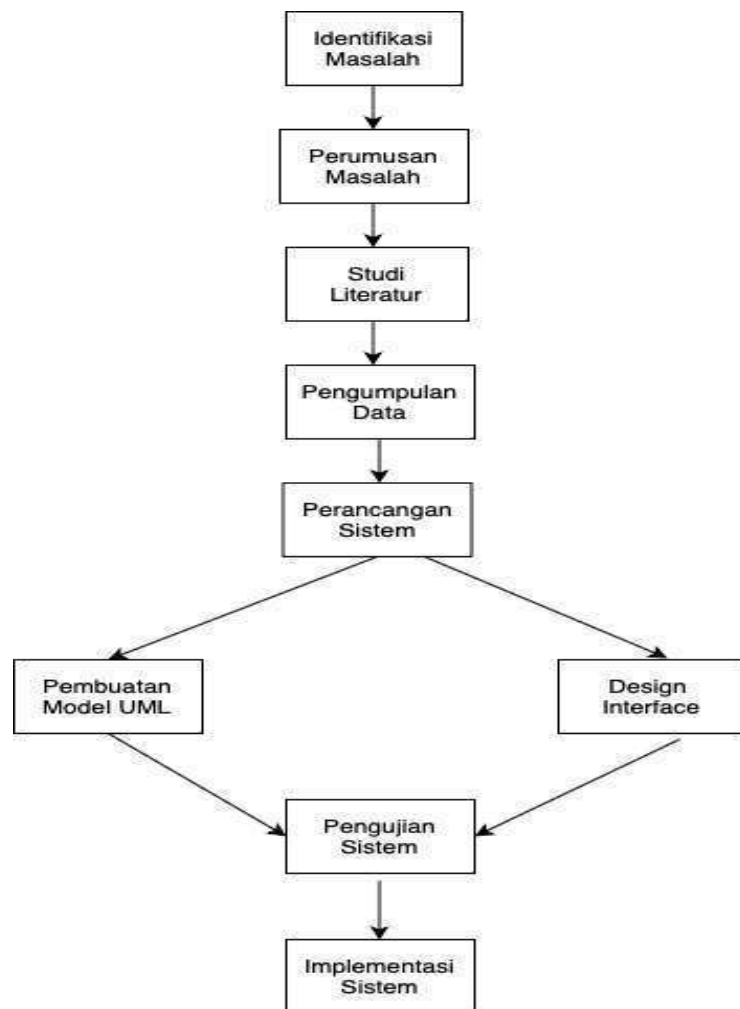


BAB III

METODE PENELITIAN

3.1 Desain Penelitian

Metode atau pendekatan peneliti untuk mengumpulkan, memeriksa, dan menafsirkan data.



Gambar 3. 1 Desain Penelitian

Sumber: Penelitian (2024)

Metodologi ini menetapkan kerangka kerja sistematis untuk mengatasi kesulitan penelitian dan mencapai tujuan penelitian.

Berikut ini penjelasan desain penelitian pada Gambar 3.1, sebagai berikut:

1. Identifikasi Masalah

Pada titik ini, para ilmuwan mulai mengidentifikasi masalah dengan penyelidikan menganalisis kemampuan algoritma regresi logistik dalam memprediksi penyakit diabetes berdasarkan dataset medis.

2. Perumusan masalah

Tahap berikutnya bagi para peneliti adalah untuk secara akurat dan ringkas merumuskan tantangan studi setelah mereka telah menentukan masalah saat ini seperti prediksi penyakit diabetes menggunakan algoritma regresi logistik, sehingga dengan adanya aplikasi yang diciptakan ini nantinya akan membantu pengguna mengetahui apakah ia terkena diabetes atau tidak.

3. Studi Literatur

Untuk memahami teori, konsep, prosedur, dan kesimpulan dari studi sebelumnya, tahap ini melibatkan pencarian dan membaca literatur yang relevan mengenai prediksi penyakit diabetes menggunakan algoritma regresi logistik.

4. Pengumpulan Data

Saat ini, pengguna yang telah menguji program akan memberikan data kepada peneliti melalui kuesioner.

5. Perancangan Sistem

Para peneliti awalnya akan membangun model UML sebelum bergerak ke desain antarmuka.

6. Pengujian Sistem

Pengujian beta digunakan untuk menguji, di mana para peneliti menggunakan survei untuk menilai hasilnya. Setelah itu, akan ditentukan apakah pengguna aplikasi mengalami manfaat apa pun.

7. Implementasi Sistem

Program ini akan diintegrasikan ke dalam lingkungan kerja nyata ketika telah diuji.

3.2 Teknik Pengumpulan Data

3.2.1 Observasi

Teknik untuk mengumpulkan data lewat mengamati orang dan lingkungan di mana fenomena penelitian terjadi dari tangan pertama. Melalui pengamatan, para peneliti dapat menyaksikan interaksi sosial, perilaku, dan situasi yang relevan dengan fenomena yang mereka pelajari (Ardiansyah et al., 2023).

3.2.2 Studi Pustaka

Pada penelitian ini, peneliti memanfaatkan sumber dari beberapa macam referensi seperti buku, jurnal nasional atau internasional seperti dari pengindeks *Google Scholar*, Garuda, DOAJ, *Scopus* dan lain sebagainya untuk memperoleh data penelitian.

3.3 Variabel Penelitian

Untuk mengetahui variabel operasional yang digunakan dalam prediksi penyakit diabetes, maka peneliti harus menentukan yang menjadi variabel dependen (Y) dan variabel independen (X).

a. Variabel Dependen (Y) Penyakit Diabetes

Penyakit ini ditandai dengan adanya tingkat glukosa darah yang tinggi karena tubuh tidak mampu memproduksi atau menggunakan insulin secara efektif. Status diabetes diukur dengan diagnosis medis menggunakan parameter seperti $HbA1c \geq 6.5\%$, glukosa plasma puasa ≥ 126 mg/dL, atau tes toleransi glukosa dengan glukosa plasma dua jam setelah makan ≥ 200 mg/dL. Variabel ini dinyatakan dalam bentuk biner (0 = tidak diabetes dan 1 = diabetes).

b. Variabel Independen (X)

- Glukosa
- Tekanan darah
- Ketebalan kulit
- Insulin
- BMI
- DPF
- Usia

Berikut ini definisi operasional variabel berdasarkan *dataset* yang digunakan oleh peneliti ditampilkan dalam tabel 3.1.

Tabel 3. 1 Definisi Operational Variable

Variabel	Definisi Konseptual	Definisi Operasional	Skala Pengukuran
Penyakit Diabetes	Kondisi dengan glukosa darah tinggi	Diagnosis medis (0 = tidak diabetes, 1 = diabetes)	Biner
Glukosa (<i>Glucose</i>)	Jumlah glukosa dalam plasma darah	Kadar glukosa dalam mg/dL	Numerik
Tekanan darah (<i>Blood Pressure</i>)	Tekanan darah terhadap dinding arteri	Tekanan sistolik dalam mmHg	Numerik
Ketebalan kulit (<i>Skin Thickness</i>)	Jumlah lemak subkutan yang diukur melalui lipatan kulit	Ketebalan kulit pada triceps dalam mm	Numerik
Insulin	Jumlah gula yang dikonsumsi sehari-hari	Asupan gula dalam gram per hari	Numerik
BMI (<i>Body Mass Index</i>)	Rasio berat badan terhadap tinggi badan	Berat badan (kg) / tinggi badan (m^2)	Numerik
DPF (<i>Diabetes Pedigree Function</i>)	Adanya riwayat diabetes di antara anggota keluarga	0 = tidak ada, 1 = ada	Biner
Usia (<i>Age</i>)	Lamanya waktu hidup seseorang	Umur dalam tahun	Numerik

3.4 Teknik Pengambilan Sampel

Dataset yang digunakan di ambil dari situs *kaggle* dengan nama *Pima Indians Diabetes Database*. Dengan menggunakan prosedur sampling acak dasar, setiap anggota populasi memiliki kesempatan yang sama untuk dipilih untuk sampel. Berikut ini langkah-langkah pengambilan sampel menggunakan teknik *simple random sampling* di *google colab*:

1. *Import* pustaka yang dibutuhkan terlebih dahulu.

```
[1] import pandas as pd
import numpy as np
```

2. Muat *dataset* ke dalam *google colab*

```
# Memuat dataset file CSV lokal
df = pd.read_csv('diabetes.csv')
column_names = ['Glucose', 'BloodPressure', 'SkinThickness', 'Insulin', 'BMI', 'DiabetesPedigreeFunction', 'Age', 'Outcome']
df = pd.read_csv(url, header=None, names=column_names)

# Menampilkan beberapa baris pertama dari dataset
print(df.head())
```

	Glucose	BloodPressure	SkinThickness	Insulin	BMI	\
6	148	72	35	0	33.6	
1	85	66	29	0	26.6	
0	183	64	0	0	23.3	
1	89	66	23	94	28.1	
0	137	40	35	168	43.1	

	DiabetesPedigreeFunction	Age	Outcome
6	0.627	50	1
1	0.351	31	0
0	0.672	32	1
1	0.167	21	0
0	2.288	33	1

3. Mengambil sampel dengan *simple random sampling*, misalnya di sini kita harus mengambil 100 data dari *dataset*.

```
# Menentukan ukuran sampel
sample_size = 100

# Mengambil sampel dengan simple random sampling
sample = df.sample(n=sample_size, random_state=42)

# Menampilkan beberapa baris pertama dari sampel
print(sample.head())
```

	Glucose	BloodPressure	SkinThickness	Insulin	BMI	\
6	98	58	33	190	34.0	
2	112	75	32	0	35.7	
2	108	64	0	0	30.8	
8	107	80	0	0	24.6	
7	136	90	0	0	29.9	

	DiabetesPedigreeFunction	Age	Outcome
6	0.430	43	0
2	0.148	21	0
2	0.158	21	0
8	0.856	34	0
7	0.210	50	0

4. Ketika kita ingin mengambil sampel dengan proporsi tertentu. Misalnya ingin mengambil 10% dari *dataset* berarti $frac=0.1$.

```
[9] # Menentukan proporsi sampel (misalnya 10% dari dataset)
sample_fraction = 0.1

# Mengambil sampel dengan simple random sampling berdasarkan proporsi
sample = df.sample(frac=sample_fraction, random_state=42)

# Menampilkan beberapa baris pertama dari sampel
print(sample.head())
```

