

BAB II
TINJAUAN PUSTAKA

2.1 Penelitian Terkait

Penelitian sebelumnya yang menjadi latar belakang pada penelitian ini, dijabarkan pada tabel dibawah ini :

Tabel 2.1 Penelitian Terdahulu

No	Peneliti	Tahun Publish	Judul	Metode	Hasil
1	Ali Amran, Anita Desiani, MS Hasibuan	2017	Detection Learning Style Vark For Out Of School Children (OSC)	<i>E-Learning</i> , VARK, dan OSC	<i>E-learning</i> harus lebih memilih gaya belajar siswa. Gaya belajar yang paling disukai dari OSC adalah pembelajaran kinestetik. Pembelajaran kinestetik membutuhkan gerakan tubuh, interaktivitas, dan kontak langsung dengan bahan belajar, hal-hal ini bisa sulit untuk diterapkan dalam <i>e-Learning</i> , tetapi <i>E-learning</i> harus dapat mengadopsi gaya belajar apa pun yang fleksibel dalam hal waktu, periode, kurikulum, pedagogi, lokasi, dan bahasa.
2	Dr. P. Nagarajan dan Dr. G. Wiselin Jiji	2010	Online Educational System (E-Learning)	<i>E-Learning</i> , <i>Information</i> dan <i>Communication Technology</i>	Konstruksi yang rumit dari sistem pendidikan online terletak dalam tiga kegiatan utama:

No	Peneliti	Tahun Publish	Judul	Metode	Hasil
					Desain, Implementasi dan Evaluasi Post-Implementasi yang tepat.
3	Ambrose Leung, Maureen McGregor, David Sabiston, Stefano Vriliotis	2014	<i>VARK LEARNING STYLES AND STUDENT PERFORMANCE IN PRINCIPLES OF MICRO- VS. MACRO-ECONOMICS</i>	<i>VARK, learning styles, teaching, economics</i>	Tujuan dari studi ini adalah untuk mengidentifikasi apakah gaya belajar siswa yang berbeda terkait dengan kinerja siswa dalam prinsip-prinsip mikro vs makro-ekonomi.
4	Hadi Peyman, Jamil Sadeghifar, Javaher Khajavikhan, Hemati Karim	2014	<i>Using VARK Approach for Assessing Preferred Learning Styles of First Year Medical Sciences Students: A Survey from Iran</i>	<i>VARK, medical science students, learning styles</i>	Sebuah studi lintas bidang yang menggunakan kuesioner gaya belajar VARK dilakukan pada 141 siswa ilmu kedokteran tahun pertama di Ilam University of Medical Sciences pada tahun 2010. Dari mereka, 17,7% lebih memilih gaya Aural, 17% lebih memilih <i>Reading</i> dan <i>Writing</i> , 6,4% lebih suka gaya Kinesthetic dan 0,7% lebih menyukai gaya Visual.
5	Thomas F. Hawk, Amit J. Shah	2007	<i>Using Learning Style Instruments to Enhance Student Learning</i>	VARK, model gaya belajar	Peneliti telah meninjau lima instrumen gaya belajar (<i>Kolb Learning Style</i>)

No	Peneliti	Tahun Publish	Judul	Metode	Hasil
					<i>Indicator</i> , Gregorc Style Delineator, Felder-Silverman <i>Index of Learning Styles</i> , VARK <i>Questionnaire</i> , dan <i>Dun and Dunn Productivity Environmental Preference Survey</i>) dalam artikel ini untuk menggambarkan mode gaya belajar atau dimensi yang diukur dalam instrumen; menemukan ukuran umum dan perbedaan; melaporkan penelitian tentang validitas instrumen,

2.2 VARK

Gaya belajar VARK adalah model yang dikembangkan oleh Neil D. Fleming pada tahun 1987 terhadap model yang sudah ada, yakni VAK (*Visual, Aural, Kinesthetic*). Fleming dengan latar belakang sebagai seorang penilik sekolah di New Zealand dan kemudian berpindah menjadi peneliti di Lincoln University, mengkaji secara longitudinal preferensi modalitas belajar guru dan siswa berdasarkan persoalan konkret pembelajaran di sekolah-sekolah. Gaya belajar VARK ini sangat populer pada tahun 1980an dan berpengaruh dalam dunia pendidikan. Selanjutnya, Fleming membedakan preferensi modalitas Visual dengan *Read/Writing* (R) karena di antara keduanya memiliki kecenderungan yang berbeda. Dari apa yang dia baca dan amati, tampak jelas bahwa beberapa siswa

memiliki preferensi yang berbeda untuk kata-kata tertulis (grafis) sementara yang lainnya lebih suka informasi simbolis (gambar) seperti dalam peta, diagram, dan grafik. Kedua preferensi tidak selalu ditemukan pada orang yang sama. Untuk selanjutnya, gaya belajar model Fleming memiliki 4 preferensi modalitas, yakni Visual, Aural atau *Auditory*, *Read/Write*, dan Kinesthetic dan disingkat menjadi VARK. Masing-masing memiliki karakteristik seperti yang diungkapkan oleh Robertson, dkk (2011) dalam Tabel 2.2 berikut.

