

## BAB II TINJUAN PUSTAKA

### 2.1 Data Mining

Data mining merupakan suatu metode untuk mengumpulkan informasi yang belum ada sebelumnya dari suatu basis data dengan menganalisis pola dan karakteristik di masa mendatang. Data mining mengeksplorasi pengetahuan dan pola dalam data melalui statistik matematika dan *machine learning* (Sekar Setyaningtyas et al., 2022).

Data mining sebenarnya merupakan rangkaian proses pencarian pengetahuan dalam *database* yang dikenal dengan *Knowledge Discovery in Database* (KDD). *Knowledge Discovery in Database* (KDD) berkaitan dengan teknik integrasi dan penemuan ilmiah. Interpretasi dan visualisasi pola hasil dari kumpulan data (Zulfa et al., 2020). *Knowledge Discovery in Database* (KDD) didefinisikan sebagai ekstraksi informasi potensial, implisit dan tidak diketahui dari kumpulan data. Proses *Knowledge Discovery In Database* (KDD) melibatkan hasil dari proses data mining (proses pengestrak kecenderungan dalam suatu pola data), kemudian mengubah hasilnya secara akurat menjadi informasi yang mudah dipahami.

Terdapat empat tugas utama data mining, yaitu sebagai berikut (Adani et al., n.d.):

1. *Predictive Modelling*

*Predictive modelling* (pemodelan prediktif) digunakan untuk membangun model untuk target variabel sebagai fungsi dari *explanatory variable* (variabel penjelas). *Explanatory variable* atau variabel penjelas dalam hal ini adalah semua atribut yang digunakan untuk melakukan prediksi, sedangkan variabel target merupakan atribut yang nilainya akan diprediksi. *Predictive modelling* dibagi menjadi dua tipe yaitu: *Classification* (klasifikasi) digunakan untuk memprediksi nilai dari variabel target yang *discrete* (diskrit) dan *regression* (regresi) digunakan untuk memprediksi nilai dari target variabel yang *continue* (berkelanjutan).

## 2. *Association Analysis*

*Association analysis* (analisis asosiasi) digunakan untuk menemukan aturan asosiasi yang memperlihatkan atau menunjukkan kondisi dari nilai atribut yang sering muncul secara bersamaan dalam suatu kumpulan data.

## 3. *Cluster Analysis*

Berbeda dengan klasifikasi yang menganalisis kelas objek data yang berisi label. *Clustering* menganalisis objek data tanpa memeriksa label kelas yang diketahui. Label kelas disertakan dalam data training atau pelatihan, karena tidak diketahui sebelumnya. *Clustering* adalah proses pengelompokan sekumpulan objek yang sangat mirip.

## 4. *Anomaly Detection*

*Anomaly detection* (deteksi *anomaly*) adalah metode deteksi suatu data dimana tujuannya yaitu menemukan objek yang berbeda dari kebanyakan objek lain. *Anomaly* dapat dideteksi menggunakan uji statistik yang menerapkan model distribusi atau probabilitas ke data.

## 2.2 *Clustering*

*Clustering* merupakan bagian dari data mining yang bersifat tanpa arahan. *Clustering* adalah proses pembagian data kedalam kelas-kelas atau disebut *cluster* berdasarkan tingkat kesamaannya. Dalam *clustering*, data yang memiliki kesamaan atau kemiripan akan digabungkan menjadi satu *cluster* yang sama, begitu pula sebaliknya data yang tidak memiliki kesamaan akan dimasukkan ke dalam *cluster* yang berbeda (Divisive et al., n.d.). Tujuan dari *clustering* yaitu untuk mengelompokkan karakteristik yang sama kedalam area yang sama dan data karakteristik yang tidak sama untuk menerima kelompok objek yang memiliki karakteristik yang sama (Luh et al., 2022).

