

BAB II

LANDASAN TEORI

2.1 Stok Barang

Menurut Stice dan Skousen, stok barang merujuk pada item atau produk yang diperjualbelikan dalam aktivitas operasional rutin perusahaan, baik yang sudah siap dijual maupun yang masih menjadi bagian dari proses produksi sebelum akhirnya dipasarkan [6]. Stok merupakan bagian dari aset yang cepat dicairkan dalam sebuah perusahaan dan memiliki peranan yang sangat penting dalam mendatangkan keuntungan. Secara umum, kata stok digunakan untuk merujuk pada produk-produk yang dimiliki untuk dijual lagi atau untuk digunakan dalam proses pembuatan produk yang akan dipasarkan. Dalam perusahaan yang bergerak di bidang perdagangan, stok barang adalah items yang didapatkan atau dibeli dengan maksud untuk dijual kembali tanpa melakukan perubahan pada barang tersebut [7].

Stok barang adalah bagian yang sangat penting dalam operasional bisnis, terutama bagi yang mengutamakan penjualan. Seringkali, masalah muncul ketika ada ketidaksesuaian antara jumlah barang yang tersedia dan permintaan dari konsumen. Hal ini dapat mengganggu efisiensi perusahaan dan juga dapat menimbulkan ketidakpuasan bagi pelanggan. Jika barang yang ada di gudang tidak cukup untuk memenuhi permintaan, peluang untuk menjual akan hilang, lebih buruk lagi, ini dapat mengurangi kesetiaan dari konsumen. Sebaliknya, jika persediaan barang terlalu banyak, perusahaan akan menanggung biaya penyimpanan yang tinggi dan risiko usang atau kerusakan barang. Oleh karena itu, pengelolaan stok barang yang efektif menjadi suatu kebutuhan yang harus dipenuhi

untuk menjaga keseimbangan antara ketersediaan stok dan permintaan [8]. Dalam melakukan stok barang, barang dikelompokkan berdasarkan atribut-atribut yang memberikan preferensi bagi konsumen, antara lain sebagai berikut:

1. Model: Model barang mengacu pada desain atau gaya utama dari sebuah produk, seperti casual, formal atau skinny.
2. Jenis Bahan: Jenis bahan merupakan atribut yang paling penting untuk menentukan kualitas, kenyamanan, dan harga produk. Contohnya katun, denim, brokat, fleece, dan poliester.
3. Ukuran: Ukuran merupakan atribut yang memainkan peran penting dalam memenuhi kebutuhan beragam konsumen, misalnya ukuran s, m, l, xl dan seterusnya.
4. Warna: Warna pakaian mencerminkan selera konsumen dan tren mode. Contohnya warna hitam, merah, biru, coklat dan sebagainya.
5. Jenis Kelamin atau Target Konsumen: Jenis kelamin atau target konsumen merupakan atribut yang mengacu pada pria, wanita, dan unisex yaitu produk yang dapat digunakan oleh pria atau wanita.
6. Harga: Harga merupakan atribut yang juga paling penting bagi konsumen, seperti harga dengan barang terjangkau, sedang, premium atau barang dengan harga diskon.
7. Musim atau Kebutuhan: Musim atau kebutuhan merupakan mengelompokkan barang sesuai dengan musim atau kebutuhan tertentu, seperti musim panas, musim dingin, atau acara spesifik.

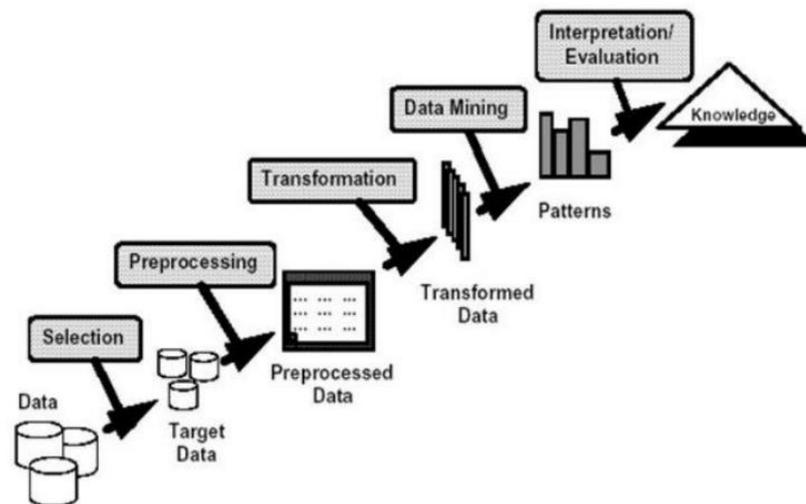
8. Popularitas atau Permintaan: Popularitas atau permintaan merupakan mengelompokkan barang berdasarkan tingkat popularitas atau permintaan di pasar seperti barang populer atau barang dengan permintaan khusus.
9. Kategori Acara: Kategori acara merupakan atribut yang mengacu pada pakaian yang akan dipakai oleh konsumen, seperti pakaian kantor, santai, olahraga atau pesta.
10. Jenis Produk: Jenis produk merupakan mengelompokkan barang berdasarkan tipe atau jenis produk tertentu dalam kategori fashion seperti atasan, bawahan, aksesoris dan sebagainya.

