

## **BAB II**

### **LANDASAN TEORI**

#### **1.1. Penelitian Terdahulu**

Pada penelitian yang dilakukan oleh Annisa Paramitha Fadillah membahas mengenai prediksi pola kelulusan studi mahasiswa dalam menempuh mata kuliah. Dalam penelitian tersebut, Annisa Paramitha Fadillah menggunakan metode CRISP-DM sebagai tahap penelitiannya dan untuk proses modelingnya menggunakan algoritma C4.5 yang digunakan untuk membangun pohon keputusan. Data yang digunakan dalam penelitian ini adalah data hasil kuesioner yang dilakukan pada 40 responden acak yang terdapat di prodi sistem informasi. Hasil dari penelitian ini adalah metode CRISP-DM dapat diterapkan untuk prediksi pola keputusan studi mahasiswa dalam menempuh mata kuliah. [4]

Berdasarkan penelitian yang dilakukan oleh Elsa Putri Molatina Eka Suci dan Milkhatun, membahas mengenai analisis rekam medis pasien risiko perilaku kekerasan dengan menggunakan algoritma C4.5 di RSJD Atma Husada Mahakam Samarinda. Dalam penelitian tersebut, Elsa Putri Molatina Eka Suci dan Milkhatun menggunakan penelitian deskriptif kuantitatif dengan jumlah data rekam medis pasien risiko perilaku kekerasan sebanyak 219 data yang diolah menggunakan algoritma c4.5 menggunakan *software Rapid Miner*. Penelitian tersebut diambil dari observasi dan studi dokumentasi data sekunder pasien risiko perilaku kekerasan skizofrenia. Hasil penelitian tersebut diketahui bahwa dari total 219 pasien yang paling dominan adalah pasien dengan jenis

kelamin laki-laki, dan kategori usia yang paling dominan adalah dewasa muda. Dari keseluruhan pasien risiko perilaku kekerasan yang berjumlah 219 orang terdapat gejala yang paling signifikan yaitu gejala emosi. [5]

Penelitian selanjutnya yang dilakukan oleh Muhammad Soleh, Sundari Retno Andani dan Hendry Qurniawan. Dalam penelitian tersebut membahas mengenai klasifikasi tindak kejahatan menggunakan algoritma c4.5, data yang digunakan merupakan data laporan masyarakat terkait tindak kriminal yang diperoleh dari Unit Reskrim Polsek Siantar Barat tahun 2018-2019. Data dianalisis menggunakan algoritma c4.5 yang bertujuan untuk melakukan klasifikasi data pada kelas kejahatan tertentu. Dalam penelitian tersebut menggunakan aplikasi Microsoft Excel 2010 dan Rapid Miner versi 8.1 sebagai alat untuk menganalisis data. Berdasarkan hasil penelitian yang dilakukan menggunakan algoritma C4.5 dalam aplikasi Rapid Miner menyatakan tindak kejahatan yang paling tinggi yaitu atribut kerugian. [6]

Persamaan ditemukan dalam penggunaan algoritma C4.5 untuk analisis dan klasifikasi data oleh Annisa Paramitha Fadillah, Elsa Putri Molatina Eka Suci dan Milkhatun, Muhammad Soleh, Sundari Retno Andani, dan Hendry Qurniawan. Tetapi ada beberapa perbedaan yang signifikan di antara mereka. Pada penelitian yang dilakukan oleh Annisa Paramitha Fadillah menggunakan metode CRISP-DM dan data dari 40 responden acak untuk memprediksi pola kelulusan studi mahasiswa dalam menempuh mata kuliah di program studi sistem informasi. Sementara itu, Elsa Putri Molatina Eka Suci dan Milkhatun menganalisis rekam medis 219 pasien di RSJD Atma Husada Mahakam Samarinda yang rentan terhadap perilaku kekerasan. Mereka

menggunakan data rekam medis dari 219 pasien. Selain itu, mereka menggunakan metode deskriptif kuantitatif dan menggunakan alat analisis Rapid Miner. Sebaliknya, penelitian yang dilakukan oleh Muhammad Soleh, Sundari Retno Andani, dan Hendry Qurniawan mempelajari klasifikasi tindak kejahatan dengan menggunakan data laporan masyarakat dari Unit Reskrim Polsek Siantar Barat tahun 2018-2019. Analisis data dilakukan dengan menggunakan program Microsoft Excel 2010 dan Rapid Miner.

