

## **BAB II**

### **TINJAUAN PUSTAKA**

#### **2.1 Landasan Teori**

##### **2.1.1 Pemilihan Umum**

Pemilihan umum atau pemilu merupakan momen krusial dalam sebuah proses partisipasi politik dimana rakyat memegang peranan penting dalam menentukan pemimpin suatu negara (Galuh Larasati et al., 2023). Menurut Nur Wardhani (2018) pemilihan umum merupakan sarana demokrasi serta wujud dari kedaulatan rakyat untuk meligitimasi pemimpin sebagai wakil rakyat yang bertanggung jawab demi kesejahteraan rakyat. Sedangkan menurut Undang-Undang No.7 Tahun 2017 tentang Pemilihan Umum pasal 1 ayat 1 berbunyi “Pemilihan Umum, selanjutnya disingkat Pemilu, adalah sarana pelaksanaan kedaulatan rakyat yang diselenggarakan secara langsung, umum, bebas, rahasia, jujur, dan adil dalam Negara Kesatuan Republik Indonesia berdasarkan Pancasila dan Undang-Undang Dasar Negara Republik Indonesia Tahun 1945.”

Menurut hukum Indonesia, pemilihan umum diadakan setiap lima tahun dan melibatkan partisipasi warga negara dalam memilih pejabat pemerintah di tingkat nasional maupun daerah (Muzakkir et al., 2021). Sistem pemilu di Indonesia sangat menekankan prinsip-prinsip sering disebut dengan asas Luber Jurdil yaitu langsung, umum, bebas, rahasia, jujur, dan adil. Dalam penyelenggaraannya dibentuk suatu lembaga khusus yaitu Komisi Pemilihan Umum (KPU) yang mempunyai tugas melaksanakan pemilu (UU No.15 Tahun 2011 Pasal 1).

Menurut UU No.7 tahun 2017, Pemilih adalah warga negara Indonesia yang telah berusia minimal 17 tahun, baik sudah menikah maupun belum pernah menikah. Dalam pemilu, pemilih biasanya dibagi menjadi tiga kategori pemilih yaitu pemilih tetap, pemilih tambahan dan pemilih khusus (Nugraha, 2019).

- a. Pemilih tetap adalah pemilih yang sudah terdaftar di KPU dan terdata dalam DPT (daftar pemilih tetap). KPU melengkapi pemilih kategori ini dengan surat undangan memilih atau C6.
- b. Pemilih tambahan adalah kategori pemilih yang pindah memilih ke TPS lain dari TPS yang sudah ditentukan. Sesuai Pasal 210 Undang-Undang Nomor 7 Tahun 2017, pemilih tambahan harus mendaftar paling lambat 30 hari sebelum pemungutan suara. Pada saat mencoblos, pemilih tambahan membawa Daftar Pemilih (A5), KTP, dan dokumen identitas lainnya.
- c. Pemilih khusus adalah pemilih yang tidak terdaftar dalam DPT (Daftar Pemilih Tetap) dan DPTb (Daftar Pemilih Tambahan). Pemilih khusus dapat memilih dengan membawa KTP atau tanda pengenal lainnya ke TPS. Hak memilih diberikan oleh KPPS dengan mempertimbangkan adanya surat suara di TPS.

### **2.1.2 Pemilih Abstain atau Golongan Putih**

Pemilih abstain atau yang lebih dikenal dengan golongan putih (Golput) dalam sejarah politik Indonesia telah menarik perhatian berbagai pihak selama puluhan tahun. Golput sejak tahun 1970 hingga saat ini turut meramaikan kelangsungan politik Indonesia (Nyarwi, 2009). Abstain atau Golongan putih (disingkat golput) adalah istilah politik ketika seorang pemilih dalam proses pemungutan suara tidak memilih atau tidak memilih calon atau seorang pemilih

datang ke TPS tetapi tidak memilih sampai selesainya proses pemilu (Nyarwi, 2009).

Adapun jenis-jenis Golput yang ada di Indonesia menurut (Nyarwi, 2009) sebagai berikut:

- a. Golput Teknis adalah mereka yang tidak menggunakan hak pilihnya, contohnya tidak dapat datang ke tempat pemungutan suara (TPS) karena suatu alasan, misalnya salah mencoblos sehingga suaranya dinyatakan tidak sah, atau namanya tidak terdaftar sebagai Daftar Pemilih Tetap (DPT) karena kesalahan penyelenggara.
- b. Pemilih hantu atau *ghost voter* adalah seseorang yang tidak memenuhi syarat sebagai pemilih karena berbagai alasan. Misalnya, nama yang terdaftar ternyata sudah meninggal, atau nama pemilih ternyata terdaftar di dua TPS dan sudah mencoblos di tempat lain.
- c. Golput Ideologis adalah mereka yang tidak memilih karena tidak percaya dengan sistem ketatanegaraan yang berlaku. Kelompok ini memandang negara sebagai korporasi yang dikuasai sejumlah elit dan tidak memiliki kedaulatan rakyat secara mutlak.
- d. Golput Pragmatis adalah kelompok yang tidak memilih karena menganggap Pemilu tidak memberi manfaat langsung bagi pemilih. Golongan jenis ini beranggapan bahwa memilih atau tidak, mereka tidak akan merasakan dampak ataupun perubahan sedikit pun.

- e. Golput Politis adalah orang-orang yang percaya pada negara dan pemilu, tetapi kelompok ini tidak memilih karena merasa kandidat dalam Pemilu tidak bisa mempertimbangkan kepentingan serta preferensi politik mereka.

### 2.1.3 *Clustering*

*Clustering* atau pengelompokan data merupakan suatu proses yang dianggap penting dalam menemukan kesamaan data dan mengelompokkan data yang sejenis. *Clustering* dianggap sebagai metode pembelajaran tanpa pengawasan yang paling penting yang bertujuan menemukan pola dalam kumpulan data tidak berlabel. *Clustering* membagi kumpulan data menjadi beberapa kelompok dimana dimana satu kelompok tertentu mempunyai kemiripan yang lebih besar dibandingkan kelompok lainnya (Kamila et al., 2019).