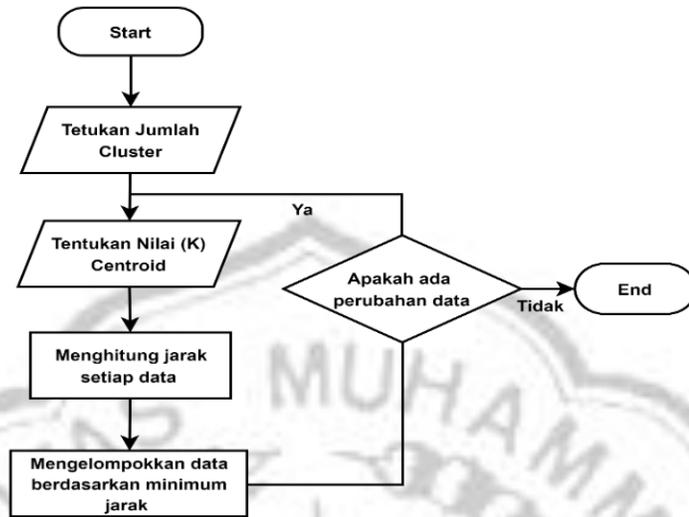
Terdapat dua jenis metode *clustering* yang digunakan dalam pengelompokan data, yaitu *hierarchical clustering* (pengelompokan hirarki) dan *non-hierarchical* (non hirarki) (Ong, n.d.).

*Hierarchical clustering* merupakan metode pengelompokan data yang dimulai dengan mengelompokkan dua atau lebih objek yang memiliki kesamaan terdekat. Kemudian proses diteruskan ke objek lain yang memiliki kedekatan kedua. Demikian seterusnya sehingga *cluster* tersebut akan membentuk semacam pohon dimana terdapat hirarki (level atau tingkatan) yang jelas antar objek, dari yang paling mirip sampai yang paling tidak mirip, secara logika semua objek pada akhirnya hanya akan membentuk *cluster*. Dendogram biasanya digunakan untuk membantu memperjelas proses hierarkis (Ong, n.d.).

Berbeda dengan metode *hierarchical clustering*, metode *non-hierarchical clustering* dimulai dengan terlebih dahulu menentukan jumlah *cluster* yang diinginkan (dua *cluster*, tiga *cluster*, atau seterusnya). Setelah jumlah *cluster* diketahui, selanjutnya dilakukan proses *cluster* tanpa mengikuti proses hirarki. Metode ini biasa disebut *K-Means Clustering* (Ong, n.d.).

### **2.3 Algoritme K-Means**

*K-Means* merupakan salah satu Algoritme *clustering* yang digunakan dalam proses data mining. Algoritme *K-Means* bekerja dengan cara mempartisi pengelompokan yang memisahkan data kedalam kelompok-kelompok tertentu dengan meminimalkan jarak rata-rata setiap klaster data (Triyandana et al., 2022). Metode *K-Means* berupaya mengelompokkan data yang ada menjadi beberapa kelompok, dimana data pada satu kelompok memiliki karakteristik yang sama antara satu sama lain dan memiliki karakteristik yang berbeda dengan data pada kelompok lainnya (Parsaoran Tamba & Toknady Kesuma, 20



**Gambar 2. 1** Flowchart Algoritme *K-Means*

**Gambar 2.1** menjelaskan tahapan Algoritme *K-Means* dimana langkah pertama adalah menentukan jumlah *cluster* ( $k$ ) yang akan digunakan. Setelah itu menentukan *centroid* awal yang diambil secara acak dari dataset. Langkah kedua yakni menentukan nilai yang akan dijadikan *centroid*. Langkah selanjutnya adalah menghitung jarak setiap titik data ke setiap *centroid* menggunakan *Manhattan*. Langkah selanjutnya yaitu mengelompokkan data berdasarkan jarak terdekat antara data dengan *centroid*. Kemudian menentukan nilai *centroid* baru dengan menghitung rata-rata *cluster*. Setelah itu, ulangi langkah menghitung jarak menggunakan teori *Manhattan*. hingga tahap penentuan nilai *cluster* baru. Tahap atau fase pengulangan ini akan dihentikan jika tidak ada perubahan data pada suatu *cluster*. Untuk menentukan nilai *centroid* baru dapat dilakukan dengan menggunakan persamaan 2.1.

$$C_k = \frac{1}{n_k} \sum_{i=1}^n d_i \quad (2.1)$$

Keterangan:

$n_k$  = jumlah semua data yang ada didalam *cluster* (k).