	Glucose	BloodPressure	SkinThickness	Insulin	BMI	\
6	98	58	33	190	34.0	
2	112	75	32	0	35.7	
2	108	64	0	0	30.8	
8	107	80	0	0	24.6	
7	136	90	0	0	29.9	

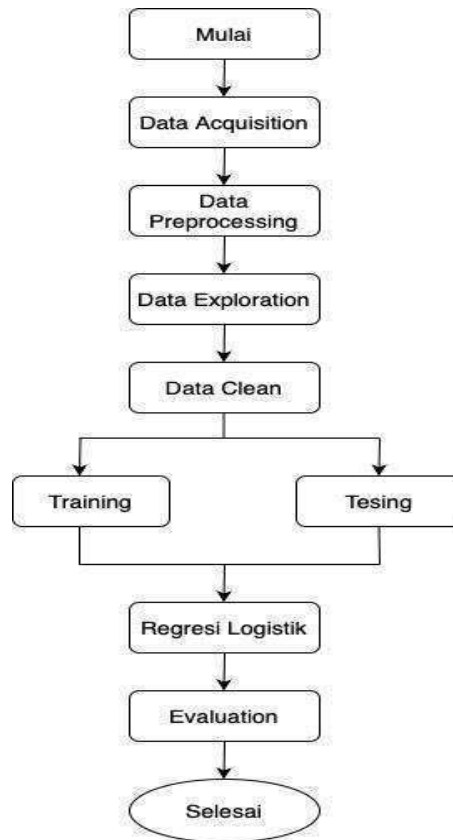
	DiabetesPedigreeFunction	Age	Outcome
6	0.430	43	0
2	0.148	21	0
2	0.158	21	0
8	0.856	34	0
7	0.210	50	0

Dengan mengikuti prosedur yang disebutkan di atas, peneliti dapat memastikan bahwa setiap potongan data dalam dataset memiliki probabilitas yang sama untuk dipilih untuk sampel.

3.5 Teknik Analisis Data

3.5.1 Proses Analisis Data

Pada penelitian ini teknik analisis data untuk memprediksi penyakit diabetes dapat dilihat pada gambar 3.2.



Gambar 3. 2 Proses Analisis Data

Sumber: Penelitian (2024)

Berikut ini keterangan proses analisis *data* prediksi penyakit diabetes:

1. *Data Acquisition*

Prosedur untuk mengumpulkan dan memeriksa data, termasuk nilai untuk setiap variable dan variable itu sendiri.

2. *Data Pre-processing*

Ini adalah prosedur yang mengubah data yang tidak diproses menjadi pengumpulan data yang mudaf ditafsirkan.

3. *Data Exploration*

Ini adalah tahap pemahaman tentang isi data.

4. *Data Clean*

Prosedur untuk menghapus atau memperbaiki data yang tidak akurat, salah, duplikat, atau hilang dari kumpulan data.

5. *Training dan Testing*

Pengujian adalah langkah yang dilakukan setelah validasi untuk menunjukkan apakah model itu akurat, sedangkan pelatihan adalah proses menggunakan data pelatihan untuk menemukan parameter yang tepat.

6. Regresi Logistik

Artinya menggunakan pemodelan analisis variabel regresi untuk memprediksi set data dengan dua hasil yang mungkin, seperti ya atau tidak.

7. *Evaluation*

Proses ini melibatkan menghitung akurasi metrik, presisi, pengingat, dan nilai skor f1. Jenis pengukuran ini digunakan untuk menilai seberapa baik model ini bekerja.

Berikut ini adalah penjelasan dari proses analisis data untuk prediksi penyakit diabetes:

1. *Data Acquisition*

Ini adalah tahap di mana peneliti mengumpulkan data yang diperlukan. *Dataset* dari format *Pima Indians Diabetes* yang diunduh dari situs *kaggle* yang digunakan dalam penelitian ini. Tabel 3.1 memberikan informasi tentang karakteristik variabel atau atribut dalam kumpulan data.

Setelah melihat karakteristik variable pada Tabel 3.1 menunjukan lima data teratas dari dataset dengan tujuh variabel (*glucose, blood pressure, skin thickness,*

insulin, BMI (*body mass indec*), *diabetes pedigree function* (DPF), dan *age*) dan satu variabel dependen adalah *outcome*. Setiap variable memiliki rentang nilai yang berbeda-beda. Rentang nilai tiap variable dapat dilihat pada Gambar 3.3.

Dataset Pima Indians Diabetes:

	Glucose	BloodPressure	SkinThickness	Insulin	BMI	\
6	148	72	35	0	33.6	
1	85	66	29	0	26.6	
8	183	64	0	0	23.3	
1	89	66	23	94	28.1	
0	137	40	35	168	43.1	

	DiabetesPedigreeFunction	Age	Outcome
6	0.627	50	1
1	0.351	31	0
8	0.672	32	1
1	0.167	21	0
0	2.288	33	1

Rentang setiap variabel dalam dataset:

Glucose	199.000
BloodPressure	122.000
SkinThickness	99.000
Insulin	846.000
BMI	67.100
DiabetesPedigreeFunction	2.342
Age	60.000
Outcome	1.000

dtype: float64

Gambar 3.3 Rentang Nilai Tiap Variabel

Sumber: Penelitian (2024)

Gambar 3.3 menunjukkan bahwa setiap variabel memiliki rentang nilai yang berbeda-beda. Berikut ini adalah deskripsi dari kisaran nilai setiap variabel pada gambar 3.3:

Variabel *glucose* memiliki nilai yang rendah 0.000 dan tinggi hingga 199.00

Variabel *blood pressure* memiliki nilai 0.000 yang terendah dan tinggi hingga 122.00

Variabel *skin thickness* memiliki nilai *minimum* 0.000 dan nilai *maximum* 99.00

Variabel insulin berikisar dari 0.000 sebagai nilai terkecil hingga 846.00 sebagai harga maksimum

Variabel BMI (*Body mass index*) memiliki nilai *minimum* 0.000 dan nilai *maximum* 67.10

Variabel *Diabetes pedigree function* memiliki nilai *minimum* 0.078 dan nilai *maximum* 2.42

Variabel *age* memiliki nilai *minimum* 21.000 dan nilai *maximum* 81.00 Variabel *outcome* memiliki nilai *minimum* 0.000 dan nilai *maximum* 1.00

2. Preprocessing

Sebelum melakukan hal lain, data yang sudah diambil dari situs Kaggle harus dikonfirmasi untuk menghapus duplikat, nilai yang hilang, dan outliers. Ada banyak nol (0) pada variabel glukosa, tekanan darah, ketebalan kulit, insulin, dan BMI untuk dimasukkan ke dalam nilai-nilai yang hilang, tetapi pada 768, tidak ada satu nilai yang hilang dalam data. Setelah itu, rata-rata digunakan untuk mengisi informasi yang hilang dengan perhitungan sebagai berikut:

$$\text{Nilai rata-rata} = \frac{\sum f_i x_i}{\sum f_i}$$

Keterangan:

f_i : frekuensi kelas ke-i

x_i : nilai tengah kelas ke-i

Seperti yang terlihat pada Gambar 3.4.

```
[18] # Mengisi nilai yang hilang dengan rata-rata setiap kolom
df_filled = df.fillna(means)

# Menampilkan beberapa baris pertama dari DataFrame setelah mengisi nilai yang hilang
print("\nDataset setelah mengisi nilai yang hilang dengan rata-rata:")
print(df_filled.head())
```

Dataset setelah mengisi nilai yang hilang dengan rata-rata:

	Glucose	BloodPressure	SkinThickness	Insulin	BMI	
6	148	72	35	0	33.6	
1	85	66	29	0	26.6	
8	183	64	0	0	23.3	
1	89	66	23	94	28.1	
0	137	40	35	168	43.1	

	DiabetesPedigreeFunction	Age	Outcome
6	0.627	50	1
1	0.351	31	0
8	0.672	32	1
1	0.167	21	0
0	2.288	33	1

Gambar 3. 4 Analisa *Missing Value*

Sumber: Penelitian (2024)

Kita dapat melihat bahwasanya nilai 0 ada pada variabel *skin thickness* dan *insulin*. Seperti terlihat padat Tabel 3.2

Tabel 3. 2 *Missing Value* pada Tabel

Column1	Column	Column3	Column4	Column	Column	Column7	Column	Column
	Glucose	Blood Pressure	Skin Thickness	Insulin	BMI	Diabetes Pedigree Function	Age	Outcome
0	148	72	35	0	33.6	0.627	50	1
1	85	66	29	0	26.6	0.351	31	0
2	183	64	0	0	23.3	0.672	32	1
3	89	66	23	94	28.1	0.167	21	0
4	137	40	35	168	43.1	2.288	33	1

Lalu angka 0 kemudian diganti dengan nilai rata-rata dari kolom yang bersangkutan dalam *dataset*. Hasilnya akan terlihat seperti Gambar 3.6

```
# Mengganti nilai nol dengan NaN
df[['Glucose', 'BloodPressure', 'SkinThickness', 'Insulin', 'BMI']] = df[['Glucose', 'BloodPressure', 'SkinThickness', 'Insulin', 'BMI']].replace(0, pd.NA)

# Menghitung rata-rata setiap kolom
means = df.mean()

# Mengisi nilai yang hilang dengan rata-rata setiap kolom
df_filled = df.fillna(means)

# Menampilkan beberapa baris pertama dari DataFrame setelah mengisi nilai yang hilang
print("\nDataset setelah menggantikan nilai nol dan mengisi nilai yang hilang dengan rata-rata:")
print(df_filled.head())
```

Dataset setelah menggantikan nilai nol dan mengisi nilai yang hilang dengan rata-rata:

	Glucose	BloodPressure	SkinThickness	Insulin	BMI	
6	148.0	72.0	35.000000	155.548223	33.6	
1	85.0	66.0	29.000000	155.548223	26.6	
8	183.0	64.0	29.15342	155.548223	23.3	
1	89.0	66.0	23.000000	94.000000	28.1	
0	137.0	40.0	35.000000	168.000000	43.1	

	DiabetesPedigreeFunction	Age	Outcome
6	0.627	50	1
1	0.351	31	0
8	0.672	32	1
1	0.167	21	0
0	2.288	33	1

Gambar 3. 5 Analisa Nilai rata-rata Kolom

Sumber: Penelitian (2024)

Maka pada tabel 3.3 akan terlihat nilai 0 sudah diganti menjadi NaN.