Sementara perbedaan yang ditemukan dari ke tiga penelitian tersebut yaitu pada penelitian yang dilakukan oleh Annisa Paramitha Fadillah melakukan penelitian tentang prediksi pola kelulusan studi mahasiswa, Elsa Putri Molatina Eka Suci dan Milkhatun melakukan penelitian tentang risiko perilaku kekerasan pada pasien skizofrenia, dan Muhammad Soleh, Sundari Retno Andani, dan Hendry Qurniawan melakukan penelitian tentang klasifikasi tindak kejahatan. Selain itu, berbagai jenis data digunakan dalam setiap penelitian, mulai dari kuesioner hingga laporan masyarakat. Selain itu, metode yang digunakan dalam penelitian berbeda, penelitian yang dilakukan oleh Annisa Paramitha Fadillah menggunakan CRISP-DM, sedangkan dua penelitian lainnya menggunakan metode deskriptif kuantitatif dan analisis data dengan Rapid Miner.

Persamaan dari ke tiga penelitian terdahulu dan penelitian yang dilakukan yaitu pada metode analisis menggunakan algoritma C4.5 untuk klasifikasi data. Sedangkan perbedaannya yaitu pada penelitian ini menggunakan metode KDD untuk tahapan *data mining* nya, dan implementasi menggunakan *website* untuk mempermudah pengolahan

data, pada penelitian dahulu menggunakan metode CRISP-DM, metode deskriptif kuantitatif dan analisis data menggunakan Rapid Miner.

## **1.2. Teori Pendukung**

Teori pendukung adalah istilah dalam penelitian yang digunakan sebagai acuan untuk memahami penelitian yang dilakukan oleh penulis.

### **1.2.1. Tindak Kekerasan**

Kekerasan adalah suatu kondisi di mana kenyataan fisik dan mentalnya lebih buruk daripada kemungkinannya. Kekerasan didefinisikan secara luas dan mencakup lebih dari hanya kekerasan fisik, seperti penganiayaan dan pembunuhan. Ini juga mencakup hal-hal seperti pembohongan, indoktrinasi, ancaman, tekanan, dan bahkan penelantaran yang dilakukan untuk mencegah seseorang mengembangkan kemampuan mental dan intelektual mereka. [7]

Penyebab utama tindak kekerasan dapat dibagi menjadi dua kategori yaitu, faktor individu dan faktor sosial. Faktor individu terkait dengan kecenderungan individu untuk bertindak kekerasan, sedangkan faktor sosial adalah lingkungan di mana seseorang bertindak kekerasan. [8]

Tindakan kekerasan adalah tindakan fisik yang dilakukan dengan sengaja atau tidak sengaja terhadap seseorang yang dapat menyebabkan cedera, depresi, kerugian psikologis, atau bahkan kematian. Setiap perbuatan yang dapat menyebabkan kesengsaraan dan penderitaan baik secara fisik, mental, seksual,

atau psikologis, termasuk perlakuan yang merendahkan martabat anak, disebut kekerasan terhadap anak.[8]

Kekerasan adalah fenomena kompleks yang tidak hanya terbatas pada tindakan fisik seperti kekerasan atau pembunuhan, tetapi juga mencakup bentuk-bentuk lain seperti pembohongan, indoktrinasi, ancaman, tekanan, dan penelantaran yang menghambat pertumbuhan mental dan intelektual seseorang.

