BAB II
LANDASAN TEORI

2.1. *Data mining*

Meningkatnya volume data yang tersimpan dalam basis data di berbagai industri, termasuk bisnis, telah menjadikan penambangan data semakin krusial. Menjelajahi data dalam basis data dan menghasilkan informasi baru yang bermanfaat merupakan tujuan dari proses penambangan data, yang juga disebut sebagai KDD (*knowladge discovery in database*). Selain itu, karena penambangan data bertujuan untuk menemukan pola yang sudah ada dalam kumpulan data, proses ini terkadang disebut sebagai pengenalan pola. Oleh karena itu, proses menganalisis dan menemukan informasi berharga dengan menerapkan metode matematika, statistika, kecerdasan buatan, dan pembelajaran mesin pada kumpulan data yang sangat besar dapat digolongkan sebagai penambangan data.(Muhamad Rizki, 2023)**.**

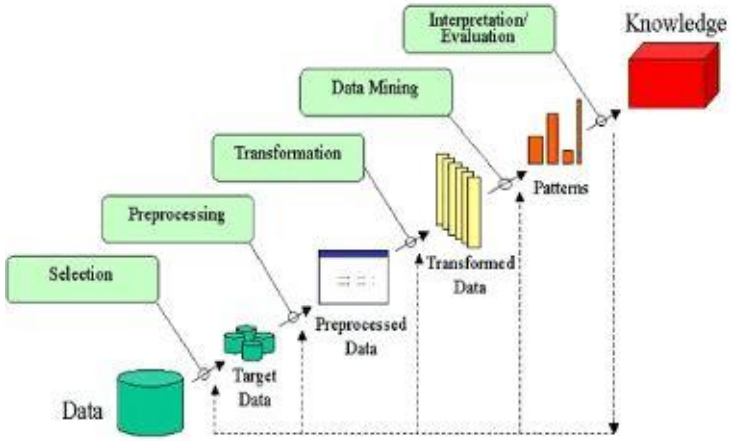
Menurut (Rafi Nahjan et al., 2023) Menemukan informasi berharga dan mengubahnya ke dalam format yang mudah dipahami merupakan tujuan penambangan data. Untuk mengidentifikasi pola dalam data yang dapat dimanfaatkan untuk meningkatkan pengambilan keputusan, penambangan data digunakan dalam berbagai domain, termasuk bisnis, sains, teknologi informasi, dan lainnya.

2.2.1. Tahap Peroses *Data mining*

Knowledge Discovery in Database (KDD)

KDD adalah proses yang tidak sepele yang digunakan untuk mengidentifikasi

validitas data, potensi, guna, dan pada akhirnya menghasilkan pola data yang dapat dimengerti (Haris Kurniawan et al., 2020) . KDD berhubungan dengan teknik integrasi dan penemuan ilmiah, interprestasi dan visualisasi dari pola-pola sejumlah kumpulan data. *Data mining* sendiri adalah bagian dari tahapan proses KDD yang diilustrasikan seperti pada gambar di bawah ini :



Gambar 2.1 Peroses KDD

Sumber : https://journal.isas.or.id/index.php/JACOST/article/view/102/33

1. Data *Selection*

   Proses memilih sekumpulan data target, kualitas, atau indikator dari kumpulan data besar dikenal sebagai pemilihan data.

2. *Pre-Processing/Cleaning*

   Proses memperbaiki masalah data, menghilangkan data duplikat, dan mencari anomali seperti kesalahan ketik dikenal sebagai pembersihan data.

3. Data Tranformasi

   Proses mengubah atau mengodekan data dikenal sebagai transformasi data. Untuk mempermudah pemrosesan data dan menyiapkan data untuk penambangan data, transformasi data dilakukan.

4. Data Mining

   Penambangan data adalah praktik penerapan alat atau pendekatan tertentu untuk mengidentifikasi informasi atau pengelompokan yang menarik/berguna dalam kumpulan data besar.

5. Interpretation atau Evalution

   Interpretation atau Evalution merupakan proses menampilkan pola/informasi/pengetahuan dari hasil prosesdata miningke dalam bentuk yang mudah dimengerti oleh pihak-pihak membutuhkan.

Tabel 2. 1 Tahapan *Data mining* dan Deskripsinya

Tahapan	**Deskripsi**
Pembersihan Data	Menghapus data yang tidak relevan, duplikasi, atau hilang.
Integrasi Data	Menggabungkan data dari berbagai sumber untuk membentuk dataset yang konsisten.
Transformasi Data	Mengubah format data agar sesuai untuk proses mining.
Data mining	Menerapkan algoritma untuk menemukan pola atau tren dalam data.
Evaluasi Pola	Menilai relevansi dan validitas pola yang ditemukan dalam proses mining.

2.2.2. Penerapan *Data mining* Dalam Bisnis *Riteil*

Menurut Nisa (2021) menegaskan bahwa pengambilan keputusan bisnis dapat diuntungkan dari penggunaan data transaksi penjualan. Data penjualan terus bertambah sebagai hasil dari aktivitas penjualan harian. Sebagian besar data transaksi penjualan tidak pernah digunakan lagi atau hanya disimpan dalam arsip dan digunakan untuk laporan penjualan. Salah satu ilmu yang dapat digunakan untuk memecahkan masalah seperti ini adalah penambangan data. Teknik penambangan data dapat digunakan untuk mengekstrak dan memproses ulang transaksi penjualan yang kurang dimanfaatkan menjadi informasi yang berharga. Algoritma Apriori adalah salah satu teknik penambangan data yang dapat digunakan untuk memproses ulang data transaksi penjualan dan menghasilkan pengelompokan pembelian pelanggan. Pada akhirnya, kompilasi pembelian pelanggan ini akan membantu pemilik bisnis dalam membuat keputusan.

