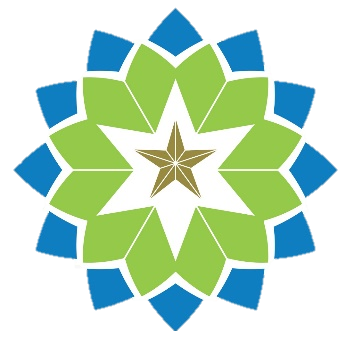
**LAPORAN MAKALAH**

**KLASIFIKASI PENYAKIT STROKE**

**MENGGUNAKAN METODE *SUPPORT VECTOR MACHINE***

**Disusun Untuk Memenuhi Tugas Mata Kuliah Kecerdasan Buatan**



Disusun Oleh :

Muhammad Fadli (11210910000079) UIN Jakarta

Muhammad Alif Ferdiansyah (11210910000103) UIN Jakarta

Hilda Zaqya Elnaz Putri (200601110112) UIN Malang

Ihza Zhafran Ramadhan (1217010038) UIN Sunan Gunung Djati Bandung

# DAFTAR ISI

[**DAFTAR ISI 2**](#_Toc150381563)

[**BAB I PENDAHULUAN 3**](#_Toc150381564)

[1.1. Latar Belakang 3](#_Toc150381566)

[1.2. Rumusan Masalah 4](#_Toc150381567)

[1.3. Tujuan Penelitian 5](#_Toc150381568)

[**BAB II KAJIAN TEORI 1**](#_Toc150381571)

[2.1 Data Mining 1](#_Toc150381573)

[2.2 Algoritma SVM (Support Vector Machine) 1](#_Toc150381574)

[2.3 Penyakit Stroke 1](#_Toc150381575)

[**BAB III METODE PENELITIAN 4**](#_Toc150381576)

[3.1 Jenis Penelitian 4](#_Toc150381578)

[3.2 Teknik Analisis Data 5](#_Toc150381579)

[3.2.1 *Exploratory Data Analysis* (EDA) 5](#_Toc150381580)

[3.2.2 *Data Selection* 6](#_Toc150381581)

[3.2.3 *Data Preprocessing* 7](#_Toc150381582)

[3.2.4 *Transformation Data* 7](#_Toc150381583)

[3.2.5 Klasifikasi 7](#_Toc150381584)

[3.2.6 Evaluasi Model 8](#_Toc150381585)

# BAB I

# PENDAHULUAN

## **Latar Belakang**

Stroke adalah kondisi medis serius yang terjadi ketika pasokan darah ke otak terganggu, menyebabkan kerusakan otak. Ada dua jenis stroke: iskemik (penyumbatan darah) dan hemoragik (pembuluh darah pecah). Faktor risiko termasuk tekanan darah tinggi, merokok, diabetes, dll. Gejala stroke meliputi kesulitan berbicara, kelemahan, dan gangguan penglihatan. Deteksi dan pengobatan dini sangat penting. Pencegahan melalui gaya hidup sehat seperti diet seimbang dan olahraga adalah kunci untuk mengurangi risiko stroke. Stroke bisa mengakibatkan kecacatan fisik dan mental serta kematian jika tidak ditangani dengan cepat.

Deteksi dini penyakit stroke sangat penting karena pengobatan yang cepat dan tepat dapat mengurangi kerusakan otak, menghindari komplikasi serius, dan bahkan menyelamatkan nyawa pasien. Deteksi dini memungkinkan untuk memberikan terapi seperti trombolisis dan intervensi bedah dalam waktu yang sangat penting, yang bisa mengatasi penyumbatan pembuluh darah otak. Hal ini juga meningkatkan peluang pemulihan dan mengurangi risiko kecacatan fisik, gangguan berbicara, dan masalah kognitif yang serius yang dapat terjadi akibat stroke. Jadi, kesadaran dan deteksi dini gejala stroke adalah langkah penting untuk menjaga kesehatan dan keselamatan pasien.

