

**BAB 2**  
**LANDASAN TEORI**

**2.1 Perbandingan dengan Penelitian Terdahulu**

Penelitian terdahulu bertujuan sebagai perbandingan untuk menghindari kesamaan isi pada penelitian. Oleh karena itu, pada Bab 2 Landasan Teori dicantumkan sebanyak empat penelitian terdahulu untuk dijadikan sebagai perbandingan. Perbandingan dengan penelitian terdahulu dapat dilihat pada Tabel 2.1.

**Tabel 2.1** Perbandingan dengan Penelitian Terdahulu

No.	Peneliti	Judul	Persamaan	Perbedaan
1	Jumeilah & Pratama, 2017	Klasterisasi Penduduk Lanjut Usia Provinsi Sumatera Selatan Menggunakan <i>K-Modes</i>	<b>Metode:</b> <i>K-Modes Clustering</i>	<ul style="list-style-type: none"> <li>• Data yang digunakan, yaitu sensus penduduk yang usianya &gt; 60 tahun dengan total data sampel 47.358 jiwa.</li> <li>• Menghasilkan 4 <i>cluster</i> lansia.</li> </ul>
2.	Indriani & Budiman, 2017	<i>K-Modes Clustering</i> Untuk Mengetahui Jenis Masakan Daerah Yang Populer Pada <i>Website</i> Resep <i>Online</i> (Studi Kasus: Masakan Banjar Di <i>Cookpad.Com</i> )	<b>Metode:</b> <i>K-Modes Clustering</i>	<ul style="list-style-type: none"> <li>• Data yang digunakan sebanyak 355 data resep masakan Banjar</li> <li>• Menghasilkan 2 alternatif klasterisasi, yaitu k=4 untuk <i>cluster</i> yang lebih umum, dan k=8 untuk <i>cluster</i> yang lebih spesifik.</li> <li>• Pengumpulan data menggunakan <i>website</i>.</li> </ul>

Tabel 2.1 Perbandingan dengan Penelitian Terdahulu (Lanjutan)

No.	Peneliti	Judul	Persamaan	Perbedaan
3	Yuliana, 2019	Penerapan Algoritma <i>K-Modes Clustering</i> Untuk Pengelompokan Desa Rawan Kebakaran Di Provinsi Riau	<b>Pengujian Cluster:</b> Menggunakan <i>Davies Bouldin Index</i> <b>Metode:</b> <i>K-Modes Clustering</i>	<ul style="list-style-type: none"> <li>• Data yang digunakan sebanyak 16.167 data.</li> <li>• Menghasilkan 3 <i>cluster</i> desa rawan kebakaran,</li> <li>• Hasil klasterisasi terbaik yaitu 2 <i>cluster</i>, dibuktikan dengan pengujian DBI yang menghasilkan nilai validitas 0,84.</li> </ul>
4	Sari & Utamajaya, 2022	Sistem Pendukung Keputusan Penerima Bantuan Langsung Tunai Dana Desa Menggunakan Metode Algoritma <i>K-Means Clustering</i>	<b>Kasus yang diselesaikan:</b> BLT	<ul style="list-style-type: none"> <li>• Data yang digunakan sebanyak 336 data.</li> <li>• variabel yang digunakan sebanyak 4 variabel.</li> <li>• Menghasilkan 3 <i>cluster</i>, yaitu tepat sasaran, dipertimbangkan, dan tidak tepat sasaran.</li> </ul>
5	Pranoto & Vendyansyah, 2022	Aplikasi Penentuan Penerima Bantuan Langsung Tunai Menggunakan <i>K-Modes</i>	<b>Metode:</b> <i>K-Modes Clustering</i> <b>Kasus yang diselesaikan:</b> BLT	<ul style="list-style-type: none"> <li>• Data yang digunakan sebanyak 10 data sampel.</li> <li>• variabel yang digunakan sebanyak 4 variabel.</li> <li>• Seleksi menggunakan aplikasi menghasilkan 8</li> </ul>

**Tabel 2.1** Perbandingan dengan Penelitian Terdahulu (Lanjutan)

No.	Peneliti	Judul	Persamaan	Perbedaan
				orang dikelompok penerima BLT, dan 2 orang dikelompok tidak menerima BLT. Seleksi secara manual memperoleh hasil 7 orang di kelompok penerima BLT, dan 3 orang dikelompok tidak menerima BLT.