2.2. *Clustering*

Proses pengelompokan melibatkan penempatan rekaman, observasi, atau objek dengan atribut serupa (A. Nugraha dkk., 2022). Berbeda dengan klasifikasi, pengelompokan tidak melibatkan variabel target. Nilai target tidak dapat diprediksi, diestimasi, atau diklasifikasikan menggunakan pengelompokan. Pengelompokan digunakan untuk memisahkan seluruh kumpulan data ke dalam kelompok-kelompok terkait.

Pengelompokan partisi merupakan teknik penambangan data tanpa pengawasan, Menurut Aulia (2021). Membagi n klaster menjadi k klaster merupakan ide dasar pengelompokan partisi. Dengan mengelompokkan objek, teknik ini berusaha memperkecil jarak antara setiap objek dan pusat klaster.

2.2.1. Aplikasi *Clustering* Dalam Analisis Penjulan Pulsa

Clustering merupakan salah satu metode dalam *data mining* yang digunakan untuk mengelompokkan data berdasarkan kemiripan tertentu. Dalam konteks analisis penjualan pulsa, *clustering* memiliki peran penting dalam membantu pemilik usaha untuk memahami pola dan karakteristik data penjualan yang kompleks (Tiara Alifa et al., 2024). Berikut adalah beberapa aplikasi *clustering* yang relevan:

1. Segmentasi Pelanggan: Mengelompokkan pelanggan berdasarkan pola pembelian pulsa, seperti frekuensi dan nominal transaksi, memungkinkan penjual untuk menargetkan promosi yang sesuai dengan setiap segmen pelanggan.(H. S. Nugraha et al., 2023)
2. Analisis Pengelompokan Penjualan: Dengan *clustering*, penjual dapat mengidentifikasi produk atau layanan yang paling diminati, sehingga dapat mengoptimalkan stok dan penawaran.(Wulandari & Farida, 2022)
3. Identifikasi Produk Laris: Dengan mengelompokkan produk berdasarkan tingkat penjualan, penjual dapat fokus pada produk yang paling diminati dan mengurangi stok produk yang kurang laku(Nurdiyansyah et al., 2018)

2.2.2. Tantangan Dalam Penerapan *Clustering*

Penerapan *clustering* dalam analisis penjualan pulsa menghadapi beberapa tantangan yang perlu diperhatikan untuk menghasilkan analisis yang akurat dan bermanfaat. Berikut adalah beberapa tantangan utama:

1. Penentuan Jumlah *Cluster*

   Menentukan jumlah *cluster* yang tepat sangat penting untuk menghindari overfitting atau *underfitting* dalam model *clustering*. Pendekatan seperti

metode Elbow dan *Silhouette Score* sering digunakan untuk menentukan jumlah *cluster* yang optimal.(Nurdiyansyah et al., 2018)

2. Kualitas dan *Pra-pemerosesan Data*

   Data penjualan pulsa sering kali tidak lengkap, redundan, atau mengandung noise. Langkah *pra-pemrosesan* seperti normalisasi, penanganan data hilang, dan penghapusan duplikasi diperlukan untuk memastikan hasil *clustering* yang akurat. Proses ini membutuhkan waktu dan keahlian yang memadai.(Rafi Nahjan et al., 2023)

3. Sensivitas terhadap *Outlier*

   Outlier dalam data dapat secara signifikan memengaruhi hasil *clustering*, terutama pada algoritma seperti *K-means* yang sangat sensitif terhadap data pencilan. Kehadiran outlier dapat menarik pusat *cluster (centroid)* ke arah yang salah, sehingga menghasilkan pengelompokan yang tidak akurat dan tidak relevan secara bisnis.(Muhamad Rizki, 2023)

2.2.3. Evaluasi Hasil *Clustering*

1. *Evaluasi Berbasis Silhoutte Score*

   Silhouette Score mengukur seberapa baik data dalam *cluster* tertentu dikelompokkan. Skor ini berkisar antara -1 hingga 1, dengan nilai lebih tinggi menunjukkan *clustering* yang lebih baik.(Dbscan & Hasan, 2024)

2. *Evaluasi Dengan Davies-Bouldin Index*

   Davies-Bouldin Index mengevaluasi jarak antar *cluster* dibandingkan dengan variasi dalam *cluster*. Nilai yang lebih kecil menunjukkan *clustering* yang lebih baik.(Dbscan & Hasan, 2024)

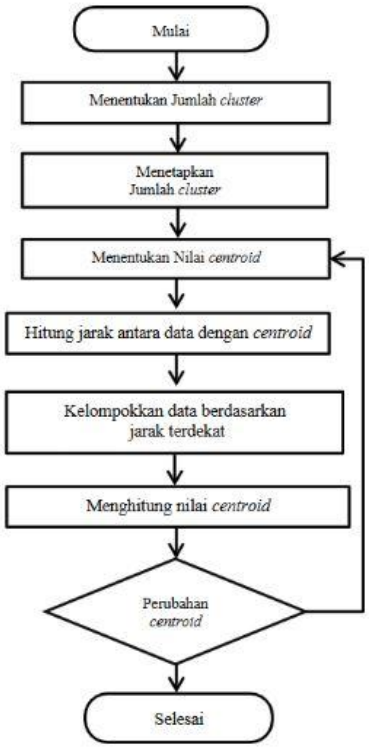
3. *Calinski-Harabasz Index*

Calinski-Harbasz Index Mengukur rasio varians antar *cluster* terhadap varians dalam cluster. Nilai yang lebih tinggi menunjukkan *cluster* yang lebih baik.(Harani et al., 2020)