Teknologi, terutama metode komputer, memainkan peran penting dalam deteksi penyakit stroke. Ini termasuk penggunaan algoritma seperti Support Vector Machine (SVM) untuk mengklasifikasikan data medis. Berikut adalah bagaimana teknologi membantu dalam mendeteksi stroke:

* Pencitraan Medis: Teknologi pencitraan medis seperti CT scan dan MRI memungkinkan dokter melihat dengan rinci kondisi otak pasien dan mendeteksi tanda-tanda stroke.
* Analisis Data: Metode komputer, termasuk SVM, digunakan untuk menganalisis data medis. Mereka dapat memproses data klinis, parameter medis, dan hasil pencitraan untuk mengidentifikasi tanda-tanda stroke.
* Pengenalan Pola: Teknologi komputer dapat mengenali pola dalam data pasien, yang dapat membantu dalam diagnosis dan prediksi risiko stroke.
* Telemedicine: Telemedicine memanfaatkan teknologi untuk memantau pasien yang berisiko tinggi mengalami stroke secara jarak jauh, memungkinkan pengambilan tindakan darurat jika diperlukan.
* Sistem Pendukung Keputusan: Sistem komputer dapat memberikan panduan kepada dokter dalam membuat keputusan diagnosis dan perencanaan perawatan.

Penggunaan teknologi dalam deteksi stroke membantu dokter mengidentifikasi kasus-kasus stroke lebih cepat dan dengan akurasi yang lebih besar, memungkinkan perawatan yang tepat waktu dan efektif. SVM dan algoritma lainnya membantu dalam mengklasifikasikan data medis, membantu dalam mengenali tanda-tanda stroke dan mengarahkan perawatan yang sesuai. Hal ini memiliki dampak positif pada prognosis pasien dan pengurangan komplikasi serius.

Penelitian terdahulu mungkin memiliki keterbatasan seperti sampel kecil, data yang tidak lengkap, fokus tunggal, atau metode analisis yang terbatas. Penelitian kami akan mengatasi beberapa dari keterbatasan tersebut dengan mengambil sampel yang lebih besar, mengumpulkan data yang lebih akurat, menggabungkan metode klasifikasi yang beragam, menggunakan teknik analisis yang canggih, memperluas cakupan usia pasien, dan melakukan validasi eksternal yang kuat. Dengan demikian, penelitian kami akan lebih solid dan relevan.

## **Rumusan Masalah**

Penelitian dan upaya telah dilakukan dalam pengembangan metode deteksi stroke, masih ada tantangan besar yang perlu diatasi. Penelitian-penelitian terdahulu mungkin memiliki keterbatasan seperti ukuran sampel yang kecil, data yang tidak lengkap, dan fokus yang terlalu spesifik. Dalam konteks ini, masalah utama yang kami hadapi adalah bagaimana mengembangkan metode klasifikasi yang akurat untuk mendeteksi penyakit stroke dengan segera.

Tujuan utama dari penelitian ini adalah mengatasi masalah ini dengan menggunakan metode Support Vector Machine (SVM) untuk mengklasifikasikan data medis pasien dan mengidentifikasi kasus-kasus yang memerlukan perhatian segera. Kami akan memperluas sampel data, mengumpulkan data yang lebih akurat, dan memperhitungkan berbagai faktor risiko dan gejala yang terkait dengan stroke. Selain itu, kami akan menggabungkan teknik analisis canggih dan melakukan validasi eksternal yang kuat untuk meningkatkan akurasi model.

Dengan menghadapi masalah utama ini, kami bertujuan untuk memberikan kontribusi yang signifikan dalam meningkatkan deteksi dini penyakit stroke, yang pada gilirannya akan membantu dalam mengurangi dampak serius yang terkait dengan penyakit ini dan menyelamatkan nyawa pasien.