Para ahli berpendapat tentang definisi kekerasan. SR Sianturi mendefinisikan kekerasan sebagai tindakan terhadap orang lain dengan menggunakan kekuatan yang menimbulkan kerugian atau mengagetkan korban. Selain itu, SR Sianturi menjelaskan ancaman kekerasan sebagai suatu perbuatan yang dapat membuat orang lain merasa ketakutan karena adanya ancaman yang merugikan dirinya dengan menggunakan kekerasan. [9]

### **1.2.2. Data Mining**

*Data mining* adalah proses penggalian data dari tumpukan *database* yang berukuran besar yang digunakan untuk menemukan *knowledge* berupa informasi penting dan bermanfaat. Secara umum kegunaan *data mining* dapat dibagi menjadi dua, yaitu deskriptif dan prediktif. Deskriptif berarti *data mining* digunakan untuk mencari pola-pola yang dapat dipahami manusia yang menjelaskan karakteristik data. Sedangkan prediktif berarti *data mining* digunakan untuk membentuk sebuah model pengetahuan yang akan digunakan untuk melakukan prediksi. [10]

*Data mining* adalah proses semi otomatis. yang menggunakan statistik, matematika, AI, dan pembelajaran mesin untuk mengekstraksi dan mengidentifikasi data pengetahuan yang mungkin dan bermanfaat yang berguna yang disimpan dalam *database* yang besar. Tujuan dari informasi Mining adalah proses mengidentifikasi pengetahuan dan informasi yang dapat digunakan untuk meningkatkan pengambilan keputusan. *Data mining* sering disebut juga sebagai *Discovery of Knowledge in Databases (KDD)*. KDD adalah aktivitas yang mencakup pengumpulan, penggunaan, dan analisis data secara historis untuk menemukan hubungan, pola, atau keteraturan dalam set data yang sangat besar. [11]

*Data mining* biasanya digunakan untuk dua tujuan utama yaitu deskriptif dan prediktif. Tujuan deskriptif adalah untuk menemukan dan menemukan pola-pola yang dapat dipahami manusia untuk menjelaskan karakteristik data. Sementara itu, tujuan prediktif adalah untuk membuat model pengetahuan yang dapat digunakan untuk memprediksi data baru atau masa depan.

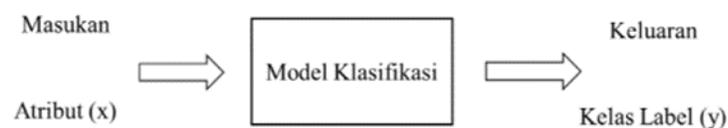
Dengan demikian, *data mining* memainkan peran penting dalam membantu organisasi memahami data mereka lebih baik dan membuat keputusan berdasarkan informasi yang diperoleh dari analisis data yang mendalam.

#### **1.2.2.1. Teknik Data Mining**

*Data mining* memiliki beberapa teknik berdasarkan tugas yang dapat dilakukan, berikut beberapa teknik dalam *data mining*.

### 1. *Clasification*

Teknik ini digunakan untuk mengelompokkan data ke dalam kelas atau kategori tertentu berdasarkan karakteristik atau atribut tertentu. Di dalam klasifikasi diberikan sejumlah *record* yang dinamakan *training set*, yang terdiri dari beberapa atribut.



**Gambar 2. 1 Blok Siklus Teknin Klasifikasi**

(sumber : Buku Data Mining [12])

Metode klasifikasi yang sering digunakan yaitu, *Support Vector Machine*, *Multilayer Perceptron*, *Naïve Bayes*, *ID3*, dll.

#### a. *Support Vector Machine (SVM)*

*Support Vector Machine (SVM)* adalah metode klasifikasi yang bekerja dengan cara menemukan *hyperplane* dengan margin terbesar, yang merupakan garis batas pemisah data antar-kelas. Data terdekat dengan *hyperplane* pada masing-masing kelas disebut *support vector machine*. [12]

*Support Vector Machine (SVM)* dikembangkan dari teori *Structural risk minimalization*. Dengan menggunakan trik kernel untuk memetakan sampel pelatihan dari ruang *input* ke ruang fitur yang berdimensi tinggi. Konsep SVM secara sederhana sebagai usaha

untuk mencari *hyperplane* terbaik yang berfungsi sebagai pemisah dua *class* pada *input space*.

b. *Multi Layer Perceptron*

*Multi Layer Perceptron* (MPL) merupakan turunan algoritma *Artificial Neural Network* (ANN) dari *perceptron*, berupa *ANN feed forward* dengan satu atau lebih dari *hidden layer*.

