

## **BAB II**

### **LANDASAN TEORI**

#### **2.1 Kemiskinan**

Kemiskinan adalah masalah sosial yang terus ada dalam kehidupan masyarakat. Masalah kemiskinan sangatlah lama, dan untuk waktu yang lama. Unsur utama permasalahannya adalah menyangkut berbagai macam bentuk atau karakteristik kehidupan manusia. Dengan kata lain kemiskinan merupakan masalah kehidupan yang bersifat global artinya masalah kemiskinan telah menjadi perhatian dunia, dan hal tersebut ada di semua negara meskipun dampak kemiskinan sangat bervariasi (Nurwati, 2008).

Badan Pusat Statistik menggunakan konsep kemampuan dalam memenuhi kebutuhan dasar untuk mengukur kemiskinan suatu wilayah. Dengan pendekatan ini, kemiskinan merupakan ketidakmampuan dari sisi ekonomi yang diukur dari sisi pengeluaran dalam memenuhi kebutuhan dasar makanan dan nonmakanan. Sehingga dikatakan penduduk miskin adalah penduduk yang mempunyai rata-rata pengeluaran perkapita perbulan di bawah garis kemiskinan (BPS, 2016).

#### **2.2. DTKS (Data Terpadu Kesejahteraan Sosial)**

Data Terpadu Kesejahteraan Sosial (DTKS) yaitu memuat data dan informasi yang memiliki status ekonomi dari 40% rumah tangga termiskin yang dapat digunakan untuk perencanaan, penganggaran, monitoring, dan mengevaluasi program-program perlindungan sosial. Tujuan dari DTKS yaitu agar

penyelenggaraan kesejahteraan sosial dapat dilaksanakan secara terarah, terpadu, dan berkelanjutan oleh pemerintah. DTKS dapat dianalisis secara sektoral dan spasial ditingkat desa atau kelurahan yang dapat mendukung proses perumusan kebijakan berbasis data atau bukti di daerah (Kemensos, 2020).

Menurut Permensos Nomor 5 Tahun 2019 tentang Pengelolaan Data Terpadu Kesejahteraan Sosial pada Pasal 2 Ayat 2, disebutkan bahwa DTKS meliputi:

1. Pemerlu pelayanan kesejahteraan sosial (PPKS) seperti fakir miskin dan anak terlantar
2. Penerima bantuan dan pemberdayaan sosial seperti keluarga penerima program keluarga harapan (KPM PKH), keluarga penerima program (KPM Sembako)
3. Potensi dan sumber kesejahteraan sosial seperti tenaga kesejahteraan sosial kecamatan (TKSK), Lembaga kesejahteraan sosial (LKS).

Maka dari itu pemerintah mengambil 40% dari populasi rumah tangga yang berada di seluruh Indonesia berdasarkan DTKS, kemudian menjadi desil 1 sampai dengan desil 4 yang lalu dirangking menjadi status kesejahteraan sebagai berikut:

1. Desil 1 yaitu rumah tangga/individu dengan kondisi kesejahteraan 10% terendah di Indonesia atau termiskin.
2. Desil 2 yaitu rumah tangga/individu dengan kondisi kesejahteraan antara 11%-20% terendah di Indonesia atau hampir miskin.

3. Desil 3 yaitu rumah tangga/individu dengan kondisi kesejahteraan antara 21%-30% terendah di Indonesia atau hampir miskin.
4. Desil 4 yaitu rumah tangga/individu dengan kondisi kesejahteraan antara 31%-40% terendah di Indonesia atau rentan miskin (puspensos.kemensos.go.id, 2021).

### **2.3 Data**

Data merupakan fakta-fakta mengenai manusia, tempat, kejadian, dan sesuatu yang sangat penting diorganisasikan (Bernard, 2012). Berdasarkan definisi tersebut maka dapat disimpulkan bahwa dalam penggunaan sehari-hari data adalah kumpulan fakta, catatan atau pernyataan, uraian atau pernyataan tersebut diterima apa adanya, dan dapat berupa angka, teks atau gambar. Terdapat jenis-jenis data yang dapat diuraikan sebagai berikut:

#### **1. Data Kuantitatif**

Data kuantitatif adalah studi berdasarkan *filsafat positivisme*, yang bertujuan untuk memeriksa populasi atau sampel tertentu dan pengambilan sampel secara acak melalui penggunaan alat untuk mengumpulkan data, dan analisis data bersifat statistik (Sugiyono, 2015).

#### **2. Data Kualitatif**

Data kualitatif didasarkan pada penelitian *filsafat postpositivisme* dan digunakan untuk kondisi objek yang alamiah. Peneliti adalah alat utamanya dan pengambilan sampel sumber data dilakukan melalui *purposive sample*,

yaitu pengambilan sampel dilakukan dengan memberikan karakteristik tertentu yang sesuai dengan tujuan penelitian(Sugiyono, 2015).

Terdapat 2 jenis sumber data yaitu data primer dan data sekunder, Berikut merupakan uraian dari data primer dan data sekunder:

### **1. Data Primer**

Data primer merupakan sumber data yang memberikan data langsung ke pengumpul data. Data ini tidak dapat disediakan dalam bentuk kompilasi atau tabel. Pencarian data harus sesuai dengan sumber atau istilah teknis dari responden, yaitu orang yang menjadi objek penelitian(Sugiyono, 2015).