## **Tujuan Penelitian**

### Tujuan Utama Penelitian

Tujuan utama dari penelitian ini adalah meningkatkan deteksi dini penyakit stroke dengan menggunakan metode Support Vector Machine (SVM) untuk mengklasifikasikan data medis pasien. Penelitian ini akan berfokus pada mengatasi beberapa tantangan yang ada dalam pengembangan metode deteksi stroke, termasuk keterbatasan dalam penelitian-penelitian terdahulu seperti ukuran sampel yang kecil, data yang tidak lengkap, dan fokus yang terlalu spesifik. Tujuan utama meliputi:

1. Mengembangkan model klasifikasi penyakit stroke yang akurat dengan menggunakan SVM.
2. Meningkatkan akurasi model dengan memperluas sampel data, mengumpulkan data yang lebih akurat, dan memperhitungkan berbagai faktor risiko dan gejala yang terkait dengan stroke.
3. Menggabungkan teknik analisis canggih untuk meningkatkan kemampuan model dalam mengidentifikasi kasus-kasus yang memerlukan perhatian segera.
4. Melakukan validasi eksternal yang kuat untuk memastikan keandalan model klasifikasi.
5. Memberikan kontribusi yang signifikan dalam meningkatkan deteksi dini penyakit stroke, yang pada gilirannya akan membantu dalam mengurangi dampak serius yang terkait dengan penyakit ini dan menyelamatkan nyawa pasien.

### Tujuan Khusus Penelitian

Ada sejumlah tujuan yang diprioritaskan untuk mencapai tujuan utama penelitian ini, salah satunya adalah pengembangan metode klasifikasi penyakit stroke yang menggunakan Support Vector Machine (SVM). Pertama, kami ingin mendapatkan data klinis dan medis yang lengkap, akurat, dan representatif tentang pasien. Usia, jenis kelamin, riwayat medis, gejala, hasil pencitraan medis, hasil tes laboratorium, dan faktor risiko tambahan yang mungkin berhubungan dengan stroke adalah semua elemen yang akan kami kumpulkan.

Selanjutnya, kami akan menggunakan metode SVM untuk mengklasifikasikan data dengan akurasi tinggi dan memeriksa pengaruh masing-masing komponen terhadap deteksi penyakit stroke. Tujuan tambahan adalah mengembangkan model klasifikasi yang dapat membantu profesional medis membuat keputusan yang tepat tentang diagnosis dan perencanaan perawatan.

Kami juga akan memastikan bahwa model yang dikembangkan dapat digunakan dalam pengaturan medis untuk meningkatkan deteksi dini dan manajemen penyakit stroke yang efektif. Hasil penelitian ini diharapkan akan meningkatkan pemahaman dan pengelolaan penyakit stroke secara signifikan. Selain itu, hasilnya akan menjadi landasan untuk penelitian lebih lanjut tentang pencegahan dan deteksi dini penyakit stroke. Tujuan khusus penelitian ini mencakup aspek pengumpulan data, penggunaan metode SVM, pengembangan model, dan konsekuensi praktis yang diharapkan dari pengelolaan penyakit.

Dengan mencapai tujuan utama dan tujuan khusus ini, penelitian ini diharapkan dapat berkontribusi positif dalam meningkatkan pengidentifikasian dini penyakit stroke, memungkinkan perawatan yang tepat waktu, dan pada akhirnya, meningkatkan prognosis dan kualitas hidup pasien yang terkena penyakit ini.

# BAB II

# KAJIAN TEORI

Kajian teori dalam penelitian "Klasifikasi Penyakit Stroke Dengan Metode Support Vector Machine" melibatkan beberapa konsep utama, termasuk Data Mining, Algoritma SVM (Support Vector Machine), dan pengetahuan tentang penyakit stroke. Berikut adalah penjelasan singkat tentang masing-masing kajian teori:

## **Data Mining**

Data Mining adalah proses ekstraksi pengetahuan atau informasi yang bermanfaat dari data yang tersimpan dalam berbagai bentuk. Ini melibatkan analisis data yang mendalam untuk mengidentifikasi pola, tren, dan hubungan yang tidak selalu terlihat dengan jelas.

