

BAB II

LANDASAN TEORI

2.1. Penelitian Terdahulu

Berdasarkan penelitian yang dilakukan oleh Nugroho Widyanto, Jin-Whan Kim, dan Agus Nursikuwagus, Jurnal ini berjudul "*Best Classifier over Feature Selection and Delta Error in Time Series Dataset*" dan membahas metode untuk memilih pengklasifikasi terbaik dalam data deret waktu dengan menggunakan teknik pemilihan fitur dan pengurangan kesalahan delta. Penelitian ini mengusulkan kombinasi metode *SelectKBest* dan *Principal Component Analysis* (PCA) untuk pemilihan fitur, serta menggunakan beberapa algoritma pengklasifikasi seperti *Regresi Linear*, *Regresi Bayesian*, *Support Vector Machine* (SVM), dan *Stochastic Gradient Descent* (SGD) *Regressor*. Dengan menghitung matrik kesalahan seperti *Mean Square Error* (MSE), *Mean Absolute Error* (MAE), dan *Root Mean Square Error* (RMSE), serta menganalisis kesalahan delta, penelitian ini menemukan bahwa *SelectKBest* lebih efektif dalam mengurangi kesalahan dibandingkan PCA. Hasil menunjukkan bahwa *Regresi Linear* dan *Regresi Bayesian* memiliki nilai R2 tertinggi, menjadikannya pengklasifikasi yang paling sesuai untuk dataset deret waktu ini[5].

Persamaan penelitian yang dilakukan oleh peneliti dengan penelitian yang dilakukan oleh Nugroho Widyanto, Jin-Whan Kim, dan Agus Nursikuwagus adalah pertama, menggunakan teknik pemilihan pengklasifikasian terbaik menggunakan *SelectKBest*. Kedua, menggunakan evaluasi matrik MAE, MSE, RMSE dan R2.

Perbedaan penelitiannya yaitu Nugroho Widyanto, Jin-Whan Kim, dan Agus Nursikuwagus dalam penelitiannya yang pertama yaitu menggunakan tambahan beberapa algoritma pengklasifikasi seperti *Regresi Linear*, *Regresi Bayesian*, *Support Vector Machine* (SVM), dan *Stochastic Gradient Descent* (SGD) *Regressor*, sedangkan peneliti menggunakan perbandingan model *Lasso Regression*, *Ridge Regression*, dan *ElasticNet Regression*. Kedua menggunakan *Principal Component Analysis* (PCA) yang dikombinasikan dengan *SelectKBest*.

Penelitian selanjutnya yang telah dilakukan oleh Redi Sebastian Situmorang dalam skripsinya membahas mengenai Penerapan Data Mining Untuk Memprediksi Jumlah Produksi Kelapa Sawit Menggunakan Metode *Regresi Linear* Berganda. Metode pengumpulan data melalui studi pustaka dan observasi langsung pada pihak terkait. Dalam implementasi sistemnya menggunakan *python* dan ditambah dengan metode *Mean Absolute Error* (MAE) untuk mengukur tingkat keakuratan model. Fokus utama dari penelitian yang dilakukan oleh Redi Sebastian Situmorang adalah membantu perusahaan dalam memprediksi produksi permintaan perusahaan[6].

Persamaan penelitian yang dilakukan oleh peneliti dengan penelitian yang dilakukan oleh Redi Sebastian Situmorang yaitu dalam implementasi sistemnya menggunakan *python* dan ditambah dengan evaluasi *Mean Absolute Error* (MAE) untuk mengukur tingkat keakuratan model peramalan. Perbedaannya peneliti sebelumnya hanya menggunakan 1 evaluasi yaitu *Mean Absolute Error* (MAE), sedangkan peneliti melakukan evaluasi dengan 4 macam, yaitu *Mean Absolute Error* (MAE), *Mean Squared Error* (MSE), *Root Mean Squared Error* (RMSE), dan *R-Squared* (R^2).