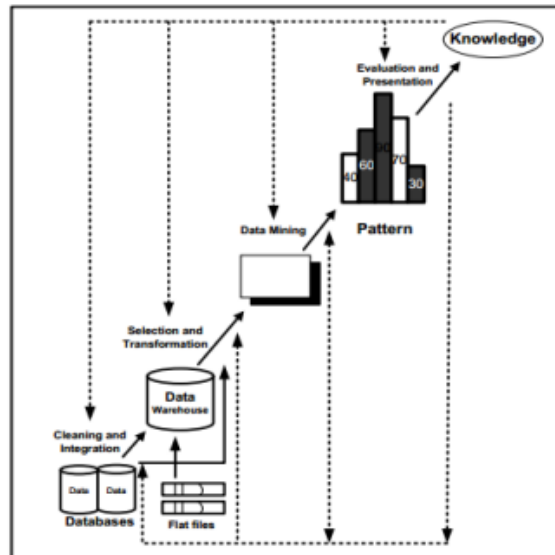
### **2. Data Sekunder**

Data sekunder merupakan sumber yang tidak secara langsung memberikan data ke pengumpul data. Data sekunder adalah data yang mendukung kebutuhan data primer(Sugiyono, 2015).

## **2.4 Data Mining**

*Data mining* merupakan rangkaian proses yang dapat menggali nilai tambah dari suatu kumpulan data berupa pengetahuan, dan pengetahuan tidak diperoleh secara manual dari kumpulan data tersebut. Definisi lain dari *data mining* adalah proses untuk mendapatkan informasi yang berguna dari gudang *database* yang besar. *Data mining* juga didefinisikan sebagai mengekstraksi informasi baru dari sejumlah besar data yang membantu pengambilan keputusan. Istilah *data mining* kadang disebut juga *knowledge discovery* (Buaton et al., 2018). Adapun langkah-

langkah untuk melakukan data mining dapat dibagi menjadi beberapa tahapan pada *Knowledge Discovery in Database* (KDD) yang ditunjukkan di Gambar 2.1.



**Gambar 2.1** Tahap-tahap KDD (Han & Kamber, 2006)

Tahap-tahap *Knowledge Discovery in Database* (KDD) ada 6 yaitu:

1. Pembersihan Data (*Data Cleaning*)

Pembersihan data adalah proses menghilangkan noise dan data yang tidak konsisten atau data tidak relevan. Seringkali data yang diperoleh baik dari database memiliki input yang tidak sempurna seperti data yang hilang, data yang tidak valid atau hanya sekedar salah ketik.

2. Integrasi Data (*Data Integration*)

Integrasi data adalah konsolidasi data dari banyak *database* yang berbeda ke dalam satu *database* baru. Tidak jarang data yang diperlukan untuk penambahan

data tidak hanya berasal dari satu *database*, tetapi juga berasal dari beberapa *database*.

### 3. Seleksi Data (*Data Selection*)

Data dalam *database* seringkali tidak semuanya digunakan, sehingga hanya data yang sesuai untuk analisis yang akan diambil dari *database*.

### 4. Transformasi Data (*Data Transformation*)

Data yang digabungkan ke dalam format yang sesuai untuk diproses dalam penambangan data. Beberapa metode penambangan data memerlukan format data khusus sebelum dapat diterapkan. Misalnya, beberapa metode standar seperti analisis asosiasi dan clustering hanya dapat menerima input data kategorikal maka diperlukan proses transformasi data.

### 5. Proses *Mining*

Merupakan proses utama ketika metode diterapkan untuk menemukan pengetahuan yang berharga dan tersembunyi dari data.

### 6. Evaluasi Pola (*Pattern Evaluation*)

Mengidentifikasi pola-pola menarik kedalam pengetahuan berbasis penemuan. Pada langkah ini, hasil dari teknik *data mining* berupa model tipikal dari prediktif prediksi dievaluasi untuk menentukan apakah hipotesis yang ada benar-benar terpenuhi.

### 7. Presentasi pengetahuan (*Knowledge Presentation*)

Merupakan visualisasi dan penyajian pengetahuan tentang metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna (Han & Kamber, 2006).

## **2.5 Clustering**

*Clustering* atau analisis pengelompokan merupakan metode pengelompokan berdasarkan kesamaan. Dalam *clustering*, terjadi proses pembagian data dari himpunan tertentu menjadi beberapa kelompok yang memiliki karakteristik data yang mirip dengan data yang dikelompokkan lainnya (Ali Hasymi et al., 2021). *Clustering* adalah metode tanpa arahan (*unsupervised*) dan bertujuan untuk mencari karakteristik pola data secara umum dalam *database*. Pendekatan yang dapat digunakan untuk mengembangkan *clustering* adalah metode *partitioning*. Metode *partitioning* mengalokasikan data ke sebuah *cluster* sehingga setiap *cluster* berisi setidaknya satu bagian data dan setiap data termasuk pada salah satu kelompok (Cahyati et al., 2017).