Dalam penelitian ini, Data Mining digunakan untuk menganalisis data medis pasien yang terkait dengan penyakit stroke. Tujuan utamanya adalah mengidentifikasi pola atau fitur-fitur yang dapat digunakan untuk mengklasifikasikan kasus-kasus stroke dengan metode SVM.

## **Algoritma SVM (Support Vector Machine)**

SVM adalah salah satu algoritma pembelajaran mesin yang digunakan untuk klasifikasi dan regresi. SVM mencari pemisah optimal dalam bentuk hyperplane dalam ruang fitur data.

Dalam penelitian ini, SVM digunakan sebagai metode klasifikasi untuk membedakan antara kasus stroke dan non-stroke. SVM bekerja dengan mencari hyperplane terbaik yang memisahkan dua kelas dengan margin maksimal, sehingga memungkinkan klasifikasi yang akurat.

## **Penyakit Stroke**

Stroke adalah kondisi medis serius yang disebabkan oleh gangguan aliran darah ke otak. Terdapat dua jenis utama stroke, yaitu iskemik (penyumbatan pembuluh darah) dan hemoragik (pembuluh darah pecah).

Penting untuk memahami dasar-dasar penyakit stroke, termasuk gejala, faktor risiko, dan dampaknya pada pasien. Informasi ini membantu dalam mengidentifikasi fitur-fitur penting yang digunakan dalam klasifikasi menggunakan SVM.

Dengan pemahaman yang kuat tentang Data Mining, SVM, dan penyakit stroke, penelitian ini dapat merancang dan mengimplementasikan model klasifikasi yang efektif untuk mendeteksi penyakit stroke secara dini berdasarkan data medis pasien. Kajian teori ini memberikan dasar yang kokoh untuk pendekatan penelitian yang berhasil.

Variabel-variabel yang diperlukan dalam penelitian "Klasifikasi Penyakit Stroke Dengan Metode Support Vector Machine" dapat mencakup berbagai aspek data medis dan klinis yang relevan untuk tujuan klasifikasi penyakit stroke. Beberapa contoh variabel yang mungkin diperlukan adalah:

1. Usia Pasien: Variabel ini penting karena risiko stroke dapat berhubungan dengan usia, dan usia dapat memengaruhi gejala dan pengobatan yang tepat.
2. Jenis Kelamin: Data jenis kelamin pasien dapat berguna dalam menganalisis apakah ada perbedaan gejala atau faktor risiko yang berhubungan dengan jenis kelamin.
3. Riwayat Medis: Informasi riwayat medis, termasuk riwayat stroke sebelumnya, penyakit kronis (seperti diabetes atau tekanan darah tinggi), dan pengobatan yang sedang berlangsung, dapat memberikan konteks penting dalam klasifikasi.
4. Gejala Stroke: Variabel ini mencakup gejala-gejala klinis yang dialami pasien, seperti kelemahan, kesulitan berbicara, dan gangguan penglihatan.
5. Hasil Pencitraan Medis: Data dari CT scan, MRI, atau metode pencitraan medis lainnya akan mencakup informasi tentang kondisi otak dan tanda-tanda penyumbatan atau pendarahan.
6. Hasil Tes Laboratorium: Hasil tes darah dan tes laboratorium lainnya mungkin diperlukan untuk mengidentifikasi faktor risiko seperti kolesterol tinggi atau koagulasi darah.
7. Faktor Risiko Tambahan: Variabel-variabel lain seperti tekanan darah, kadar gula darah, indeks massa tubuh (BMI), dan riwayat merokok juga harus dipertimbangkan sebagai faktor risiko potensial.
8. Hasil Diagnosa: Ini adalah variabel target yang akan diklasifikasikan, yaitu apakah pasien didiagnosis dengan stroke atau tidak.
9. Hasil SVM: Jika menggunakan metode Support Vector Machine, hasil dari model klasifikasi SVM juga merupakan variabel yang relevan dalam penelitian ini.