2.2 Data Mining

Data mining adalah hasil perpaduan antara beberapa bidang ilmu yang menggabungkan teknik-teknik dari *Machine Learning*, pengenalan pola, statistik, manajemen basis data, serta visualisasi. Tujuan utamanya adalah mengatasi permasalahan dalam mengambil informasi berharga dari kumpulan data besar dalam sebuah database yang signifikan [9].

Data mining menelusuri data pada database untuk membangun model dan menggunakannya untuk mengenali pola data lain yang tidak tersimpan dalam basis data. Data mining juga memiliki tujuan lain yaitu memanfaatkan data suatu basis data dengan mengolahnya sehingga mendapatkan informasi baru yang berguna. Secara sederhana, data mining biasa dikatakan sebagai proses penyaring atau "menambang" pengetahuan dari sejumlah data yang besar. Istilah lain untuk data mining adalah *Knowledge Discovery in Database (KDD)*. Walaupun data mining

sendiri adalah bagian dari tahapan proses *KDD* seperti yang terlihat pada gambar dibawah ini [10].



Sumber: [10]

Gambar 2.1 Tahapan Proses *Knowlegde Discovery in Database (KDD)*

Berikut merupakan penjelasan tahapan dari proses *Knowlegde Discovery in Database (KDD)* [11]:

1. Seleksi atau Pemilihan Data

Pada tahap ini, data yang relevan untuk analisis diidentifikasi dan dipilih. Pemilihan data yang tepat sangatlah penting karena kualitas dan relevansi data mempengaruhi hasil akhir dari proses *Knowlegde Discovery in Database (KDD)*.

2. Pemrosesan dan Pembersihan Data

Pemrosesan dan pembersihan data adalah dua tahapan krusial dalam siklus *Knowledge Discovery in Database (KDD)*. Proses ini bertujuan untuk memastikan bahwa data yang digunakan dalam analisis atau pemodelan tidak hanya lengkap dan konsisten, tetapi juga sesuai dan siap digunakan

dalam model analisis. Tanpa pemrosesan dan pembersihan data yang tepat, hasil analisis atau model yang dibangun mungkin menyesatkan atau tidak valid.

3. Transformasi

Setelah data dibersihkan, langkah selanjutnya adalah mentransformasikan data. Transformasi data bertujuan untuk mengubah data yang telah diseleksi dan dibersihkan ke dalam format atau bentuk yang lebih sesuai untuk dianalisis lebih lanjut. Pada tahap ini, menggunakan teknik tertentu, seperti data mining.

4. Data Mining

Data mining adalah langkah utama *KDD*. Pada fase ini diterapkan teknik analisis algoritma untuk menemukan pola atau informasi tersembunyi dalam data. Data mining melibatkan penggunaan berbagai teknik statistik dan pembelajaran mesin untuk menganalisis data dan menemukan hubungan atau pola yang sebelumnya tidak diketahui.

5. Interpretasi/Evaluasi

Setelah pola atau model sudah ditemukan maka langkah selanjutnya adalah evaluasi. Pada fase ini model yang ditemukan diuji untuk mengevaluasi efektivitasnya dalam mendeskripsikan data hasil yang diinginkan. Evaluasi bertujuan untuk mengetahui apakah pola atau model yang ditemukan cukup valid dan layak digunakan dalam pengambilan keputusan.

2.3 Clustering

Salah satu metode yang populer dalam penambangan data adalah pengelompokan, di mana sekelompok data atau objek dikelompokkan menjadi beberapa *cluster*. Tujuannya adalah untuk memastikan setiap *cluster* memiliki data yang serupa sebanyak mungkin dan berbeda dari *cluster* lainnya [12]. *Clustering* merupakan teknik untuk mengelompokkan data ke dalam dua kelompok atau lebih, dengan tujuan agar data dalam satu kelompok memiliki kemiripan yang tinggi dibandingkan dengan data pada kelompok lain. Teknik ini termasuk dalam kategori pembelajaran tanpa pengawasan (*unsupervised learning*), karena tidak memerlukan label atau kategori yang sudah ditentukan sebelumnya.

Dengan kata lain, *clustering* adalah proses pengelompokan data menggunakan pendekatan *unsupervised learning*, yaitu tanpa melalui tahap pelatihan atau pemberian label sebelumnya. Teknik ini bertujuan untuk membagi data ke dalam beberapa kelompok atau cluster berdasarkan kesamaan atribut yang dimiliki oleh data di dalam kelompok tersebut [13]. *Clustering* memainkan peran penting dalam analisis data besar, terutama dalam pengelolaan stok barang. *Clustering* memungkinkan untuk mengelompokkan barang menjadi beberapa kategori, seperti produk dengan tingkat permintaan rendah, sedang, dan tinggi. Dengan *clustering*, toko dapat mengoptimalkan pengelolaan stok dan mengurangi biaya penyimpanan barang yang tidak laku.

2.4 K-Means Clustering

2.4.1 Pengertian K-Means Clustering

K-Means merupakan algoritma *clustering* dengan menentukan sejumlah data untuk di *cluster* dalam kesamaan karakteristik dan memaksimalkan perbedaan antar *cluster*. *K-Means* adalah salah satu algoritma *clustering* yang paling banyak orang pakai untuk mengkategorikan sebuah produk. Yang dimaksud *K* disini yaitu sebagai konstanta merupakan total *cluster* yang diperlukan, sedangkan *Means* berarti nilai rata-rata dari suatu kategori data yang dalam hal ini didefinisikan sebagai *cluster*, sehingga *K-Means Clustering* adalah suatu metode menganalisa data atau metode data mining yang melakukan proses pemodelan tanpa supervisi (*unsupervised*) dan merupakan salah satu metode yang melakukan pengkategorian data dengan sistem partisi [14].