## **2.6 K-Means**

Terdapat dua jenis *clustering* yang biasa digunakan dalam proses pengelompokan data yaitu Hierarchical dan non-Hierarchical. K-Means merupakan salah satu metode data *clustering* non- Hierarchical atau *Partitioning clustering*. Algoritma K-Means mencoba untuk mengelompokkan data yang ada ke dalam beberapa kelompok, sehingga data dalam satu kelompok mempunyai karakteristik yang sama dan berbeda dengan data pada kelompok lain (Perdana Windarto, 2017). Huruf K diartikan sebagai jumlah *cluster* yang akan dibuat. Selain itu, meskipun nilai rata-rata adalah nilai sementara (juga disebut *centroid*) dalam *cluster*, nilai K ditentukan secara acak. Setiap data menggunakan rumus Euclidean untuk

menghitung jarak ke semua centroid ini untuk mendapatkan jarak terdekat ke centroid dari setiap data (Sari et al., 2020). Berikut merupakan proses Algoritma metode K-Means (Ali Hasymi et al., 2021):

1. Tentukan jumlah *cluster* yang dibutuhkan.
2. Tentukan centroid awal (pusat *cluster*), yang biasanya dipilih secara acak.
3. Gunakan centroid awal yang ditentukan dengan rumus jarak Euclidean untuk menghitung jarak terdekat tiap objek dan tentukan jarak terdekat objek dengan centroid.

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

Keterangan:

$x_i$ : Objek x ke-i

$y_i$ : Objek y ke-i

n: Banyaknya objek

4. Tentukan centroid baru dengan menghitung rata-rata pada masing-masing *cluster* menggunakan rumus  $ckl = x_{1l} + x_{2l} + \dots + x_{pl} / p$ .

Keterangan:

$C_{kl}$ : Pusat *cluster* baru

$x_{1l}$ : Anggota *cluster*

p: Banyaknya data



5. Hitung pada setiap objek menggunakan *cluster* baru. Jika objek tidak berpindah *cluster*, maka proses *clustering* selesai. Atau ulangi langkah 3 sampai pusat *cluster* tidak berubah.

## 2.7 Davies Bouldin Index (DBI)

Davies Bouldin Index (DBI) merupakan metode yang digunakan untuk menentukan evaluasi *cluster*. Davies Bouldin Index (DBI) merupakan metode evaluasi internal yang mengukur evaluasi *cluster* dalam metode *clustering* berdasarkan nilai kohesi dan separasi. Dalam *clustering*, kohesi didefinisikan sebagai jumlah dari kedekatan data terhadap centroid dari *cluster* yang diikuti. Sedangkan separasi didasarkan pada jarak antar centroid dari *cluster* (Nabila et al., 2021). Davies-bouldin index (DBI) diperkenalkan oleh David L. Davies dan Donald W. Bouldin (1979) yang tujuannya untuk melakukan evaluasi *cluster*. Validitas internal yang dicapai adalah seberapa baik melakukan *clustering* dengan menghitung kuantitas dan kriteria atau karakteristik yang diperoleh dari kumpulan data (Tempola et al., 2020).

*Sum of square within cluster* (SSW) merupakan persamaan yang digunakan untuk mengetahui matrik kohesi dalam sebuah *cluster* ke-i yang dirumuskan sebagai berikut:

$$SSW = \frac{1}{m_i} \sum_{j=i}^{m_i} d(x_j, c_j)$$

Keterangan:

m = Jumlah data dalam *cluster* ke-i

c = Centroid *cluster* ke-i

$d$  = Jarak

*Sum of square between cluster* (SSB) merupakan persamaan yang digunakan untuk mengetahui separasi antar *cluster* yang dihitung menggunakan persamaan:

$$SSB = d(c_i, c_j)$$

Keterangan:

$d$  = Jarak

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

Setelah nilai kohesi dan separasi diperoleh, kemudian dilakukan pengukuran rasio ( $R_{ij}$ ) untuk mengetahui nilai perbandingan antara *cluster* ke- $i$  dan *cluster* ke- $j$ . *Cluster* yang baik adalah *cluster* yang memiliki nilai kohesi sekecil mungkin dan separasi yang sebesar mungkin. Nilai rasio dihitung menggunakan persamaan sebagai berikut:

$$RASIO = \frac{SSW_i + SSW_j}{SSB_{ij}}$$

Keterangan:

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

$SSW_j$  = *Sum of Square Within- Cluster* pada centroid  $j$

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

Nilai rasio yang diperoleh tersebut digunakan untuk mencari nilai davies bouldin index (DBI) dari persamaan berikut:

$$DBI = \frac{1}{K} \sum_{i=1}^K \max_{i \neq j} (R_{ij})$$

Keterangan :

$K$  = Jumlah *cluster* yang digunakan

R = Rasio Maksimal

Dari persamaan tersebut,  $k$  merupakan jumlah *cluster* yang digunakan. Semakin kecil nilai DBI yang diperoleh (non-negatif  $\geq 0$ ), maka semakin baik *cluster* yang diperoleh dari pengelompokan K-means yang digunakan (Nabila et al., 2021).