Variabel-variabel ini harus dikumpulkan secara cermat dari pasien dan digunakan dalam analisis untuk mengembangkan model klasifikasi yang akurat. Pengumpulan data yang lengkap dan akurat akan memainkan peran kunci dalam keberhasilan penelitian Anda dalam mengklasifikasikan penyakit stroke dengan metode SVM.

# BAB III

# METODE PENELITIAN

# Jenis Penelitian

Jenis penelitian yang digunakan adalah penelitian kuantitatif. Penelitian ini menggunakan data kuantitatif untuk menguji data sekunder berupa data pasien yang memiliki penyakit strok dari Madrid, Spanyol. Data ini diambil dari *website* *Kaggle*. Akses kedalam data dapat diklik pada *link* dibawah ini:

<https://drive.google.com/drive/u/1/folders/1gah1BkVaLC1brNLXkE2pLBFQOSfArb1G>

Total data yang digunakan sebanyak 5110 data. Atribut yang digunakan dalam data ini yaitu sebagai berikut:

**Tabel 3.1** Atribut-Atribut Pada Data

|  |  |  |
| --- | --- | --- |
| **No.** | **Atribut** | **Deskripsi Atribut** |
| 1 | Id | Kode unik dari masing-masing pasien |
| 2 | Jenis Kelamin | Laki-laki, perempuan dan lainnya |
| 3 | Umur | Umur dari masing-masing pasien. |
| 4 | Penyakit Hipertensi | 0 : Tidak memiliki penyakit hipertensi.  1 : Pasien yang memiliki penyakit hipertensi. |
| 5 | Penyakit Jantung | 0 : Tidak memiliki penyakit jantung.  1 : Pasien yang memiliki penyakit jantung. |
| 6 | Status Menikah | Iya atau Tidak. |
| 7 | Tipe Pekerja | Wiraswata, *private*, tidak pernah bekerja atau PNS. |
| 8 | Tipe Tempat Tinggal | Pedesaan atau perkotaan. |
| 9 | Level Rata-rata Glukosa | Level rata-rata glukosa di dalam darah. |
| 10 | BMI (*Body Mass* *Index*) | Berat badan masing-masing pasien. |
| 11 | Status Merokok | Tidak pernah merokok, merokok, tidak diketahui atau sebelumnya merokok. |
| 12 | Struk | 0: Pasien tidak memiliki penyakit struk  1: Pasien memiliki penyakit struk |

## **Teknik Analisis Data**

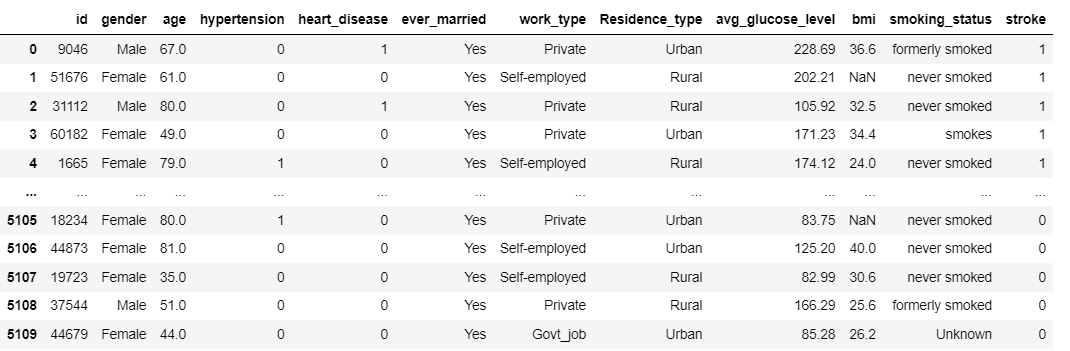
### ***Exploratory Data Analysis* (EDA)**

Tahap ini merupakan sebuah proses kritis dalam melakukan investigasi awal pada data dengan tujuan menemukan pola, anomali, menguji hipotesis dan memeriksa asumsi menggunakan visualisasi data. Pada tahapan ini menggunakan bahasa *python* yang dijalankan di *Jupyter Notebook*.