$d_i$  = jumlah dari setiap *cluster*

#### 2.4 Algoritme *K-Means++*

Metode *K-Means++* digunakan untuk mengatasi masalah dengan memilih pusat *cluster* awal secara acak yang meningkatkan jumlah iterasi. Tujuan utama dari algoritme ini adalah untuk mengambil poin data tersebut sebagai pusat awal yang berjarak sejauh mungkin antara satu sama lain (Sri Fastaf & Yamasari, 2022b). Metode *K-Means++* memastikan bahwa titik pusat awal ditentukan lebih cerdas untuk meningkatkan kualitas pengelompokan. Berikut merupakan langkah-langkah untuk metode *K-Means++* yaitu:

1. Menentukan titik *centroid* secara acak.
2. Menghitung jarak semua titik dalam *dataset* dari *centroid* yang dipilih. Jarak titik  $x_i$  dari *centroid* terjauh dapat dihitung menggunakan persamaan 2.2.

$$d_i = \max_{(j:1 \rightarrow m)} \|x_i - C_j\|^2 \quad (0.2)$$

3. Menjadikan titik  $x_i$  sebagai *centroid* baru yang memiliki probabilitas maksimum.
4. Ulangi langkah 2-3 sampai titik *centroid* pada setiap *cluster*(k) terpenuhi.
5. Setelah mendapatkan *centroid*, langkah selanjutnya sama dengan metode *K-Means*.

#### 2.5 *Euclidean Distance*

*Euclidian Distance* merupakan salah satu metode perhitungan jarak yang digunakan untuk mengukur jarak dari 2 (dua) buah titik yang dalam euclidian space (meliputi bidang euclidian dua dimensi, tiga dimensi atau bahkan lebih) (Pribadi et al., 2022).

Dalam pengukuran jarak *Manhattan Distance* digunakan sebagai Notasi sebagai berikut:

$$d(x_i, y_i) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (2.2)$$

Keterangan :

$D(x_i, y_i)$  = jarak antara *clustering x* dengan titik *centroid y* pada data ke-*i*  
 $x_i$  = bobot ke-*i* pada *cluster* yang ingin dicari jaraknya  
 $y_i$  = bobot data ke-*i* pada titik *centroid*  
 $n$  = merupakan jumlah data

## 2.6 *Davies Bouldin Index (DBI)*

*Davies Bouldin Index (DBI)* digunakan untuk menentukan evaluasi *cluster*. *Davies Bouldin Index* adalah salah satu metode evaluasi internal yang digunakan untuk mengukur evaluasi *cluster* yang didapatkan pada suatu metode pengelompokan yang didasarkan pada nilai kohesi dan separasi. Separasi didasarkan pada jarak antar *centroid* dari setiap *clusternya*. Sedangkan kohesi didefinisikan sebagai jumlah pengelompokan (Nabila et al., 2021)

*Sum of square within cluster* merupakan matrik kohesi dalam sebuah *cluster* Separasi dengan *Sum-of-square-between-cluster (SSB)* dengan mengukur jarak antara *centroid*  $C_i$  dan  $C_j$ .  $R_{i,j}$  adalah ukuran rasio seberapa baik nilai perbandingan antara *cluster* ke-*i* dan *cluster* ke-*j*.