### **2.8 State of The Art Penelitian Terkait**

Tabel 2.1 menunjukkan perbandingan penelitian sebelumnya yang berhubungan dengan fokus penelitian komparasi algoritma *clustering*. Terdapat beberapa kesamaan serta perbedaan dari masing-masing penelitian. Hal ini dilihat dari data serta penggunaan metode dan algoritmanya.

**Tabel 2.1** *State of The Art* Penelitian Terkait

No	Nama Pengarang	Judul	Tahun	Isi Ringkasan	Hasil
1.	Elly Muningsih, Ina Maryani, Vembria Rose Handayani	Penerapan Metode K-Means dan Optimasi Jumlah Cluster dengan Index Davies Bouldin untuk Clustering Provinsi Berdasarkan Potensi Desa	2021	Potensi Desa atau Podes sejak tahun 1980 sudah dilakukan pendataan oleh Badan Pusat Statistik (BPS). Sejak saat itu, Podes dilaksanakan secara rutin sebanyak 3 kali dalam kurun waktu sepuluh tahun untuk mendukung kegiatan Sensus Penduduk, Sensus Pertanian, ataupun Sensus Ekonomi. Dengan demikian, fakta penting terkait ketersediaan infrastruktur dan potensi yang dimiliki oleh setiap wilayah dapat dipantau perkembangannya secara berkala dan terus menerus. Penelitian kali ini akan mengaplikasikan Indeks Davies Bouldin (DBI) sebagai salah satu cara optimasi jumlah cluster untuk mengelompokkan provinsi berdasar potensi desa dengan	Hasil pengolahan data didapatkan bahwa K-Means dengan evaluasi nilai DBI menghasilkan optimasi jumlah cluster. Hasil clustering menghasilkan pengelompokan propinsi berdasarkan jumlah industri sebagai potensi desanya dengan kategori sedikit, sedang dan banyak.

No	Nama Pengarang	Judul	Tahun	Isi Ringkasan	Hasil
				banyaknya jenis industri yang dimiliki wilayahnya.	
2.	Parjito, Permata	Penerapan Data Mining untuk Clustering Data Penduduk Miskin Menggunakan Metode K-Means	2021	Desa SukaBhakti Kecamatan Gedung Aji Baru Kabupaten Tulang Bawang Provinsi Lampung. Penanggulangan kemiskinan telah dilakukan berbagai cara oleh pemerintah seperti halnya pemberian bantuan kepada masyarakat miskin. Tetapi kenyataannya masalah kemiskinan belum dapat ditanggulangi. Penggunaan data mining dengan Teknik cluster K-Means dapat mengelompokkan data masyarakat untuk mendapatkan cluster yang seharusnya dapat menerima bantuan sosial dari pihak pemerintah.	Berdasarkan hasil perhitungan validitas yang dilakukan, maka didapatkan hasil untuk jumlah cluster 1 (130 data) dan cluster 2 (682 data). Pengujian data ini dilakukan dengan menggunakan metode Davies Bouldin Index (DBI) dihasilkan jumlah cluster yang cocok adalah 2 cluster dengan nilai 0,1643 yang berarti cluster tersebut cocok untuk digunakan mengelompokkan data masyarakat.
3.	Arief Wibowo, Moh Makruf, Inge Virdyana, Farah Chikita Venna	Penentuan laster Koridor TransJakarta dengan Metode	2021	Pandemi Covid-19 menyebabkan perubahan pola pada aktivitas masyarakat. Salah satunya Pembatasan Sosial Berskala Besar	Perhitungan menggunakan tools Microsoft Excel, Rapid Miner, dan Bahasa pemrograman Python. Hasil yang didapatkan menggunakan DBI

No	Nama Pengarang	Judul	Tahun	Isi Ringkasan	Hasil
		Majority Voting pada Algoritma Data Mining		(PSBB) diimplementasikan guna menekan angka penularan virus tersebut. Hal ini berpengaruh terhadap moda transportasi. Moda transportasi membuat aturan baru untuk mengurangi jumlah kapasitas penumpang disetiap armada salah satunya layanan TransJakarta. Penelitian ini akan melakukan clustering untuk pengelompokan sebelum dan selama masa pandemi Covid-19. Menggunakan metode K-Means dan K-Medoids untuk menghasilkan perhitungan yang akurat.	0,184, dan selama masa pandemi, jumlah cluster optimal adalah 2 cluster dengan nilai DBI 0,188. Sementara itu, bila menggunakan algoritma K-Medoids sebelum masa pandemi, jumlah cluster optimal adalah 3 cluster dengan nilai DBI 0,200, dan selama masa pandemi, jumlah cluster optimal adalah 4 cluster dengan nilai DBI 0,190. Penentuan cluster akhir menggunakan pendekatan metode majority voting dari semua tools yang digunakan.
4.	Zulfa Nabila, Auliya Rahman Isnain, Permata, Zaenal Abidin	Analisis Data Mining untuk Clustering Kasus Covid-19 di Provinsi Lampung dengan	2021	Dinas Kesehatan Provinsi Lampung dan Badan Perencanaan Pembangunan Daerah Provinsi Lampung merincikan lima kriteria zonasi berdasarkan warna. Namun belum dilakukan perhitungan secara umum dan detail hanya berdasarkan pengumpulan data dan	Clustering data kasus Covid-19 di Provinsi Lampung dengan menggunakan algoritma K-Means. dengan atribut Kabupaten/Kota, Suspek, Probable, Konfirmasi Positif, Selesai Isolasi, dan Kematian yang digunakan dalam proses perhitungan dan membagi