Tabel 3. 3 Nilai 0 Diganti dengan NaN

Column1	Column	Column3	Column4	Column	Column	Column7	Column	Column
	Glucose	Blood Pressure	Skin Thickness	Insulin	BMI	Diabetes Pedigree Function	Age	Outcome
0	148.0	72.0	35.00000	155.548223	33.6	0.627	50	1
1	85.0	66.0	29.00000	155.548223	26.6	0.351	31	0
2	183.0	64.0	29.15342	155.548223	23.3	0.672	32	1
3	89.0	66.0	23.00000	94.000000	28.1	0.167	21	0
4	137.0	40.0	35.00000	168.000000	43.1	2.288	33	1

Setelah nilai yang hilang telah diatasi, periksa outlier dan hapus dengan merujuk pada skor Z yang ditunjukkan dalam tabel 3.4

$$Z\text{-score} : \frac{x-\mu}{\sigma}$$

Keterangan:

μ = rata-rata populasi

σ = simpangan

baku populasi

X = skor

mentah

Berikut ini perintah untuk menghitung *z-score*:

```

[28] # Hitung Z-Score
z_scores = df.iloc[:, :-1].apply(zscore, nan_policy='omit')

[29] # Identifikasi Outlier
outliers = (z_scores > 3) | (z_scores < -3)

[30] # Hitung jumlah Outlier untuk setiap Fitur
print(outliers.sum())

0

# Menggunakan Z-score
import pandas as pd
from scipy.stats import zscore

# Muat dataset
df = pd.read_csv("diabetes.csv")

# Hitung z-score
z_scores = df.iloc[:, :-1].apply(zscore, nan_policy='omit')
print(z_scores.head())

# Identifikasi outlier
outliers = (z_scores > 3) | (z_scores < -3)
print(outliers.sum())

Series([], dtype: float64)
0

```

Gambar 3. 6 Penghitungan *Z-score*

Sumber: Penelitian (2024)

Z-score mengasumsikan bahwa data terdistribusi normal jika nilainya 0 atau berada pada posisi grafik yang tegak lurus. Berikut ini pengecekan *outlier* menggunakan *z-score*:

```

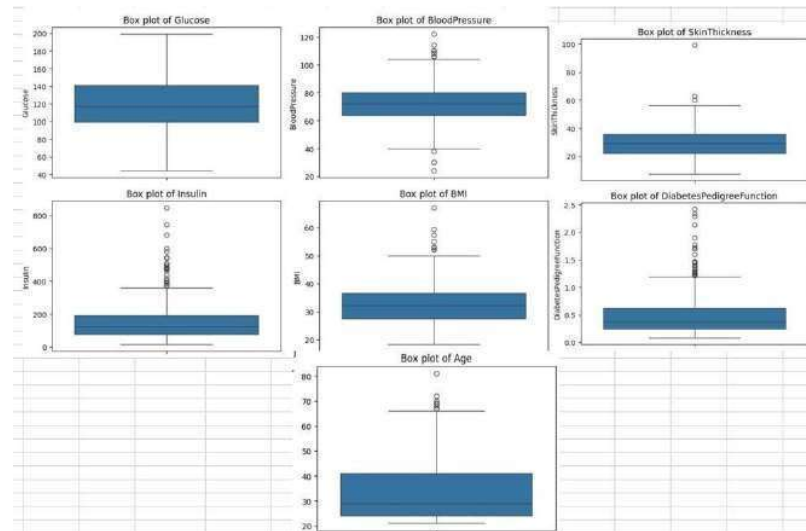
import matplotlib.pyplot as plt
import seaborn as sns

# Membuat box plot untuk setiap variabel
plt.figure(figsize=(15, 10))
for i, column in enumerate(df.columns[:-1], 1): # Exclude the 'Outcome' column
    plt.subplot(3, 3, i)
    # Reset index to avoid potential duplicate index issues
    sns.boxplot(y=df.reset_index()[column])
    plt.title(f'Box plot of {column}')
plt.tight_layout()
plt.show()

```

Gambar 3. 7 Perintah Pengecekan *Outlier*

Sumber: Penelitian (2024)



Gambar 3. 8 Gambar *Outlier*

Sumber: Penelitian (2024)

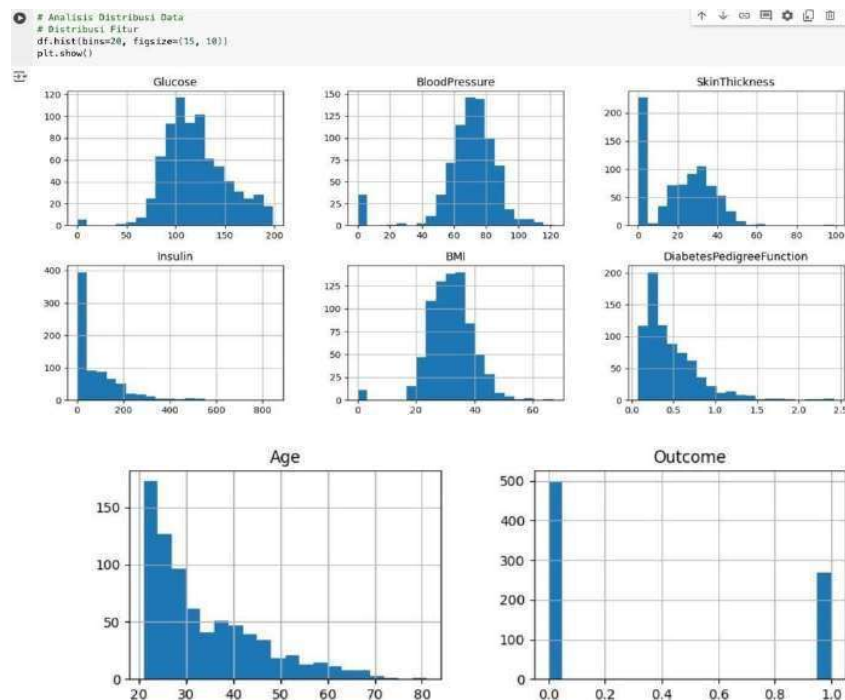
Dari 8 variabel yang ada terlihat bahwasanya setiap variable memiliki *outlier*, kecuali *Glucose*. *Outlier* dihapus menggunakan Z-Score. Jumlah data menghasilkan 768 sampai 718 data karena data keempat dihapus karena nilai dari skor Z. Selanjutnya, gunakan korelasi antara variabel untuk menentukan variabel mana yang termasuk dalam pemodelan.

3. *Data Exploration*

Setelah tahapan di atas, Data exploration, langkah berikutnya dalam proses, dimaksudkan untuk membantu dengan pemahaman, membersihkan dan menyiapkan data untuk dianalisis lebih lanjut. Dalam penelitian ini, tahap *data exploration* sangat penting dalam memastikan seluruh data siap untuk digunakan

dalam model prediksi. Berikut ini adalah tahapan dalam proses eksplorasi data yaitu:

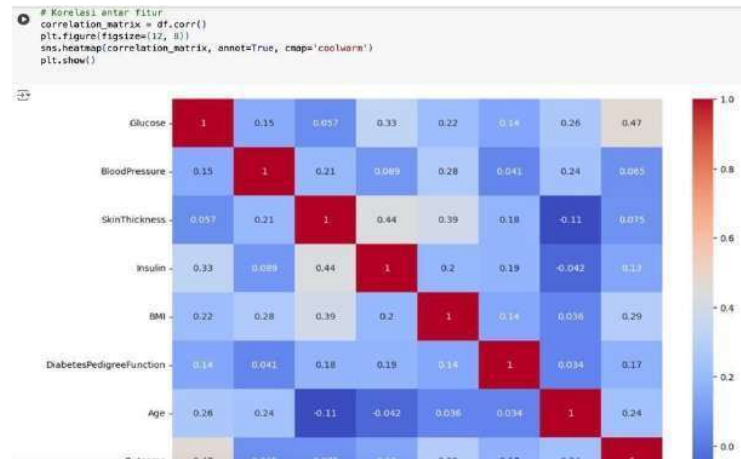
- Memahami struktur data
- Memeriksa dan menangani *missing values*
- Memeriksa *outliers*
- Analisis distribusi data



Gambar 3. 9 Analisis Fitur

Sumber: Penelitian (2024)

- Feature engineering*



Gambar 3. 10 Korelasi antar Fitur

Sumber: Penelitian (2024)

3. *Modelling dan Evaluation*

Proses ini membuat data dibagi menjadi data pelatihan dan pengujian dengan rasio 70:30, yang berarti ada 718 data total, 502 data total pelatihan, dan 216 data tes total yang digunakan untuk pemodelan. Karena korelasi dekat antara variabel yang diamati secara umum, model ini menggunakan semua variabel independen.

3.5.2 Pemilihan Atribut (Selection Features)

Pemilihan atribut pada prediksi penyakit diabetes menggunakan teknik korelasi Pearson atau Spearman. Yang mana teknik ini digunakan untuk mengukur hubungan antara fitur dan target label. Berikut ini rumus korelasi *pearson*:

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n \sum x^2 - (\sum x)^2][n \sum y^2 - (\sum y)^2]}}$$

Dengan ketentuan sebagai berikut:

- r = Koefisien pearson

- n = Jumlah data
- $\sum xy$ = Jumlah data yang berpasangan
- $\sum x$ = Jumlah dari nilai x
- $\sum y$ = Jumlah dari nilai y
- $\sum x^2$ = Jumlah dari nilai x^2
- $\sum y^2$ = Jumlah dari nilai y^2

Sedangkan rumus untuk korelasi *spearman* sebagai berikut:

$$r_R = 1 - \frac{6\sum d_i^2}{n(n^2 - 1)}$$

Di Sini,

n = jumlah titik data kedua variabel

d_i = selisih rangking elemen “ith”.

Koefisien Spearman, ρ , dapat bernilai antara +1 hingga -1 dengan,

- Nilai $\rho +1$ berarti asosiasi peringkat yang sempurna
- Nilai $\rho 0$ berarti tidak ada asosiasi peringkat
- Nilai $\rho -1$ berarti hubungan negatif sempurna antar peringkat.

