

BAB III

METODE PENELITIAN

3.1. Bahan Penelitian

Bahan penelitian menggunakan data pasien gejala Strok yang telah terkumpul pada situs Kaggle untuk dijadikan acuan penelitian. Data akan diklasifikasikan menggunakan metode Algoritma SVM untuk diprediksi apakah Algoritma SVM mendapatkan hasil terbaik dalam prediksi penyakit Strok. Data penyakit strok dapat dilihat pada tabel 3.1.

Tabel 3.1 Tabel Bahan Data Penyakit Strok

<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>	<i>7</i>	<i>8</i>	<i>9</i>	<i>10</i>	<i>11</i>
<i>Male</i>	55	0	0	<i>Yes</i>	<i>Private</i>	<i>Urban</i>	89.17.00	31.05.00	<i>Never Smoked</i>	0
<i>Female</i>	42	0	0	<i>No</i>	<i>Private</i>	<i>Urban</i>	98.53.00	18.05	<i>Never Smoked</i>	0
<i>Female</i>	24	0	0	<i>No</i>	<i>Private</i>	<i>Urban</i>	97.55.00	26.02.00	<i>Never Smoked</i>	0
<i>Female</i>	33	0	0	<i>Yes</i>	<i>Private</i>	<i>Rural</i>	86.97	42.02.00	<i>Never Smoked</i>	0
<i>Female</i>	20	0	0	<i>No</i>	<i>Private</i>	<i>Rural</i>	84.07.00	27.06.00	<i>Smokes</i>	0

Dengan keterangan variabel kolom pada tabel 3.2.

Tabel 3.2 Tabel keterangan Data Penyakit Strok

Variabel Kolom	Nomor	Keterangan
<i>Gender</i>	1	Jenis Kelamin
<i>Age</i>	2	Umur
<i>Hypertension</i>	3	Riwayat Hipertensi
<i>Heart Disease</i>	4	Riwayat Jantung
<i>Ever Married</i>	5	Pernah Menikah
<i>Work Type</i>	6	Jenis Pekerjaan
<i>Residence Type</i>	7	Tipe Perumahan
<i>Average Glucose Level</i>	8	Tingkat Gula
<i>BMI</i>	9	Indeks Massa Tubuh
<i>Smoking Status</i>	10	Status Perokok
<i>Stroke</i>	11	Diagnosa Strok

3.2. Peralatan Penelitian

Perangkat keras yang digunakan dalam penelitian menggunakan Laptop, Processor Intel Core i5 2520M vPro 2.5GHz L3 Cache 3MB, RAM 4 GB, HDD 500 GB dan Chipset Intel HM65. Perangkat lunak yang digunakan dalam penelitian ini adalah: *Operating System Microsoft Windows 2016 Professional* 64bit, *Microsoft Excel* 2016, Windows 10 dan Google Colab

3.3. Lokasi dan Waktu Penelitian

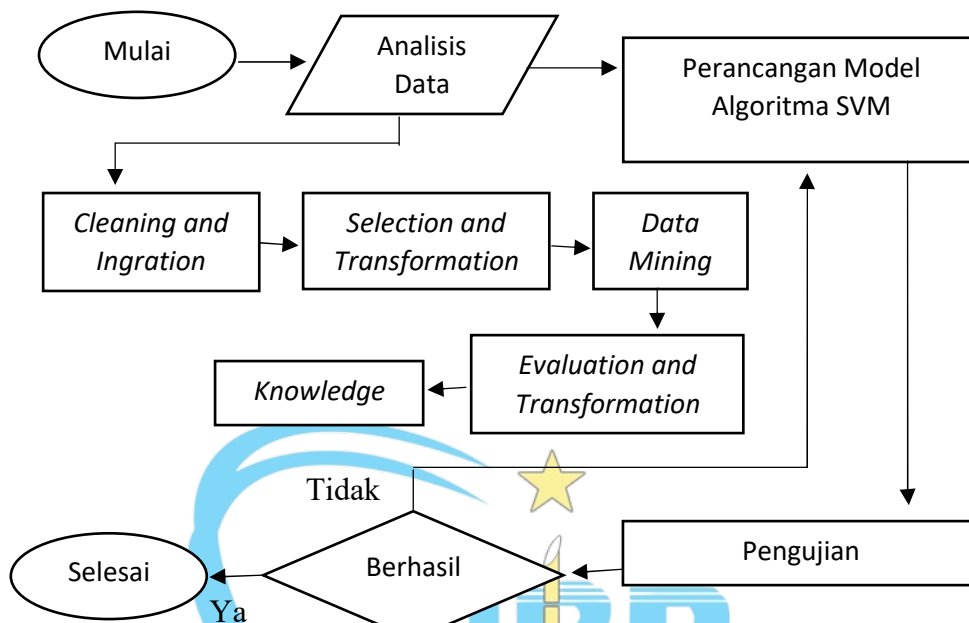
Lokasi yang digunakan dalam penelitian ini dilaksanakan di Lab Riset Universitas Buana Perjuangan Karawang. Dilaksanakan sejak bulan Februari sampai dengan April 2022 pada tabel 3.3.