No	Nama Pengarang	Judul	Tahun	Isi Ringkasan	Hasil
		Algoritma K-Means		kajian secara maupun analisis dari tim pakar Gugus Tugas Nasional penentuan zona. Dilakukan penerapan data mining pada kasus Covid-19 di Provinsi Lampung menggunakan algoritma K-Means Clustering.	data ke dalam 4 cluster yang dikategorikan sebagai Zona Merah, Zona Orange, Zona Kuning dan Zona Hijau. Serta validasi menggunakan Davies-Bouldin Index (DBI). Hasil Validasi menggunakan manual lebih baik dan memiliki perbedaan dengan tools Rapid Miner, tetapi hasil kedua perhitungan tersebut sama-sama mendekati 0 yang berarti klaster yang dievaluasi menghasilkan klaster yang baik.
5.	Suhartini, Ria Yuliani.	Penerapan Data Mining Untuk Mengeluster Data Penduduk Miskin Menggunakan Algoritma K-Means Di Dusun Bagik Endep Suk	2021	Kemiskinan yang terjadi pada masyarakat salah satunya Dusun Endep Desa Sukamulia Timur disebabkan oleh suatu kondisi ketidakmampuan kepala keluarga secara ekonomi. Berdasarkan kondisi tersebut diperlukan pengklasteran untuk membantu Pemerintah desa dalam pengelompokan keluarga miskin	Data yang digunakan adalah data Penduduk Sukamulia Timur pada tahun 2019 yang berjumlah 200 data dengan 9 atribut yaitu nama penduduk, pekerjaan, penghasilan/bulan, jumlah anak yang sekolah SD, jumlah anak yang sekolah SMP, jumlah anak yang sekolah SMA, jumlah anak yang kuliah, jumlah anak yang tidak sekolah serta jumlah anggota

No	Nama Pengarang	Judul	Tahun	Isi Ringkasan	Hasil
				dengan menggunakan metode K-Means	keluarga. Berdasarkan hasil pengujian yang dilakukan dengan menerapkan algoritma K-Means didapatkan hasil dengan Cluster 1 berjumlah 18 penduduk dengan kriteria Penduduk ekonomi tinggi, Cluster 2 berjumlah 72 Penduduk dengan kriteria Penduduk ekonomi sedang, dan Cluster 3 berjumlah 110 penduduk dengan kriteria Penduduk ekonomi rendah.
6.	Irmanita Nasution, Agus Perdana Windarto, M. Fauzan.	Penerapan Algoritma K-Means Dalam Pengelompokan Data Penduduk Miskin Menurut Provinsi	2020	Menurut Badan Pusat Statistik (BPS) mencatat bahwa pada September 2018 mencatat bahwa jumlah penduduk miskin di Indonesia termasuk ke dalam kategori negara dengan penduduk miskin terbesar di dunia. Banyak indikator yang menyebabkan data tingkat kemiskinan tiap provinsi berbeda. Diperlukan pengclusteran untuk mengetahui provinsi mana yang termasuk dengan cluster	Jumlah record yang digunakan sebanyak 34 provinsi dengan menghasilkan 2 cluster yakni cluster tinggi sebanyak 8 provinsi dan cluster rendah sebanyak 26 provinsi. Berdasarkan hasil pengujian k-means untuk kasus persentase data penduduk miskin menggunakan tools RapidMiner versi 5.3 diperoleh hasil yang sama dengan analisis perhitungan algoritma dimana diperoleh 8 provinsi dengan cluster tinggi yang menjadi pusat



No	Nama Pengarang	Judul	Tahun	Isi Ringkasan	Hasil
				provinsi tingkat kemiskinan tinggi dan rendah.	perhatian bagi pemerintah dalam melakukan sosialisasi dan pemetaan dalam pemberian bantuan pada provinsi tersebut.
7.	Herlawati, Rahmadya Trias Handayanto	Penggunaan Matlab dan Python dalam Klasterisasi Data	2020	Organisasi harus mengeksplorasi data lewat proses klasterisasi data, baik data historis maupun data dari internet. Terkadang data perlu dilakukan klasterisasi ulang untuk mencocokkan dengan kondisi yang sebenarnya. Oleh karena itu, perlu disiapkan perangkat pendukung klasterisasi. Dalam penelitian, ini metode K-Means dipilih untuk membandingkan dua bahasa komputasi teknis yaitu Matlab dan Python yang saat ini banyak diminati para peneliti dan dapat digunakan oleh organisasi yang membutuhkan proses klasterisasi.	Berdasarkan data eksperimen yang diperoleh di dalam penelitian, kedua Bahasa pemrograman tersebut mampu mengklasterisasi data. Matlab dan Python keduanya memiliki cukup pustaka (library) dan toolbox untuk membantu pengguna mengklasterisasi data dan mempresentasikan grafik.
8.	Nisa Hanum Harani, Cahyo	Segmentasi Pelanggan Produk Digital Sevice	2020	Pada penelitian ini dilakukan analisa karakteristik pelanggan sebagai dasar penetapan	Hasil simulasi terbaik berdasarkan evaluasi cluster dengan presentasi data train 50% dan data test 50%