K-Means Clustering digunakan dalam analisis data untuk mengelompokkan data menjadi beberapa kelompok berdasarkan kemiripan karakteristik. Metode ini sangat berguna dalam memahami struktur data dan mengidentifikasi pola yang ada. *K-Means* berusaha untuk mengelompokkan data sedemikian rupa sehingga data dalam satu kelompok memiliki kesamaan yang tinggi, sementara kelompok lainnya memiliki karakteristik yang berbeda [15].

2.4.2 Langkah-Langkah K-Means Clustering

Adapun langkah-langkah dalam algoritma *K-Means Clustering* [16]:

1. Tentukan nilai *k* sebanyak jumlah *cluster* atau kelompok yang diinginkan.
2. Pilih sebanyak *k* data dari set data sebagai pusat *cluster* (centorid) secara random.

3. Menghitung jarak antara objek dengan masing-masing centroid dengan menggunakan rumus jarak *Euclidean Distance*. *Euclidean Distance* adalah metode umum dalam mengukur jarak antara data point dan centroid. Rumusnya sebagai berikut:

$$d(x, y) = \|x - y\| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad ; i = 1, 2, 3, \dots n$$

Keterangan:

X_i : objek x ke-i

Y_i : data y ke-i

n: banyaknya objek

4. Mengelompokkan objek berdasarkan jarak terdekat dengan centroid.
5. Menentukan centroid baru dengan menggunakan rumus sebagai berikut:

$$v = \frac{\sum_{i=1}^n x_i}{n} \quad ; i = 1, 2, 3, \dots n$$

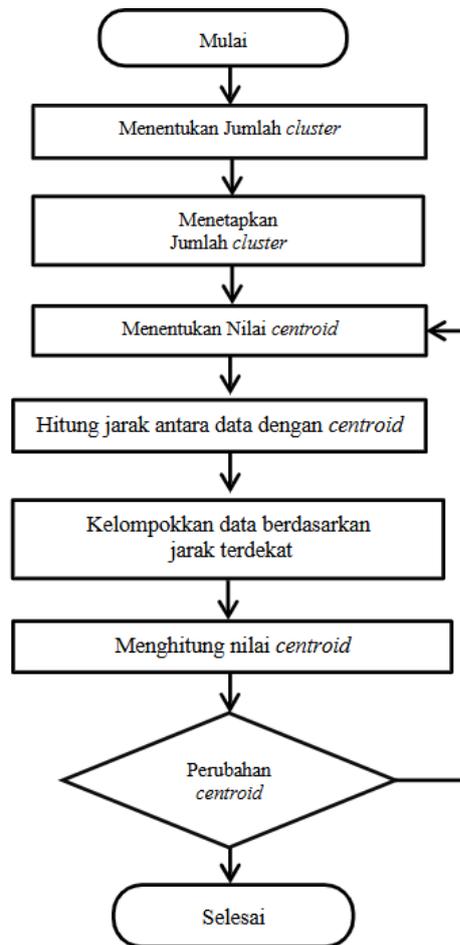
Keterangan:

V: centroid pada *cluster*

X_i : objek ke-i

n: banyaknya objek/jumlah objek yang menjadi anggota *cluster*

6. Ulangi langkah 3 dan 4 hingga tidak ada lagi objek yang berpindah *cluster*.



Sumber: [17]

Gambar 2.2 Flowchart Algoritma *K-Means Clustering*

Adapun kelebihan dari algoritma *K-Means Clustering*, antara lain:

- A. Sederhana dan Mudah Diimplementasikan: *K-Means* memiliki struktur algoritma yang relatif sederhana dan tidak memerlukan parameter atau struktur yang kompleks, sehingga cocok untuk mahasiswa, peneliti pemula, atau implementasi pada sistem dengan sumber daya terbatas.
- B. Cepat dan Efisien: Algoritma ini memiliki kompleksitas waktu $O(nkt)$ (n = jumlah data, k = jumlah klaster, t = iterasi), yang artinya *K-Means* sangat

cepat dalam menangani data berskala besar, terutama jika dibandingkan dengan metode *clustering* lain seperti *hierarchical clustering*.

- C. Dapat Diaplikasikan pada Berbagai Bidang: *K-Means* telah banyak digunakan dalam bidang pemasaran, pengelompokan pelanggan, analisis stok, segmentasi citra, dan masih banyak lagi.
- D. Dapat Diintegrasikan dengan Teknologi *Machine Learning*: Sangat mudah diintegrasikan dengan bahasa pemrograman seperti *Python* menggunakan *library Scikit-learn*, serta bisa digunakan dalam *pipeline machine learning*.
- E. Hasil Klaster yang *Interpretable*: Setiap data akan diklasifikasikan ke dalam klaster tertentu berdasarkan centroid terdekat, yang membuat hasilnya mudah diinterpretasikan dalam bentuk visual (misalnya *scatter plot*).