3.5.3 Data Training dan Data Testing

Pada penelitian ini menggunakan *data training* berupa *dataset* dari Institut Nasional Diabetes dan Penyakit Pencernaan dan Ginjal menggunakan metode Metrik Evaluasi yang menghitung adanya *accuracy* (mengingat (% dari kasus positif benar yang diakui), akurasi (perbandingan prediksi positif yang benar dengan semua prediksi negatif), dan rasio prediksi yang benar terhadap prediksi total) terhadap semua kasus positif aktual), dan *F1-Score* (mean dari *precision* dan *recall*, ketika terdapat keseimbangan kelas).

NO.	Pregnancy	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
1	6	148	72	35	0	33,6	0,627	50	1
2	1	85	66	29	0	26,6	0,351	31	0
3	8	183	64	0	0	23,3	0,672	32	1
4	1	89	66	23	94	28,1	0,167	21	0
5	0	137	40	35	168	43,1	2,288	33	1
6	5	116	74	0	0	25,6	0,201	30	0
7	3	78	50	32	88	31,0	0,248	26	1
8	10	115	0	0	0	35,3	0,134	29	0
9	2	197	70	45	543	30,5	0,158	53	1
10	8	125	96	0	0	0,0	0,232	54	1
11	4	110	92	0	0	37,6	0,191	30	0
12	10	168	74	0	0	38,0	0,537	34	1
13	10	139	80	0	0	27,1	1,441	57	0
14	1	189	60	23	846	30,1	0,398	59	1
15	5	166	72	19	175	25,8	0,587	51	1
16	7	100	0	0	0	30,0	0,484	32	1
17	0	118	84	47	230	45,8	0,551	31	1
18	7	107	74	0	0	29,6	0,254	31	1
19	1	103	30	38	83	43,3	0,183	33	0
20	1	115	70	40	96	34,6	0,529	32	1
21	3	126	88	41	235	39,3	0,704	27	0
22	8	99	84	0	0	35,4	0,388	50	0
23	7	196	90	0	0	39,8	0,451	41	1
24	9	119	80	35	0	29,0	0,263	29	1

Gambar 3. 11 Dataset Training dan Data Testing

Sumber: Penelitian (2024)

3.5.4 Menentukan Nilai Logistik Regression

Persamaan regresi logistik digunakan untuk memprediksi probabilitas dari kejadian tertentu (apakah seseorang menderita diabetes atau tidak). Persamaan utamanya adalah:

$$\pi(x)_i = \frac{\exp(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_p X_{pi})}{1 + \exp(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_p X_{pi})}$$

Dimana:

$\pi(x_i)$: Peluang terjadinya kategori variabel respon

X_{pi} : Variabel prediktor ke-j

p : Banyaknya variabel prediktor

β_0 : Intersep

$\beta_1, \beta_2, \beta_p$: Koefisien regresi untuk setiap variabel prediktor

i : 1, 2, ..., n

Untuk mempermudah interpretasi dan pendugaan parameter, peluang pada persamaan diatas dilakukan transformasi logit sehingga didapatkan fungsi logit sebagai berikut:

$$g(x_i) = \text{logit } \pi(x_i) = \ln \left(\frac{\pi(x_i)}{1 - \pi(x_i)} \right),$$

$$\text{Jika } \pi(x_i) = \frac{\exp(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_p X_{pi})}{1 + \exp(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_p X_{pi})}$$

$$g(x_i) = \ln \left(\frac{\pi(x_i)}{1 - \pi(x_i)} \right)$$

$$\begin{aligned} &= \ln \left(\frac{\frac{\exp(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_p X_{pi})}{1 + \exp(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_p X_{pi})}}{\frac{\exp(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_p X_{pi})}{1 + \exp(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_p X_{pi})}} \right) \\ &= \ln (\exp (\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_p X_{pi})) \\ &= \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_p X_{pi} \end{aligned}$$

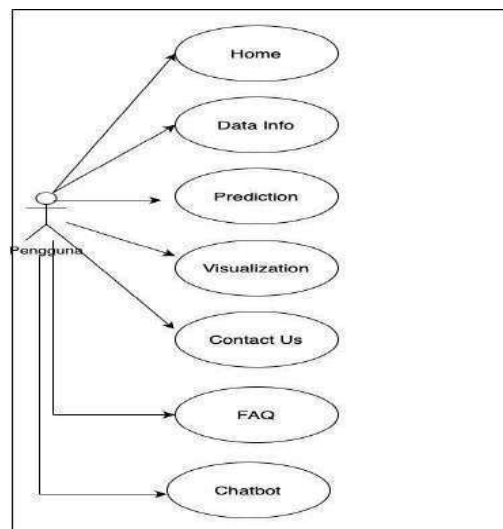
$$g(x_i) = \beta_0 + \sum_{j=1}^p \beta_j X_{ji}$$

3.6 Perancangan Sistem

Adalah proses merancang atau menciptakan suatu sistem yang bertujuan untuk memenuhi kebutuhan atau tujuan tertentu. Proses ini melibatkan berbagai tahapan dan metodologi untuk memastikan bahwa sistem yang dirancang efektif, efisien, dan memenuhi kebutuhan pengguna. Seperti yang ditunjukkan pada Gambar 3.10.

3.6.1 Perancangan UML

1. *Diagram Usecase*



Gambar 3. 12 Diagram *Usecase*

Sumber: Penelitian (2024)

Salah satu cara untuk menggambarkan bagaimana sistem dan pihak-pihak yang berurusan dengannya berinteraksi adalah melalui diagram kasus penggunaan. Instansi penggunaan, sektor terkait, dan interkoneksi semuanya ditampilkan secara visual dalam grafik ini (UNhelkar, 2019).

Pada gambar 3.2, dijelaskan bahwasanya semua orang dapat mengakses halaman *web* ini tanpa perlu halaman *login*. Ketika pengguna masuk dalam *web* ini akan muncul menu *Home*. Pada menu ini akan muncul halaman pertama yang mengucapkan “*Welcome to the Diabetes Predictor!*”. Di halaman ini pengguna akan menekan tombol *get started*. Hasil dari menekan tombol tersebut pengguna akan dibawa ke halaman selanjutnya yaitu *Data Info*.

Di menu *data info* ini pengguna akan memasukan data Kesehatan pribadi seperti usia, kadar insulin tubuh, jumlah kelahiran, *body mass index*, glukosa, ketebalan lipatan kulit, tekanan darah dan *diabetes pedigree function* (Riwayat diabetes dalam keluarga). Untuk nilai di masing-masing parameter tersebut pengguna bisa dengan bebas mengisi nilai dari setiap parameter itu disesuaikan dengan kondisi dari pengguna itu sendiri.

Di menu *data info* menampilkan data dari dataset yang digunakan. Dengan ketentuan range nilainya sebagai berikut: Untuk usia di mulai dari rentang usia 21-81 tahun, kadar insulin dari 0-846, jumlah kelahiran dari 0-17, BMI dari 0-67.1, glukosa dari 0-199, tekanan darah dari 0-122 dan DPF yang memiliki dua input yaitu 1 (ada riwayat keluarga diabetes) dan 0 (tidak ada riwayat keluarga diabetes). Dalam menu *prediction*, pengguna akan memasukan data Kesehatan pribadi seperti usia, kadar insulin tubuh, jumlah kelahiran, *body mass index*, glukosa, ketebalan lipatan kulit, tekanan darah dan *diabetes pedigree function* (riwayat diabetes dalam keluarga). Untuk nilai di masing-masing parameter tersebut pengguna bisa dengan bebas mengisi nilai dengan menyesuaikan kondisi tubuhnya, Setelah

menginput datanya langkah selanjutnya pengguna menekan tombol prediksi dan hasil prediksinya akan keluar saat itu juga.

Untuk menu *visualization* akan menampilkan grafik prediksi dari tiap parameter pada dataset prediksi diabetes yang digunakan. Sehingga jikalau dataset yang digunakan berbeda maka hasil dari prediksi grafiknya juga akan berbeda.

Pada menu *contact us*, pengguna dapat berkomunikasi seputar prediksi penyakit diabetes ini kepada pengembang *website*. Baik itu dalam bentuk saran, penambahan informasi, ucapan terimakasih dan sebagainya. Nantinya pengguna akan menginput nama, *email* dan juga pesan yang ingin disampaikan dan terakhir *klik submit*.

Di menu FAQ berisi seputar pertanyaan yang sering ditanyakan oleh pengguna seputar penyakit diabetes. Contohnya ap aitu DPF, ap aitu BMI dan sebagainya. Dengan adanya menu ini pengguna bisa mendapatkan informasi seputar penyakit diabetes.

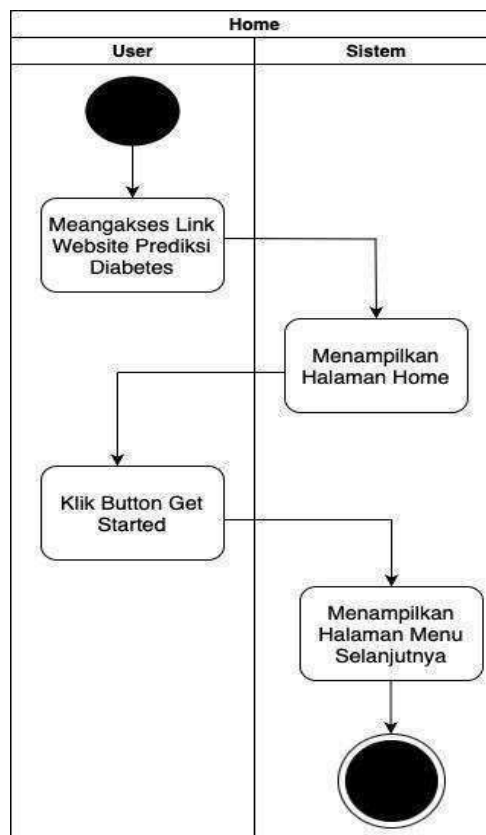
Pada menu *chatbot* di halaman *web* berupa *ikon* berbentuk pesan yang menggunakan teknologi kecerdasan buatan, yang mana pengguna dapat bertanya seputar penyakit diabetes dan cara mengatasinya.