Penggunaan algoritma *clustering* bergantung pada jenis data yang tersedia untuk tujuan dan aplikasi tertentu. Jika analisis klaster digunakan sebagai alat deskriptif atau eksplorasi, beberapa jenis algoritma dapat diuji pada data yang sama untuk mendapatkan informasi yang terkandung dalam data tersebut. Secara umum, metode *Clustering* dapat diklasifikasikan ke dalam beberapa kategori, salah satunya adalah kategori metode partisi. Metode partisi ini didasarkan pada penentuan awal jumlah kelompok dan dilanjutkan dengan penempatan ulang objek secara iteratif untuk menemukan kelompok yang terletak di suatu titik. Salah satu algoritma yang terkenal dalam metode partisi ini adalah algoritma *K-Means* dan algoritma *K-Medoids* (Kamila et al., 2019).

## 1. **Persyaratan *Clustering***

Menurut (Agarwal, 2013) untuk melakukan suatu analisa *Clustering* memiliki beberapa syarat atau ketentuan yang harus dipenuhi oleh algoritma *Clustering*, yaitu sebagai berikut:

### a. **Skalabilitas (*Scalability*)**

Skalabilitas mengacu pada kemampuan suatu proses akumulasi dalam menangani data dalam jumlah besar, karena database yang besar tidak hanya terdiri dari ratusan data, tetapi bisa mencapai jutaan objek. Oleh karena itu, diperlukan algoritma *Clustering* yang dapat secara efisien menangani skala yang besar.

### b. **Kemampuan analisa beragam bentuk data (*Ability to deal with different types of attributes*)**

Beberapa algoritma pengelompokan telah dikembangkan untuk analisis numerik. Namun, dalam eksplorasi data, penting untuk menangani berbagai jenis data, seperti data biner, data ordinal, atau kombinasi keduanya.

### c. **Menentukan klaster dengan bentuk yang tak terduga (*Discovery of clusters with arbitrary shape*)**

Banyak algoritma *Clustering* menggunakan metrik Euclidean atau Manhattan yang menghasilkan klaster berbentuk bola. Namun, hasil *Clustering* tidak selalu berbentuk bulat. Klaster dapat memiliki bentuk yang aneh dan berbeda satu sama lain. Oleh karena itu, diperlukan kemampuan untuk menganalisis klaster dengan bentuk yang beragam dalam algoritma *Clustering*.

### d. **Kemampuan untuk menangani noise (*Ability to deal with noisy data*)**

Dalam kenyataannya, data seringkali mengandung gangguan, kesalahan, ketidakjelasan, atau hilang sehingga tidak selalu dalam kondisi yang baik. Oleh karena itu, suatu algoritma *Clustering* harus dapat mengatasi data yang bising. Beberapa algoritma *Clustering* sangat sensitif terhadap data yang bising, yang dapat menghasilkan klaster dengan kualitas yang rendah.

**e. Sensitivitas terhadap perubahan input (*Incremental Clustering and insensitivity to input order*)**

Input data yang berbeda dapat mengakibatkan perubahan total pada klaster atau merubah klaster yang sudah ada, bahkan dapat menghasilkan perubahan yang signifikan jika menggunakan algoritma *Clustering* dengan tingkat sensitivitas yang rendah. Hal ini disebabkan karena algoritma clustering tidak merespon perubahan urutan data masukan. Oleh karena itu diperlukan algoritma *clustering* yang tidak memperhatikan urutan data masukan.

**f. Mampu melakukan klasterisasi data dimensi tinggi (*Capability of Clustering high-dimensionality data*)**

Suatu kelompok data dapat memiliki banyak dimensi atau atribut. Sebagian besar algoritma *Clustering* dapat menangani data dengan ukuran lebih kecil. Oleh karena itu, diperlukan algoritma yang dapat menangani data berdimensi tinggi.

**g. Interpretasi dan kegunaan (*Interpretability and usability*)**

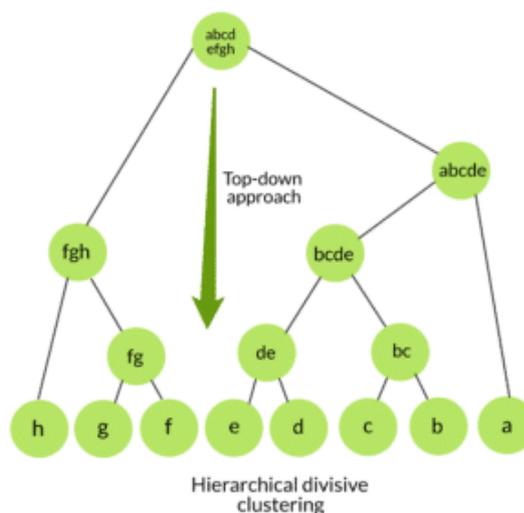
Hasil dari *Clustering* harus dapat diinterpretasikan dan berguna. Pengguna ingin hasil *Clustering* yang mudah diinterpretasikan, dimengerti, dan manfaat. Artinya *Clustering* harus memenuhi beberapa persyaratan yang sesuai dengan

kebutuhan pengguna, dan tentunya hal ini mempengaruhi pemilihan metode *Clustering* yang akan digunakan.

## 2. Metode *Clustering*

### a. *Hierarchical Clustering*

Menurut (Aggarwal, 2015), dalam *Hierarchical Clustering*, data dikelompokkan berdasarkan representasi hierarkis, dimana pada setiap iterasi dua kelompok terdekat digabungkan atau seluruh data dibagi menjadi beberapa kelompok. Contoh metode *Hierarchical Clustering* meliputi *Single Linkage*, *Complete Linkage*, *Average Linkage*, dan *Average Group Linkage*.



Gambar 2. 1 Hierrarchical Clustering (Aggarwal, 2015)

### b. *Partitional Clustering*

Menurut (Aggarwal, 2015), *Partitional Clustering* adalah metode pengelompokan data menjadi beberapa kelompok tanpa struktur hierarki. Pada metode *Partitional Clustering*, setiap kelompok memiliki pusat kelompok (*Centroid*), dan tujuan utamanya adalah meminimalkan jarak (*dissimilarity*) dari

setiap data ke pusat kelompoknya. Contoh metode *Partitional Clustering* meliputi *K-Means*, *Fuzzy -Means*, dan *Mixture Modelling*. Metode ini merupakan analisis partisi klaster yang paling sederhana dan mendasar, dimana objek dalam satu himpunan dikelompokkan menjadi beberapa kelompok atau klaster.