Algoritma pembelajaran mesin yang dikenal sebagai *perceptron multilayer* terdiri dari lapisan *input*, satu atau lebih lapisan tersembunyi, dan lapisan *output*. Setiap lapisan jaringan memiliki *node*, dengan setiap *node* terhubung ke lapisan berikutnya dan memiliki berat dan bias tertentu. Untuk masalah pembelajaran yang diawasi, MLP menggunakan fungsi aktivasi untuk menempatkan *output* nya pada label yang dinyatakan sebelumnya. [13]

c. *Naïve Bayes*

*Naïve Bayes* adalah algoritma klasifikasi untuk pembelajaran mesin yang didasarkan pada *Teorema Bayes*, yang menetapkan bahwa setiap pasangan fitur dalam data pelatihan saling independen. Setelah memproses data pelatihan, algoritma ini menghasilkan model klasifikasi, yang dapat digunakan untuk memprediksi label kelas yang tepat untuk data yang belum dilihat sebelumnya. [14]

Metode pengklasifikasian dengan menggunakan metode probabilitas dan *statistic*, yaitu dengan memprediksi peluang di masa

depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai *Teorema Bayes*.

d. Pohon Keputusan (ID3)

Pohon keputusan adalah struktur yang dapat digunakan untuk membagi kumpulan data yang besar menjadi himpunan catatan yang lebih kecil dengan menerapkan serangkaian aturan keputusan. Ini adalah salah satu metode klasifikasi yang paling populer karena mudah diinterpretasikan oleh manusia. Pohon keputusan memiliki *node* pohon yang menunjukkan atribut yang telah diuji, dan setiap cabangnya menunjukkan pembagian hasil uji. Selain itu, *node* daun, atau cabang, menunjukkan kelompok kelas tertentu. [15]

Teknik pohon keputusan, juga dikenal sebagai *Decision Tree*, dapat digunakan sebagai bagian dari kriteria seleksi dan juga untuk membantu dalam penggunaan dan pemilihan data tertentu di antara struktur keseluruhan. Pohon keputusan biasanya dimulai dengan pertanyaan sederhana dengan dua atau lebih jawaban. Setiap jawaban akan menggiring ke pertanyaan berikutnya, yang membantu mengidentifikasi dan mengklasifikasikan data sehingga dapat dikategorikan atau diprediksi berdasarkan jawaban.

2. Regresi (*Regression*)

Teknik *regression* digunakan untuk mencari pola dan menentukan nilai numerik. Regresi linear, atau regresi linear

seederhana, melihat pengaruh antara variabel bebas (independen) dan variabel tidak bebas. Dalam regresi linear sederhana, hanya ada satu variabel bebas dan satu variabel tidak bebas, sedangkan dalam regresi linear berganda, lebih dari satu variabel bebas. Salah satu manfaat analisis regresi linear adalah untuk menemukan seberapa besar dan bagaimana variabel independen mempengaruhi variabel dependen. Variabel yang sering mempengaruhi disebut variabel dependen atau tidak bebas, dan variabel yang mempengaruhi variabel lain disebut variabel independen atau variabel bebas. [16]

### 3. Klastering (*Clustering*)

Salah satu teknik penting dalam pengolahan data adalah *clustering*, yang digunakan untuk mengelompokkan data menjadi kelompok yang sama berdasarkan pola atau karakteristik tertentu. Tujuan utama dari *clustering* adalah untuk menemukan struktur yang tersembunyi dalam data, yang dapat membantu kita memahami kelompok atau kategori yang ada di dalamnya. [17]

*Clustering* sering digolongkan sebagai metode *unsupervised learning*. Algoritma *k-means* dan *clustering hierarchical* adalah dua contoh metode klastering.