2. Diagram Aktivitas

Diagram aktivitas mereplikasi proses dan aliran dari sistem operasi. Oleh karena itu, pemodelan aliran dilakukan dalam kasus penggunaan, pada tingkat proses bisnis, dan kadang-kadang di antara kasus penggunaan. baik di tingkat perusahaan maupun di tingkat teknis. Diagram aktivitas menampilkan perilaku proses bisnis

internal. Diagram ini biasanya digunakan sebagai diagram konteks untuk menjelaskan bagaimana operasi bisnis yang berbeda terkait satu sama lain. Kemampuan diagram ini untuk menunjukkan ketergantungan interaktif adalah fitur kunci. Menggunakan diagram aktivitas untuk mengalokasikan tugas ke aktor sistem yang sesuai juga berguna. Selain itu, karena kemampuan untuk menyelesaikan beberapa tugas dalam satu sistem sekaligus (UNhelkar, 2019).

a. Home



Gambar 3. 13 Diagram Aktivitas *Home*

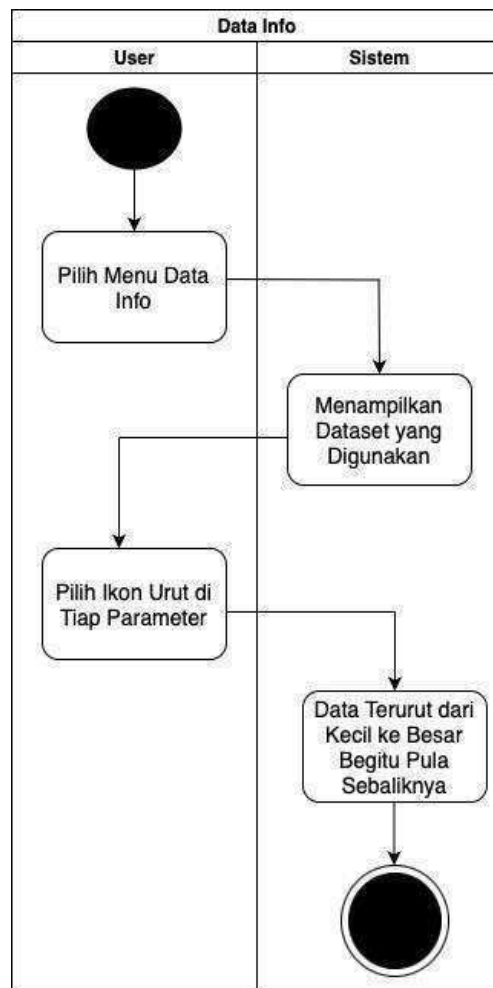
Sumber: Penelitian (2024)

Gambar 3.11 menjelaskan bahwa ketika pengguna pertama kali membuka

website prediksi penyakit diabetes, maka sistem akan menampilkan halaman *home*.

Setelah itu, pengguna akan mengklik tombol mulai dan sistem akan meresponnya dengan menampilkan halaman menu selanjutnya.

b. Data Info



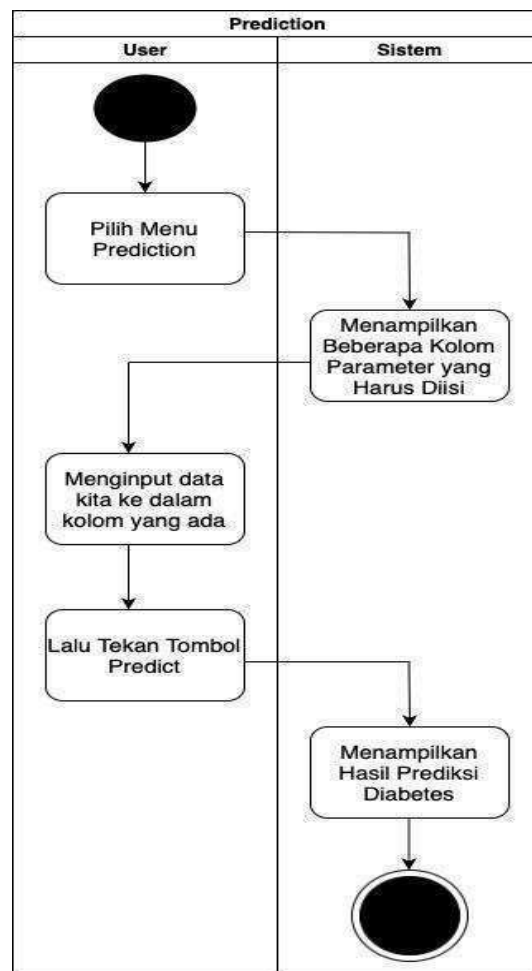
Gambar 3. 14 Diagram Aktivitas *Data Info*

Sumber: Penelitian (2024)

Penjelasan Gambar 3.12 yaitu pengguna akan memilih menu data info, lalu sistem akan menampilkan *dataset* yang digunakan untuk prediksi penyakit diabetes.

Sesudah itu, pengguna akan memilih ikon urutan untuk setiap parameter atau variabel yang dapat mengurutkan nilai dari yang terkecil ke terbesar begitu pula sebaliknya. Nantinya data yang terurut pun akan tampil.

c. *Prediction*



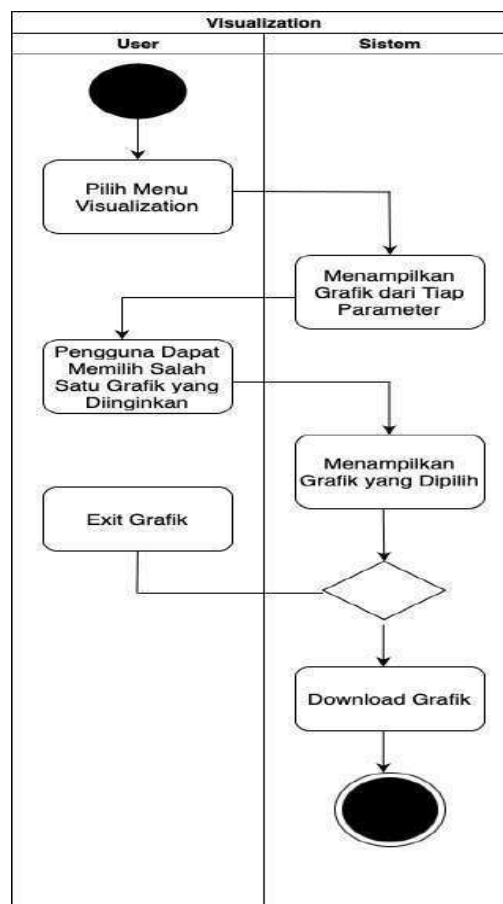
Gambar 3. 15 Diagram Aktivitas *Prediction*

Sumber: Penelitian (2024)

Pada gambar di atas menjelaskan Ketika pengguna memilih menu *prediction*, sistem akan menampilkan beberapa kolom yang berisikan variabel dalam prediksi

penyakit diabetes. Seperti glukosa, insulin, DPF, tekanan darah, BMI, ketebalan kulit dan sebagainya. Pengguna akan menginput data ke dalam tiap kotak yang berisikan masing-masing variabel tersebut. Setelah itu tekan tombol *predict* maka sistem akan menampilkan hasilnya, apakah terkena penyakit diabetes atau tidak.

d. *Visualization*



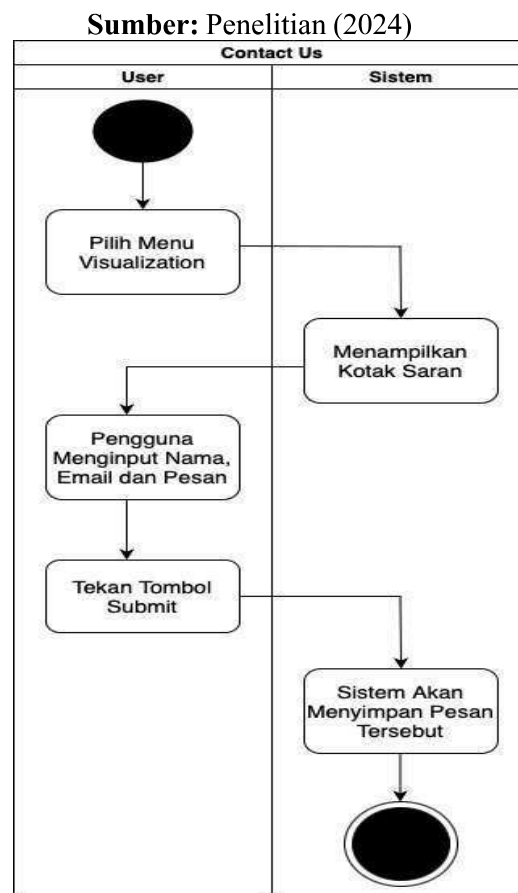
Gambar 3. 16 Diagram Aktivitas *Visualization*

Sumber: Penelitian (2024)

Awal mulanya ketika pengguna memilih menu *visualization*, sistem akan menampilkan grafik dari tiap variabel prediksi penyakit diabetes. Grafik ini

dihasilkan dari penghitungan *outlier* menggunakan *python*. Pengguna dapat memilih salah satu grafik yang diinginkan dari tiap variabel. Dan sistem akan menampilkan grafiknya. Pada sistem ada dua kemungkinan yang ditampilkan, yaitu pengguna memilih keluar dari grafik atau mendownload tampilan grafik tersebut. Setelah memilih di antara dua pilihan itu, proses pun selesai.

e. Contact Us

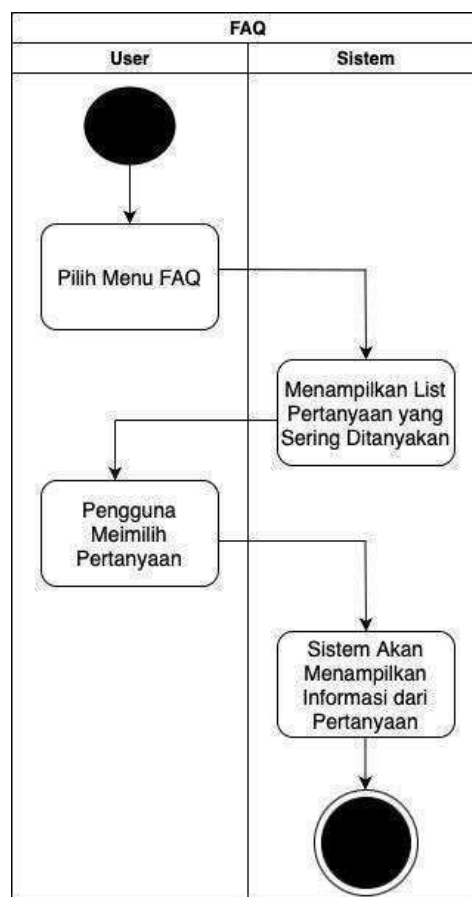


Gambar 3. 17 Diagram Aktivitas *Contact Us*

Gambar 3.15 menunjukkan pengguna memilih menu *contact us*, nantinya

sistem akan menampilkan kotak saran yang terdiri dari nama pengguna, email, dan juga pesan yang ingin disampaikan. Selanjutnya, pengguna pun mengisi data yang dibutuhkan di kotak saran dan tekan tombol submit. Dan sistem akan menyimpan data tersebut.