**c. *Density-Based***

Menurut (Aggarwal, 2015), metode partisi dan hierarkis umumnya dirancang untuk menemukan klaster dengan bentuk bola (*spherical*). Namun, mereka memiliki kesulitan dalam menemukan klaster yang bentuknya sembarang seperti "S" atau oval. Sebagai alternatif, metode berbasis kepadatan digunakan untuk menemukan klaster dengan bentuk sembarang dengan memodelkan klaster sebagai daerah dengan kepadatan tinggi yang dipisahkan dari daerah dengan kepadatan rendah. Metode ini dapat menemukan klaster dengan bentuk *non-spherical*.

**d. *Grid-Based***

Menurut (Aggarwal, 2015), metode *Clustering* yang telah dibahas sejauh ini membagi objek sesuai dengan sebaran objek pada ruang yang tertanam (*embedding space*). Pendekatan *Grid-Based Clustering* menggunakan struktur data *grid multiresolusi*. Pendekatan ini membagi ruang objek menjadi sejumlah grid, di mana operasi *Clustering* dilakukan. Pendekatan ini memiliki keunggulan kecepatan pemrosesan yang cepat, yang umumnya bergantung pada jumlah objek data dan jumlah sel di setiap dimensi ruang terkuantisasi.

### 2.1.4 Algoritma K-Means

*K-Means* adalah salah satu metode analisis kluster yang bertujuan untuk mempartisi objek ke dalam  $k$  kluster, dimana setiap objek dalam kluster diperoleh melalui rata-rata terdekat. Algoritma ini merupakan salah satu algoritma yang terkenal dan mudah dipelajari dalam menyelesaikan masalah pengelompokan data. *K-Means* adalah algoritma evolusioner yang namanya menggambarkan cara kerjanya. Algoritma ini mengelompokkan observasi ke dalam  $k$  kluster, dengan  $k$  sebagai parameter masukan. Kemudian, setiap titik data distribusikan ke kluster berdasarkan kedekatannya dengan rata-rata kluster. Rata-rata kluster dihitung beberapa kali selama proses awal (Kamila et al., 2019)

Langkah-langkah untuk melakukan *Clustering K-Means* (Solichin and Khairunnisa, 2020) adalah sebagai berikut:

- a. Menentukan jumlah  $k$ , dimana  $k$  merupakan jumlah *cluster*.
- b. Pilih secara acak nilai  $k$  sebagai pusat awal *cluster*.
- c. Setiap titik pada dataset dibagi menjadi beberapa kelompok  $k$  cluster antara setiap titik dengan pusat kluster yang diperoleh sesuai dengan jarak *Euclidean*.

Rumus untuk menghitung jarak *Euclidean* ditunjukkan pada Persamaan 2.1 berikut.

$$d(x, y) = \sqrt{\sum_{i=0}^n (x_i - y_i)^2} \quad (2.1)$$

Dimana,

$D(x, y)$  = jarak data ke  $x$  ke pusat kluster  $y$

$x_i$  = data  $x$  pada observasi ke- $i$

$y_i$  = titik pusat ke- $y$  observasi ke- $i$

$n$  = banyaknya observasi

- d. Mengelompokan semua data berdasarkan kluster terdekat.
- e. Menghitung pusat kluster baru dengan menghitung rata-rata jarak data dengan titik pusat kluster menggunakan Persamaan 2.2 berikut.

$$c_{ij} = \frac{\sum_{i=1}^p x_{ij}}{p} \quad (2.2)$$

Dimana,

$c_{ij}$  = *Centroid* terbaru pada iterasi  $k$

$x_{ij}$  = Anggota kluster ke  $k$

$p$  = banyaknya anggota kluster  $k$

- f. Melakukan perulangan langkah 2-5 hingga kondisi konvergen tercapai. Setiap anggota kluster tidak mengalami perubahan letak kluster.

### 2.1.5 *Elbow Method*

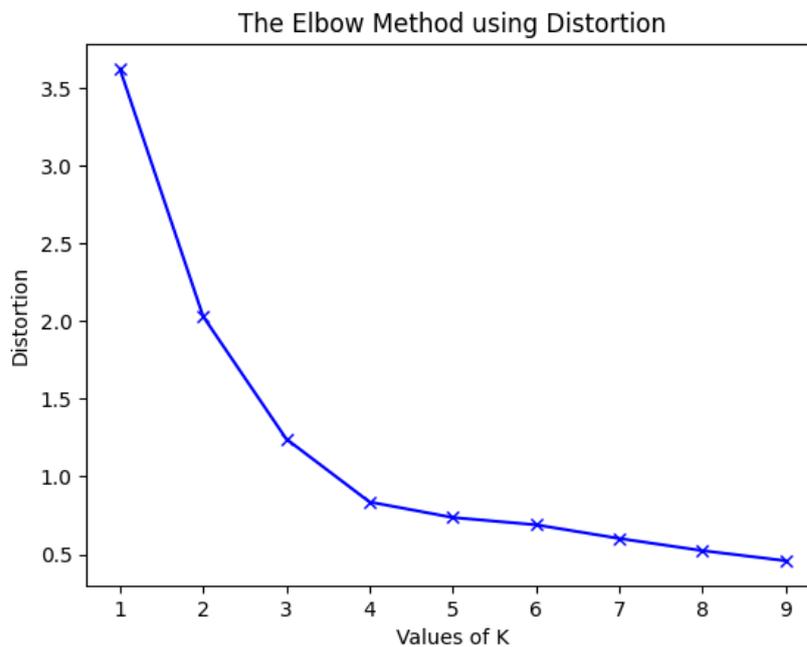
*Elbow method* adalah metoda yang sering digunakan untuk menentukan jumlah *cluster* yang akan digunakan pada *k-means clustering*. *Elbow method* menghitung selisih kuadrat dari nilai  $k$  yang berbeda. Ketika nilai  $k$  meningkat, tingkat distorsi rata-rata menurun. Jumlah sampel di setiap kategori berkurang dan sampel mendekati pusat. Ketika nilai  $k$  meningkat, nilai  $k$  yang sesuai dengan siku, efek peningkatan derajat distorsi, paling menurun (Mengyao Cui, 2020).