Langkah-langkah perhitungan *Davies Bouldin Index* adalah sebagai berikut (JNKTI\_Vol3No1April2020\_FullVersion, n.d.):

### 1. *Sum Of Square Within-Cluster (SSW)*

Merupakan persamaan yang digunakan untuk mengetahui matrik kohesi dalam sebuah *cluster* ke-*i*. Persamaan yang digunakan untuk memperoleh nilai *Sum of Square Within cluster* adalah sebagai berikut:

$$SSW = \frac{1}{N} \sum_{i=1}^N \|x_i - c_{pi}\|^2 \quad (2.3)$$

Keterangan:

$m$  = Jumlah data dalam *cluster* ke- $i$

$c$  = *Centroid cluster* ke- $i$

$d$  = Jarak

## 2. *Sum Of Square Between Cluster (SSB)*

Merupakan persamaan yang digunakan untuk mengetahui separasi antar *cluster*. Persamaan yang digunakan untuk menghitung nilai *Sum of Square Between cluster* adalah sebagai berikut:

$$SSB = \frac{2}{M(M-1)} \sum_{i=1}^M \sum_{j=1, j \neq i}^M \|c_i - c_j\|^2 \quad (2.4)$$

Keterangan:

$d$  = Jarak

$c$  = *Centroid cluster* ke- $i$

## 3. *Ratio (Rasio)*

Perhitungan rasio ( $R_{i,j}$ ) ini bertujuan untuk mengetahui nilai perbandingan antara *cluster* ke- $i$  dan *cluster* ke- $j$  untuk menghitung nilai rasio yang dimiliki oleh masing-masing *cluster*. *Cluster* yang baik adalah *cluster* yang memiliki nilai kohesi sekecil mungkin dan separasi yang sebesar mungkin. Nilai rasio dihitung menggunakan persamaan sebagai berikut:

$$R_{i,j} = \frac{SSW_i + SSW_j}{SSB_i} \quad (2.5)$$

Keterangan:

$SSWi$  = Sum Of Square Within-Cluster pada centroid  $i$

$SSBi,j$  = Sum of Square Between Cluster data ke  $i$  dengan  $j$  pada cluster yang berbeda

#### 4. *Davies Bouldin Index (DBI)*

Nilai yang didapatkan dari perhitungan rasio nantinya akan digunakan untuk mencari nilai DBI yang menggunakan persamaan sebagai berikut:

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} (R_{i,j,\dots,k}) \quad (2.6)$$

Dari perhitungan *Davies Bouldin Index (DBI)* dapat disimpulkan bahwa jika semakin kecil nilai *Davies Bouldin Index (DBI)* yang diperoleh (non negatif  $\geq 0$ ) maka cluster tersebut semakin baik.

#### 2.7 *Silhouette Coefficient*

*Silhouette Coefficient* adalah pengabungan dari metode *cohesion* dan *separation* yang merupakan metode evaluasi untuk *Cluster*. Jarak antara data dapat dihitung menggunakan rumus *Euclidian Distance* dan *Manhattan Distance* (Pribadi et al., 2022). Adapun langkah-langkah untuk menghitung nilai *Silhouette Coefficient* sebagai berikut.

1. Menghitung rata-rata jarak suatu data dengan data lain dalam suatu *cluster* yang sama menggunakan persamaan 2.8.

$$a(i) = \frac{1}{|A| - 1} \sum_{j \in A, j \neq i} d(i,j) \quad (0.3)$$

Keterangan :

$a(i)$  = Perbedaan rata-rata pada data ( $i$ ) ke semua data lain di *cluster A*

$d(i,j)$  = Jarak antara data  $i$  dan data  $j$

2. Menghitung rata-rata jarak data tersebut dengan semua data di *cluster* lain menggunakan persamaan 2.9.

$$d(i, C) = \frac{1}{|C|} \sum_{j \in C} d(i, j) \quad (0.4)$$

Keterangan :

$d(i, C)$  = Perbedaan rata-rata pada data ( $i$ ) ke seluruh data  $C$

$C$  = *Cluster* lain selain  $A$

3. Memilih nilai jarak yang paling minimum menggunakan persamaan 2.10.

$$b(i) = \min_{c \neq A} d(i, j) \quad (0.5)$$

Keterangan :

$b(i)$  = Nilai minimum jarak rata-rata data ke- $i$  dengan semua data di *cluster* berbeda.