2.3. Algoritma *K-means*

Algoritma K-means merupakan algoritma yang paling populer dan paling banyak digunakan, menurut Putri Adilah Asih dkk., 2024. Algoritma ini didasarkan pada konsep yang sederhana. Pertama, tentukan berapa banyak klaster yang perlu dibentuk. Pusat klaster, atau titik kontrol, dapat berupa objek apa pun atau elemen awal klaster. Langkah-langkah ini diulang oleh algoritma K-means hingga mencapai stabilitas (hingga objek tidak dapat lagi bergerak). Jenis produk untuk konsumsi bulanan dibuat menggunakan hasil pengelompokan data mining, dan ini dapat menjadi panduan untuk perencanaan inventaris tahun berikutnya.

2.3.1 Langkah-Langkah Melakukan *K-means Clustering*



Gambar 2.2 Flowchart Algoritma *K-means*

Menurut (Purba et al., 2019) Berikut langkah-langkah yang terlibat dalam pengelompokan menggunakan metode K-means:

1. Dengan menggunakan K sebagai jumlah klaster, tentukan jumlah klaster yang ingin Anda buat.
2. Identifikasi sentroid awal, atau pusat klaster. Jumlah *centroid* awal sama dengan jumlah klaster, dan dipilih secara acak dari data yang tersedia.
3. Setelah menentukan *centroid* awal, maka setiap data akan menemukan centroid terdekatnya yaitu dengan menghitung jarak setiap data ke masing-masing *centroid* menggunakan rumus korelasi antar dua obyek yaitu *Euclidean Distance.*

4. Pengelompokan data berdasarkan jarak minimumnya merupakan langkah selanjutnya setelah menentukan jarak antara data dan pusatnya. Anggota klaster dengan jarak terpendek dari pusat klasternya akan dianggap sebagai item data.
5. Setelah pengelompokan ini, rata-rata data di setiap klaster dihitung untuk menentukan pusat baru berdasarkan keanggotaan di setiap klaster.
6. Kembali ke langkah 3.
7. Ketika tidak ada data yang dipindahkan, iterasi berakhir.

2.3.2 Keunggulan dan Kelemahan Algoritma *K-means*

Penggunaan Algoritma *K-means* memiliki beberapa keunggulan dan kelemahan yaitu.

Keunggulan :

1. Sederhana dan Mudah Dipahami

   Algoritma *K-means* mudah diterapkan dan dipahami, sehingga sangat populer untuk berbagai aplikasi *clustering*, termasuk dalam analisis penjualan pulsa.

2. Efisien untuk Dataset Besar

   K-means cukup efisien dalam menangani dataset besar, menjadikannya ideal untuk pengelompokan data penjualan pulsa dalam jumlah besar.

3. Fleksibilitas

   Algoritma ini dapat diterapkan untuk berbagai jenis data dan memberikan hasil yang memadai dalam berbagai konteks, seperti analisis pengelompokan pembelian pulsa.

4. Skalabilitas yang Baik

   K-means dapat mengelola data yang lebih besar atau lebih kompleks dengan baik, sehingga cocok untuk transaksi penjualan pulsa dalam skala besar.

Kelemahan:

1. Ketergantungan pada Pemilihan Jumlah *Cluster*

   Pemilihan jumlah *cluster* (K) yang tepat sangat penting. Pemilihan yang tidak tepat dapat menghasilkan hasil yang tidak akurat atau tidak berguna dalam analisis.

2. Sensitif terhadap *Outliers*

   Outliers dapat mempengaruhi hasil *clustering*, karena *K-means* sensitif terhadap posisi centroid yang dipengaruhi oleh data yang jauh dari pusat *cluster.*

3. Bergantung pada Initial *Centroid*

   Hasil *clustering* sangat dipengaruhi oleh pemilihan *centroid* awal. Pemilihan yang buruk dapat menyebabkan hasil yang kurang optimal atau bahkan salah interpretasi.

4. *Cluster* dengan Bentuk *Non-Sferis* Tidak Teridentifikasi dengan Baik

   K-means lebih efektif dalam mengelompokkan data dengan bentuk *cluster sferis* dan kurang efektif untuk *cluster* dengan bentuk yang lebih kompleks.

2.3.4. Aplikasi Algoritma *K-means* Dalam Analisis Penjualan Pulsa

Algoritma *K-means* memiliki banyak aplikasi dalam analisis penjualan pulsa, di mana data penjualan dapat dikelompokkan berdasarkan pembelian dan

karakteristik tertentu. Berikut adalah beberapa aplikasi utama algoritma *K-means* dalam konteks ini:

1. Segmentasi Pelanggan

   Dengan menggunakan *K-means*, data pelanggan dapat dikelompokkan berdasarkan kebiasaan atau preferensi mereka dalam membeli pulsa. Misalnya, pelanggan dapat dikelompokkan berdasarkan frekuensi pembelian, nominal pembelian, atau waktu pembelian.(Savitri et al., 2018)

2. Prediksi Penjualan Pulsa

   K-means dapat digunakan untuk mengelompokkan transaksi penjualan pulsa yang serupa dan menganalisis pwmbelian transaksi dari waktu ke waktu. Misalnya, data transaksi bulanan dapat dikelompokkan untuk menemukan pengelompokan penjualan yang berulang.(Butsianto & Mayangwulan, 2020)