f. FAQ

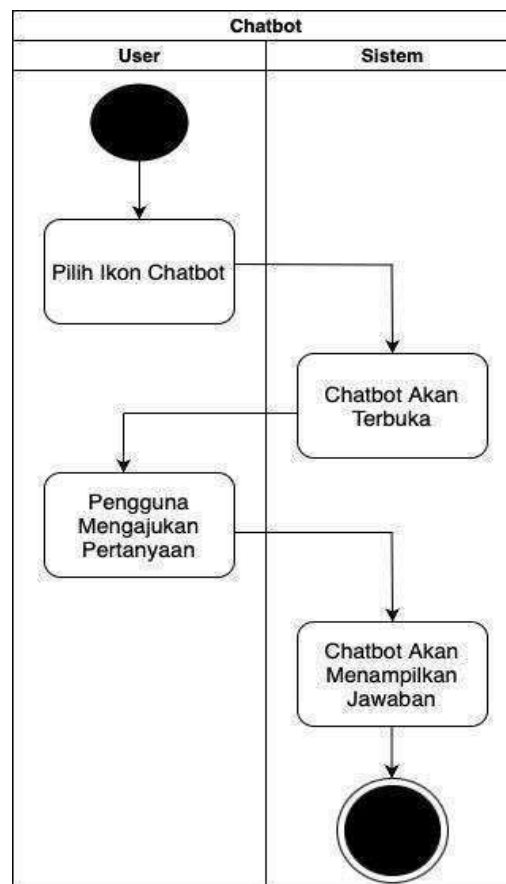


Gambar 3.16 Diagram Aktivitas FAQ
Sumber: Penelitian (2024)

Dalam Gambar 3.16 pengguna memilih menu FAQ, lalu sistem akan menampilkan list pertanyaan yang akan sering ditanyakan oleh pengguna.

Pengguna akan memilih dari beberapa pertanyaan itu sesuai dengan kebutuhan dan kondisinya. Dan sistem akan menampilkan informasi dari pertanyaan tersebut.

g. Chatbots



Gambar 3. 18 Diagram Aktivitas *Chatbot*

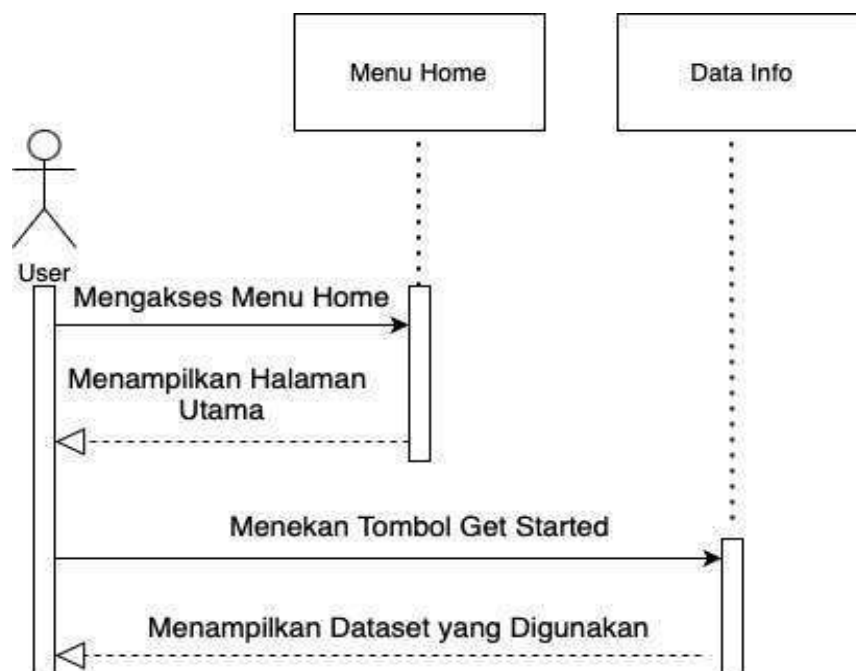
Sumber: Penelitian (2024)

Untuk Gambar 3.17 pengguna akan memilih ikon *chatbot* yang berada dibawah dan sistem akan membuka *chatbot* tersebut. Pengguna dapat mengajukan pertanyaan pada *chatbot*, dan *chatbot* akan menampilkan jawabannya pada sistem.

3. Diagram *Sequence*

Sejak diperkenalkan oleh Jacobson sebagai alat untuk dokumentasi perilaku kasus penggunaan, diagram urutan telah mendapatkan popularitas. Karena secara visual menggambarkan situasi (atau contoh) dalam kasus penggunaan, diagram urutan juga dikenal sebagai diagram sekuensi dalam penggunaan sebelumnya (UNhelkar, 2019).

a. *Home*

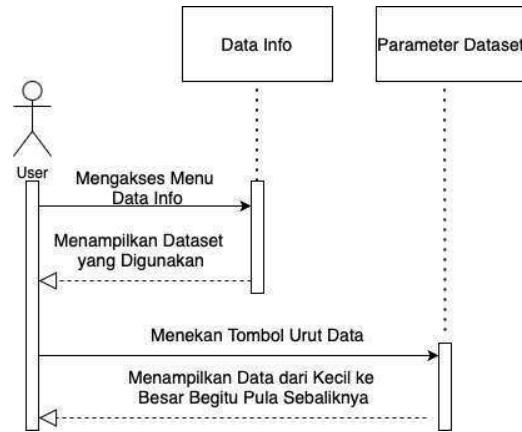


Gambar 3. 19 Diagram Urutan Menu *Home*

Sumber: Penelitian (2024)

Pada Gambar 3.18 pengguna akan mengakses menu *home* sehingga halaman utama akan ditampilkan. Lalu, pengguna menekan tombol *get started* sehingga *dataset* pun ditampilkan.

b. Data Info

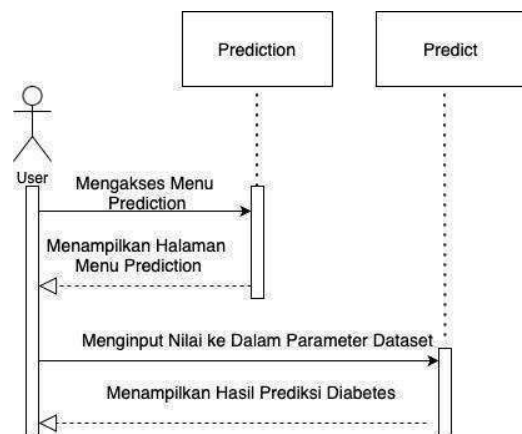


Gambar 3. 20 Diagram Urutan Menu *Data Info*

Sumber: Penelitian (2024)

Gambar 3.19 menunjukkan pengguna waktu mengakses menu data info, maka *dataset* akan ditampilkan oleh sistem. Ketika pengguna menekan tombolurut data, maka data pun akan terurut.

c. Prediction

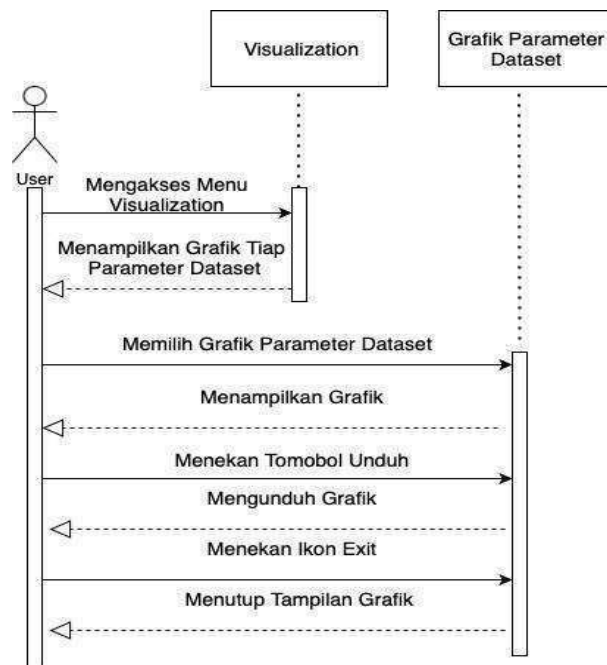


Gambar 3. 21 Diagram Urutan Menu *Prediction*

Sumber: Penelitian (2024)

Gambar 3.20 menampilkan ketika pengguna ingin mengakses menu *prediction*, maka halaman dari menu *prediction* akan ditampilkan. Selanjutnya, pengguna menginput data ke dalam tiap kolom yang merupakan parameter dari tiap variabel, sehingga hasil prediksi penyakit diabetes pun akan ditampilkan.

d. Visualization

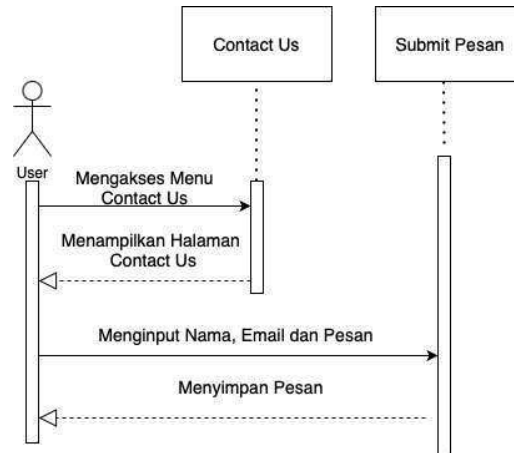


Gambar 3. 22 Diagram Urutan Menu *Visualization*

Sumber: Penelitian (2024)