Adapun beberapa langkah-langkah dalam EDA, yaitu:

1. Mengamati kumpulan data yang ada

Pertama-tama, untuk mengamati kumpulan data yang ada *import* terlebih dahulu data yang digunakan dalam penelitian menggunakan *library pandas* untuk melihat isi dari data yang digunakan.



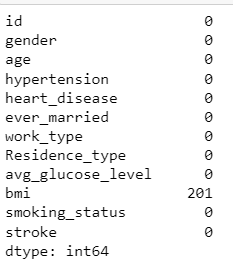
Gambar 3.4 Tampilan Isi Data

1. Mengamati deskriptif statistik

Setelah meng*import* data yang sudah ada, lalu dapat melakukan pencarian deskriptif statistik seperti *mean,* median, *mode, quartile* dan *standard deviation*.

1. Mengecek apakah terdapat *missing value* dari masing-masing atribut

Setelah dicek, terdapat *missing value* pada atribut ‘BMI’ sebanyak 201 data. Untuk perbaikan terhadap *missing value* akan dilakukan di tahap *Data Preprocessing.*

**

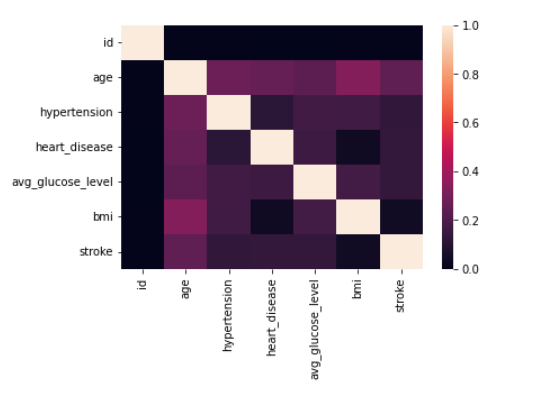
Gambar 3.5 pengecekan *missing value* pada masing-masing atribut

1. Kategorisasi data yaitu data kategorikal dan numerikal

Data kategorikal terdiri dari atribut *gender* (jenis kelamin)*, ever\_married* (status menikah), *residence\_type* (tipe tempat tinggal), dan *smoking\_status* (status merokok). Sedangkan untuk data numerikal teridiri dari atribut *hypertension* (hipertensi). *Heart\_disease* (riwayat penyakit jantung), *stroke* (strok), *age* (umur), *avg\_glucose\_level*(level rata-rata glukosa), dan bmi(indeks berat badan).

1. Mengidentifikasi hubungan antar variabel

Untuk mengidentifikasi hubungan antar variabel, data target yang akan dipakai adalah atribut *stroke* yang nantinya masing-masing atribut akan diidentifikasi hubungannya dengan atribut *stroke.* Masing-masing atribut kecuali atribut *stroke* dinotasikan sebagai X dan atribut *stroke* dinotasikan sebagai Y.



Gambar 3.6 Visualisasi Korelasi Antar Atribut Dengan Atribut Lainnya

### ***Data Selection***

Setelah melakukan *Exploration Data Analysis* (EDA), akan dilakukan proses seleksi atau proses pemilihan data. Tahapan ini hampir sama dengan proses interupsi terhadap data agar data siap diolah. Hanya saja tahap ini tidak mengarah pada destruktif//perusakan data.

### ***Data Preprocessing***

Tahapan ini merupakan sebuah tahapan pembersihan data. Dalam klasifikasi penyakit strok, *Data preprocessing* berperan untuk menyaring atribut yang tidak digunakan dalam penelitian. Pada tahapan ini menggunakan bahasa *python* yang dijalankan di *Jupyter Notebook*. Berikut langkah-langkah dalam *Data Preprocessing* yaitu sebagai berikut:

1. *Data Cleaning*

Setelah mengetahui *missing value* melalui tahapan EDA, data-data yang memiliki *missing value* akan diubah menjadi nilai rata-rata dari atribut tersebut.