3. Optimasi Stok Pulsa

   Dalam bisnis penjualan pulsa, *K-means* dapat digunakan untuk mengelompokkan jenis pulsa yang paling banyak terjual berdasarkan lokasi, waktu, atau tipe pelanggan. Hal ini dapat membantu pengelola untuk menentukan stok pulsa yang perlu disiapkan di masing-masing konter atau lokasi.(Agung Yuliyanto Nugroho, 2022)

2.3.5. Tantangan Dalam Penerapan Algotitma *K-means*

Berikut adalah tantangan utama dalam penerapan algoritma *K-Means***,** khususnya dalam konteks *riset*, bisnis, dan aplikasi nyata seperti analisis penjualan pulsa atau segmentasi pelanggan:

1. Penentuan Jumlah *Cluster* (K)

   Dalam analisis penjualan produk digital, menentukan jumlah *cluster* (K) yang tepat sangat penting untuk menghasilkan segmentasi pelanggan atau pola pembelian yang relevan. Jika jumlah *cluster* yang dipilih terlalu sedikit, pola data yang signifikan dapat terlewatkan. Sebaliknya, jika terlalu banyak, interpretasi hasil menjadi sulit dan tidak praktis. Oleh karena itu, diperlukan pendekatan tambahan seperti *Elbow Method* atau *Silhouette Score* untuk menentukan nilai K yang optimal.(Guntara & Lutfi, 2023).

2. Sensitivitas terhadap *Outlier* pada Data Penjualan

   Data transaksi penjualan sering kali mengandung *outlier*, seperti pembelian dalam jumlah besar yang tidak biasa atau pelanggan dengan pola transaksi yang tidak konsisten. *Outlier* ini dapat memengaruhi posisi centroid secara signifikan, sehingga hasil *clustering* menjadi kurang representatif. *Preprocessing* data untuk mendeteksi dan menangani *outlier* menjadi langkah penting sebelum algoritma *K-means* diterapkan.(Kamila et al., 2019).

3. Ketergantungan pada Skala Data Penjualan

   Dalam data penjualan, fitur seperti jumlah transaksi, nominal pembelian, dan frekuensi pembelian dapat memiliki skala yang berbeda. Misalnya, frekuensi pembelian dalam satuan "kali" mungkin jauh lebih kecil

dibandingkan total nominal pembelian dalam satuan "rupiah." Ketidakseimbangan ini dapat menyebabkan bias pada hasil *clustering*. Oleh karena itu, normalisasi atau standarisasi data perlu dilakukan untuk memastikan setiap fitur memiliki kontribusi yang setara dalam proses *clustering.*

2.4. Penjualan Produk *Digital* (Pulsa)

Pulsa adalah saldo prabayar yang digunakan untuk mengakses layanan telekomunikasi, seperti panggilan telepon, SMS (*Short Message Service*), dan penggunaan data internet. Pulsa ini berfungsi sebagai metode pembayaran bagi pengguna layanan prabayar, yang perlu membeli dan mengisi ulang saldo pulsa sebelum mereka dapat menggunakan layanan telekomunikasi tersebut.

Dalam konteks yang lebih luas, pulsa tidak hanya digunakan untuk komunikasi telepon dan pesan, tetapi juga dapat dipakai untuk berbagai layanan digital lainnya, seperti pembelian paket data internet, layanan premium, pembayaran tagihan, pembelian token listrik, dan bahkan transaksi digital lainnya. Dengan berkembangnya teknologi, pulsa juga mulai digunakan sebagai metode pembayaran untuk aplikasi atau game digital.

2.4.1. Krakteristik Produk Pulsa

Karakteristik Kategori (Dini Anggraini, 2019) menjelaskan bahwa berikut ini adalah karakteristik kategori produk: keterlibatan, perbedaan persepsi merek, fitur hedonis, kekuatan preferensi, frekuensi pembelian, dan urutan pilihan merek. Sangat penting untuk mempertimbangkan fitur produk saat ini saat mengkaji hubungan antara konsumen dan produk. Telah terbukti bahwa beberapa karakteristik ini memengaruhi kesuksesan merek atau produk. Meskipun

batasannya tidak dapat sepenuhnya dipisahkan, beberapa faktor ini secara langsung berkontribusi pada keinginan untuk mencoba, sementara yang lain dapat meningkatkan loyalitas merek dan mendorong keinginan untuk mencoba.

2.4.2. Tantangan Dalam Penjualan Pulsa

Penjualan pulsa di sektor ritel menghadapi beberapa tantangan berikut:

1. Persaingan ketat : Penyedia layanan digital seperti platform media sosial dan aplikasi pesan instan, yang memungkinkan komunikasi tanpa menggunakan biaya kredit tradisional, merupakan pesaing selain penyedia layanan telekomunikasi lainnya.(Khadijah, 2024).
2. Permintaan yang Tidak Stabil: Preferensi pelanggan dapat berubah dengan cepat, terutama saat ada promosi atau kebijakan baru dari operator telekomunikasi.
3. Manajemen Stok : Banyak permasalahan yang sering terjadi bilamana admin yang tidak teliti akan mengalami kesalahan pencatatan sehingga laporan yang diberikan tidak sesuai dengan stok barang awal dan penjualan, sehingga menyulitkan admindan karyawan untuk dapat mencari kesalahan catat maupun kekeliruan yang dihadapinya.(Masgo & Santoso, 2022).