4. Menghitung nilai *Silhouette Coefficient* dengan menggunakan persamaan 2.11.

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (0.6)$$

Keterangan :

$a(i)$  = Perbedaan rata-rata pada data ( $i$ ) ke semua data lain di *cluster*  $A$

$b(i)$  = Perbedaan rata-rata pada data ( $i$ ) ke semua data pada *cluster* berbeda

$s(i)$  = Nilai *Silhouette Coefficient*

*Range* nilai  $s(i)$  yakni antara -1 dan 1, interpretasi nilai tersebut yaitu:

$s(i) = -1$ : data ke-  $i$  digolongkan lemah (dekat pada *cluster* B daripada A)

$s(i) = 0$ : data ke-  $i$  berada di tengah dua kluster (A dan B)

$s(i) = 1$ : data ke-  $i$  digolongkan baik

Intepretasi nilai *Silhouette Coefficient* juga ditunjukkan dalam **Tabel 2.1**.

**Tabel 2. 1** Interpretasi Nilai *Silhouette Coefficient*

<i>Silhouette Coefficient</i>	Interpretasi
$\leq 0.25$	Tidak terstruktur
0.26-0.50	Hasil struktur lemah
0.51-0.70	Hasil struktur baik
0.71-1.00	Hasil struktur kuat

## 2.8 Metode *Waterfall*

Metode *waterfall* biasanya disebut juga dengan metode air terjun ini menggambarkan pendekatan yang sistematis dan juga berurutan pada pengembangan perangkat lunak, dimulai dengan spesifikasi kebutuhan pengguna lalu dilanjutkan dengan tahapan perencanaan (*planing*), permodelan (*modeling*), konstruksi (*contruction*), serta penyerahan sistem ke para pengguna (*depoyment*), lalu diakhiri dengan dukungan perangkat lunak.

Model *waterfall* pertama kali diperkenalkan oleh *Winston Royce* sekitar tahun 1970 sehingga sering dianggap kuno, tetapi merupakan model yang paling banyak dipakai didalam *Software Engineering* (SE) (Wahid, n.d.).

Tahapan yang digunakan dalam metode *waterfall* adalah sebagai berikut:



**Gambar 2. 2** Ilustrasi Metode *Waterfall*

## 2.9 *Black Box Testing*

Metode *Black Box Testing* adalah sebuah metode yang dipakai untuk menguji sebuah software tanpa harus memperhatikan detail software. Proses *Black Box Testing* dengan cara mencoba program yang telah dibuat dengan mencoba memasukkan data pada setiap formnya (Made et al., 2021).

*Black Box Testing* berkerja dengan mengabaikan struktur kontrol sehingga perhatiannya difokkas pada informasi domain. Memungkin pengembang *software* untuk mmbuat himpunan kodisi input yang akan melatih seluruh syarat-syarat fungsional.

Keuntungan penggunaan metode *Blac Box* adalah penguji tidak perlu memiliki pengetahuan tentang bahasa pemrograman tertentu. Pengujian dilakukan dari sudut pandang pengguna, ini membantu untuk mengungkapkan ambiguitas atau inkonsistensi dalam spesifikasi persyaratan. Programer dan tester keduanya saling bergantung satu sama lain.

Kekurangan dari metode *Black Box Testing* adalah Uji kasus sulit desain tanpa spesifikasi yang jelas. Kemungkinan memiliki pengulangan tes sudah dilakukan oleh programmer. Beberapa bagian tidak diuji sama sekali.

## 2.10 Daftar Pustaka

Berikut beberapa referensi pembelajaran dari beberapa contoh kasus yang hampir sama dengan permasalahan yang dihadapi dan dijadikan penulis sebagai referensi dalam penelitian kali ini, referensi yang digunakan dapat dilihat pada **Tabel 2.2**.