No	Nama Pengarang	Judul	Tahun	Isi Ringkasan	Hasil
	Prianto, Fikri Aldi Nugraha	Indihome Menggunakan Algoritma K-Means Berbasis Python		segmentasi pelanggan dan customer profiling pelanggan produk digital service add on Indihome menggunakan Algoritma K-Means, penentuan jumlah cluster terbaik dilakukan menggunakan metode Elbow dan diperoleh nilai $K = 3$ . Data pelanggan dikelompokkan menjadi tiga segmen yaitu data train dan data test 80% - 20%, 70% - 30% dan 50% - 50%. Data yang digunakan sebanyak 1392 record. Evaluasi cluster dilakukan menggunakan metode Silhouette Index, Davies Bouldin Index dan Calinski Harabasz Index	dimana customer profiling dilihat dengan menganalisis anggota masing-masing cluster dari simulasi ketiga dimana cluster 0 memiliki anggota 396 pelanggan dengan kategori pelanggan dengan keuntungan terbesar bagi perusahaan, cluster 1 memiliki anggota 286 pelanggan dengan kategori memiliki potensi besar dalam memberikan keuntungan bagi perusahaan, dan cluster 2 memiliki anggota 14 pelanggan dengan kategori pelanggan yang memberikan keuntungan lebih sedikit daripada biaya untuk memberikan pelayanan.
9.	Rofika Julianti Hablum, Amal Khairan, Rosihan	Clustering Hasil Tangkap Ikan di Pelabuhan Perikanan Nusantara (PPN) Ternate Menggunakan	2019	PPN Ternate adalah salah satu Pelabuhan Perikanan Nusantara (PPN), banyak ragam jenis ikan hasil tangkapan di PPN Ternate yang belum diketahui jenis ikan yang unggul setiap tahunnya, disamping itu juga belum	Hasil yang didapatkan menggunakan metode DBI (Davies Bouldin Index) dengan menggunakan cluster 2 sampai cluster 5 untuk menentukan perbandingan. Hasil yang didapat pad cluster 2 tahun 2015 = 0,395,

No	Nama Pengarang	Judul	Tahun	Isi Ringkasan	Hasil
		Algoritma K-Means		dikelompokkan jenis ikan apa yang paling banyak dan sedikit hasil tangkapnya. Maka dibutuhkan clustering hasil tangkap ikan di PPN Ternate menggunakan Algoritma K-Means	tahun 2016 = 0,276, dan tahun 2017 = 0,54. Hasil penelitian menggunakan 18 data ikan dengan menghasilkan 2 cluster, dimana ada 16 data pada k=1 dan 2 data pada k=2.
10.	Firman Tempola, Achmad Fuad Assagaf	Clustering of Potency of Shrimp in Indonesia with K-Means Algorithm and Validation of Davies-Bouldin Index	2018	Di antara negara-negara ASEAN, Indonesia adalah yang pertama peringkat produsen udang. Udang di Indonesia adalah salah satu produsen utama di Kawasan ASEAN, ternyata potensi masing-masing provinsi belum dijelaskan secara rinci, sehingga perlu dikaji lebih lanjut terkait produksi udang di setiap provinsi di Indonesia. Penelitian ini akan melakukan pemetaan terhadap potensi udang per provinsi di Indonesia. Metode diterapkan dalam melakukan pemetaan adalah dengan menggunakan K-Means Clustering algoritma, dengan	Hasil penelitian menunjukkan bahwa posisi clustering berada pada iterasi keenam untuk dua cluster, cluster 1 menjadi yang terbaik cluster dengan nilai SSW 0.13507 dan nilai DBI 0,96. Sedangkan untuk pembagian 3 cluster hanya terjadi satu kali dan iterasi, cluster pertama menjadi cluster terbaik dengan nilai SSW 0,095 dan nilai DBI 0,76. Jadi, dari distribusi jumlah cluster yang terbentuk kemudian pembagian menggunakan pembagian 3 cluster lebih baik dari pada pembagian 2 cluster.