Adapun kekurangan dari algoritma *K-Means Clustering*, antara lain:

- A. Harus Menentukan Jumlah Klaster (K) di Awal: Nilai K harus ditentukan sebelum proses klasterisasi, dan pemilihan nilai K yang kurang tepat dapat menghasilkan klaster yang tidak optimal. Oleh karena itu, sering diperlukan metode tambahan seperti *Elbow Method*, *Davies Bouldin Index (DBI)* atau *Silhouette Coefficient (SC)*.
- B. Sensitif terhadap Nilai Awal Centroid: pemilihan awal yang tidak tepat bisa membuat hasil *clustering* tidak akurat atau tidak stabil.
- C. Peka terhadap *Outlier*: Jika ada data yang nilainya sangat jauh (menyimpang) dari yang lain (*outlier*), maka itu bisa mengganggu posisi pusat klaster (centroid) dan membuat hasil pengelompokan menjadi tidak akurat.

2.5 Evaluasi Model *Clustering*

Evaluasi model *clustering* adalah langkah penting dalam data mining untuk menilai kualitas pengelompokan data yang dihasilkan oleh algoritma *clustering*. Tujuan evaluasi ini adalah memastikan bahwa *cluster* yang terbentuk memiliki kesamaan tinggi dalam *cluster* dan perbedaan signifikan antar *cluster*. Evaluasi ini membantu menentukan efektivitas algoritma *clustering* dan memungkinkan perbandingan antar metode untuk memilih model yang paling sesuai dengan kebutuhan analisis data.

2.5.1 *Elbow Method*

Metode *Elbow* adalah sebuah konsep yang digunakan untuk memberikan panduan dalam memilih nilai *cluster* yang optimal. Ini dilakukan dengan mengamati presentasi hasil perbandingan antara jumlah *cluster* yang membentuk siku pada titik tertentu. Selain itu, presentasi perhitungan yang diperoleh juga menjadi tolok ukur antara jumlah *cluster* yang ditambahkan. Perbedaan dalam presentasi antara nilai *cluster* pertama dan nilai *cluster* kedua akan menunjukkan sudut dalam grafik yang menunjukkan nilai *cluster* terbaik. Untuk mendapatkan perbandingan tersebut, perlu menghitung *SSE* (*Sum of Square Error*) untuk masing-masing nilai *cluster*. Karena, ketika jumlah cluster k meningkat, nilai *SSE* akan semakin kecil [18].

2.5.2 *Davies Bouldin Index (DBI)*

Davies Bouldin Index (DBI) merupakan salah satu metode evaluasi internal yang digunakan untuk menilai kualitas hasil pengelompokan. Penilaian ini didasarkan pada dua aspek utama, yaitu tingkat kedekatan antar data dalam satu

cluster (kohesi) dan jarak antar pusat *cluster* (separasi). Dalam konteks ini, kohesi menunjukkan seberapa dekat data terhadap pusat kelompoknya, sementara separasi menggambarkan sejauh mana antar *cluster* saling terpisah. Semakin kecil nilai DBI (minimal 0 dan tidak bernilai negatif), maka semakin baik kualitas pengelompokan yang dihasilkan oleh algoritma *K-Means* [19].

2.5.3 *Silhouette Coefficient (SC)*

Silhouette Coefficient (SC) berfungsi untuk mengevaluasi seberapa baik dan kuat suatu *cluster*. *SC* menilai seberapa dekat dan terpisahnya objek dalam *cluster* tersebut. Semakin tinggi nilai *SC*, semakin baik struktur *cluster* dan semakin kuat hubungan antar objek dalam *cluster* tersebut [20].

Tabel 2.1 Rentang Nilai *Silhouette Coefficient* dan Interpretasi

Rentang Nilai	Interpretasi
0.71 – 1.0	Struktur kuat
0.51 – 0.70	Struktur yang beralasan
0.26 – 0.50	Struktur lemah
<0.25	Tidak ditemukan struktur yang substansial

Sumber: [21]

Tabel diatas memberikan informasi tentang rentang nilai dan interpretasi dengan ketentuan *cluster* dengan nilai *SC* mendekati 1 dianggap sebagai hasil *cluster* yang baik.

2.6 *Machine Learning*

Machine Learning merupakan bagian dari kecerdasan buatan yang memungkinkan komputer berperilaku seperti manusia, di mana *Machine Learning*

dapat didefinisikan sebagai algoritma yang bertujuan untuk menemukan dan mengaplikasikan pola-pola di dalam data. Dalam *Machine Learning*, algoritma memanfaatkan metode statistik untuk mengidentifikasi pola-pola. Pola yang dicari dalam data biasanya berukuran besar [22].

Machine Learning dapat didefinisikan sebagai kemampuan sistem komputer untuk belajar secara otomatis dari data tanpa harus diberikan instruksi pemrograman secara langsung. Sistem ini menggunakan data sebagai dasar untuk membentuk model yang nantinya digunakan dalam proses pengambilan keputusan [23]. Pada penelitian ini, *Machine Learning* digunakan untuk meningkatkan efisiensi dan akurasi pengelolaan stok barang. Dengan menganalisis data historis, teknologi ini memungkinkan perencanaan inventaris yang lebih baik, mengurangi risiko kelebihan atau kekurangan stok, dan membantu sebuah bisnis merespon kebutuhan pasar secara lebih cepat.