### 4. Asosiasi (*Association*)

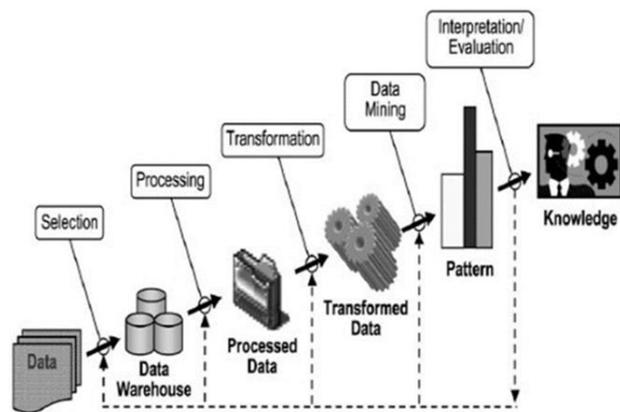
Metode ini digunakan untuk menemukan hubungan atau asosiasi antara item atau atribut dalam dataset yang besar. Dalam proses ini,

tujuan dari *association rule mining* adalah untuk menemukan aturan asosiasi yang memiliki makna atau pola transaksi yang sering terjadi secara bersamaan dalam data. Menurut aturan, jika item A ada dalam suatu transaksi atau himpunan item, maka item B juga cenderung ada dalam transaksi atau himpunan tersebut. [18] Beberapa metode yang umum digunakan untuk menerapkan *association rule mining* adalah Algoritma Apriori dan FP-Growth.

### 1.2.3. *Knowledge Discovery in Database*

*Knowledge Discovery in Database* (KDD) adalah proses pengambilan informasi yang tersembunyi, di mana informasi sebelumnya tidak dikenal. KDD mengumpulkan pengetahuan dari *database* yang ada. Informasi yang dihasilkan didapatkan dari suatu data yang besar atau dikenal dengan tambang data yang disimpan dalam basis data yang awalnya belum diketahui dan menghasilkan suatu data yang potensial bermanfaat. [19]

*Knowledge Discovery in Database* (KDD) adalah proses menggali dan menganalisis kumpulan data yang sangat besar dengan bantuan komputer, menghasilkan pengetahuan dan informasi yang bermanfaat. *Data mining* merupakan salah satu langkah dari serangkaian proses KDD. Berikut merupakan proses tahapan dari *Knowledge Discovery in Database* (KDD).



**Gambar 2. 2 Proses Tahapan KDD**

(sumber : Data Mining dan Penerapan Algoritma)

Berikut penjelasan mengenai proses tahapan dalam *Knowledge Discovery in Database*.

### 1. *Data Selection*

Pembersihan data adalah fase pertama dari proses KDD. Untuk mendapatkan kumpulan data yang bersih untuk digunakan dalam fase *data mining*, persyaratan pertama ini harus dipenuhi saat memilih setiap aspek dari kumpulan data untuk mendapatkan nilai yang relevan tanpa redundansi atau nilai yang hilang. Data yang tidak memenuhi syarat termasuk data yang tidak memiliki nilai, tidak akurat, dan tidak konsisten selama kompilasi. [20]

Sebelum tahapan penggalian informasi dalam *Knowledge Discovery in Database* dimulai, seleksi atau pemilihan data harus dilakukan dari himpunan data operasional. Sehingga tidak semua data digunakan dalam proses *data mining*.

### 2. *Pre-processing (Cleaning)*

Sebelum proses *data mining* dimulai, data terfokus dari KDD harus dibersihkan. Proses ini mencakup beberapa tindakan seperti menghilangkan duplikat data, memeriksa data yang dinilai tidak konsisten, mencari dan memperbaiki kesalahan data seperti kesalahan cetak (*typography*).

### 3. *Transformation*

Tahap transformasi adalah proses mencari berbagai fitur yang dapat digunakan untuk mempresentasikan data sesuai dengan tujuan yang ingin dicapai. Tahap *coding* ini juga merupakan proses transformasi data yang telah dipilih agar dapat digunakan dalam proses penggalian data, dengan mengubah data ke dalam model analitis dan membuat model sesuai dengan analisis yang diharapkan dan diperlukan oleh algoritma penggalian data. Proses ini mempengaruhi pola dan jenis informasi yang dicari dalam *database*.