No	Nama Pengarang	Judul	Tahun	Isi Ringkasan	Hasil
				validasi clustering yang digunakan adalah Davies-Bouldin indeks.	
11.	M. Nanda Variestha Waworuntu, Muhammad Faisal Amin	Penerapan Metode K-Means Pemetaan Calon Penerima JAMKESDA	2018	Seksi Kesejahteraan Sosial (KESSOS) di Kelurahan Kemuning terdapat pelayanan masyarakat miskin untuk menerima bantuan Jaminan Kesehatan (JAMKESDA). Selama ini bagian Seksi KESOS di Kelurahan Kemuning belum dilakukannya pengelompokan tingkat kemiskinan agar penerima bantuan tepat sasaran. Oleh sebab itu, perlu dilakukan pengelompokan kemiskinan menggunakan metode K-Means untuk mencegah ketidaktepatan sasaran penerima JAMKESDA.	Penerapan metode K-Means diimplementasikan pada aplikasi RapidMiner yang dibuat dengan 2 kluster. Penelitian ini menggunakan sebanyak 440 sampel data. Dari hasil perhitungan Davies Bouldin Index (DBI) diperoleh nilai penentuan jumlah kluster dengan nilai k=2 (0.243), k=3 (0.256), dan k=4 (0.275). Nilai yang digunakan adalah k=2 karena nilai tersebut mendekati 0.
12.	Falentino Sembiring, Saeful Bahri Rizqi, M. Abdul Aziz, Dado Firmasyah	Analisis Pemetaan Tingkat Pengangguran di Pulau Jawa dan	2017	Tingkat pengangguran di Indonesia dari tahun ke tahun meningkat dan menurun seiring pertambahan jumlah penduduk. Tingkat pengangguran tertinggi di	Hasil pengelompokan tingkat pengangguran pada pengujian koding bahasa pemrograman python adalah 1 wilayah tingkat pengangguran rendah (C1), 3

No	Nama Pengarang	Judul	Tahun	Isi Ringkasan	Hasil
		Bali dengan Metode K-Means		Indonesia dari tahun 2014 hingga 2019 berada di pulau Jawa tepatnya di provinsi Banten. Kemudian tingkat pengangguran terendah dari tahun 2014 hingga 2019 berada provinsi Bali. Penelitian ini menggunakan metode K-Means Clustering dan implementasi program dengan bahasa pemrograman python. Dengan adanya penelitian ini diharapkan dapat memberikan solusi untuk pemetaan tingkat pengangguran di pulau Jawa dan Bali.	wilayah tingkat pengangguran sedang (C2) dan 3 wilayah lagi dikategorikan sebagai tingkat pengangguran tinggi (C3).
13.	Purnami Yuli S., Baharudin Machmud, M. Iswahyudi Subroto, Zhazhen Regina P., Yunita Hilda S., Edy Widodo.	Analisis Cluster Hirarki dan Pemetaan Kemiskinan Daerah Istimewa Yogyakarta Tahun 2015	2016	Penelitian ini memakai data kemiskinan yang terbagi menjadi 4 jenis yaitu Desil 1, Desil 2, Desil 3, dan Desil 4 di D.I.Yogyakarta tahun 2015. Dengan memakai analisis cluster hirarki dan pemetaan untuk mengetahui pengelompokkan kemiskinan dan visualisasi sebaran kemiskinan pada desa yang menjadi objek	Pada data kemiskinan yang terdapat di D.I.Yogyakarta dari Desil 1, Desil 2, Desil 3, dan Desil 4 mendapatkan hasil kelompok tertinggi pada desil 3 sebanyak 100.343 rumah tangga dan data kemiskinan tersebut dikelompokkan menjadi 5 cluster yaitu prioritas 1,

No	Nama Pengarang	Judul	Tahun	Isi Ringkasan	Hasil
				penelitian karena pemerintah saat ini telah memfokuskan pembangunannya ke daerah-daerah.	prioritas 2, prioritas 3, prioritas 4, dan prioritas 5.

Pada *state of the art* tersebut belum ada penelitian yang meneliti data kemiskinan yang terjadi di Kota Tasikmalaya dengan menggunakan Davies Bouldin Index (DBI) untuk optimasi *cluster* terbaik yang akan digunakan untuk menentukan *cluster* daerah kemiskinan di tingkat kelurahan menggunakan atribut desil 1 – desil 4 pada penelitian sebelumnya.

## 2.9 Matriks Penelitian

**Tabel 2.2** Tabel Matriks Penelitian

No	Penulis/ Tahun	Judul	Ruang Lingkup								
			Algoritma/Metode			Tujuan		Objek			
			K- Means	Average Lingked	Fuzzy C- Means	Klasifi kasi	Clustering	Python	Rstudio	Weka	Rapid Miner
1.	(Yuli et al., 2016)	Analisis Cluster Hirarki dan Pemetaan Kemiskinan Daerah Istimewa Yogyakarta Tahun 2015	-	√	-	-	√	-	√	-	-
2.	(Sembiring et al., 2019)	Analisis Pemetaan Tingkat Pengangguran Di Pulau Jawa Dan Bali Dengan Metode K-Means	√	-	-	-	√	√	-	-	-
3.	(Rifa'i et al., 2019)	Penggunaan Metode K-Means Pada Analisa Dan Klasifikasi Capres 2019 Di Twitter	√	-	-	√	-	√	-	-	-

No	Penulis/ Tahun	Judul	Ruang Lingkup								
			Algoritma/Metode			Tujuan		Objek			
			K- Means	Average Lingked	Fuzzy C- Means	Klasifi kasi	Clustering	Python	Rstudio	Weka	Rapid Miner
4.	(Harani et al., 2020)	Segmentasi Pelanggan Produk Digital Service Indihome Menggunakan Algoritma K-Means Berbasis Python	√	-	-	-	√	√	-	-	-
5.	(Wardoy o & Tripuspi ta, 2020)	Penentuan Cluster Optimum pada Tingkat Pengangguran dan Tingkat Kemiskinan di Jawa Timur Menggunakan Algoritma Fuzzy C- Means	-	-	-	-	-	-	√	-	-
6.	(Parjito & Permata, 2021)	Penerapan Data Mining untuk Clustering Data Penduduk Miskin	√	-	-	-	√	-	-	√	-