Adapun kelebihan yang dimiliki *Machine Learning*, antara lain:

- A. Belajar dari Data: *Machine Learning* bisa belajar dan memahami pola dari data yang ada. Jadi, jika kita punya banyak data, sistem bisa menemukan pola yang kita mungkin tidak bisa lihat dengan mudah.
- B. Mengerjakan Tugas Secara Otomatis: Dengan *Machine Learning*, banyak tugas yang biasanya dilakukan manusia bisa dikerjakan secara otomatis oleh komputer, seperti mengenali gambar, menganalisis sentimen dari teks, atau memprediksi tren bisnis.

- C. Semakin Akurat Seiring Waktu: Semakin banyak data yang dimiliki, semakin pintar sistem *Machine Learning* dalam memberikan prediksi atau keputusan. Jadi, modelnya bisa terus berkembang menjadi lebih baik.
- D. Bisa Digunakan di Banyak Bidang: *Machine Learning* bisa diterapkan di berbagai macam bidang, seperti kesehatan, keuangan, pemasaran, hingga teknologi. Misalnya, di bidang kesehatan untuk mendeteksi penyakit atau di sektor keuangan untuk memprediksi nilai saham.
- E. Mampu Menyesuaikan Diri dengan Perubahan: *Machine Learning* bisa menyesuaikan diri jika ada perubahan dalam data yang diterima. Misalnya, dalam mendeteksi penipuan online, model *Machine Learning* bisa terus diperbarui untuk mengenali pola penipuan yang baru.

Adapun kekurangan yang dimiliki *Machine Learning*, antara lain:

- A. Butuh Banyak Data: *Machine Learning* membutuhkan data yang banyak agar bisa bekerja dengan baik. Jika data yang digunakan sedikit, hasilnya bisa tidak akurat atau bahkan salah.
- B. Sulit Dipahami: Banyak model *Machine Learning* yang bekerja seperti kotak hitam (*black box*), artinya meskipun hasil prediksi atau keputusan yang diberikan bisa tepat, cara model tersebut mengambil keputusan sering kali sulit dimengerti.
- C. Tergantung Kualitas Data: Jika data yang digunakan berkualitas buruk atau banyak yang hilang, hasil dari model *Machine Learning* bisa buruk. Data yang buruk bisa membuat model jadi salah paham dan memberi hasil yang tidak tepat.

- D. Masalah *Overfitting* dan *Underfitting*: *Overfitting* terjadi jika model terlalu "terlalu menghafal" data yang ada, sehingga tidak bisa bekerja dengan baik pada data yang baru. Sedangkan *underfitting* terjadi jika model terlalu sederhana dan tidak bisa menangkap pola yang ada di data. Mencari keseimbangan yang tepat antara keduanya itu sulit.
- E. Butuh Pengaturan yang Tepat: Hasil dari *Machine Learning* juga sangat tergantung pada cara kita mengatur parameter dalam model. Pengaturan yang salah bisa membuat hasilnya buruk, dan proses untuk mencari pengaturan yang tepat memerlukan waktu dan percakapan yang banyak.

2.7 Alat Bantu Pemrograman dan *Tools* Pendukung

2.7.1 *Python*

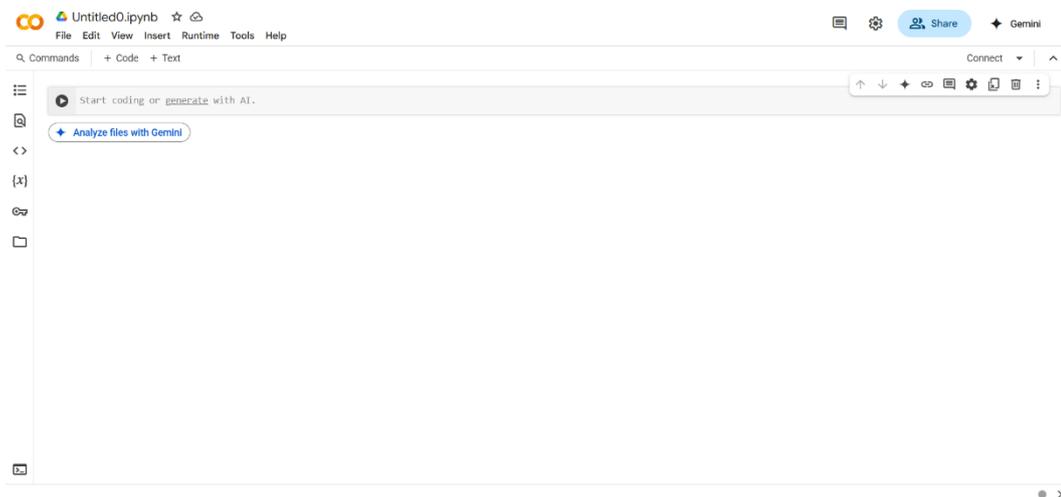
Python merupakan sebuah bahasa pemrograman yang memiliki tingkat kesulitan yang tinggi, bersifat dinamis, interpretatif, serta memiliki kemampuan yang sangat bermanfaat dalam berbagai keperluan. *Python* juga termasuk bahasa pemrograman yang dapat dengan mudah dipelajari dan digunakan. *Python* merupakan salah satu bahasa pemrograman yang paling banyak digunakan karena memiliki sintaks yang sederhana dan mudah dipahami.