Tabel 3.3 Tabel Waktu Penelitian

No.	Kegiatan	Bulan 1				Bulan 2				Bulan 3			
		1	2	3	4	1	2	3	4	1	2	3	4
1.	Studi Literatur												
2.	Penulisan Proposal												
3.	Analisis Data												
4.	Implementasi Model												
5.	Evaluasi												
6.	Pengujian												
7.	Penulisan Laporan												
8.	Tugas Akhir												
9.	Yudisium												

3.4. Prosedur Penelitian

Prosedur Penelitian ini dimulai dengan analisis *dataset* yang telah diterbitkan oleh situs kaggle, dengan rincian penelitian sebagai berikut:



Gambar 3.1 Gambar Prosedur Penelitian

3.5. Analisis Data

Data yang digunakan pada penelitian ini yaitu rekapitulasi data pasien yang diperoleh dari situs Kaggle dengan 3.426 baris dan 10 kolom variabel yaitu jenis kelamin, umur, hipertensi, penyakit jantung, status menikah, jenis pekerjaan, jenis tempat tinggal, tingkat gula darah, indeks massa tubuh, status merokok yang dijadikan sebagai data *testing*, dan 3.426 baris dengan 11 kolom variabel sama yang ditambahkan diagnosis penyakit Strok. Data yang didapat masih bersifat inkonsisten, sehingga data banyak yang belum sesuai untuk dilakukan proses. oleh karena itu data penelitian ini harus melalui tahap *preprocessing* agar mendapatkan format data yang dibutuhkan. *Preprocessing* data adalah tahapan penting yang wajib dilakukan dalam proses *data mining*, yang mengubah suatu *dataset* atau atribut menjadi seragam (Rerung, 2018).

3.6. Implementasi Algoritma

Algoritma yang digunakan dalam penelitian ini menggunakan *Support Vector Machine* (SVM). Implementasi melalui proses data dan klasifikasi menjadi

beberapa bagian. Penelitian yang dilakukan yaitu untuk mengidentifikasi data yang digunakan untuk menghasilkan diagnosis pasien apakah positif stroke atau negatif stroke.

3.7. Pengujian

Tahap pengujian *dataset* menggunakan Algoritma SVM, diuji melalui situs pemrograman *Python Google Colab*. Beberapa tahapan proses pengujian dibagi menjadi dua bagian metode yaitu:

3.7.1 Metode Algoritma SVM dengan *Relief-f*

Metode *Relief-f* memiliki kinerja yang baik dalam memprediksi kecacatan perangkat lunak. *Dataset* prediksi Stroke memiliki atribut yang tidak relevan sehingga data menurunkan kinerja akurasi prediksi *machine learning*. Oleh karena itu penelitian ini menggunakan teknik filter variabel yang tidak diperlukan dalam proses penelitian.

3.7.2 Metode *Confusion Matrix*

Data yang akan diuji tingkat akurasinya menggunakan *Confusion Matrix*. Metode *Confusion Matrix* digunakan sebagai pengukur proses suatu metode klasifikasi. *Confusion Matrix* mengandung empat istilah yaitu *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), *False Negative* (FN) (Karsito & Susanti, 2019). Tabel 3.7 merupakan dasar untuk melakukan perhitungan *Confusion Matrix* (Pratiwi, Handayani, & Sarjana, 2020).

Tabel 3.4 Tabel Metode *Confusion Matrix*

		<i>Observed</i>	
		<i>True</i>	<i>False</i>
<i>Predicted Class</i>	<i>True</i>	<i>True Positive (TP)</i>	<i>False Positive (FP)</i>
<i>Predicted Class</i>	<i>False</i>	<i>False Negative (FN)</i>	<i>True Negative (TN)</i>

Penjelasan pada tabel metode *confusion matrix*,

True Positive (TP) adalah akumulasi data positif yang terklasifikasi dengan tepat oleh sistem. Kemudian *True Negative* (TN) adalah akumulasi

data negatif yang terklasifikasi dengan tepat oleh sistem. *False Negative* (FN) adalah akumulasi data negatif yang terklasifikasi salah oleh sistem. *False Positive* (FP) adalah akumulasi data positif yang terklasifikasi salah oleh sistem. Untuk menghitung jumlah klasifikasi, dapat dilakukan dengan cara perhitungan Akurasi. Akurasi merupakan hasil dari suatu proses yang menunjukkan nilai data antara data yang terklasifikasi dengan benar dari keseluruhan data dalam bentuk persentase untuk memprediksi diagnosa penyakit Strok. Rumus akurasi dapat digunakan dengan persamaan 10.

$$Akurasi = \frac{TN + TP}{TP + TN + FP + FN} \quad (13)$$