2.4.3. Bisnis Penjualan Pulsa

Menurut (Khadijah, 2024) Pentingnya diversifikasi layanan bagi konter ponsel di era digital tercermin dalam industri pulsa ponsel. Margin keuntungan yang rendah dan persaingan harga yang ketat merupakan masalah umum bagi konter ponsel yang hanya menjual kartu kredit. Oleh karena itu, menciptakan layanan baru seperti transfer uang, pembayaran tagihan, isi ulang e-wallet, dan

layanan digital lainnya dapat menjadi sangat menguntungkan. Peningkatan aliran pendapatan merupakan salah satu keuntungan utama dari diversifikasi layanan. Dengan mengenakan biaya transaksi untuk layanan tambahan yang mereka berikan kepada pelanggan, konter ponsel dapat menghasilkan lebih banyak keuntungan. Hal ini meningkatkan profitabilitas perusahaan dan mengurangi ketergantungan pada penjualan pulsa ponsel.

Dalam konteks bisnis penjualan pulsa, terdapat beberapa model distribusi yang umum, yaitu:

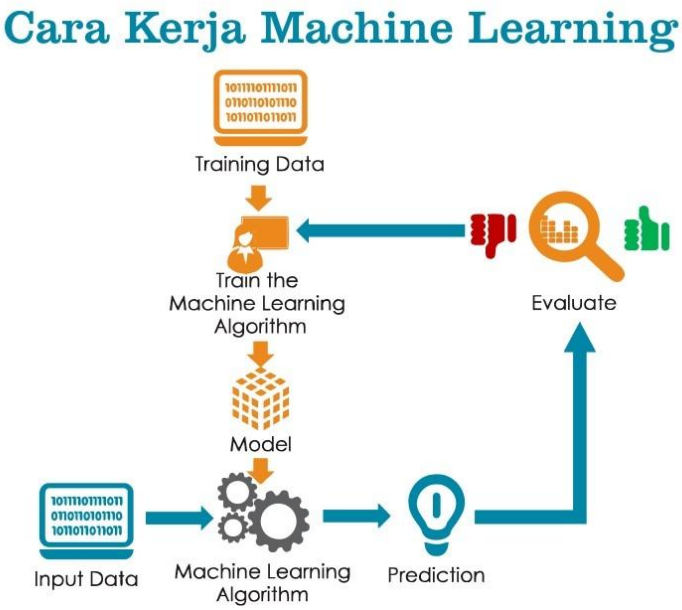
1. Konter Pulsa: Usaha fisik di mana penjualan pulsa dilakukan secara langsung kepada konsumen. Model ini umum di daerah-daerah perkotaan dan pedesaan, dengan keuntungan langsung dari setiap penjualan.
2. Agen Pulsa: Seseorang yang menjadi perantara antara distributor pulsa dan konsumen akhir, di mana mereka menjual pulsa melalui perangkat elektronik atau aplikasi.
3. Penjualan Pulsa Online: Penjualan melalui platform digital seperti aplikasi seluler, *marketplace*, atau situs web. Model ini memberikan kemudahan bagi konsumen yang ingin melakukan transaksi tanpa harus datang ke konter fisik.

Bisnis penjualan pulsa juga dikenal dengan margin keuntungan yang tipis per transaksi, sehingga volume penjualan yang tinggi sangat diperlukan untuk mencapai keuntungan yang signifikan. Namun, dengan tingginya permintaan akan layanan komunikasi dan internet di Indonesia, bisnis ini masih dianggap menjanjikan.

2.5. *Machine Learning*

Metode kecerdasan buatan (AI) yang disebut pembelajaran mesin digunakan untuk mensimulasikan dan menggantikan kemampuan pemecahan masalah manusia. Singkatnya, pembelajaran mesin adalah kemampuan mesin untuk mempelajari dan menjalankan tugas tanpa bantuan manusia.(Wijoyo A et al., 2024).

Dalam konteks ML penelitian ini penjualan produk *digital* seperti pulsa, metode ini dapat dimanfaatkan untuk mengelompokkan pelanggan atau produk berdasarkan karakteristik tertentu, seperti frekuensi pembelian, nominal transaksi, atau periode waktu transaksi.



Gambar 2.3 Diagram Alur kerja *Machine Learning*

2.6. *Evaluasi Model Clustering*

Metrik evaluasi untuk algoritma *clustering*, seperti *K-means*, membantu menentukan kualitas pengelompokan yang dihasilkan. Beberapa metrik yang umum digunakan adalah:

1. *Silhoutte Score*

   Silhouette Score mengevaluasi seberapa baik objek dalam *cluster* dibandingkan dengan objek di *cluster* lain. Nilai mendekati 1 menunjukkan bahwa objek lebih dekat dengan *cluster* mereka sendiri dibandingkan dengan *cluster* lainnya.(Daffa Rachman & Voutama, 2024)

Rumus perhitungan *Silhoutte Score* :

$$s(i) = (b(i) - a(i)) / \max(a(i), b(i))$$

A. (i): rata-rata jarak objek iii ke semua objek lain dalam cluster yang sama.

B. (i): rata-rata jarak objek iii ke semua objek dalam cluster terdekat.

2. *Davies-Bouldin Index*

   Davies-Bouldin index (DBI) digunakan untuk menilai keluaran model pengelompokan. Nilai kohesi dan pemisahan digunakan oleh DBI untuk mengkuantifikasi hasil pengelompokan. Hasil pengelompokan pemisahan didasarkan pada jarak antar sentroid klaster, sedangkan hasil pengelompokan kohesi adalah jumlah kedekatan data dengan titik pusat (sentroid) hasil pengelompokan.(Sholeh & Aeni, 2023).