## 2.2 Clustering

*Clustering* merupakan suatu teknik atau metode untuk mengelompokkan data. Pengelompokan data berupa proses pembagian satu set objek data ke dalam himpunan bagian yang di sebut *cluster*. Objek dalam *cluster* memiliki kemiripan karakteristik antara satu sama lainnya dan berbeda dengan *cluster* lain. Pembagian objek tidak dilakukan secara manual tetapi menggunakan algoritma *clustering*. (Irwansyah & Faisal, 2015).

*Clustering* dibagi menjadi 2 metode, yaitu *hierarchival* dan *non-hierarchival*. *Hierarchival* merupakan metode pengelompokan data yang dimulai dengan mengelompokkan beberapa objek yang memiliki kesamaan paling dekat, kemudian dilanjutkan ke objek lain yang memiliki kedekatan lainnya. Sehingga *cluster* akan membentuk pohon dari yang paling mirip sampai yang tidak mirip. Metode *Hierarchival* digunakan apabila belum diketahui jumlah *clusternya*. Sedangkan metode *non-hierarchival* dimulai dengan menentukan dahulu jumlah *cluster* kemudian dilakukan proses *clustering*. Metode *non-hierarchival* bertujuan mengelompokkan objek ke dalam k buah *cluster* (Jumeilah & Pratama, 2017).

## 2.3 Data Mining

*Data Mining* merupakan proses yang digunakan untuk menangani masalah dalam pengambilan informasi dari *database* menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning*. *Data mining* terdapat dalam

KDD (*Knowledge Discovery in Database*) digunakan untuk mengidentifikasi informasi yang bermanfaat dan pengetahuan terkait *database*. *Data mining* bertujuan untuk menemukan, atau menggali pengetahuan dari data atau informasi yang dimiliki. Tahapan *data mining* dibagi menjadi tujuh bagian, yaitu: (Wahyudi, Masitha, Saragih, & Solikhun, 2020).

1. Pembersihan data (*data cleaning*), sebelum adanya proses *data mining* perlu dilakukan proses pembersihan data antara lain membuang duplikasi data, memeriksa data yang tidak konsisten, dan memperbaiki kesalahan data seperti adanya kesalahan cetak.
2. Integrasi data (*data integration*), merupakan penggabungan data dari basis data lain ke dalam basis data baru, atau menggabungkan data dari beberapa sumber. Data yang diperlukan untuk *data mining* tidak jarang hanya berasal dari satu sumber tetapi juga berasal dari beberapa sumber.
3. Seleksi data (*data selection*), data pada basis data tidak semua digunakan, hanya data yang sesuai untuk dianalisis yang akan diambil pada basis data.
4. Transformasi data (*data transformation*), data diubah atau digabung ke dalam format yang sesuai untuk proses *data mining*. Beberapa metode *data mining* membutuhkan format data yang sesuai sebelum dapat digunakan. Proses tersebut sering disebut transformasi data.
5. *Data mining*, proses *mining* merupakan proses penerapan metode yang digunakan untuk mencari informasi yang tersembunyi dari data.
6. Evaluasi pola (*pattern evaluation*), merupakan identifikasi pola menarik pada *data mining*.
7. Presentasi pengetahuan (*knowledge presentation*), merupakan penyajian mengenai metode yang digunakan untuk mendapatkan informasi dari pengguna.