### 4. *Data Mining*

*Data mining* adalah proses menemukan pola dan informasi yang unik dalam data yang telah diseleksi sebelumnya menggunakan Teknik tertentu. Metode atau algoritma yang digunakan sangat bervariasi, pemilihan algoritma ditentukan sesuai dengan proses serta tujuan dari KDD.

### 5. *Interpretation / Evaluation*

Pada tahapan ini dilakukan proses terjemah dari beberapa pola yang didapatkan dari *data mining*. Pola informasi yang dihasilkan harus

ditampilkan ke dalam bentuk yang lebih mudah dipahami oleh pihak yang berkepentingan.

#### 1.2.4. Algoritma C4.5

Algoritma C4.5 yaitu merupakan sebuah pohon keputusan atau yang dikenal dengan decision tree, untuk mengubah data menjadi pohon keputusan yang menghasilkan aturan keputusan. Tujuan algoritma ini adalah untuk meningkatkan keakuratan hasil prediksi dengan membuat beberapa aturan dan pohon keputusan. [21] Induksi pohon keputusan hanya dapat dilakukan pada fitur bertipe kategorikal (nominal atau ordinal). Perbedaan lain yang membedakan algoritma C4.5 dari ID3 adalah bahwa algoritma C4.5 juga menggunakan kriteria gain, dan dapat memotong pohon keputusan (*pruning*) dan menurunkan set peraturan (*deriving*).

Langkah pertama dalam algoritma C4.5 untuk membangun pohon keputusan adalah memilih atribut sebagai akar. Selanjutnya, untuk tiap nilai di dalam akar tersebut, dibuat cabang dan kemudian kasus dibagi ke dalam cabang-cabang tersebut. Setelah itu, ulangi prosedur untuk setiap cabang sampai kasus di setiap cabang memiliki kelas yang sama.

Algoritma C4.5 memiliki beberapa konsep seperti *Entropy* dan konsep Gain. Formula dalam perhitungan *Entropy* adalah sebagai berikut :

$$Entropy(S) = \sum_{i=1}^n - p_i * \log_2 p_i \quad \dots\dots\dots\text{persamaan (2.1)}$$

Keterangan :

S : Himpunan Kasus

A : Fitur

n : jumlah partisi S

pi : proposi dari Si terhadap S

sementara untuk menghitung Gain dapat dilihat pada formula berikut :

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \dots\dots\dots\text{persamaan (2.2)}$$

Keterangan :

S : Himpunan Kasus

A : Atribut

n : jumlah partisi atribut A

|Si| : jumlah kasus pada partisi ke-i

|S| : jumlah kasus dalam SS

Sebelum melakukan proses klasifikasi perlu dilakukan pengolahan data yang diantaranya meliputi tahap *cleaning*, dan *data transformation* sehingga menghasilkan data baru untuk di proses ke tahap selanjutnya. Selanjutnya melakukan pembagian *data training* untuk proses pembentukan pohon dan *data testing* untuk mengukur kemampuan dari pohon klasifikasi yang terbentuk.

### 1.3. Piranti Pendukung

Piranti pendukung merupakan alat yang digunakan untuk membantu penulis dalam menyelesaikan penelitian. Berikut beberapa piranti yang digunakan :

### 1.3.1. Anaconda Navigator

Anaconda navigator adalah perangkat lunak gratis yang menyediakan toolkit untuk penelitian dan ilmu pengetahuan, memungkinkan untuk pengkodean dalam *Python* atau R. Anaconda navigator dapat digunakan untuk mengakses ke berbagai lingkungan pengembangan terintegrasi (IDE).

IDE adalah platform yang memfasilitasi penulisan, pengeditan, visualisasi, dan pemeriksaan data, serta kolaborasi proyek. Meskipun setiap IDE memiliki kelebihan dan kekurangan, pengkodean *Python* tetap konsisten di semua IDE, sehingga keterampilan pemrograman dapat dengan mudah ditransfer dari satu IDE ke IDE lainnya.[22]

### 1.3.2. VsCode (*Visual Studio Code*)

*Vs Code* adalah proyek yang dikembangkan oleh Microsoft untuk berbagai sistem operasi. *Visual Studio Code* adalah kode editor yang berbasis *open source* di bawah lisensi MIT. Dalam perangkat lunak ini pengembang dapat menambahkan bahasa baru seperti *python* menggunakan *Vs Code*.