Penelitian selanjutnya dilakukan oleh Elvrida Jesika Purba, membahas mengenai Penerapan Data Mining Untuk Prediksi Stok Produk Susu Pada PT.PS Maju Bersama Menggunakan Metode *Regresi Linear* Berganda. Dalam penelitian ini, Elvrida Jesika Purba melakukan proses analisis terhadap data pengolahan produk susu dengan tujuan untuk melakukan prediksi stok produk susu yang siap dipasarkan menggunakan metode Regresi Linear Berganda. Dalam proses pengumpulan data, Elvrida Jesika Purba melakukan observasi dan wawancara langsung pada perusahaan. Dalam pengujian sistemnya, menggunakan data prediksi di aplikasi *RapidMiner Studio 5.3*, diperoleh tingkat akurasi perhitungan sebesar 100% dan membuktikan bahwa sistem yang dirancang mampu digunakan untuk memprediksi stok produk susu sehingga tidak akan terdapat kesalahan data di masa yang akan datang. Fokus utama dari penelitian yang dilakukan oleh Elvrida Jesika Purba adalah mempermudah Perusahaan dalam mengestimasi produk yang akan dipasarkan.[7]

Persamaan dalam penelitian ini adalah dalam tahapan metode penelitian yang digunakan yaitu melakukan observasi dan wawancara langsung pada pihak terkait. Perbedaan dalam penelitian yaitu Elvrida Jesika Purba dalam pengujian sistemnya menggunakan aplikasi prediksi *RapidMiner Studio 5.3* sedangkan peneliti dalam implementasi sistemnya menggunakan *python* dan ditambah dengan metode evaluasi *Mean Absolute Error (MAE)*, *Mean Squared Error (MSE)*, *Root Mean Squared Error (RMSE)*, dan *R-Squared (R^2)*.

2.2. Teori Pendukung

Teori pendukung dapat disebut dengan istilah dalam penelitian yang digunakan sebagai acuan untuk memahami penelitian yang dilakukan oleh peneliti.

2.2.1. Kelompok Tani

Kelompok tani merupakan sekelompok petani yang memiliki kepentingan dan tujuan yang sama dan ingin meningkatkan kinerja usaha mereka. Dalam proses pengembangannya, usaha tani seringkali mendapatkan kendala dalam permodalan. Dengan demikian, dilakukan pemberian pinjaman usaha oleh Gabungan Kelompok Tani (GAPOKTAN). Namun dalam peminjamannya seringkali dinilai kurang objektif, tidak tepat sasaran, dan tidak transparan[8]. Dalam konteks pembangunan pertanian dan pedesaan, kelompok tani sering menjadi elemen penting dalam pelaksanaan program-program pemerintah atau organisasi non-pemerintah yang bertujuan untuk meningkatkan produktivitas pertanian, mengurangi kemiskinan di pedesaan, dan mendorong pembangunan berkelanjutan.

2.2.2. Data Mining

Data mining merupakan serangkaian proses untuk menggali nilai tambah dari kumpulan data yang belum diketahui secara manual, *data mining* dapat memecahkan masalah dengan menganalisis data yang telah ada dalam *database*. *Data mining* mempunyai proses yang didalamnya terdapat pengetahuan yang menarik seperti pola, struktur signifikan, perubahan, dan anomali dari beberapa jumlah besar data atau *repository*[9]. *Data mining* juga dapat diartikan sebagai proses yang melibatkan analisis statistik, pembelajaran mesin, dan kecerdasan buatan untuk menemukan informasi yang tersembunyi dalam data mentah.

Tujuan dari proses ini adalah untuk menemukan pola, hubungan, atau tren yang dapat membantu perusahaan membuat keputusan yang lebih baik atau mendapatkan wawasan yang berharga. Ini memungkinkan perusahaan untuk menemukan pola yang dapat digunakan untuk memprediksi perilaku di masa depan, mengetahui preferensi pelanggan, mengoptimalkan proses bisnis, dan membuat keputusan yang lebih cerdas berdasarkan data. *Data mining* juga merupakan sekumpulan prosedur yang berguna untuk mengeksplorasi dan menemukan nilai dalam bentuk informasi serta hubungan kompleks yang telah tersimpan di dalam basis data. Proses penggalian pola informasi termasuk mengubah data menjadi informasi yang lebih berguna dan mengekstraksi pola-pola yang berharga atau menarik dari data yang ada di dalam *database*.