**Tabel 2. 2** Penelitian – Penelitian Terdahulu

PENELITI	JUDUL	METODE	OBJEK PENELITIAN	HASIL
(Triyandana et al., 2022)	Penerapan Data Mining Pengelompokan Menu Makanan dan Minuman	Algoritme <i>K-Means</i> .	Menu Makanan dan Minuman Berdasarkan Tingkat	Penjualan di Dpom Coffee menghasilkan 3 cluster yaitu penjualan rendah,

	Berdasarkan Tingkat Penjualan Menggunakan Metode <i>K-Means</i> .		Penjualan di Dpom Coffe.	penjualan sedang, dan penjualan tinggi.
(Ramadanti & Muslih, 2022)	Penerapan Data Mining Algoritma <i>K-Means Clustering</i> Pada Populasi Ayam Petelur Di Indonesia.	Algoritme <i>K-Means Clustering</i> .	Data pengelompokan populasi ayam petelur di Indonesia.	Hasil dari ketiga perhitungan yaitu 1 provinsi untuk <i>cluster</i> populasi tinggi yaitu Jawa Timur. Untuk <i>cluster</i> populasi sedang ada 3 provinsi di antaranya Sumatera Utara, Jawa Barat dan Jawa Tengah. Sedangkan untuk <i>cluster</i> populasi rendah ada 30 provinsi.
(Annisa et al., 2022)	Penerapan Data Mining Pengelompokan Data Pengguna Air Bersih Berdasarkan Keluhannya Menggunakan Metode <i>Clustering</i> .	<i>Clustering</i> .	Pengelompokan data pengguna air bersih berdasarkan keluhannya.	Dari hasil <i>cluster</i> terdapat kesamaan hasil yaitu jenis keluhan pada <i>cluster</i> 1 dengan <i>cluster</i> 2 yaitu kode 5 jenis keluhan pipa bocor dengan penanganan kerusakan menyambung pipa air (gibout join).
(Gustrianda & Mulyana, 2022)	Penerapan Data Mining Dalam Pemilihan Produk Unggulan dengan Metode Algoritma <i>K-</i>	Algoritme <i>K-Means</i> dan <i>K-Medoids</i> .	Data penjualan produk pada perusahaan bisnis PT. XYZ	Terdapat pengelompokan produk yang menjadi produk unggulan, produk yang peminatnya

	<i>Means</i> Dan <i>K-Medoids</i> .			sedang, dan produk yang kurang diminati.
(Normah et al., 2021)	Penerapan Data Mining Metode <i>K-Means Clustering</i> Untuk Analisa Penjualan Pada Toko Fashion Hijab Banten.	<i>K-Means Clustering</i> .	Data penjualan pada Toko Helai Fashion Hijab Banten.	Terdapat 3 macam <i>cluster</i> yaitu, sangat laris, laris, dan kurang laris.
(Nasir, 2021b)	Penerapan Data Mining <i>Clustering</i> Dalam Mengelompokkan Buku Dengan Metode <i>K-Means</i> .	<i>K-Means Clustering</i> .	Pengelompokan buku pada perpustakaan yang berada di Kota Batam.	Terdapat pada <i>cluster</i> 1 sebanyak 9 item, buku yang paling sedikit dipinjam terdapat pada <i>cluster</i> 2 sebanyak 15 item, buku yang cukup banyak dipinjam terdapat pada <i>cluster</i> 0 sebanyak 12 item.
(Wahyuni & Saepudin, 2021)	Penerapan Data Mining <i>Clustering</i> Untuk Mengelompokkan Berbagai Jenis Merk Mesin Cuci.	<i>K-Means Clustering</i> .	Pengelompokan berbagai jenis merks mesin cuci.	Terdapat 3 kelompok berbagai jenis merk Mesin cuci yaitu untuk kelompok 1 berjumlah 5 data, kelompok 2 berjumlah 8 data, dan kelompok 3 berjumlah 12 data.
(Ordila et al., 2020)	Penerapan Data Mining Untuk Pengelompokan Data Rekam Medis Pasien Berdasarkan Jenis	Algoritme <i>Clustering</i> .	Data rekam medis pasien berdasarkan jenis penyakit pada Poli Klinik PT. Inecda.	Pengelompokan berdasarkan wilayah, jenis penyakit dan umur.