*Visual Studio Code* dapat dijalankan di semua perangkat desktop secara gratis, bahkan hampir mendukung semua sistem operasi seperti Windows, Mac Os, Linux, dan lainnya. Perangkat lunak ini memiliki ekosistem ekstensi yang kaya untuk Bahasa lain (seperti C ++, C, C#, *Java*, *Python*, *PHP*, *Go*) dan *runtime* (seperti *.NET* dan *Unity*).[23]

### 1.3.3. *Website*

Pada dasarnya, *website* adalah kepanjangan dari *Word Wide Web* (WWW). Data WWW disimpan pada *web server*, dan jaringan *browser* terlebih dahulu, seperti *Internet Explorer* dan *Firefox*, dapat mengaksesnya.

*website* adalah kumpulan semua halaman web yang fungsinya untuk menampilkan berbagai informasi dalam bentuk tulisan, gambar dan suara dari sebuah domain yang terbentuk dalam suatu rangkaian yang saling terkait. Semua halaman web yang terhubung satu sama lain disebut "*hyperlink*", dan teks yang terhubung ke teks lain disebut "*hypertext*". *Website* dapat diakses oleh siapa pun menggunakan jaringan internet.[24]

### 1.3.4. *Python*

*Python* adalah bahasa pemrograman interpretatif multiguna dengan filosofi perancangan yang berfokus pada tingkat keterbacaan kode. Dikatakan sebagai bahasa yang memiliki banyak kemampuan dan kemampuan dengan sintaksis kode yang sangat jelas, dan memiliki banyak fitur pustaka standar yang besar dan komprehensif. [25]

*Python* merupakan bahasa pemrograman yang memiliki manfaat untuk mendukung pemrograman yang berorientasi objek. *Python* merupakan bahasa yang lebih sederhana dari Bahasa pemrograman yang sudah ada seperti C, C++ , tetapi memiliki kemampuan sintak yang lebih baik. Hal ini dikarenakan oleh dukungan bahasa *python* terhadap *library-library* yang didalamnya menyediakan

fungsi analisis data dan fungsi *machine learning*, data *pre-processing tools*, dan visualisasi data.



**Gambar 2. 3 Logo Python**

**(Sumber : python.org)**

### **1.3.5. Streamlit**

*Streamlit* adalah kerangka kerja web yang ditujukan untuk menyebarkan model dan visualisasi dengan mudah menggunakan bahasa *Python*. *Widget* bawaan yang dapat digunakan oleh pengguna dapat mencakup pengunggahan gambar, penggeser, *input* teks, dan elemen *hypertext markup language* (HTML) lain yang sudah dikenal, seperti tombol *checkbox* dan tombol radio.[25]

Aplikasi *Streamlit* dapat berjalan pada editor Anaconda dan bahasa *Python* seri 3.7 ke atas, tetapi tidak mendukung editor *Jupyter Notebook*, demikian pula, tampilan beranda aplikasi dapat dibagi menjadi dua bagian, tombol untuk memilih menu dan tampilan grafik visual. Pada *streamlit* diperlukan *library NumPy* dan *Pandas* diperlukan untuk menampilkan gambar.

### 1.3.6. Flowchart

Flowchart adalah gambaran urutan kegiatan dalam sistem perangkat produksi dari awal hingga akhir. Oleh karena itu, *flowchart* adalah bagan simbol tertentu yang menjelaskan secara rinci urutan proses serta hubungan antara masing-masing proses dalam program.[26]

Dengan cara ini, *flowchart* dapat membantu menyelesaikan masalah yang mungkin muncul saat membangun sistem. Pada dasarnya, *flowchart* digambarkan dengan menggunakan simbol-simbol. Setiap simbol mewakili suatu proses tertentu. Sedangkan untuk menghubungkan satu proses ke proses selanjutnya digambarkan dengan menggunakan garis penghubung.