2.2.2.1. Teknik Data Mining

Beberapa Teknik dan sifat data mining yaitu sebagai berikut:

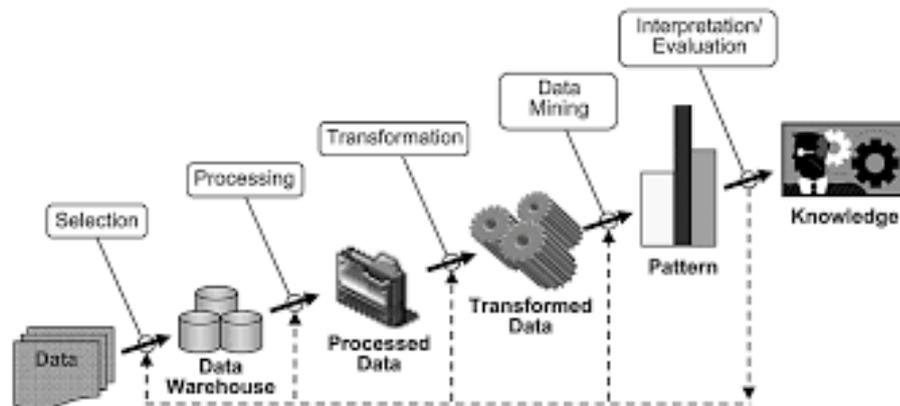
1. *Clustering* (Pengelompokan) proses membagi dataset menjadi beberapa kelompok atau subnet yang lebih kecil sehingga elemen dalam kelompok tertentu memiliki set fitur yang sama, dengan tingkat kesamaan yang tinggi dalam kelompok yang lebih kecil “*unsupervised learning*”.
2. *Classification* (Klasifikasi) adalah menempatkan daftar data baru ke salah satu dari beberapa kategori (kelas) yang telah didefinisikan sebelumnya. Ini juga disebut sebagai “*supervised learning*”. Teknik ini yang mengklasifikasikan data ke dalam kategori atau kelas berdasarkan karakteristik tertentu, menggunakan pola dari data latihan.

3. *Regression* (Regresi) yaitu teknik yang menggunakan model ketergantungan linier atau non-linier untuk memprediksi nilai variabel kontinu berdasarkan nilai variabel lain. Ini memungkinkan prediksi nilai berdasarkan data saat ini dengan memahami hubungan antara variabel input dan output.
4. *Associate Rule* (Asosiasi) adalah metode untuk mengidentifikasi kombinasi pola atau item dari data transaksi yang telah digunakan sebelumnya hanya untuk melaporkan hasil penjualan setiap bulan dan kemudian tidak digunakan lagi[10]. Teknik ini mengidentifikasi bagaimana variabel atau item dalam dataset berhubungan satu sama lain, seperti dalam analisis belanja konsumen.

2.2.3. Knowledge Discovery in Database

Knowledge Discovery in Database (KDD) adalah keseluruhan proses penggalian dan analisis kumpulan data yang sangat besar untuk menemukan pola, informasi, dan pengetahuan yang tersembunyi.