	Penyakit Dengan Algoritma <i>Clustering</i> (Studi Kasus : Poli Klinik PT. Inecda).			
(Ardiansyah et al., 2020)	Penerapan Data Mining Menggunakan Metode <i>Clustering</i> untuk Menentukan Status Provinsi di Indonesia 2020.	Algoritme <i>K-Means Clustering</i> .	Pengelompokan wilayah di Indonesia menjadi zona merah atau kuning atau hijau mengenai COVID-19.	Hasil pengelompokan daerah yang masuk zona merah adalah Jakarta untuk yang masuk zona kuning yaitu Banten. Dan sebanyak kurang dari 600 pasien yang terinfeksi dapat dikategorikan sebagai zona hijau.
(Magdalena & Fahrudin, 2020)	Penerapan Data Mining Untuk Koperasi Se-Jawa Barat Menggunakan Metode <i>Clustering</i> pada Kementerian Koperasi dan UKM.	<i>K-Means Clustering</i> .	menegelompokan data koperasi di Jawa Barat.	Hasil pengelompokan menggunakan metode <i>K-Means</i> berdasarkan nilai Modal Sendiri, Modal Luar dan Volume Usaha, maka dihasilkan 3 klaster dengan nilai klaster tinggi, sedang dan rendah.
(Annur, 2019)	Penerapan Data Mining Menentukan Strategi Penjualan Variasi Mobil Menggunakan Metode <i>K-Means</i>	<i>K-Means Clustering</i> .	Strategi penjualan mobil pada Toko Luxor Gorontalo.	Mengelompokkan data produk yang terjual untuk mengetahui data yang memiliki potensi atau kecenderungan

	<i>Clustering</i> (Studi Kasus Toko Luxor Variasi Gorontalo).			pelanggan dalam membeli barang tersebut.
(Irawan, 2019)	Penerapan Data Mining Untuk Evaluasi Data Penjualan Menggunakan Metode <i>Clustering</i> Dan Algoritma <i>Hirarki Divisive</i> .	<i>Clustering</i> dan Algoritma <i>Hirarki Divisive</i> .	Data penjualan pelayanan jasa dan penjualan mesin dan bahan-bahan kebutuhan digital printing di Media World Pekanbaru.	pengetahuan berupa customer loyal yang ada di perusahaan Media World Pekanbaru dan ditemukan bahwa customer loyal berada pada <i>cluster</i> 3 dengan range 117-358.
(Tamba et al., 2019)	Penerapan Data Mining Untuk Menentukan Penjualan Sparepart Toyota Dengan Metode <i>K-Means Clustering</i> .	<i>K-Means Clustering</i> .	Data penjualan sparepart Toyota di CV Terang Jaya.	Penjualan di CV Terang Jaya terdapat 3 macam <i>cluster</i> yaitu, sangat laris, laris, dan kurang laris.
(Novianto & Goeiranto, 2019)	Penerapan Data Mining Menggunakan Algoritma <i>K-Means Clustering</i> untuk Menganalisa Bisnis Perusahaan Asuransi.	Algoritma <i>K-Means Clustering</i> .	Pengelompokan data asuransi pada Perusahaan keuangan non Bank yang bergerak di bidang Jasa Asuransi.	Pengelompokan data nilai pertanggungan, premi dan claim berdasarkan <i>clustering</i> dengan nilai terendah sedang dan tertinggi. Dengan pengelompokan sebanyak 3 <i>cluster</i> dimana memiliki batasan nilai yang berbeda.
(Budiman & Rudianto, 2019)	Penerapan Data Mining Untuk Menentukan Lokasi Promosi Penerimaan	<i>K-Means Clustering</i> .	Pengelompokan lokasi atau daerah promosi penerimaan mahasiswa baru	Terdapat tiga <i>cluster</i> . <i>Cluster</i> pertama kategori tinggi, <i>cluster</i> kedua kategori

	Mahasiswa Baru Pada Universitas Banten Jaya (Metode <i>K-Means Clustering</i> )		pada Universitas Banten Jaya.	rendah dan <i>cluster</i> tiga kategori sedang.
--	---------------------------------------------------------------------------------	--	-------------------------------	-------------------------------------------------