Kelebihan dari *python* adalah penggunaan yang cukup mudah dalam pengoprasian kode-kode yang ada. Selain itu, *Python* juga sangat fleksibel untuk berbagai macam keperluan pemrograman. *Python* dapat digunakan untuk membuat aplikasi desktop, web, dan mobile yang beragam. *Python* juga populer digunakan

untuk melakukan *Data Science*, *Machine Learning*, dan berbagai macam analisis data lainnya [24].

2.7.2 Google Colaboratory

Google Colaboratory atau biasa disebut *Google Colab* adalah sebuah *platform* penelitian dan pengembangan yang disediakan oleh *google* secara gratis. *Google Colab* memberikan lingkungan kerja berbasis *cloud* yang memungkinkan pengguna untuk menulis dan mengeksekusi kode *Python* dalam lingkungan *Jupyter Notebook* tanpa memerlukan konfigurasi atau instalasi tambahan. Keuntungan utama dari *Google Colab* adalah kemampuannya untuk berbagi dan mengakses sumber daya komputasi berbasis *cloud google*, termasuk unit pemrosesan grafis (*GPU*) dan unit pemrosesan tensor (*TPU*). Ini sangat bermanfaat untuk proyek-proyek yang memerlukan daya komputasi tinggi, seperti *Machine Learning* dan *Deep Learning* [25].



Gambar 2.3 Tampilan Pada Google Colab

Gambar diatas merupakan tampilan awal dari *Google Colab*, di mana *Google Colab* memiliki fitur-fitur utama seperti [26]:

1. *Cloud Python*: Anda bisa menulis serta menjalankan kode *Python* langsung di browser tanpa harus menginstal *Python* atau pustaka di komputer Anda.
2. *Gratis*: *Google Colab* adalah platform yang ditawarkan oleh *Google* secara gratis. Anda bisa mengaksesnya tanpa biaya.
3. *GPU Tanpa Biaya*: *Google Colab* memberikan akses gratis ke GPU (*Graphics Processing Unit*), yang sangat membantu dalam melatih model *Machine Learning* yang membutuhkan banyak daya komputasi.
4. *Penyimpanan Google Drive*: Anda bisa dengan mudah mengakses dan menyimpan file di *Google Drive*, yang membuatnya praktis untuk berbagi dan menyimpan pekerjaan Anda.
5. *Notebook Interaktif*: *Google Colab* menggunakan format "notebook", yang memungkinkan Anda menggabungkan kode, teks, gambar, dan hasil dalam satu dokumen yang interaktif. Ini bermanfaat untuk dokumentasi dan membagikan analisis.
6. *Ketersediaan Pustaka*: *Google Colab* menawarkan banyak pustaka populer seperti *NumPy*, *Pandas*, *TensorFlow*, *PyTorch*, dan lainnya. Anda dapat dengan mudah mengimpor pustaka-pustaka ini ke dalam lingkungan Colab.
7. *Kolaborasi Tim*: Anda bisa berbagi notebook Colab dengan rekan tim dan bekerja sama secara langsung.
8. *Mudah dan Fleksibel*: Meskipun memiliki banyak fitur, *Google Colab* dirancang untuk pemula. Ini adalah cara yang baik untuk memulai belajar pemrograman *Python* dan melakukan eksplorasi.

Adapun cara menggunakan *Google Colab* adalah sebagai berikut [27] :

1. Login pada akun *Google* dan masuk *Google Drive*.
2. Setelah masuk, buat folder pada *Google Drive*.
3. Membuat notebook *Google Colab*.
4. Menuliskan kode pemrograman *Python*.
5. Menjalankan kode pemrograman *Python* sampai kode berhasil dijalankan dengan tanda bercentang hijau.