   Rumus perhitungan *Davies-Bouldin Index :*

$$DBI = (1 / k) * \Sigma (i=1 \text{ to } k) \max (j \neq i) ((s\_i + s\_j) / d(C\_i, C\_j))$$

A. Si Adalah rata-rata jarak antara titik data dalam cluster Ci dan pusat cluster Ci.

B. d(Ci,Cj) adalah jarak antara pusat cluster Ci dan Cj.

C. k adalah jumlah *cluster*

3. *Calinski-Harabasz Index*

   Calinski-Harabasz Index (CHI) adalah metrik evaluasi yang mengukur rasio antara varians antar cluster dan varians dalam *cluster.* Indeks ini sering digunakan untuk menilai separasi cluster dan konsistensi data dalam setiap

cluster. Nilai CHI yang lebih tinggi menunjukkan bahwa *cluster* memiliki tingkat kepadatan yang tinggi dan separasi yang baik(Harani et al., 2020).

Rumus perhitungan *Calinski-Harabasz Index* :

$$CHI = (SSB / SSW) * ((n - k) / (k - 1))$$

a. SSB: jumlah kuadrat antar *cluster.*

b. SSW: jumlah kuadrat dalam *cluster.*

c. n: jumlah total data.

d. k: jumlah *cluster*.

2.6.1. Perbandingan Matrik Evaluasi *Clustering*

Setiap metrik evaluasi *clustering* memiliki karakteristik dan tujuan yang berbeda, tergantung pada kebutuhan analisis. Tabel berikut menyajikan perbandingan antara metrik *silhouette score, Davies-Bouldin Index,* dan *Calinski-Harabasz Index.*

Tabel 2. 2 Perbandingan Matrik Evaluasi *Clustering*

Matrik	Tujuan	Nilai yang Diinginkan
Silhouette Score	Mengukur seberapa mirip objek dalam *cluster* dibandingkan dengan objek di *cluster* lain.	Nilai mendekati 1 menunjukkan *clustering* baik.
Davies-Bouldin Index (DBI)	Mengevaluasi seberapa mirip *cluster* satu sama lain (rasio antara dispersi dalam *cluster* dan jarak antar *cluster*).	Nilai lebih kecil menunjukkan *clustering* lebih baik.

Matrik	Tujuan	Nilai yang Diinginkan
Calinski-Harabasz Index	Mengukur rasio antara dispersi antar *cluster* dengan dispersi dalam-*cluster.*	Nilai lebih tinggi menunjukkan *clustering* lebih baik.

2.7. Kerangka Teoritis

Kerangka teoritis merupakan landasan penting dalam penelitian untuk memahami konsep dan teori yang mendasari implementasi algoritma dalam pengelolaan data. Dalam konteks penelitian ini, algoritma *K-means* digunakan untuk menganalisis produk pulsa berdasarkan pengelompokan data. Algoritma ini berfungsi untuk mengidentifikasi pola dalam data inventaris yang besar dan kompleks, sehingga dapat mendukung pengambilan keputusan yang lebih akurat.

Menurut (Nawangsih et al., 2021) *Data mining* menggunakan metode *k-means* dapat membantu pengetahuan dan informasi mengenai klasterisasi pada penjualan produk *digital* (pulsa) dan masalah diatas penelitian ini penulis akan menggunakan metode *k-means*, maka sehubungan dengan hal tersebut.

2.8. Penelitian Terdahulu

penelitian sebelumnya yang berkaitan dengan *K-means Clustering.*

Tabel 2.3 Penelitian Terdahulu

No	Judul	Penulis	Hasil	Hubungan	Kekurangan
1	Penerapan *Data mining* Pada Penjualan	Nur Astuti, Joy Nashar Utamajaya	Penelitian ini berhasil mengelompokka	Studi ini relevan dengan topik	Penelitian ini tidak menyebutkan

No	Judul	Penulis	Hasil	Hubungan	Kekurangan
	Produk Digital Konter Leppangeng Cell Menggunakan Metode *K-means Clustering*	, Aditya Pratama	n produk digital menjadi tiga kategori: sangat laris (114 produk), kurang laris (5 produk), dan tidak laku (14 produk).	yang diminta karena menerapkan metode *K-means Clustering* untuk menganalisis penjualan produk digital di konter pulsa.	analisis mengenai faktor-faktor yang mempengaruhi penjualan produk dalam setiap cluster.
2	Penerapan *Data mining* Metode *K-means Clustering* Untuk Analisa Penjualan Pada Toko Fashion Hijab Banten	Normah, Siti Nurajizah, Arinda Salbinda	Penelitian ini berhasil mengelompokkan produk fashion hijab menjadi tiga kategori: sangat laris (11 artikel), laris (55 artikel), dan kurang laris (34 artikel).	Meskipun fokus pada produk fashion, metode yang digunakan serupa dan dapat diaplikasikan pada penjualan produk digital seperti pulsa.	Penelitian ini tidak membahas implementasi hasil *clustering* dalam strategi pemasaran atau pengambilan keputusan bisnis.