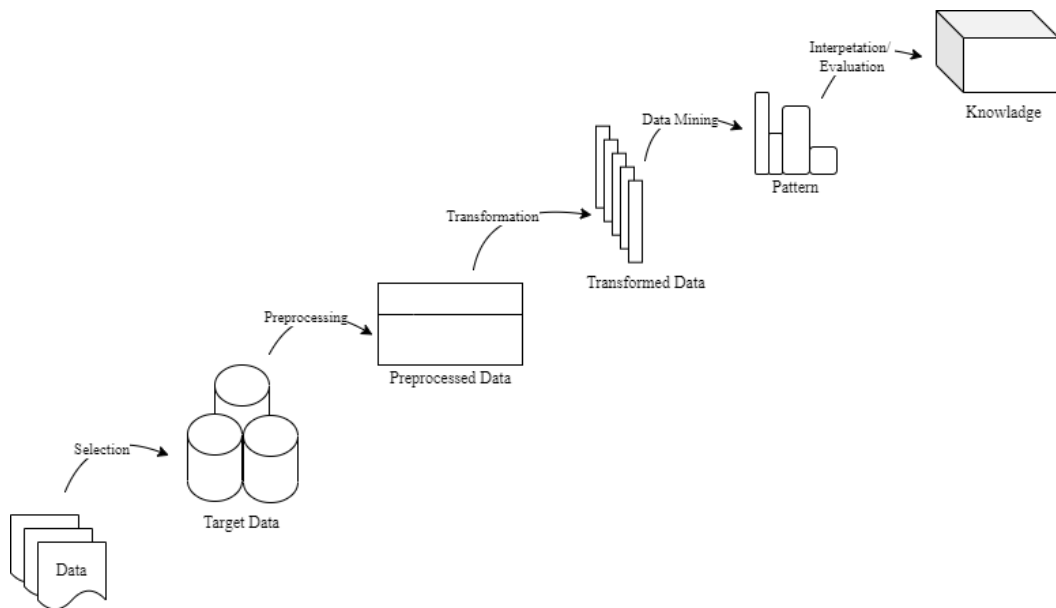
Berdasarkan dengan yang telah disebutkan di atas bahwa KDD adalah proses menemukan pengetahuan yang ada di dalam *database*. Dalam *database* terdapat tabel yang saling berhubungan atau berelasi. Hasil penemuan pengetahuan yang diperoleh dalam proses KDD dapat digunakan sebagai pengetahuan untuk pengambilan keputusan. KDD dalam *data mining* sering digunakan untuk menjelaskan proses penggalian informasi yang tersembunyi dalam *database* yang

besar (Aldo, Syawatri, Alwendi, Darmansah, & Samosir, 2021). Tahapan proses KDD adalah:

1. Data, dalam proses KDD perlu terlebih dahulu dipersiapkan data yang akan digunakan. Data yang digunakan adalah data yang sudah tersedia.
2. *Selection*, pada tahap ini proses pemilihan data dari sekumpulan data yang digunakan. Proses pemilihan data meliputi pembuatan kumpulan data target, penentuan variabel, pemilihan sampel data, dan penyimpanan data pada sebuah berkas.
3. *Pre-processing/Cleaning, pre-processing* dilakukan sebelum proses *data mining* dilakukan. Perlu terlebih dahulu dilakukan proses pembersihan pada data yang menjadi fokus KDD. Proses pembersihan meliputi membuang duplikasi data, memeriksa data yang tidak konsisten, dan memperbaiki kesalahan pada data seperti kesalahan pada pencetakan. Proses ini juga dapat menambah informasi lain yang relevan yang disebut dengan istilah *enrichment*.
4. *Transformation*, dalam *data mining* banyak terdapat algoritma yang dapat digunakan. Format data pada setiap algoritma berbeda. Data yang akan digunakan untuk proses KDD perlu diubah dan disesuaikan dengan algoritma yang digunakan dalam data mining agar dapat dilakukan dengan mudah.
5. *Data Mining, data mining* merupakan proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Pada *data mining* terdapat banyak teknik, metode, atau algoritma yang sangat bervariasi. Pemilihan metode atau algoritma yang tepat bergantung pada tujuan dan proses KDD yang dicari.
6. *Interpretation/Evaluation*, pengetahuan atau informasi yang telah dihasilkan dari *data mining* akan dipresentasikan atau ditampilkan kedalam bentuk yang mudah dimengerti. Pengetahuan atau informasi yang dihasilkan dari proses *data mining* diperiksa agar tidak bertentangan dengan hipotesis sebelumnya.
7. *Knowledge*, proses ini merupakan tujuan dari KDD. Tujuan tersebut untuk memperoleh pengetahuan atau informasi yang bermanfaat dan mudah

dimengerti. Pengetahuan atau informasi yang dihasilkan diimplementasikan sesuai dengan manfaat ataupun kegunaan pengetahuan.