Banyak fungsi yang bisa digunakan dalam *elbow method* salah satunya *Sum of Squares Error* atau SSE yang mengukur varians dalam setiap *cluster*. Semakin baik pengelompokannya, semakin rendah SSE secara keseluruhan.

Persamaan yang digunakan untuk memperoleh nilai *Sum of Squares Error* (SSE) menggunakan Persamaan 2.3 berikut.

$$SSE = \sum_{k=1}^k \| x_1 - C_k \|^2 \quad (2.3)$$

Hasil perhitungan SSE akan lebih mudah dibaca ketika dibuatkan visualisasi ke dalam bentuk grafik seperti Gambar 2.2 berikut.



Gambar 2.2 Contoh Grafik SSE (Mengyao Cui, 2020)

Berdasarkan Gambar 2.2 diatas, jumlah k yang paling optimal yaitu berada pada titik belok kurva atau lengkungan garis yang membentuk siku.

### 2.1.6 *Davies-Bouldin Index* (DBI)

*Davies-Bouldin Index* (DBI) adalah metode yang diperkenalkan oleh David L. Davies dan Donald W. Bouldin. *Davies-Bouldin Index* digunakan untuk mengevaluasi *cluster* secara umum berdasarkan kuantitas dan kedekatan antar anggota cluster. Skor didefinisikan sebagai ukuran kemiripan rata-rata dari setiap

*cluster* dengan *cluster* yang paling mirip, dimana rasio jarak dalam-*cluster* dengan jarak antar-*cluster* adalah sama. Dengan demikian, *cluster* yang terpisah lebih jauh dan kurang tersebar akan menghasilkan skor yang lebih baik (Davies dan Bouldin, 1979).

*Davies-Bouldin Index* adalah metrik validasi yang digunakan untuk mengevaluasi model pengelompokan yang dihitung sebagai ukuran rata-rata kesamaan setiap cluster dengan cluster yang paling mirip dengannya. Dalam konteks ini, kesamaan diartikan sebagai perbandingan antara jarak antar cluster dan jarak intra cluster. Oleh karena itu, indeks ini mengklasifikasikan cluster yang terpisah dengan baik dan memiliki varian lebih sedikit sebagai cluster yang memiliki skor lebih baik (Butsianto dan Saepudin, 2019).  $X = \{X_1, X_2, X_3, \dots\}$ , *Davies-Bouldin Index* untuk k jumlah cluster dapat dihitung dengan Persamaan 2.4 sebagai berikut

$$DBI = \frac{1}{k} \sum_{i=1}^k \max \left( \frac{\Delta(X_i) + \Delta(X_j)}{\delta(X_i, X_j)} \right) \quad (2.4)$$

Di mana:

$\Delta(X_i)$  adalah jarak intracuster dalam cluster  $k$ .

$\delta(X_i, X_j)$  adalah jarak antar cluster antar cluster  $X_i$  dan  $X_j$

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

## 2.2 Penelitian Terkait

### 2.2.1 *State of The Art*

Berdasarkan rumusan masalah dan tujuan penelitian yang telah dibuat, maka dilakukan penyusunan *literature review* dari penelitian sebelumnya yang berkaitan dengan penerapan *clustering* untuk pemilihan umum, Algoritma *K-Means* dan algoritma *clustering* lainnya. Beberapa penelitian sebelumnya yang dilakukan adalah, sebagai berikut:

Tabel 2. 1 *State of The Art*

| NO | Nama, Tahun Penelitian | Algoritma/Metode   | Permasalahan   | Solusi   |
|----|------------------------|--|--|--|
| 1. | (Fajriansyah, 2021)    | <i>Algoritma K-Means</i>   | Masih banyak masyarakat, baik pemula, dewasa, maupun lanjut usia, yang tidak menggunakan hak pilihnya dalam pemilihan umum.  | Pengelompokan DPT berdasarkan umur dan wilayah. Daerah penelitian adalah Kabupaten Langkapura, Rajabasa dan Kemiling.  |
| 2. | (Wahyuni dkk., 2023)   | <i>Algoritma K-Means</i><br><i>Davies-Bouldin</i><br><i>Indeks</i> | Banyaknya daftar pemilih yang tidak memilih sehingga menyebabkan pemborosan anggaran, sehingga penting untuk mengidentifikasi kluster daftar pemilih untuk mengurangi risiko anggaran pilkada. | Penelitian ini fokus pada klasterisasi pemilih yang abstains. prediksi <i>cluster</i> menggunakan algoritma K-means. Oleh karena itu, penting untuk mempelajari informasi peserta pemilu dan anggaran pemilu tahun 2024. |