2.8 Penelitian Terdahulu

No.	Judul	Nama Penulis	Tahun	Hasil
1.	Penerapan <i>Data Mining</i> Metode <i>K-Means Clustering</i> Untuk Analisa Penjualan Pada Toko <i>Yana Sport</i>	Agung Nugraha, Odi Nurdianan, Grifthera Dwilestari [28].	2022	Hasil rekomendasi penelitian ini mendapatkan informasi atau pola dari penerapan algoritma <i>k-means</i> dengan data penjualan terdapat sebanyak 99 item barang yang laris terjual dan terdapat 23 item barang yang tidak terjual sehingga pemilik dapat melakukan strategi penjualan dan pembelian ulang berdasarkan barang yang laris terjual.
2.	Rekomendasi <i>Restock</i> Barang di Toko Pojok UMKM Menggunakan Algoritma <i>K-Means Clustering</i>	Ayuni Salsabiela, Adam Prayogo Kuncoro, Pungkas Subarkah, Primandani Arsi [29].	2024	Penggunaan metode <i>K-Means</i> mendapatkan hasil <i>clustering</i> terbaik yaitu dengan membaginya kedalam 3 cluster dimana nilai $k = 3$ ditentukan dari nilai <i>Davies Bouldin Index</i> masing-masing iterasi awal dan didapat nilai <i>Davies Bouldin Index</i> dari $k = 3$ sebesar 0,436. Hasil dari <i>cluster</i> model didapat <i>cluster</i> 0 dengan 110 data merupakan <i>cluster</i> 3 di mana produk yang masuk termasuk kedalam produk yang direkomendasikan untuk pembelian dalam jumlah sedikit, sedangkan <i>cluster</i> 2 dengan 19 data merupakan produk yang direkomendasikan untuk pembelian dalam jumlah sedang dan <i>cluster</i> 1 dengan 9 data merupakan produk yang direkomendasikan dibeli dalam jumlah banyak.
3.	Implementasi <i>Data Mining</i> Untuk Klustering Stunting Gizi Pada Balita Dipuskesmas Sigambal Menggunakan Metode <i>K-Medoids</i> Dan <i>K-Means</i>	Melisa, Syaiful Zuhri Harahap, Masrizal [30].	2024	Kesimpulan dari analisis <i>clustering</i> yang dilakukan pada 116 sampel balita menunjukkan bahwa terdapat dua kelompok utama, yaitu <i>Cluster</i> C1 yang terdiri dari 48 balita dan <i>Cluster</i> C2 yang terdiri dari 68 balita.
4.	Algoritma <i>K-Means</i> Untuk Analisis Presepsi Masyarakat Labuhanbatu Dalam Promosi Produk Berbasis Digital Pasca <i>Covid-19</i>	Shabrina Rasyid Munthe, Sudi Suryadi, Muhammad Irwansyah Hasibuan, Ibnu rasyid munthe, Ronal Watrianthos [31].	2021	Masyarakat Labuhanbatu yang berbisnis dengan mempromosikan produk memanfaatkan produk digital dengan 142 pelaku usaha pasca <i>covid-19</i> dimana atribut yang digunakan adalah penggunaan internet, media sosial, dan promosi produk. Sebanyak 47 orang yang memanfaatkan, dan sebanyak 95 pelaku usaha tidak memanfaatkan peluang promosi produk.
5.	<i>Application Of Data Mining In Selecting Superior Products Using The K-Means And K-Medoids Algorithm Methods</i>	Eva Hermika, Syaiful Zuhri Harahap, Irmayanti [32].	2024	Dengan penerapan metode <i>clustering</i> maka permasalahan kekurangan stok pada produk unggulan dapat diatasi, karena perusahaan dapat melihat hasil pengelompokan produk yang permintaannya tinggi, sedang, dan rendah. Bagi yang termasuk dalam pengelompokan produk unggulan, perusahaan akan lebih memperhatikan ketersediaan, sehingga tidak terjadi kekurangan stok. Dan untuk barang yang kurang diminati perusahaan tidak perlu melakukan pengelompokan produk yang permintaannya sedang, sedang, dan rendah.

6.	Analisis Pengelompokan Stok Barang dengan Memanfaatkan Algoritma K-Means	Rifaldo Pratama, Febby Kesumaningtyas, Dhio Saputra [33].	2024	Dari pengolahan data yang digunakan dengan algoritma K-Means, dapat mengklasifikasikan produk yang laris dan yang kurang laris. Kelompok data barang yang laris adalah m3,m5,m8,m15,m22 dan m23 sedangkan kelompok data barang yang kurang laris adalah m1, m2, m4, m6, m7, m9, m10, m11, m12, m13, m14, m16, m17, m18, m19, m20, m21, m24, m25, m26, m27, m28, m29 dan m30.
7.	Implementasi Data Mining Untuk Menentukan Persediaan Stok Barang Di Mini Market Menggunakan Metode K-Means Clustering	Hani Prastiwi, Jeny Pricilis, Errissya Raswir [34].	2022	Pada hasil pengujian dengan 20 data, cluster optimal menyumbangkan 17 data untuk cluster C1 dan 3 data untuk cluster C2.
8.	Penerapan Algoritma K-Means untuk Mengetahui Pola Persediaan Barang pada Toko Raja Bekasi	Intan Safira, Ratna Salkiawati, Wowon Priatna [35].	2022	Perhitungan akurasi dalam penelitian ini menggunakan Hasil pengujian metode clustering K-Means menggunakan DaviesBouldin Index (DBI) adalah sebesar 1,856 dimana nilai DBI mendekati nol cluster sudah cukup baik. Dari akurasi yang dihasilkan dapat disimpulkan bahwa metode Clustering K-Means dapat mendukung sistem dengan baik.
9.	Implementasi Data Mining Pada Stok Penggunaan Barang di GMF AeroAsia Menggunakan Algoritma K-Means Clustering	Tri Bayu Pamungkas, Siti Maesaroh, Pebri Ardiansyah [36].	2023	Hasil analisis dan penelitian terhadap data penggunaan barang di GMF AeroAsia menggunakan algoritma k-means menunjukkan bahwa data berhasil dikelompokkan menjadi tiga klaster. Klaster pertama terdiri dari 13 data barang dan menggambarkan kelompok barang dengan tingkat kedaluwarsa yang rendah. Klaster kedua terdiri dari 5 data barang dan mencerminkan kelompok barang dengan tingkat kedaluwarsa yang sedang. Klaster ketiga terdiri dari 2 data barang dan mewakili kelompok barang dengan tingkat kedaluwarsa yang tinggi.
10.	Pengelompokan Data Barang Dengan Menggunakan Metode K-Means Untuk Menentukan Stok Persediaan Barang	Ika Anikah, Agus Surip, Nela Puji Rahayu, Muhammad Harun Al-Musa, Edi Tohidi [37].	2020	Hasil dari penelitian ini yaitu setiap cluster yaitu cluster_0 sebanyak 1282 atau sebesar 98,6%, cluster_1 sebanyak 18 atau sebesar 1,4%. Berdasarkan hal tersebut didapat kelompok stok barang yang berada pada cluster_0 memiliki rata-rata penjualan yang besar yaitu mencapai 2368 barang. Artinya bahwa kelompok barang pada cluster_0 perlu menyimpan banyak barang karena merupakan produk yang paling banyak diminati.