Tahapan proses KDD dapat dilihat pada Gambar 2.1.



**Gambar 2.1** Tahapan Proses KDD

(Bulolo, 2020)

## 2.4 Metode *K-Modes*

Metode *K-Modes* merupakan metode pengembangan dari *K-Means* yang dapat mengelompokkan data kategorikal, dan proses perhitungan untuk pembentukan *cluster* lebih singkat. Berdasarkan hal diatas, *K-Modes* melakukan modifikasi pada *K-Means* sebagai berikut (Wang, 2009).

1. Menggunakan ukuran pencocokan ketidakmiripan sederhana dari fitur data bertipe kategorikal.
2. Mengganti *mean cluster* dengan modus (nilai yang paling sering muncul).
3. Menggunakan metode berbasis frekuensi untuk mencari modus dari sekumpulan nilai.

Berikut ini adalah langkah-langkah *clustering* menggunakan *K-Modes*:

1. Pilih  $k$  data sebagai inisialisasi *centroid* (modus), satu untuk setiap *cluster*.
2. Hitung jarak antara masing-masing objek dan mode *cluster*. Selanjutnya objek ditetapkan ke *cluster* yang pusatnya memiliki jarak terdekat ke objek.

Langkah tersebut diulangi sampai semua objek ditetapkan ke kelompok. Persamaan untuk mencari jarak ditunjukkan pada Persamaan 2.1.

$$D_{(x,y)} = \sum_{j=1}^r \in (X_j, Y_j) \quad (2.1)$$

Keterangan:

$D_{(x,y)}$  = Jarak data x ke y

$X_j$  = Nilai fitur ke-j dari x

$Y_j$  = Nilai fitur ke-j dari y

r adalah jumlah fitur, untuk mencari jarak menggunakan nilai pencocokan ditunjukkan pada Persamaan 2.2.

$$\in (X_j, Y_j) = \begin{cases} 0, & X_j = Y_j \\ 1, & X_j \neq Y_j \end{cases} \quad (2.2)$$

3. Perbarui *centroid* (modus) dari setiap *cluster* dengan nilai kategori yang sering muncul pada setiap *cluster*.
4. Ulangi langkah 2 dan 3 untuk memenuhi syarat, yaitu data pada *cluster* tidak berpindah posisi *cluster* atau posisi pusat *centroid* tidak berubah.
5. Jika salah satu syarat terpenuhi maka proses *K-Modes* selesai.

## 2.5 *Davies Bouldin Index (DBI)*

DBI merupakan metode evaluasi *cluster* yang digunakan untuk mengetahui baik atau tidak hasil *cluster*. Baik atau tidak hasil *cluster* dilihat dari besar jarak antar data dan perbedaan data-data yang ada pada *cluster* (Davies & Bouldin, 1979). Langkah-langkah perhitungan *Davies Bouldin Index* sebagai berikut.

1. *Sum of Square Within Cluster (SSW)*

Digunakan untuk mengetahui jumlah dari kedekatan atau kemiripan data terhadap titik pusat *cluster* dari sebuah *cluster* yang diikuti. Persamaan untuk menghitung nilai SSW ditunjukkan pada Persamaan 2.3.

$$SSW_i = \frac{1}{m_i} \sum_{j=1}^{m_i} d(x_j, c_i) \quad (2.3)$$

Keterangan:

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

$c_i$  = *Centroid cluster* ke-i

$d(x_j, c_i)$  = Jarak setiap data ke *centroid* i

2. *Sum of Square Between Cluster (SSB)*

Digunakan untuk mengetahui jarak antar *cluster*. Persamaan untuk menghitung nilai SSB ditunjukkan pada Persamaan 2.4.

$$SSB_{i,j} = d(c_i, c_j) \quad (2.4)$$

Keterangan:

$SSB_{i,j}$  = *Sum of Square Between Cluster* data ke i dengan j pada *cluster* yang berbeda

$d(c_i, c_j)$  = Jarak *centroid*  $c_i$  dengan *centroid*  $c_j$

3. *Ratio (Rasio)*

Perhitungan rasio digunakan untuk mengetahui nilai perbandingan antara *cluster* ke-i dan *cluster* ke-j. Index i dan j merupakan jumlah *cluster*.