Gambar diatas menunjukan Ketika pengguna mengakses menu *visualization* nantinya akan ditampilkan grafik dari setiap variabel. Pengguna akan memilih grafik yang ingin dilihat, lalu grafik akan tampil. Setelah itu pengguna dapat mengunduh grafik atau menutup tampilan grafik.

e. Contact Us

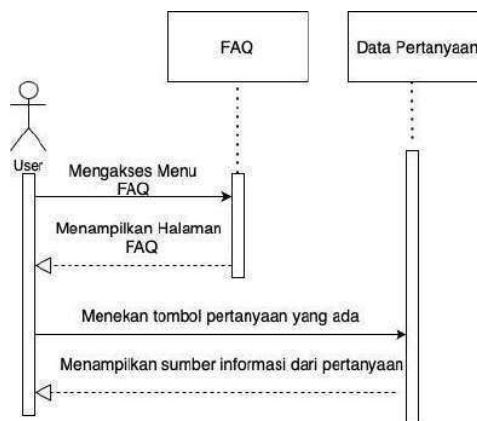


Gambar 3. 23 Diagram Urutan Menu *Contact Us*

Sumber: Penelitian (2024)

Pada Gambar 3.22 terlihat Ketika pengguna mengakses menu *contact us*, maka halaman dari *contact us* akan terlihat. Setelah itu pengguna akan menginput nama, email dan juga pesan, maka pesan akan disimpan oleh sistem

f. FAQ

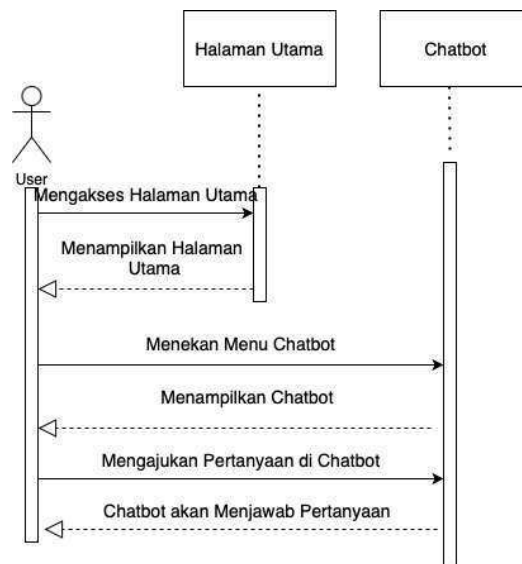


Gambar 3. 24 Diagram Urutan Menu FAQ

Sumber: Penelitian (2024)

Gambar 3.23 saat pengguna mengakses menu FAQ, sistem akan menampilkan menu tersebut. Lalu pengguna akan menekan salah satu pertanyaan yang diinginkan dan pertanyaan tersebut pun tampil.

g. Chatbot



Gambar 3. 25 Diagram Urutan Menu FAQ

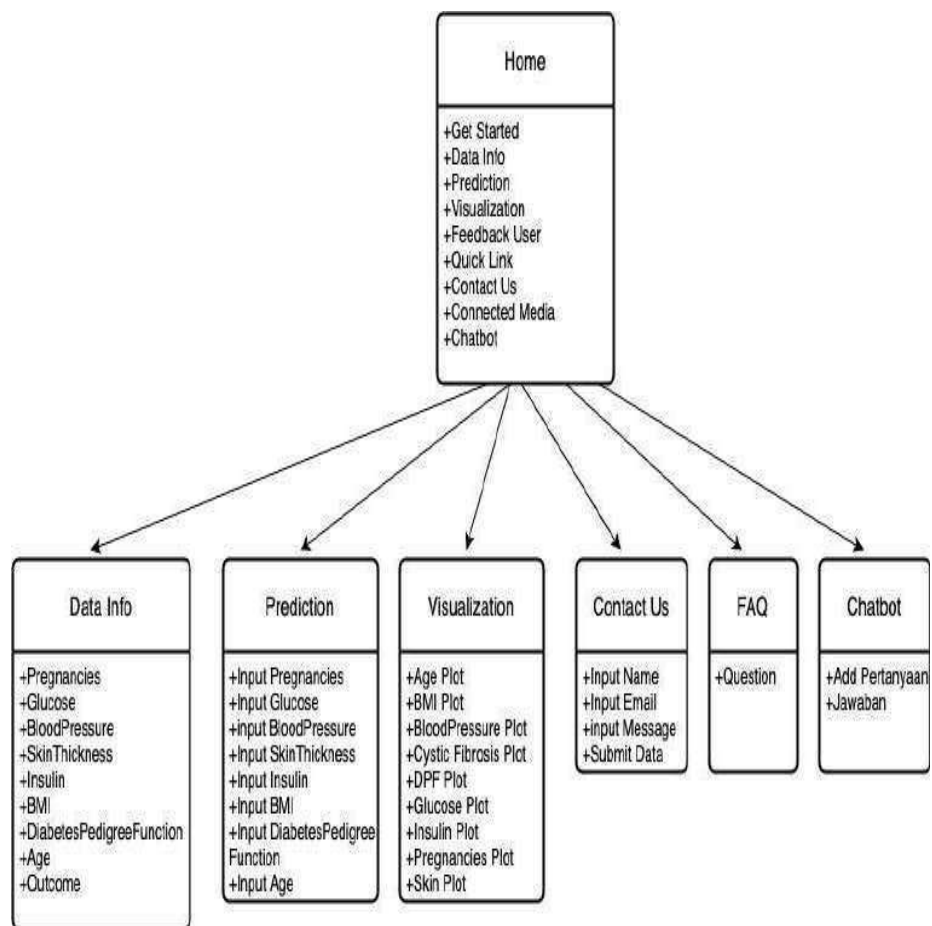
Sumber: Penelitian (2024)

Digram diatas menjelaskan saat pengguna mengakses halaman utama, yang akan dilihat oleh sistem. Setelah itu pengguna akan *klik chatbot* dan sistem akan menampilkannya. Pengguna dapat mengajukan pertanyaan dan *chatbot* akan meresponnya.

4. Diagram Class

Dalam penelitian ini, struktur suatu sistem dijelaskan dalam hal kelas yang

dibangun menggunakan diagram kelas. Entitas penting, yang sifatnya sebagian besar struktural dan statis, diwakili oleh diagram ini. Selain kelas, grafik ini menampilkan hubungan antara kelas. Gambar 3.17 akan digunakan untuk membahas seluruh deskripsi kelas atau entitas, seperti yang disebutkan sebelumnya dalam ruang masalah pada penelitian ini.



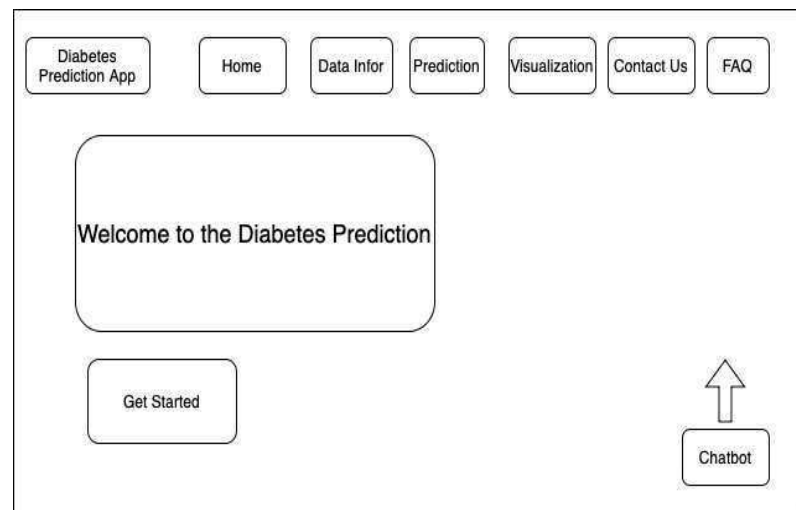
Gambar 3. 26 Diagram *Class*

Sumber: Penelitian (2024)

Diagram kelas pada penelitian ini digunakan untuk menggambarkan struktur statis dari sistem perangkat lunak serta menunjukkan kelas-kelas yang ada didalamnya.

3.6.2 Perancangan *Interface*

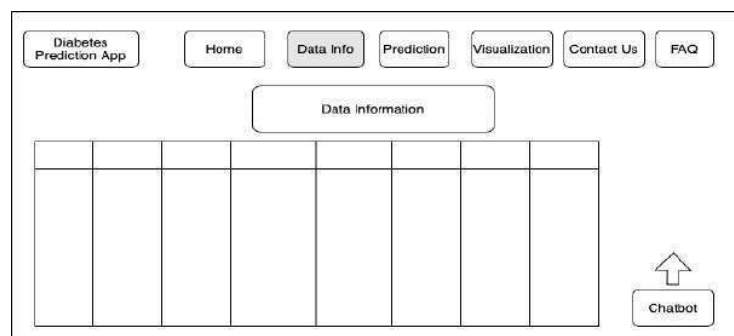
1. Tampilan Halaman *Home*



Gambar 3. 27 Halaman *Home*

Sumber: Penelitian (2024)

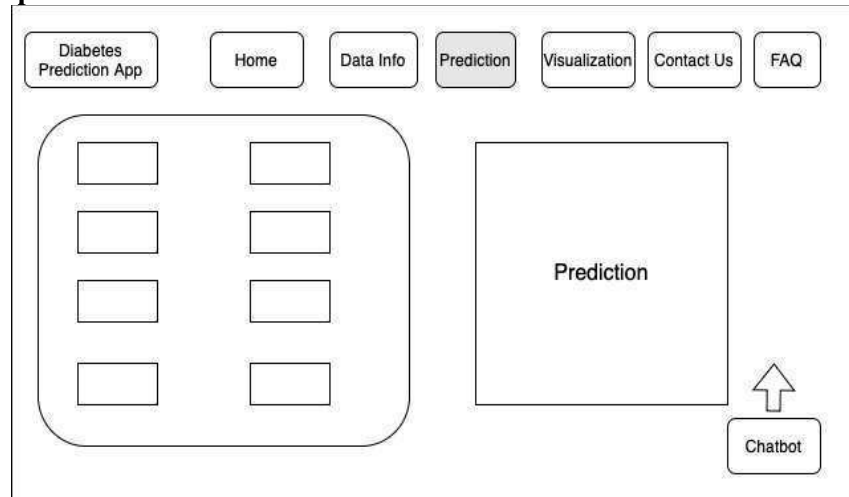
2. Tampilan Halaman *Data Info*



Gambar 3. 28 Halaman *Data Info*

Sumber: Penelitian (2024)