KDD lebih luas dari sekedar data mining karena mencakup keseluruhan proses mendapatkan pengetahuan dari data, mulai dari persiapan data hingga interpretasi hasilnya. Metode teknis yang bermanfaat untuk mencari dan mengidentifikasi pola (*pattern*) dalam data adalah penemuan pengetahuan dalam database (KDD). Pola yang telah ditemukan adalah sah dan baru, sehingga dapat bermanfaat dan dipahami[11]. Berikut ini tahapan dari proses *Knowledge Discovery in Database* (KDD) dapat dilihat pada gambar:



Gambar 2.1. Tahapan KDD
 (Sumber: JISKA (Jurnal Informatika Sunan Kalijaga) [10])

Data mining dan knowledge discovery in databases (KDD) adalah istilah yang sering digunakan untuk menjelaskan proses penggalian informasi tersembunyi dalam suatu basis data yang besar. Sebenarnya, istilah-istilah ini merujuk pada konsep yang berbeda tetapi saling terkait. *Data mining* adalah salah satu langkah dalam proses keseluruhan KDD[12]. Proses KDD garis besarnya dapat dijelaskan sebagai berikut:

1. *Data Selection*

Sebelum tahap penggalian informasi dalam KDD dimulai, data yang diperlukan harus dipilih dari sekumpulan data operasional. Data yang digunakan dalam proses penggalian informasi disimpan dalam berkas terpisah dari basis data operasional.

2. *Pre-processing*

Sebelum proses *data mining* dapat dimulai, data yang menjadi fokus KDD harus dibersihkan. Proses pembersihan mencakup hal-hal seperti menghilangkan duplikat data, memeriksa data yang tidak konsisten, dan memperbaiki kesalahan.

3. *Transformation*

Mengubah dan menggabungkan data ke dalam format tertentu adalah proses yang dikenal sebagai transformasi data. Proses ini melibatkan penggunaan teknik seperti normalisasi, pengkodean, atau *discretization* untuk mengubah data menjadi format yang lebih sesuai untuk analisis.

4. *Data Mining*

Data mining adalah istilah lain untuk proses penambangan data. Ini adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma yang digunakan dalam *data mining* sangat beragam, dan pemilihan teknik atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan. Saat ini, proses menemukan pengetahuan penting yang tersembunyi dalam data dilakukan.

5. *Interpretation/Evaluasi*

Pola informasi yang dihasilkan dari proses *data mining* harus ditampilkan dalam bentuk yang mudah dipahami oleh semua pihak yang terlibat dalam proses tersebut. Ini adalah bagian dari proses interpretasi KDD. Hasil pemrosesan dievaluasi untuk memastikan bahwa mereka relevan dan bermanfaat untuk tujuan analisis.

2.2.4. Klasifikasi

Data mining dapat mengklasifikasikan. Untuk pertama kalinya, klasifikasi tanaman digunakan untuk mengkategorikan suatu spesies tanaman tertentu[13].

Terdapat beberapa teknik yang digunakan dalam klasifikasi, yaitu:

2.2.5. Prediksi

Prediksi adalah upaya untuk mengurangi perkiraan volume penjualan atau bahkan menentukan luas pasar dan potensi penjualan. Selain itu, prediksi ini membantu perusahaan dalam perencanaan penyediaan stok karena memberikan hasil yang optimal bagi perusahaan, sehingga kesalahan perencanaan dapat diminimalkan[14].

2.2.5.1. Teknik Prediksi

Teknik prediksi dalam *data mining* mencakup berbagai metode pengembangan yang berdasarkan *data historis*. Terdapat beberapa data prediksi yang umum digunakan dalam data mining:

1. *Regression* (Regresi)

Teknik ini digunakan untuk mengestimasi berbagai nilai kontinu yang berdasarkan hubungan antara statistik dan variabel input output.

2. *Decision Tree* (Pohon Keputusan)

Teknik ini digunakan untuk membangun nilai prediktif yang berupa struktur pohon keputusan yang dapat memecahkan data menjadi kelompok-kelompok yang spesifik berdasarkan aturan Keputusan.

3. *Artificial Neural Networks* (Jaringan Syaraf Tiruan)

Teknik ini digunakan untuk membuat model yang terinspirasi dari saraf manusia, dengan *neuron-neuron* yang terhubung dan terdapat pola-pola kompleks yang dapat dipelajari dalam data tersebut.

4. *K-Nearest Neighbors* (KNN)

KNN merupakan sebuah algoritma pembelajaran mesin yang sederhana dan efektif untuk masalah klasifikasi dan regresi.