| NO | Nama, Tahun Penelitian   | Algoritma/Metode   | Permasalahan   | Solusi   |
|----|--------------------------|--|--|--|
| 3. | (Vernanda dkk., 2022)    | <i>Algoritma K-Means</i>   | Kesalahan dalam pengadimistrasian Formulir C Pemberitahuan yang akan berdampak negatif terhadap hilangnya hak pilih warga Negara.  | Data mining dapat membantu KPU menemukan informasi baru dengan mengolah data yang ada dengan metode <i>clustering</i> dan menggunakan algoritma K-Means.   |
| 4. | (Alfatah dkk., 2021)     | Algoritma K-Means  | Memetakan strategi pemilihan daerah pemilih pada Calon Legislatif Kota Bengkulu, data dikelompokkan berdasarkan karakteristik yang sama akan dimasukkan ke dalam kelompok yang sama dan set data yang dimasukkan ke dalam kelompok tidak tumpang tindih  | Penggunaan teknik data mining diharapkan dapat mempercepat proses pengambilan keputusan, dimana informasi yang terkandung dalam hasil pemilu dapat dikelola dengan variabel-variabel yang mempengaruhi hasil pemilu sehingga menjadi informasi baru. |
| 5. | (Dewi dan Pramita, 2019) | <i>K-Medoids</i><br><i>Elbow methods</i><br><i>Silhouette Coefisien</i><br><i>Davies Bouldin Index</i> | Untuk memperoleh hasil pengelompokan data atau clustering yang maksimal, perlu penentuan jumlah cluster yang tepat. Berbagai metode yang dapat digunakan untuk menentukan jumlah <i>cluster</i> yang tepat, yaitu metode elbow, koefisien <i>silhouette</i> , <i>gap statistics</i> , dan lainnya. | Membandingkan metode untuk menentukan jumlah <i>cluster</i> yaitu metode metode elbow dan koefisien <i>silhouette</i> pada proses <i>clustering</i> k-medoid menggunakan data kerajinan.   |
| 6. | (Hardiani, 2022)         | <i>K-Means</i><br><i>Cross Industry</i><br><i>Standard Process for</i>                                 | Penelitian ini bermaksud untuk menerapkan teknik penambangan data (data mining) dengan algoritma <i>clustering</i> K-Means untuk menganalisis pola   | Menerapkan algoritma K-Means guna mengelompokan tingkat penyebaran covid 19 di Indonesia yang diperoleh dari   |

| NO | Nama, Tahun Penelitian     | Algoritma/Metode  | Permasalahan  | Solusi   |
|----|----------------------------|---|---|--|
|    |                            | <i>Data-Mining</i><br>(CRISP-DM)<br><i>David Boulden index</i><br>(DBI)                   | penyebaran Covid 19 dengan mengelompokkan kasus Covid 19 di Indonesia yang didapatkan dari website dataset Kaggle. Metode penambangan data (data mining) yang digunakan yaitu <i>Cross Industry Standard Process for Data-Mining</i> (CRISP-DM)   | website dataset Kaggle. Penentuan jumlah <i>cluster</i> yang optimal menggunakan David boulden index   |
| 7. | (Sulistiyawan dkk., 2021)  | <i>K-Means</i><br><i>Elbow Method</i><br><i>Silhouette Method</i><br><i>Gap Statistic</i> | Perlu dilakukan pengelompokan sektor perikanan Indonesia berdasarkan faktor-faktor yang mempengaruhinya, dengan menggunakan analisis <i>cluster</i> untuk mengetahui karakteristik anggota kelompok yang akan dibentuk dan mengetahui metode mana yang terbaik untuk menentukan jumlah <i>cluster</i> pada perikanan Indonesia. | Mengelompokkan industri perikanan Indonesia berdasarkan faktor-faktor yang mempengaruhinya. Selain mengetahui metode mana yang terbaik untuk menentukan jumlah klaster pada sektor perikanan Indonesia |
| 8. | (Retsya Lapiza dkk., 2023) | <i>K-Medoids</i><br><i>Clustering Large Application</i><br><i>Silhouette Coefficient</i>  | Penelitian ini bertujuan untuk mengetahui kelompok daerah dan membandingkan kinerja pengelompokkan menggunakan analisis klaster.  | Analisis klaster yang digunakan dalam penelitian ini adalah metode K-Medoids dan <i>Clustering Large Application</i> (CLARA).  |
| 9. | (Rivatunisa, 2022))        | Algoritma KNN   | Penggunaan Algoritma KNN yang dapat memberikan rekomendasi untuk bakal calon DPRD Provinsi periode 2024-2029 dalam memilih daerah pemilihan yang sesuai profil dan memiliki potensi kemenangan yang lebih besar saat  | Merancang sistem yang dapat menentukan keberhasilan bakal calon legislatif pada pemilihan umum di daerah pemilihan provinsi Jambi menggunakan  |

| NO  | Nama, Tahun Penelitian       | Algoritma/Metode   | Permasalahan  | Solusi  |
|-----|------------------------------|--|---|---|
|     |                              |  | Pemilihan Umum berdasarkan data profil anggota DPRD terpilih periode sebelumnya.  | Data Mining dengan Algoritma <i>K-Nearest Neighbor</i> .  |
| 10. | (Safitri dan Fitriani, 2022) | Algoritma SVM<br><i>Kernel Gaussian RBF</i>  | Pengaruh demografi pada suatu terhadap tingkat partisipasi masyarakat dalam pemilu. Hubungan antara demografi dan partisipasi pemilih, misalnya, mayoritas penduduk di suatu wilayah mempunyai tingkat pendidikan yang tinggi, dan partisipasi pemilih juga tinggi.   | Dengan diterapkannya SVM diharapkan hasil yang diperoleh dapat memberikan kontribusi terhadap partisipasi masyarakat dalam penyelenggaraan pemilu di kota Surabaya. |
| 11  | (Saputra dkk., 2019)         | DBSCAN<br><i>Sillhoutte Index</i>  | Membandingkan hasil calon terkuat yang diambil lembaga penelitian Indonesia sebelum resmi dicalonkan sebagai calon presiden dengan hasil yang diperoleh setelah keputusan resmi Komisi Pemilihan Umum (KPU) baik positif, negatif, dan netral. Selain itu, Anda juga akan mengetahui tren pemilu Presiden 2019. | Melakukan pengklasteran dan analisis data sentimen terhadap calon presiden tahun 2019 melalui cuitan Twitter.   |
| 12  | (Aditya dan Fitriyah, 2021)  | <i>Fuzzy C-Means</i><br><i>K-Means</i><br><i>Silhouette Coefficient</i><br><i>Davies-Bouldin Index</i> | Berdasarkan survei statistik <i>e-commerce</i> tahun 2020 yang dipublikasikan bps.go.id, terdapat peningkatan aktivitas perusahaan baru setiap tahunnya. Untuk menjaga dan meningkatkan antusiasme pelanggan serta mengembangkannya dalam kaitannya dengan aktivitas perusahaan,                                | membandingkan Algoritma K-Means dan Fuzzy C-Means untuk Menentukan Segmentasi Pelanggan yang Akan Digunakan untuk Identifikasi Prospek Menggunakan Model RFM.       |