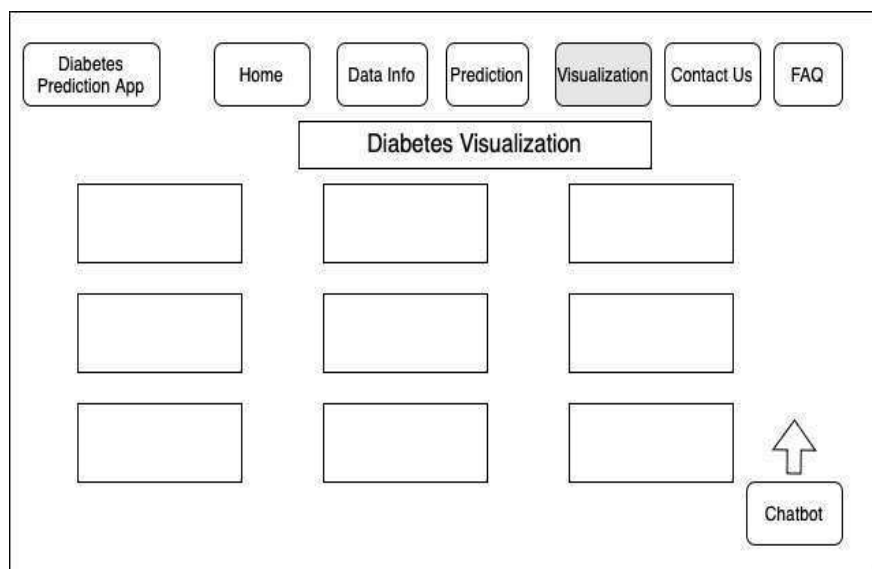
3. Tampilan Halaman *Prediction*



Gambar 3. 29 Halaman *Prediction*

Sumber: Penelitian (2024)

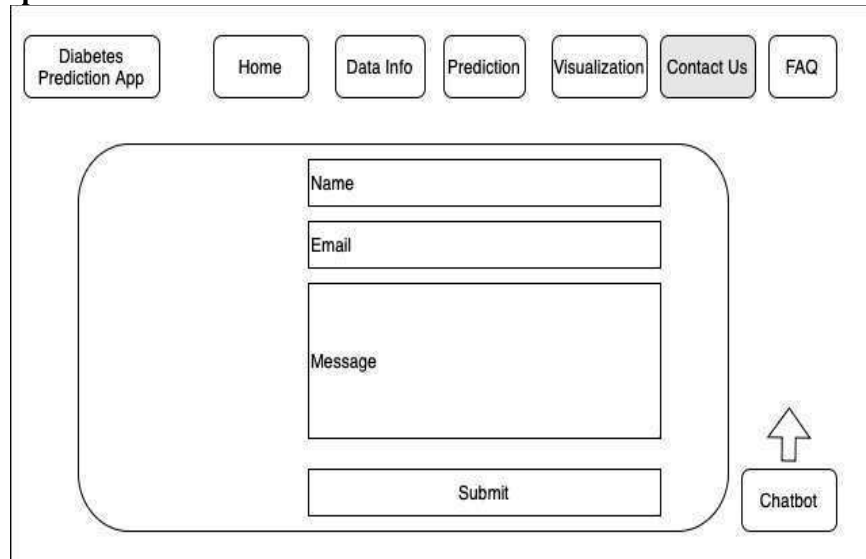
4. Tampilan Halaman *Visualization*



Gambar 3. 30 Halaman *Visualization*

Sumber: Penelitian (2024)

5. Tampilan Halaman *Contact Us*

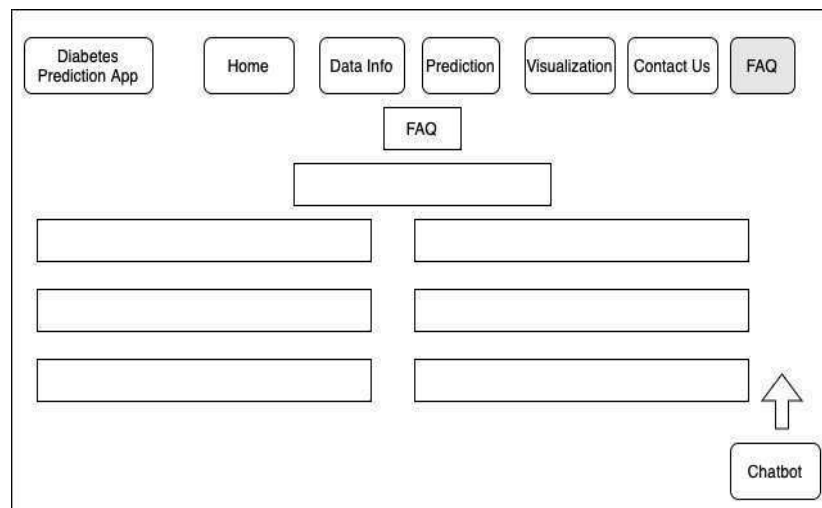


The wireframe for the 'Contact Us' page features a top navigation bar with buttons for 'Diabetes Prediction App', 'Home', 'Data Info', 'Prediction', 'Visualization', 'Contact Us' (highlighted), and 'FAQ'. The main content area contains a large rounded rectangle with three input fields labeled 'Name', 'Email', and 'Message', followed by a 'Submit' button. To the right of this form is a 'Chatbot' button with an upward-pointing arrow icon.

Gambar 3. 31 Halaman *Contact Us*

Sumber: Penelitian (2024)

6. Tampilan Halaman FAQ

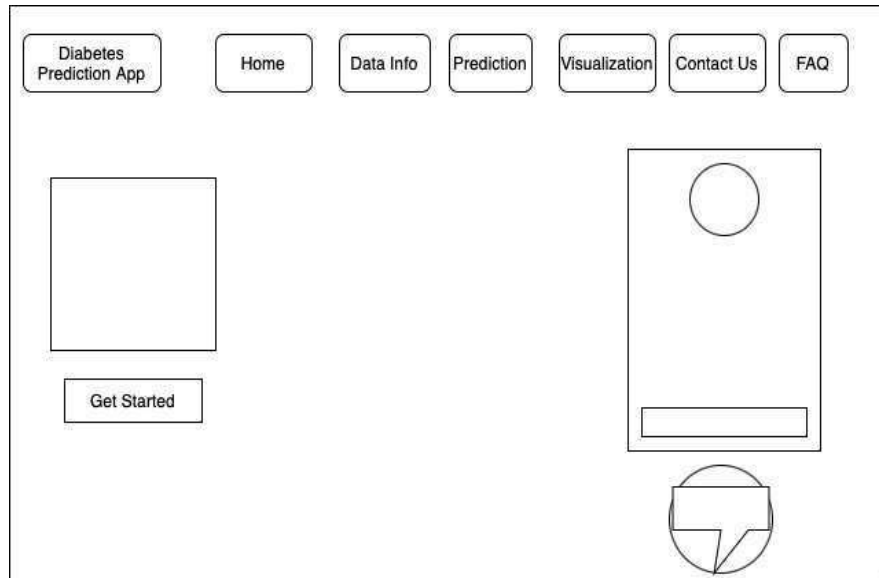


The wireframe for the 'FAQ' page features a top navigation bar with buttons for 'Diabetes Prediction App', 'Home', 'Data Info', 'Prediction', 'Visualization', 'Contact Us', and 'FAQ' (highlighted). Below the navigation bar is a sub-header 'FAQ' and a search input field. The main content area displays a grid of six question-and-answer pairs, each consisting of a question box and an answer box. A 'Chatbot' button with an upward-pointing arrow icon is located at the bottom right.

Gambar 3. 32 Halaman FAQ

Sumber: Penelitian (2024)

7. Tampilan Halaman *Chatbot*



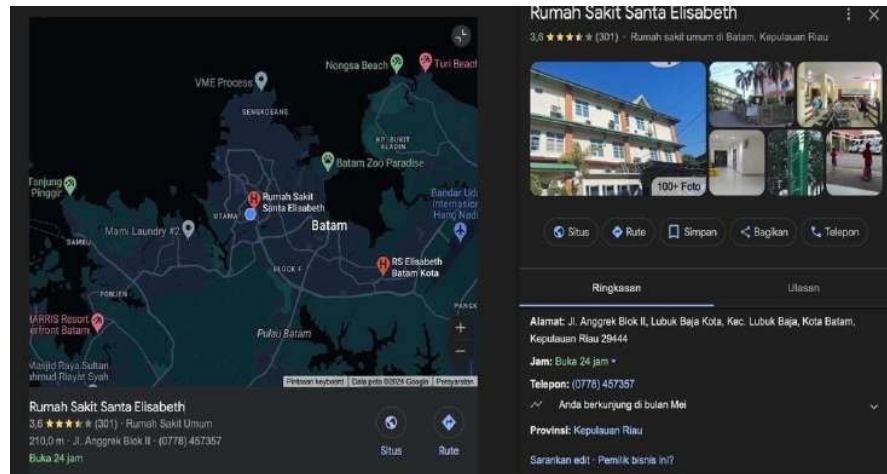
Gambar 3. 33 Halaman *Chatbot*

Sumber: Penelitian (2024)

3.7 Lokasi dan Jadwal Penelitian

a. Lokasi

Lokasi penelitian ini dilakukan di RS Elizabeth Jln. Anggrek Blok 2 Batam.



Gambar 3. 34 Lokasi Penelitian

Sumber: Penelitian (2024)

b. Jadwal Penelitian

Ini adalah pengadilan enam bulan. Rentang waktu studi ditunjukkan pada Gambar 3.26.

Nama kegiatan	Maret 2024				April 2024				Mei 2024				Juni 2024				Juli 2024			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Pengajuan judul																				
Penyusunan BAB 1																				
Penyusunan BAB 2																				
Penyusunan BAB 3																				
Penyusunan BAB 4																				
Penyusunan BAB 5																				
Penyelesaian Sikripsi																				

Gambar 3. 35 Jadwal Penelitian

Sumber: Penelitian (2024)