5. *Naive Bayes*

Merupakan Teknik yang menggunakan *Bayes* untuk memprediksi nilai atau kelas yang berdasarkan perhitungan dari berbagai fitur. Metode ini memungkinkan perhitungan anggota kelas data yang terpilih masuk kedalam kelas khusus[15].

2.2.6. Regresi Linear

2.2.6.1. Pengertian Regresi Linear

Regresi linear adalah teknik statistik yang digunakan untuk memodelkan hubungan linier antara satu atau lebih prediktor (variabel independen) dan satu variabel dependen (target). Tujuan dari metode ini adalah untuk memperkirakan koefisien regresi, atau parameter model, sehingga model dapat digunakan untuk membuat prediksi tentang nilai variabel dependen berdasarkan nilai variabel independen. Dalam regresi berganda, model ini mencoba menemukan garis lurus (atau hiper bidang). Metode ini berfungsi dengan baik untuk menganalisis dan memprediksi hubungan antara variabel yang memiliki hubungan linier satu sama lain.

2.2.6.2. Model Regresi Linear

Beberapa model regresi linear yang paling umum digunakan untuk mensimulasikan hubungan linier antara variabel independen (fitur) dan variabel dependen (target) dalam kumpulan data, adalah sebagai berikut:

1. Model Regresi Linear Sederhana

Dalam model ini, satu *variable independent* digunakan untuk memprediksi *variable dependent*, hubungan antara keduanya digambarkan sebagai garis lurus. Tujuan dari regresi linear sederhana adalah untuk mengevaluasi hubungan sebab akibat antara variabel faktor penyebab dan variabel akibatnya[16].

Persamaan regresi linear sederhana adalah: $y = b_0 + b_1x$

- y adalah variabel dependen.
- x adalah variabel independen tunggal.
- b_0 adalah *intercept* (titik potong garis regresi dengan sumbu y).
- b_1 adalah koefisien regresi, menunjukkan kemiringan garis regresi.

2. Model Regresi Linear Berganda

Regresi linear berganda adalah jenis regresi yang memiliki satu variabel *dependent* (tidak bebas) dan lebih dari satu variabel independen. Tujuan dari analisis ini adalah untuk mengetahui apakah masing-masing dari variabel independen memiliki hubungan yang positif atau negatif, dan untuk memperkirakan apakah nilai dari masing-masing dari variabel independen akan meningkat atau menurun. Jika ada lebih dari satu variabel bebas atau variabel terikat, disebut regresi linier berganda. Sebaliknya, jika hanya ada satu variabel bebas dan satu variabel terikat, disebut *regresi linear sederhana*[17].

Analisis regresi menggunakan metode regresi linear berganda untuk menjelaskan hubungan antara pengubah, atau variabel dependen, dan komponen yang mempengaruhi lebih dari satu prediktor, atau variabel independen.

Persamaan Regresi Linear Berganda adalah:

$$\text{Untuk 2 prediktor: } Y = b_1 + b_2X_1 + b_3X_2$$

$$\text{Untuk 3 prediktor: } Y = b_1 + b_2X_1 + b_3X_2 + b_4X_3$$

$$\text{Untuk n prediktor: } Y = b_1 + b_2X_1 + b_3X_2 + \dots + b_nX_n$$

Langkah-langkah dalam analisis regresi berganda adalah sebagai berikut:

1. Memasukkan nilai-nilai itu ke dalam persamaan:

$$Y = b_1n + b_2\Sigma X_1 + b_3\Sigma X_2$$

$$YX_1 = b_1\Sigma X_1 + b_2\Sigma X_1^2 + b_3\Sigma X_1X_2$$

$$YX_2 = b_1\Sigma X_2 + b_2\Sigma X_1X_2 + b_3\Sigma X_2^2$$

Bentuk umum Regresi Linear Berganda sebagai berikut:

$$Y = a + b_1X_1 + b_2X_2 + \dots + b_nX_n \dots \dots \dots (1)$$

Dimana:

Y : nilai variabel Y hasil prediksi

Y : variabel tak bebas

X : variabel bebas

a : konstanta regresi

b : koefisien regresi

Nilai konstanta a dan koefisien regresi b dapat dihitung menggunakan metode matriks, persamaan normal (substitusi), dan metode kuadrat terkecil.