1. *Data Reduction*

Setelah itu, akan dilakukan reduksi data untuk atribut-atribut yang tidak diperlukan dalam pengujian performa seperti *id*, *gender(*jenis kelamin), *Residence\_type* (tipe tempat tinggal) dan *ever\_married* (status menikah) untuk mempertahankan kualitas data.

### ***Transformation Data***

*Transformation Data* merupakan proses mengubah format tipe data pada dengan menyesuaikan tipe data yang lain. Pada tahap ini, dilakukan perubahan pada atribut *work\_type* dan *smoking\_status*. Atribut dari keseluruhan data akan diubah menjadi tipe data *integer*/*float* untuk mempermudah dalam proses *data mining* dan *evaluation*.

### **Klasifikasi**

Setelah dilakukan transformasi, maka selanjutnya masuk ke dalam proses *data mining* menggunakan algoritma *Support Vector Machine* (SVM). Pada tahap ini, proses yang dijalankan menggunakan bahasa pemrograman *python* yang dijalankan pada *Jupyter Notebook*. Prosesnya yaitu sebagai berikut:

1. Implementasi Algoritma *Support Vector Machine* (SVM)

Langkah-langkah dalam klasifikasi algoritma SVM yaitu sebagai berikut:

* 1. Menentukan fungsi kernel yang akan digunakan.

1. Menentukan parameter terbaik pada fungsi kernel yang digunakan.
2. Membangun model SVM menggunakan fungsi kernel.
3. Melakukan evaluasi akurasi untuk mengetahui keakuratan algoritma SVM menggunakan *confusion matrix.*

### **Evaluasi Model**

Pada tahap evaluasi, akan dilakukan skenario pengujian untuk menentukan nilai *accuracy, recall, precision dan F1-Score.* Pola informasi yang dihasilkan dari *data mining* akan dievaluasi menggunakan metode *confusion matrix.* Sebelum menghitung nilai tersebut, data dibagi menjadi dua yaitu *data training* dan *data testing* masing-masing rasio 90:10, 80:20 dan 70:30. *Data training* merupakan data yang digunakan sebagai data model sedangkan *data testing* digunakan sebagai uji coba model yang akan di stimulasi penggunaan model dunia nyata. Oleh sebab itu, *data training* memiliki persentase lebih tinggi daripada data *testing*. Dalam pengkategorian *confusion matrix*, dilakukan klasifikasi data yaitu jika data memiliki nilai benar, maka nilai yang diperoleh adalah *true positive* dan *true negative*. Sedangkan data yang memiliki nilai salah memperoleh nilai *false-positive* dan *false-negative.* Berikut tabel 3.3 adalah bentuk tabel *confusion matrix.*

**Tabel 3.2** Tabel *Confusion Matrix*

|  |  |  |  |
| --- | --- | --- | --- |
| Data | | Aktual | |
| *True* | *False* |
| Prediksi | *True* | *True Positive* (TP) | *False Positive* (FP) |
| *False* | *False Negative* (FN) | *True Negative* (TN) |

Keterangan:

*True Positive* (TP): Jumlah data yang bernilai positif diprediksi benar sebagai positif.

*False Positive* (FP): Jumlah data yang bernilai negatif tetapi diprediksi positif.

*False Negative* (FN): Jumlah data yang bernilai positif tetapi diprediksi sebagai negatif.

*True Negative* (TN): Jumlah data yang bernilai negatif tetapi diprediksi sebagai negatif.

Lalu pada tabel 3.3 terdapat konsep rumus untuk menghitung performa menggunakan *confusion matrix* yaitu sebagai berikut:

Tabel 3.3 Rumus Klasifikasi *Confusion Matrix*

|  |  |
| --- | --- |
| **Nilai** | **Rumus** |
| *Accuracy* |  |
| *Precision* |  |
| *Recall* |  |
| *F1-Score* |  |