No	Judul	Penulis	Hasil	Hubungan	Kekurangan
3	Aplikasi *Data mining* Berbasis Web Menggunakan Metode *K-means Clustering* Untuk Pengelompokan Penjualan Terlaris Produk Kacamata	Anggraeni Triningsih, Heru Supriyono	Penelitian ini menghasilkan aplikasi berbasis web yang mampu mengelompokkan penjualan produk kacamata menjadi beberapa cluster untuk membantu analisis penjualan.	Studi ini menunjukkan implementasi metode *K-means* dalam bentuk aplikasi web, yang dapat menjadi referensi untuk pengembangan sistem serupa pada penjualan produk digital.	Penelitian ini tidak menyebutkan uji coba aplikasi pada data penjualan produk digital seperti pulsa, sehingga perlu adaptasi lebih lanjut.
4	Penerapan *Data mining* Metode *K-means Clustering* Untuk Analisa Penjualan Pada	Agung Nugraha, Odi Nurdiawan, Gifthera Dwilestari	Penelitian ini berhasil mengelompokkan produk di Toko Yana Sport menjadi dua kategori: laris terjual (99 item)	Meskipun fokus pada produk olahraga, metode *K-means Clustering* yang	Penelitian ini tidak membahas detail mengenai interpretasi hasil *clustering*

No	Judul	Penulis	Hasil	Hubungan	Kekurangan
	Toko Yana Sport		dan tidak terjual (23 item).	digunakan dapat diterapkan pada analisis penjualan produk digital.	dalam konteks pengambilan keputusan bisnis.
5	Implementasi *Data mining* Untuk Menentukan Tingkat Penjualan Paket Data Telkomsel Menggunakan Metode *K-means Clustering*.	Suhandio Handoko, Fauziah, Endah Tri Esti Handayani	Penelitian ini menghasilkan aplikasi berbasis web yang mampu mengelompokkan data penjualan paket data Telkomsel menjadi tiga kategori: penjualan rendah, sedang, dan tinggi, dengan tingkat akurasi 100% dibandingkan dengan	Studi ini sangat relevan karena fokus pada produk digital serupa (paket data) dan menggunakan metode *K-means Clustering* untuk analisis penjualan.	Penelitian ini tidak membahas implementasi hasil *clustering* dalam strategi pemasaran atau pengembangan produk lebih lanjut.

No	Judul	Penulis	Hasil	Hubungan	Kekurangan
			clustering manual.		

2.9. Kerangka Kerja Penelitian

Kerangka kerja penelitian ini disusun berdasarkan langkah-langkah berikut:

1. Pengumpulan Data

   Langkah pertama dalam kerangka kerja ini adalah pengumpulan data transaksi, menekankan pentingnya data yang lengkap dan relevan untuk memastikan analisis yang akurat.

2. Pengolahan Data

   Setelah data terkumpul, langkah selanjutnya adalah *preprocessing* data. Proses ini mencakup pembersihan data untuk menghilangkan duplikasi atau data yang tidak relevan, *transformasi data* menjadi format yang dapat digunakan oleh algoritma, serta normalisasi untuk memastikan data memiliki skala yang seragam.

3. Penerapan Algoritma *K-means*

   Algoritma *K-means* diterapkan untuk mengidentifikasi pengelompokan yang relevan dalam data transaksi penjualan pulsa, seperti segmentasi pelanggan atau pengelompokan produk berdasarkan penjualan. Proses ini melibatkan penentuan jumlah *cluster* (k), inisialisasi pusat *cluster*, pengelompokan data berdasarkan jarak terdekat, pembaruan pusat cluster, dan iterasi hingga hasil *clustering* stabil.

4. Analisis Hasil *Clustering*

Tahap analisis hasil *clustering* bertujuan untuk memahami dan mengevaluasi pola yang terbentuk dari proses *clustering* menggunakan algoritma *K-means*. Analisis ini membantu mengidentifikasi informasi penting yang dapat digunakan untuk pengambilan keputusan strategis.

5. Penyusunan Rekomendasi Strategis

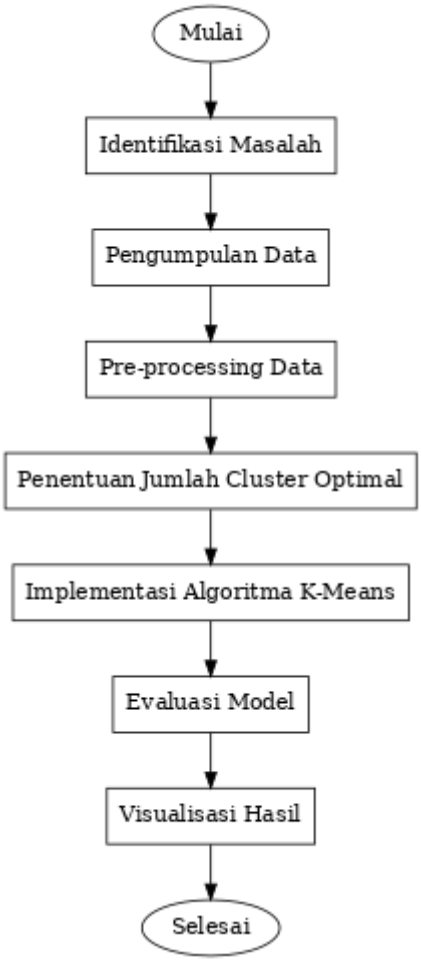
   Tahap penyusunan rekomendasi strategi bertujuan untuk mengonversi hasil klasterisasi menjadi langkah-langkah konkret yang dapat diimplementasikan untuk mendukung pengelolaan bisnis. Berdasarkan pengelompokan yang ditemukan melalui *K-means clustering*, rekomendasi ini disesuaikan dengan kebutuhan operasional dan potensi pasar.