Persamaan untuk menghitung rasio ditunjukkan pada Persamaan 2.5.

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

Keterangan:

$R_{i,j}$  = Nilai Perbandingan *cluster* i dengan *cluster* j

$SSW_i$  = *Sum of Square Within Cluster* pada *centroid* i

4. DBI

Nilai rasio yang diperoleh digunakan untuk mencari nilai DBI. Persamaan untuk menghitung DBI ditunjukkan pada Persamaan 2.6.

$$DBI = \frac{1}{K} \sum_{j=1}^K \max(R_{i,j}) \quad (2.6)$$

Keterangan:

DBI = *Davies Bouldin Index*

K = Jumlah *cluster* yang digunakan

Semakin kecil nilai DBI yang diperoleh maka kelompok semakin baik, artinya kemiripan antar data dalam satu kelompok akan semakin mirip.

## 2.6 Perangkat Lunak

### 2.6.1 Website

*Website* merupakan sebuah media yang memiliki banyak halaman yang saling terhubung, yang berfungsi memberikan informasi dalam bentuk teks, gambar, video, suara, dan animasi atau gabungan dari semuanya. *Website* dapat



bersifat statis maupun dinamis yang dapat membentuk satu rangkaian yang saling terkait dan masing-masing terhubung dengan jaringan-jaringan halaman. *Website* digunakan sebagai media yang dapat menyimpan banyak data, dan diakses menggunakan jaringan internet. Pada umumnya *website* dibangun menggunakan bahasa pemrograman seperti PHP, yang digabung dengan HTML, CSS, dan Javascript (Elgamar, 2020).

### 2.6.2 Database

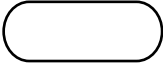
*Database* merupakan kelompok data yang saling terhubung dan diorganisasi agar dapat dimanfaatkan kembali dengan cepat dan mudah. Kelompok data dalam bentuk file/tabel/arsip yang berhubungan tersimpan dalam suatu media penyimpanan elektronis. Penyimpanan tersebut digunakan untuk memberikan kemudahan dalam pengaturan, dan pemilihan data sesuai tujuan (Rachmadi, 2020).

Basis data (*database*) yang digunakan, yaitu MySQL. MySQL merupakan *database server* gratis yang mendukung bahasa *database* SQL sebagai bahasa yang saling berhubungan dalam mengelola *database*. SQL dan MySQL merupakan dua hal yang berbeda. SQL merupakan bahasa pemrograman untuk mengolah *database*, sedangkan MySQL adalah merek *software database management system* (DBMS) untuk mengolah *database* menggunakan bahasa SQL (Fitri, 2020).

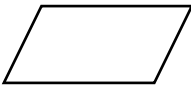
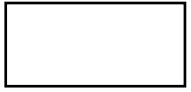
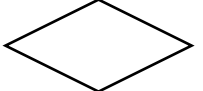
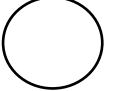
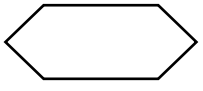
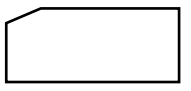



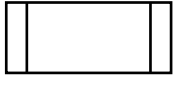

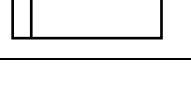
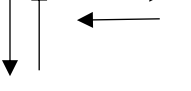
## 2.7 Flowchart

*Flowchart* atau yang disebut juga diagram alir merupakan teknik menggambarkan langkah-langkah suatu masalah menggunakan simbol-simbol tertentu. *Flowchart* akan menunjukkan alur logika dalam sebuah program. Setiap *flowchart* harus dimulai dengan satu *start* dan diakhiri dengan satu atau lebih terminal akhir (Hanief & Jepriana, 2020). Simbol-simbol *flowchart* beserta fungsinya dapat dilihat pada Tabel 2.2.