Dalam penelitian ini, peneliti menggunakan metode kuadrat terkecil, yang berasal dari persamaan 2. Nilai konstanta a dapat ditemukan dengan menggunakan rumus berikut:

$$a = \bar{Y} - b_1\bar{X}_1 + b_2\bar{X}_2 \dots \dots \dots (2)$$

Untuk \bar{Y} , \bar{X}_1 , \bar{X}_2 digunakan rumus:

$$\bar{Y} = \frac{\sum Y}{n} \dots \dots \dots (3)$$

$$\bar{X}_1 = \frac{\sum X_1}{n} \dots \dots \dots (4)$$

$$\bar{X}_2 = \frac{\sum X_2}{n} \dots \dots \dots (5)$$

b1 dan b2 dicari dengan persamaan:

$$b_1 = \frac{(\sum X_2^2)(\sum X_1 Y) - (\sum X_1 X_2)(\sum X_2 Y)}{(\sum X_1^2)(\sum X_2^2) - (\sum X_1 X_2)^2}$$

$$b_2 = \frac{(\sum X_1^2)(\sum X_2 Y) - (\sum X_1 X_2)(\sum X_1 Y)}{(\sum X_1^2)(\sum X_2^2) - (\sum X_1 X_2)^2} \text{ [18].}$$

2.3. Alat Pendukung

Terdapat beberapa alat pendukung yang digunakan untuk membantu peneliti menyelesaikan penelitian, diantaranya:

2.3.1. Python

Python adalah bahasa pemrograman berorientasi objek yang dapat digunakan untuk membuat perangkat lunak dan berjalan diberbagai sistem operasi. *Python* sangat populer di bidang *data science* dan analisis karena dukungannya terhadap *library*. *Library* ini menawarkan fungsi analisis data, *machine learning*, alat *preprocessing* data, dan *visualisasi* data[19].

Secara umum, *Python* memiliki ciri-ciri berikut:

1. Banyak *library* yang didukung
2. Bahasa yang relatif mudah dipahami
3. Aturan *layout source code* yang memudahkan pengecekan kode
4. Bisa diinterpretasi karena *coding*, *eksekusi*, dan *debugging* lebih mudah dibandingkan dengan bahasa yang dikompilasi
5. Portabel karena hanya perlu menulis satu kali untuk dijalankan di platform lain
6. *Open-source*[20].

Pada penelitian ini juga terdapat beberapa *library* populer yang digunakan untuk melakukan transformasi dataset dalam data mining. Diantaranya :

1. *NumPy*

Pustaka *Python NumPy* digunakan untuk komputasi numerik dan menawarkan *array multidimensi* serta sejumlah fungsi matematika yang kuat untuk bekerja dengan array. *NumPy* sering digunakan sebagai pondasi untuk pustaka ilmiah lainnya dalam lingkungan *Python*.

2. *Pandas*

Pandas adalah pustaka *Python* yang digunakan untuk analisis dan manipulasi data tabular. Ini menawarkan struktur data tingkat tinggi seperti *dataframe*, yang memungkinkan pengguna mengimpor, mengatur, dan menganalisis data dengan mudah. *Pandas* juga digunakan untuk memuat dan membersihkan data sebelum diterapkan pada model pembelajaran mesin atau visualisasi data.

3. *Scikit-learn*,

Scikit-learn adalah pustaka analisis data *open source* yang mencakup berbagai metode pengolahan data, seperti *clustering*, *regresi*, dan *klasifikasi*. Pustaka ini juga berfungsi sebagai standar emas untuk pembelajaran mesin (ML) dalam ekosistem *Python*[21].