| NO | Nama, Tahun Penelitian       | Algoritma/Metode                 | Permasalahan  | Solusi  |
|----|------------------------------|----------------------------------|---|---|
|    |                              |                                  | perlu dilakukan penerapan segmentasi pelanggan yang tujuannya adalah untuk mengidentifikasi pelanggan potensial.  |   |
| 13 | (Mengyao Cui, 2020)          | <i>K-Means<br/>Elbow Methods</i> | Algoritma K-means <i>clustering</i> merupakan algoritma yang umum digunakan dan juga merupakan algoritma pembelajaran tanpa pengawasan. Namun algoritma K-means juga mempunyai kekurangan tertentu.   | Ekperimen dilakukan untuk mencari nilai <i>Within-Cluster Sum-of-Squares</i> atau SSE dari setiap simulasi yang dilakukan yaitu $k = 1$ , $k=2$ , dan $k=3$ .   |
| 14 | (Ekasetya dan Jananto, 2020) | <i>K-Means<br/>Elbow Methods</i> | Kota Semarang merupakan kota metropolitan terbesar kelima di Indonesia yang banyak mengalami kecelakaan lalu lintas. Oleh karena itu, diperlukan suatu sistem yang menghasilkan model ekstraksi data untuk mengidentifikasi pola kecelakaan yang terjadi. | Penelitian ini menggunakan proses pengujian manual dan aplikasi RStudio. Perhitungan manual menggunakan jumlah sampel $K=2$ sampai $K=10$ dengan metode elbow menggunakan perhitungan nilai SSE pada setiap cluster. Pada saat yang sama, perhitungan komputer menggunakan RStudio menghasilkan perbedaan SSE |
| 15 | (Hartanti, 2020)             | <i>K-Means<br/>Elbow Methods</i> | Keberhasilan siswa dalam lulus ujian nasional (UN) terlihat dari perolehan poin pada mata pelajaran yang diujikan. Mengukur kesiapan siswa dalam menghadapi ujian negara sangat diperlukan  | Algoritma K-Means digunakan untuk mengukur tingkat kesiapan siswa. Penerapan metode Elbow untuk menentukan jumlah <i>cluster</i> optimal yang   |

| NO | Nama, Tahun Penelitian  | Algoritma/Metode  | Permasalahan   | Solusi  |
|----|-------------------------|---|--|---|
|    |                         |   | untuk keberhasilan siswa. Sangat diperlukan suatu sistem atau pedoman yang memudahkan kegiatan sekolah dan membantu pengambilan keputusan mengenai penilaian kesiapan ujian nasional.                          | kemudian digunakan dalam perhitungan cluster menggunakan algoritma K-Means.   |
| 16 | (Rahman, 2017)          | <i>K-Means</i><br><i>Elbow Methods</i>                          | PT Global Bangkit Utama sedang berusaha mencari strategi yang tepat. Untuk mengambil keputusan strategis, Perusahaan menganalisis informasi bisnis untuk menghadapi persaingan.                                | Hasil <i>clustering</i> akan dianalisis untuk memperoleh informasi yang akan membantu menentukan strategi bisnis yang diterapkan oleh PT Global Bangkit Utama.  |
| 17 | (Adhitama dkk., 2020)   | <i>X-Means</i><br><i>K-Means</i><br><i>Davies-Bouldin Index</i> | Belum adanya pengelompokan SMK berdasarkan data pokok kemendikbud di Jawa Tengah merupakan sebuah peluang untuk mengembangkan arah revitalisasi SMK menjadi lebih baik dan jelas.                              | Menerapkan algoritma X-Means dan K-Means untuk menentukan berapa jumlah <i>cluster</i> ideal untuk membagi data SMK yang ada di Jawa Tengah berdasarkan data dari kemendikbud.  |
| 18 | (Sinaga dan Yang, 2020) | <i>Unsupervised K-Means</i>                                     | Algoritma K-means dan ekstensinya selalu dipengaruhi oleh inisialisasi dengan jumlah <i>cluster</i> yang diperlukan secara apriori. Artinya, algoritma K-means bukanlah metode pengelompokan tanpa pengawasan. | Algoritma K-means yang dilakukan pada beberapa kumpulan data sintetik dan dibandingkan dengan sebagian besar algoritma yang ada, seperti R-EM, C-FS, K-means dengan bilangan sebenarnya $c$ , K-means+gap, dan X-Means. |

| NO | Nama, Tahun Penelitian | Algoritma/Metode                                     | Permasalahan   | Solusi  |
|----|------------------------|--|--|---|
| 19 | (Ahmed dkk., 2020)     | <i>algoritma K-Means dan beberapa varian K-means</i> | Algoritma pengelompokan K-means dianggap sebagai salah satu algoritma penambangan data yang paling kuat dan populer di komunitas riset. Terlepas dari popularitasnya, algoritma ini masih memiliki beberapa keterbatasan, termasuk masalah dengan inisialisasi centroid secara acak, yang menyebabkan konvergensi yang tidak dapat diprediksi. | Makalah ini berfokus pada algoritma K-means yang populer dan masalah inisialisasi serta ketidakmampuannya memproses data dengan karakteristik campuran. |
| 20 | (Syakur dkk., 2018)    | <i>K-Means Elbow Methods</i>                         | KM Indonesia mempunyai pelanggan yang berbeda-beda, namun UKM kurang memiliki pemetaan pelanggan sehingga tidak mengetahui pelanggan mana yang loyal atau sebaliknya. Pemetaan pelanggan merupakan pengelompokan profil pelanggan yang memudahkan analisis dan praktik UKM dalam produksi barang khususnya penjualan batik.                    | menggabungkan K-Means dengan metode elbow untuk menentukan jumlah sebenarnya <i>cluster</i> segmentasi profil Pelanggan di UKM.                         |
| 21 | (Arora dkk., 2016)     | <i>K-Means K-Medoids</i>                             | <i>Clustering</i> memainkan peran yang sangat penting dalam mengeksplorasi data, membuat prediksi dan mengatasi anomali pada data. <i>Cluster</i> yang berisi jaminan, karakteristik identik dalam kumpulan  | membandingkan dan mengevaluasi dua algoritma clustering paling populer K-Means dan K-Medoids pada dataset transaksi 10k KEEL.                           |