**Tabel 2.2** Simbol-Simbol pada *Flowchart*

Simbol	Nama	Fungsi
	Terminal	Simbol untuk menyatakan tanda mulai ( <i>start</i> ) dan selesai ( <i>end/stop</i> ).

Tabel 2.2 Simbol-Simbol pada *Flowchart* (Lanjutan)

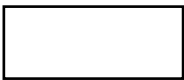

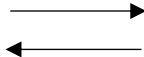

Simbol	Nama	Fungsi
	<i>Input Output</i>	Simbol yang menunjukkan proses <i>input</i> (masukan) dan <i>output</i> (keluaran).
	Proses	Simbol untuk menyatakan suatu tindakan/aksi yang telah di <i>input</i> (masukan).
	Percabangan/pilihan keputusan	Simbol untuk menyakan pemilihan suatu kondisi tertentu.
	<i>Connector</i> (tanda sambung)	Simbol untuk menyatakan sambungan dari suatu proses pada halaman yang sama.
	<i>Preparation</i> (persiapan)	Simbol yang menunjukkan persiapan awal atau pemberian nilai awal.
	<i>Punched Card</i>	Simbol untuk menyatakan masukan dari kartu atau keluaran ditulis ke kartu.
	<i>Punch Tape</i>	Simbol untuk menyatakan masukan dan keluaran dari <i>punched card</i> .
	<i>Document</i>	Simbol untuk menyatakan <i>output</i> tercetak dalam bentuk dokumen.
	Manual <i>input</i>	Simbol untuk menyatakan proses yang dilakukan manual (oleh manusia).
	<i>Subroutine</i>	Simbol untuk menyatakan proses yang telah didefinisikan sebelumnya.
	<i>Stored Data</i>	Simbol untuk menyatakan tempat penyimpanan data.
	<i>Internal Storage</i>	Simbol untuk menyatakan perangkat penyimpanan <i>internal</i> .
	Garis Arah Aliran	Simbol untuk menyatakan tanda untuk menghubungkan simbol yang satu dengan simbol lainnya.

## 2.8 Data Flow Diagram (DFD)

DFD adalah suatu model untuk menggambarkan aliran data. DFD digunakan untuk mengolah data dalam suatu sistem yang akan digunakan secara

rinci (Lise, Fatmasari, Monika, & Solikhun, 2021). Simbol-simbol pada DFD dapat dilihat pada Tabel 2.3.

**Tabel 2.3** Simbol-Simbol pada DFD

Simbol	Nama	Keterangan
	Entitas Eksternal	Simbol ini digunakan untuk menggambarkan asal atau tujuan data.
	Proses	Simbol ini digunakan untuk menggambarkan proses pengolahan atau transformasi data.
	Data <i>Flow</i>	Simbol ini digunakan untuk aliran data dari sumber ke tujuan.
	Data <i>Store</i>	Simbol ini digunakan untuk menggambarkan data <i>flow</i> yang sudah disimpan.

## 2.9 Bantuan Langsung Tunai

BLT adalah bantuan uang yang bersumber dari dana desa untuk mengurangi dampak pandemi terhadap keluarga miskin. Nilai BLT-Dana Desa adalah Rp 600.000 setiap bulan untuk setiap keluarga miskin yang memenuhi kriteria yang telah ditetapkan oleh pemerintah desa, dan diberikan selama tiga bulan. Kemudian Rp 300.000 setiap bulan untuk tiga bulan berikutnya. Bantuan BLT-dana desa ini tidak dikenai pajak. Apabila kebutuhan dana di desa melebihi ketentuan nilai maksimal dari yang telah ditetapkan, maka Kepala Desa dapat mengajukan usulan penambahan alokasi dana desa untuk penerima BLT kepada Bupati setempat. Usulan tersebut harus disertai alasan sesuai keputusan musyawarah desa (Kompak, 2020).