4. *Matplotlib*

Matplotlib adalah pustaka *Python* yang digunakan untuk visualisasi data dan menawarkan berbagai jenis plot, seperti grafik garis, *scatter plot*, dan histogram, antara lain. *Matplotlib* memungkinkan pengguna untuk membuat plot dengan kontrol yang luas terhadap semua aspek plot, termasuk warna dan label, ukuran, dan tata letak.

5. *Seaborn*

Seaborn adalah pustaka *Python* yang didedikasikan untuk visualisasi data statistik dan dibangun di atas *Matplotlib*. Antarmuka yang lebih mudah digunakan dan lebih abstrak daripada *Matplotlib* memungkinkan pengguna membuat visualisasi data yang menarik dengan sedikit kode.

6. *Pickle*

Pickle adalah modul dalam pustaka standar *Python* yang dapat digunakan untuk serialisasi dan deserialisasi objek *Python*. Serialisasi mengubah objek *Python* menjadi aliran *byte*, dan deserialisasi mengubah objek *Python* kembali ke aliran *byte*. Ini memungkinkan pengguna menyimpan objek *Python* ke dalam *file* atau mengirimkannya melalui jaringan, dan kemudian dengan mudah memulihkan objek kembali.



Gambar 2.2. Logo Python
(Sumber: www.python.org)

2.3.2. HTML (*Hyper Text Markup Language*)

Bahasa *Hyper Text Markup Language* (HTML) adalah bahasa yang digunakan untuk memformat teks dokumen pada jaringan komputer yang disebut *World Wide Web* (www). *World Wide Web Consortium* (W3C) menetapkan standar untuk penggunaan HTML. Pada tahun 1989, Tim Caillau dan Robert Berners-lee bekerja sama untuk membuat HTML. CERN di Jenewa adalah pusat penelitian fisika energi tinggi[22].



Gambar 2.3. Logo Html
(Sumber: www.w3.org/html/logo)

2.3.3. CSS (*Cascading Style Sheet*)

Cascading Style Sheets (CSS) adalah sekumpulan perintah yang digunakan untuk membuat tampilan halaman web.

CSS beroperasi melalui tag "*Style*", dan dengan adanya CSS pada HTML, menjadi lebih mudah untuk mengatur warna dan tampilan. Jika tidak ada CSS, tampilan *website* akan menjadi tidak menarik[23].



Gambar 2.4. Logo CSS
(Sumber: www.w3.org/Style/CSS/Overview.en.html)

```

1  /* CSS untuk membuat teks berwarna merah */
2  h1 {
3      color: red;
4  }
5  /* CSS untuk mengatur lebar dan tinggi gambar
6  img {
7      width: 200px;
8      height: 200px;
9  }

```

Gambar 2.5. Tampilan CSS
(Sumber: www.w3.org/Style/CSS)

2.3.4. Google Collaboratory

Google Collaboratory adalah alat editor kode yang menggunakan sistem penyimpanan awan gratis untuk penelitian. *Google collabs* dengan elemen *jupyter*, dan hampir semua masukan *library* tersedia untuk digunakan dalam penelitian, termasuk meneliti dengan *data mining*[24].

Pengguna dapat lebih mudah mengembangkan dan menjalankan proyek pembelajaran mesin dan *data science* dengan *Google Colabs*. Mereka tidak perlu khawatir tentang konfigurasi perangkat lunak dan perangkat keras.



Gambar 2.6. Logo Google Collabs
(Sumber: colab.research.google.com)

2.3.5. Streamlit

Framework ini menjadi pilihan karena kemampuannya untuk memfasilitasi pengembangan aplikasi web interaktif secara cepat dan efisien menggunakan bahasa pemrograman *python* tanpa memerlukan tingkat keahlian yang tinggi dalam pembuatan antarmuka pengguna yang kompleks. Pengembang dapat fokus pada pengembangan fitur inti sistem yang sedang diteliti sambil mempertahankan pengalaman pengguna yang responsif dan mudah diakses dengan *streamlit*[25].



Gambar 2.7. Logo Streamlit
(Sumber: streamlit.io)