| NO | Nama, Tahun Penelitian  | Algoritma/Metode | Permasalahan  | Solusi  |
|----|-------------------------|------------------|---|---|
|    |                         |                  | data dikelompokkan menggunakan teknik reiteratif.   |   |
| 22 | (Ardiansyah dkk., 2022) | <i>K-Means</i>   | Penyelenggaraan pemilihan umum diadakan terlepas dari pelanggaran. Pelanggaran-pelanggaran yang terjadi seperti hak pilih atau pelanggaran lainnya bertujuan hanya untuk memperoleh simpati atau hak pilih untuk dipilih, atau mempunyai tujuan tertentu. | Algoritma K-Means untuk mengelompokkan pelanggaran pada Pemilihan Walikota dan Wakil Walikota Binjai Tahun 2020. Variabel yang diuji pada data terdiri dari inisial pelapor, kategori pelanggaran, tahapan pelanggaran, dan sanksi pelanggaran. |



| No. | Penulis   | Judul   | Ruang Lingkup       |     |      |         |           |         |           |       |               |               |                      |                         |               |
|-----|---|---|---------------------|-----|------|---------|-----------|---------|-----------|-------|---------------|---------------|----------------------|-------------------------|---------------|
|     |   |   | Penerapan Algoritma |     |      |         |           |         |           |       |               | Metode        |                      |                         |               |
|     |   |   | Naïve Bayes         | SVM | K-NN | K-Means | K-Medoids | X-Means | U K-Means | CLARA | Fuzzy C-Means | Elbow Methods | Davies Bouldin Index | Sillhouette Coefffrenct | Gap Statistic |
| 2   | Sri Ngudi Wahyuni,<br>Nazmun Nahar<br>Khanom, Yuli Astuti<br>(2023) | <i>K-Means Algorithm Analysis For Election Cluster Prediction</i>   |                     |     |      | ✓       |           |         |           |       |               |               | ✓                    |                         |               |
| 3   | Meggy Vernanda ,<br>Relita Buaton , Husnul<br>Khair, (2022)         | Metode <i>Clustering</i> Untuk Pengelompokan Data Pengembalian Formulir C Pemberitahuan Berdasarkan Wilayah Kelurahan                       |                     |     |      | ✓       |           |         |           |       |               |               |                      |                         |               |
| 4   | Dhika Alfatah, (2021)   | <i>Application Of The K-Means Clustering Algorithm In Mapping The Regional Voter Strategy For The Legislative Candidates For The DPR RI</i> |                     |     |      | ✓       |           |         |           |       |               |               | ✓                    |                         |               |

| No. | Penulis   | Judul  | Ruang Lingkup       |     |      |         |           |         |           |       |               |               |                      |                        |               |   |
|-----|---|--|---------------------|-----|------|---------|-----------|---------|-----------|-------|---------------|---------------|----------------------|------------------------|---------------|---|
|     |   |  | Penerapan Algoritma |     |      |         |           |         |           |       |               | Metode        |                      |                        |               |   |
|     |   |  | Naïve Bayes         | SVM | K-NN | K-Means | K-Medoids | X-Means | U K-Means | CLARA | Fuzzy C-Means | Elbow Methods | Davies Bouldin Index | Sillhouette Coeffrenct | Gap Statistic |   |
| 5   | Dewa Ayu Indah Cahya Dewi , Dewa Ayu Kadek Pramita, (2019)                    | Analisis Perbandingan Metode Elbow Dan <i>Sillhouette</i> Pada Algoritma Clustering K-Medoids Dalam Pengelompokan Produksi Kerajinan Bali                |                     |     |      |         | ✓         |         |           |       |               |               | ✓                    |                        | ✓             |   |
| 6   | Tikaridha Hardiani, (2022)  | Analisis <i>Clustering</i> Kasus Covid 19 Di Indonesia Menggunakan Algoritma K-Means   |                     |     |      | ✓       |           |         |           |       |               |               |                      | ✓                      |               |   |
| 7   | Edy Sulistiyawan , Alfisyahrina Hapsery , Lucky Junita Ayu Arifahanum, (2021) | Perbandingan Metode Optimasi Untuk Pengelompokan Provinsi Berdasarkan Sektor Perikanan Di Indonesia (Studi Kasus Dinas Kelautan Dan Perikanan Indonesia) |                     |     |      | ✓       |           |         |           |       |               |               | ✓                    |                        | ✓             | ✓ |



| No. | Penulis   | Judul  | Ruang Lingkup       |     |      |         |           |         |           |       |               |               |                      |                         |               |   |  |
|-----|---|--|---------------------|-----|------|---------|-----------|---------|-----------|-------|---------------|---------------|----------------------|-------------------------|---------------|---|--|
|     |   |  | Penerapan Algoritma |     |      |         |           |         |           |       | Metode        |               |                      |                         |               |   |  |
|     |   |  | Naïve Bayes         | SVM | K-NN | K-Means | K-Medoids | X-Means | U K-Means | CLARA | Fuzzy C-Means | Elbow Methods | Davies Bouldin Index | Sillhouette Coefffrenct | Gap Statistic |   |  |
| 11  | Rifaldi Saputra,<br>Mustakim, Muhammad<br>Ridwan, Okfalisa (2019) | Menentukan Popularitas<br>Calon Presiden Dan Tren<br>Pada Pilpres 2019<br>Menggunakan Algoritma<br>Dbscan                        |                     |     |      |         |           |         |           |       |               | ✓             |                      |                         |               | ✓ |  |
| 12  | Difa Lazuardi Aditya ,<br>Devi Fitriana (2021)                    | <i>Comparative Study Of Fuzzy<br/>C-Means And K-Means<br/>Algorithm For Grouping<br/>Customer Potential In Brand<br/>Limback</i> |                     |     |      | ✓       |           |         |           |       |               |               | ✓                    |                         | ✓             | ✓ |  |
| 13  | Mengyao Cui (2020)  | <i>Introduction To The K-<br/>Means Clustering Algorithm<br/>Based On The Elbow Method</i>                                       |                     |     |      | ✓       |           |         |           |       |               |               |                      | ✓                       |               |   |  |
| 14  | Vada Annisa Ekasetya ,<br>Arief Jananto (2020)                    | Pengelompokan Data<br>Kecelakaan Lalu Lintas Di<br>Kota Semarang   |                     |     |      | ✓       |           |         |           |       |               |               |                      | ✓                       |               |   |  |