2.10. *Flowchart* Kerangka Kerja Penelitian

Flowchart adalah alat visualisasi yang dirancang untuk menyajikan kerangka kerja penelitian secara sistematis dan terstruktur. Visualisasi ini membantu peneliti dalam memahami alur logika penelitian, dari tahap awal hingga penerapan hasil. Dengan menyederhanakan informasi kompleks, *flowchart* memungkinkan peneliti untuk mengikuti langkah-langkah secara logis, memastikan konsistensi dan efisiensi dalam proses penelitian(Sundary et al., 2020). Selain itu, visualisasi seperti *flowchart* juga efektif dalam menyampaikan konsep kepada audiens non-teknis, seperti manajemen atau pemangku kepentingan lainnya.

2.10.1. Langkah-Langkah dalam *Flowchart* Penelitian

Flowchart dalam penelitian ini menggambarkan langkah-langkah utama yang mencakup proses algoritma *K-means*. Adapun langkah-langkah tersebut meliputi :



Gambar 2.4 *Flowchart* Penelitian

1. Identifikasi Masalah

   Penelitian dimulai dengan mengidentifikasi masalah utama yang akan dikaji. Masalah ini harus dijelaskan secara jelas dan spesifik agar penelitian memiliki fokus yang jelas. Misalnya, pada penelitian tentang penjualan pulsa di Konter Butet CELL, masalah yang dihadapi mungkin berkaitan dengan bagaimana data penjualan dapat dikelompokkan secara efektif untuk mendukung keputusan bisnis.

2. Pengumpulan Data

   Tahap ini melibatkan pengumpulan data yang relevan dengan penelitian. Dalam hal ini, data yang dikumpulkan mencakup transaksi penjualan pulsa,

seperti jumlah transaksi, tanggal, jenis produk yang dibeli, dan perilaku pembelian pelanggan(Syam Al ghifari et al., 2024).

3. *Pre-processing data*

   Pre-processing data merupakan langkah selanjutnya.

   Data akan dibersihkan dari kesalahan, duplikat, dan properti yang tidak berkorelasi. Sebelum langkah ini, atribut dataset yang akan digunakan harus diperiksa untuk melihat apakah ada nilai yang hilang dan apakah ada korelasi antar atribut..(Putri Adilah Asih et al., 2024)

4. Penentuan Jumlah *Cluster* Optimal

   Metode Elbow adalah teknik yang digunakan untuk menentukan jumlah cluster yang optimal dengan cara memeriksa perbandingan hasil varians antara jumlah cluster yang dibentuk(Daffa Rachman & Voutama, 2024).

5. Implentasi Algoritma *K-means*

   Algoritma *K-means* diterapkan untuk mengelompokkan produk berdasarkan karakteristik penjualannya. Langkah ini mencakup inisialisasi *centroid*, pengelompokan data berdasarkan jarak terdekat ke *centroid*, dan iterasi(Nawangsih et al., 2021).

6. EvaluasiModel

   Hasil *clustering* dievaluasi menggunakan metrik evaluasi seperti Silhouette Score dan Davies-Bouldin Index. Evaluasi ini memastikan bahwa pengelompokan data telah dilakukan secara optimal(Sholeh & Aeni, 2023).

7. Visualisasi Hasil

   Setelah algoritma *K-means* diterapkan dan model dievaluasi, hasil *clustering* divisualisasikan untuk memudahkan interpretasi dan membantu dalam

pengambilan keputusan. Visualisasi ini memainkan peran penting dalam menjelaskan pengelompokan data yang kompleks dengan cara yang mudah dipahami oleh para pemangku kepentingan.

2.11. Tinjauan Umum Objek Penelitian

Konter Butet *Celluler*, sebagai objek penelitian ini, merupakan sebuah usaha ritel yang berfokus pada penjualan Pulsa. Konter ini berlokasi di daerah Rantauperapat yang strategis dengan jumlah pelanggan yang terus meningkat, mencerminkan pentingnya peran Konter dalam memenuhi kebutuhan teknologi masyarakat setempat. Operasional konter ini tidak hanya mencakup penjualan barang, tetapi juga memberikan layanan pelanggan yang berorientasi pada kepuasan konsumen.

Struktur organisasi Konter Butet.

1. Pemilik Konter

   Berfungsi sebagai pengambil keputusan utama yang bertangung jawab dalam menetapkan strategi bisnis,termasuk ke bijakan pemasaran, manajemen stok, dan harga jual

2. Staf Penjualan

   Bertugas untuk melayani pelanggan secara langsung, memberikan informasi tentang produk, serta memastikan proses transaksi berjalan dengan lancar.

3. Staf Administrasi

   Staf administrasi mencatat data penjualan harian, memantau stok barang, dan membantu pengelolaan keuangan konter.

Dalam konteks penelitian ini, Konter Butet *Celluler* berkontribusi sebagai penyedia data transaksi yang mencerminkan pengelompokan pembelian pelanggan selama periode 2023-2024. Data tersebut mencakup atribut penting seperti ID Operator, Nominal, Jumlah Pembayaran, dan Frekuensi Pembelian. Data ini menjadi bahan utama untuk analisis menggunakan algoritma *K-means*, yang bertujuan untuk mengelompokkan pelanggan berdasarkan pengelompokan pembelian mereka.

Peran Konter Butet *Celluler* dalam penelitian ini sangat penting, karena data yang disediakan memungkinkan peneliti untuk mengidentifikasi segmen pelanggan yang homogen. Hasil akhir penelitian, berupa wawasan tentang preferensi pelanggan dan rekomendasi strategi pemasaran, akan memberikan manfaat praktis bagi konter dalam meningkatkan efektivitas operasional dan daya saing di pasar. Dengan demikian, kontribusi Konter Butet *Celluler* menjadi fondasi utama yang mendukung keberhasilan penelitian ini.