No	Penulis/ Tahun	Judul	Ruang Lingkup								
			Algoritma/Metode			Tujuan		Objek			
			K- Means	Average Lingked	Fuzzy C- Means	Klasifi kasi	Clustering	Python	Rstudio	Weka	Rapid Miner
		Menggunakan Metode K-Means									
7.	(Muning sih et al., 2021)	Penerapan Metode K-Means dan Optimasi Jumlah Cluster dengan Index Davies Bouldin untuk Clustering Provinsi Berdasarkan Potensi Desa	√	-	-	-	√	-	-	-	√
8.	(Fanny Aulia Fadillah, 2022)	Penerapan Algoritma K-Means dan Penentuan Cluster Optimum untuk Mengelompokan Daerah Kemiskinan Berdasarkan Kelurahan Di Kota Tasikmalaya	√	-	-	-	√	√	-	-	-

## 2.10 Relevansi Penelitian

**Tabel 2.3** Relevansi Penelitian

Penelitian	(Sembiring et al., 2019)	(Harani et al., 2020)	(Fanny, 2022)
Judul	Analisis Pemetaan Tingkat Pengangguran Di Pulau Jawa Dan Bali Dengan Metode K-Means	Segmentasi Pelanggan Produk Digital Service Indihome Menggunakan Algoritma K-Means Berbasis Python	Penerapan Algoritma K-Means dan Penentuan Cluster Optimum untuk Mengelompokan Daerah Kemiskinan di Kelurahan Kota Tasikmalaya
Masalah Penelitian	Melakukan analisis <i>clustering</i> menggunakan K-Means berdasarkan tingkat pengangguran di pulau Jawa dan Bali	Melakukan analisis <i>clustering</i> data pelanggan yang sebelumnya dilakukan penentuan jumlah cluster terbaik dilakukan menggunakan metode Elbow dan diperoleh nilai K = 3 sehingga data pelanggan dikelompokkan kedalam 3 simulasi dengan persentase data train dan data test 80% - 20%, 70% - 30% dan 50% - 50%.	Melakukan analisis <i>clustering</i> menggunakan K-Means <i>clustering</i> dengan Davies Bouldin Index (DBI) pada data penduduk kemiskinan berdasarkan desil 1- desil 4 menurut Kelurahan dan Kecamatan di Kota Tasikmalaya
Objek Penelitian	<i>Clustering</i> pada data pengangguran dilakukan dengan algoritma K-Means menggunakan bahasa pemrograman <i>python</i>	<i>Clustering</i> pada data pelanggan dilakukan dengan algoritma K-Means dan dilakukan evaluasi cluster dengan metode Silhouette Index, Davies Bouldin Index dan Calinski Harabasz Index menggunakan bahasa pemrograman <i>python</i>	Clustering pada data kemiskinan dilakukan dengan algoritma K-Means dan evaluasi cluster Davies Bouldin Index (DBI) perhitungan manual dan bahasa pemrograman <i>Python</i>
Algoritma /Metode	K-Means	K-Means dan evaluasi cluster dengan metode Silhouette Index, Davies	K-Means dan evaluasi cluster Davies Bouldin Index (DBI)

Penelitian	(Sembiring et al., 2019)	(Harani et al., 2020)	(Fanny, 2022)
		Bouldin Index dan Calinski Harabasz Index	
Implementasi	<p>Hasil klasterisasi tingkat pengangguran pada pengujian K-Means <i>clustering</i> menggunakan 3 <i>cluster</i> yaitu terdapat 1 wilayah tingkat pengangguran rendah (C1), 3 wilayah tingkat pengangguran sedang (C2) dan 3 wilayah lagi dikategorikan sebagai tingkat pengangguran tinggi (C3)</p>	<p>Hasil dari penelitian menunjukkan bahwa simulasi ketiga merupakan simulasi terbaik berdasarkan evaluasi cluster dengan presentasi data train 50% dan data test 50% dimana customer profiling dilihat dengan menganalisis anggota masing-masing cluster dari simulasi ketiga dimana cluster 0 memiliki anggota 396 pelanggan dengan kategori pelanggan yang memberikan keuntungan terbesar bagi perusahaan, cluster 1 memiliki anggota 286 pelanggan dengan kategori pelanggan yang tanpa disadari memiliki potensi besar dalam memberikan keuntungan bagi perusahaan, dan cluster 2 memiliki anggota 14 pelanggan dengan kategori pelanggan yang memberikan keuntungan lebih sedikit daripada biaya untuk memberikan pelayanan.</p>	<p>Hasil klasterisasi kemiskinan menggunakan algoritma K-Means <i>clustering</i> menghasilkan 2 <i>cluster</i>, 3 <i>cluster</i>, 4 <i>cluster</i>, dan 5 <i>cluster</i>. Evaluasi cluster menggunakan Davies Bouldin Index (DBI) yang paling optimal yaitu 4 <i>cluster</i></p>