| No. | Penulis  | Judul   | Ruang Lingkup       |     |      |         |           |         |           |       |               |               |                      |                        |               |  |
|-----|--|---|---------------------|-----|------|---------|-----------|---------|-----------|-------|---------------|---------------|----------------------|------------------------|---------------|--|
|     |  |   | Penerapan Algoritma |     |      |         |           |         |           |       |               | Metode        |                      |                        |               |  |
|     |  |   | Naïve Bayes         | SVM | K-NN | K-Means | K-Medoids | X-Means | U K-Means | CLARA | Fuzzy C-Means | Elbow Methods | Davies Bouldin Index | Sillhouette Coeffrenct | Gap Statistic |  |
| 15  | Ninik Tri Hartanti (2020)                                | Metode Elbow Dan K-Means Guna Mengukur Kesiapan Siswa SMK Dalam Ujian Nasional                            |                     |     |      | ✓       |           |         |           |       |               |               | ✓                    |                        |               |  |
| 16  | Aulia Tegar Rahman, Wiranto, Rini Angrainingsih (2017)   | <i>Coal Trade Data Clustering Using K-Means (Case Study Pt. Global Bangkit Utama)</i>                     |                     |     |      | ✓       |           |         |           |       |               |               | ✓                    |                        |               |  |
| 17  | Rifki Adhitama , Auliya Burhanuddin, Ridho Ananda (2020) | Penentuan Jumlah Cluster Ideal Smk Di Jawa Tengah dengan Metode X-Means Clustering Dan K-Means Clustering |                     |     |      | ✓       |           | ✓       |           |       |               |               |                      | ✓                      |               |  |
| 18  | Kristina P. Sinaga, Miin-Shen Yang (2020)                | <i>Unsupervised K-Means Clustering Algorithm</i>  |                     |     |      |         |           |         | ✓         |       |               |               |                      |                        |               |  |



| No. | Penulis             | Judul  | Ruang Lingkup       |     |      |         |           |         |           |       |               |               |                      |                         |               |
|-----|---------------------|--|---------------------|-----|------|---------|-----------|---------|-----------|-------|---------------|---------------|----------------------|-------------------------|---------------|
|     |                     |  | Penerapan Algoritma |     |      |         |           |         |           |       |               | Metode        |                      |                         |               |
|     |                     |  | Naïve Bayes         | SVM | K-NN | K-Means | K-Medoids | X-Means | U K-Means | CLARA | Fuzzy C-Means | Elbow Methods | Davies Bouldin Index | Sillhouette Coefffrenct | Gap Statistic |
|     |                     | Badan Pengawas Pemilihan Umum Kota Binjai)   |                     |     |      |         |           |         |           |       |               |               |                      |                         |               |
| 23  | Ace Hermawan (2023) | Klasterisasi Pemilih Abstain Pada Penyelenggaraan Pemilihan Umum Menggunakan Algoritma K-Means |                     |     |      | ✓       |           |         |           |       |               | ✓             | ✓                    |                         |               |

### 2.2.3 Relevansi Penelitian

Relevansi penelitian terkait algoritma, metode dan topik yang diambil dimana terdapat pembaharuan dari penelitian sebelumnya, relevansi penelitian bisa dilihat pada Tabel 2.3 berikut ini.

Tabel 2. 3 Relevansi Penelitian

|                           |  |  |
|---------------------------|--|--|
| <b>Peneliti</b>           | Sri Ngudi Wahyuni, Nazmun Nahar Khanom, Yuli Astuti (2023)   | Ace Hermawan (2023)  |
| <b>Judul</b>              | <i>K-Means Algorithm Analysis For Election Cluster Prediction</i>  | Klasterisasi Pemilih Abstain Pada Penyelenggaraan Pemilihan Umum Menggunakan Algoritma K-Means   |
| <b>Masalah Penelitian</b> | Berdasarkan data pemilu terbaru yang diambil dari Margokaton, Yogyakarta, Indonesia, banyak orang yang mencoblos di tahun 2021 namun angka golputnya tinggi. Hal tersebut menyebabkan pemborosan anggaran, sehingga pentingnya melakukan identifikasi klaster daftar pemilih untuk mengurangi resiko anggaran pilkada. | Sekitar 34,75 juta orang yang tidak menggunakan hak pilihnya atau abstain dalam Pemilu 2019. Sementara itu, data pemilih yang tidak menggunakan hak pilih di Kota Tasikmalaya sebanyak 43,6 ribu jiwa atau 10,04 % berdasarkan data dari KPU Kota Tasikmalaya. Jumlah pemilih abstain yang tinggi menyebabkan pemborosan anggaran. |
| <b>Objek Penelitian</b>   | Pemilih abstain pada Pilkada Yogyakarta tahun 2021   | Pemilih abstain pada Pemilu 2019 di Kota Tasikmalaya   |
| <b>Algoritma/Metode</b>   | Algoritma K-Means dan <i>Davies Bouldin Index</i>  | Algoritma K-Means, <i>Elbow Method</i> , dan <i>Davies Bouldin Index</i>   |
| <b>Dataset</b>            | Data hasil Pilkada Yogyakarta yang diambil dari website resmi Kementerian Dalam Negeri RI Tahun 2021   | Data hasil Pemilu di Kota Tasikmalaya tahun 2019 yang diambil dari KPU Kota Tasikmalaya  |
| <b>Implementasi</b>       | Melakukan <i>clustering</i> dengan variabel yang digunakan adalah daftar pemilih, surat suara tidak terpakai, dan jumlah Abstain   | Melakukan <i>clustering</i> dengan variabel yang digunakan adalah daftar pemilih, surat suara tidak sah, dan gap antara jumlah daftar pemilih dengan jumlah suara untuk 5 jenis surat suara.   |