11.	Optimalisasi Stok Barang Melalui Algoritma K-Means Clustering Analisis untuk Manajemen Persediaan Dalam Konteks Bisnis Modern	Risma'ruf Samsudin, Martanto, Umi Hayati [38].	2024	Hasil analisis menunjukkan bahwa algoritma clustering berhasil mengelompokkan data dengan kualitas yang baik distribusi data yang terpusat dengan baik dan nilai Davies Bouldin yang rendah menunjukkan bahwa klaster-klaster yang dihasilkan valid dan dapat digunakan untuk pedoman manajemen dalam mengambil keputusan penyediaan stock barang produk.
12.	Klasterisasi Stok Produk Retail untuk Menentukan Pergerakan Kebutuhan Konsumen dengan Algoritma K-Means	Niko Suwaryo, Arif Rahman, Dewi Mariani Umi Atmaja, Amat Basri [39].	2023	Metode klasterisasi tersebut di proses dengan algoritma K-Means yang dimana hasilnya juga menunjukkan sebuah wawasan baru yaitu mengelompokkan produk berdasarkan 3 cluster. Cluster 1 merupakan kategori produk dengan ketersediaan rendah atau Low yaitu 939 dari 1000 kategori ketersediaan berdasarkan jumlah produk yang diuji, kemudian kategori ketersediaan berdasarkan jumlah produk yang diuji, kemudian cluster 2 adalah kategori produk dengan ketersediaan sedang atau Medium yaitu 51 dari 1000 kategori ketersediaan berdasarkan jumlah produk yang diuji, dan terakhir adalah cluster 3 merupakan kategori produk dengan ketersediaan cukup tinggi atau High yaitu 10 dari 100 kategori ketersediaan berdasarkan jumlah produk yang diuji.
13.	Penerapan Algoritma K-means Clustering untuk Pengelompokan Data Pasien Rehabilitasi Narkoba	Ega Yolanda, Suhardi [40].	2023	Hasil dari penelitian ini adalah penggunaan algoritma K-Means clustering dalam pengelompokan data pasien rehabilitasi narkoba berhasil menghasilkan tiga kelompok (cluster) berdasarkan karakteristik mereka. Cluster pertama merupakan kelompok umur 26-45 tahun dengan lama penggunaan 1->5 bulan dan dengan jenis narkoba yang digunakan adalah sabu, ganja, inex, dan extacy. Cluster kedua adalah kelompok umur 55 tahun dengan lama penggunaan 3->5 bulan dan dengan jenis narkoba yang digunakan adalah sabu, ganja, dan extacy.
14.	Pengelompokan Data Penyakit THT Menggunakan Algoritma K-Means Clustering	Minta Ito Hutagalung, Sriani [41].	2024	Hasil penelitian menunjukkan bahwa pasien dapat dibagi menjadi tiga kelompok (3 cluster) berdasarkan usia dan jenis penyakit, dengan cluster 0 terdiri dari pasien usia lanjut yang lebih rentan terhadap penyakit seperti sinusitis dengan total 10 data, kemudian cluster 1 terdiri dari pasien yang lebih muda dengan diagnosis penyakit yang lebih ringan dengan total 19 data. Sedangkan cluster 2 mencakup pasien dengan rentang umur yang bervariasi dan diagnosis yang lebih beragam, seperti OMSK dan Rhinitis Alergi dengan total 22 data.
15.	Analisa Tingkat Penjualan Makanan dan Minuman dengan Klasterisasi menggunakan Algoritma K-Means	Era Serly Ana, Syaiful Anwar, Mohammad Haddiel Fuad [42].	2024	Hasil dari penelitian yang dilakukan ini mendapatkan 3 cluster yang terdiri dari cluster 1(C0) tingkat penjualan tinggi sebanyak 34 menu salah satunya Kakigori 2, cluster 2(C1) tingkat penjualan sedang sebanyak 14 menu salah satunya Strawberry Smoothies, dan cluster 3(C2) tingkat penjualan rendah sebanyak 25 menu salah satunya Latte Shake. Sehingga hasil pengelompokkan ini memungkinkan restoran untuk membuat keputusan yang lebih akurat mengenai strategi penjualan.

2.9 Kelebihan Penelitian

Penelitian ini memiliki keunggulan dalam konteks praktis dan teknis dibandingkan penelitian terdahulu. "Implementasi *K-Means Clustering* menggunakan *Machine Learning* dengan alat bantu dan *tools* yang digunakan seperti *Python* dan *Google Colab* memungkinkan analisis data skala besar dengan akurasi tinggi, relevan dengan kebutuhan manajemen stok yang lebih efisien." Penelitian ini juga menawarkan integrasi hasil *clustering* dengan strategi bisnis yang aplikatif, sesuatu yang sering tidak dibahas dalam penelitian sebelumnya.