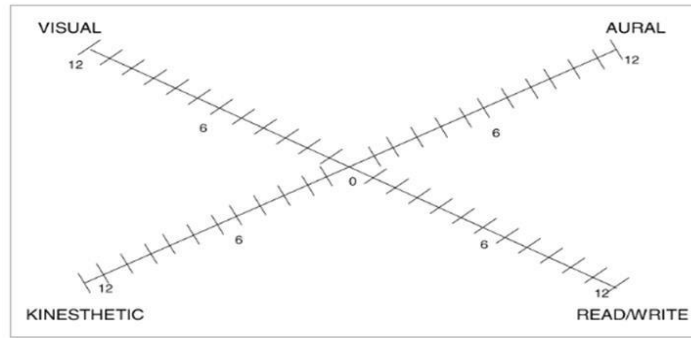
Tabel 2.2 Gaya Belajar VARK

<i>Learning Style</i>	<i>Charateristics</i>
<i>Visual</i>	<i>Prefrence for using visual resources such as diagram. Pictures and videos . Like to see people in action.</i>
<i>Auditory</i>	<i>Need to talk about situations and with a range of people, enjoy hearing stories from other.</i>
<i>Reader/Writer</i>	<i>Prolific note-taker text books are importan extensive use of journals to write down the facts and stories.</i>
<i>Kinesthetic</i>	<i>Prefrence for hands on experience within a 'real' setting and for global learning.</i>

Dengan demikian, menurut Robertson, dkk (2011) empat tipe gaya belajar siswa, yakni siswa visual, siswa aural atau *auditory*, siswa *read/write*, dan siswa kinestetik, memiliki karakteristik yang khas dan berbeda antara satu dengan yang lain, seperti tertuang dalam Tabel 2. 1 di atas. Menurut Marcy (2001), model gaya belajar VARK ini dikembangkan dalam upaya untuk meningkatkan pengembangan fakultas dan membantu siswa menjadi pelajar yang lebih baik. Lebih lanjut, VARK, menurut Fleming (2006), mengacu pada kategori preferensi komunikasi. Ini berkaitan dengan cara seseorang mengambil dan memberikan informasi. Selama

bertahun-tahun Fleming mengkaji preferensi modalitas sensorik seseorang dengan menggunakan kuesioner VARK dan hasilnya adalah sebagai berikut. Pertama, seorang siswa mungkin memiliki preferensi untuk satu modalitas atau lebih dari satu modalitas (multimodal). Kedua, modalitas belajar yang disukai tersebut mempengaruhi perilaku individu, termasuk belajarnya. Ketiga, preferensi gaya belajar tidak tetap, namun stabil dalam jangka menengah. Ketiga, baik guru dan siswa, dipercaya keduanya dapat mengidentifikasi dan memberikan contoh-contoh penggunaan modalitas preferensi dalam belajar. Keempat, informasi yang diakses menggunakan strategi yang selaras dengan preferensi modalitas siswa lebih mungkin dipahami dan dapat memotivasi. Kelima, pencocokan strategi pembelajaran dengan preferensi modalitas juga cenderung mengarah ketekunan tugas-tugas belajar, pendekatan lebih mendalam untuk belajar, dan metakognisi aktif dan efektif. Keenam, pengetahuan tentang, dan bertindak atas, preferensi modal seseorang adalah kondisi penting untuk meningkatkan belajar seseorang.

Hawk dan Shah (2007) menyajikan matriks yang diadaptasi dari Fleming (2001) tentang empat mode persepsi, dengan individu yang memiliki preferensi untuk manapun dari satu sampai empat. Menurut Hawk dan Shah (2007) masing-masing siswa memiliki satu rentangan preferensi relatif dari empat mode persepsi yang ada, akan tetapi siswa juga dapat belajar untuk menggunakan mode lainnya, seperti yang tergambar dalam Gambar 2.1 berikut.



Gambar 2.1 Empat Preferensi Mode Belajar

1. Gaya Belajar Visual

Siswa dengan gaya belajar visual mengandalkan indera mata atau penglihatan dalam proses menangkap informasi sebelum akhirnya memahami informasi tersebut. Siswa dengan gaya ini lebih mudah mengingat apa yang mereka lihat daripada apa yang mereka dengar, baca, dan lakukan. Preferensi yang mengandalkan indera mata ini lebih mudah, lebih menarik, manakala menangkap dan memahami informasi yang berasal peta, gambar, desain, diagram laba-laba, grafik, flow chart, diagram berlabel, semua anak panah simbolik, lingkaran, hirarki, foto, power point, film, demonstrasi guru, dan sebagainya yang digunakan orang untuk menyajikan informasi sebagai pengganti wujud kata-kata.

Untuk menguji seorang siswa memiliki kecenderungan gaya belajar visual, berikut ini beberapa pertanyaan yang biasa diujikan. Pertama, untuk memperoleh dan mengingat suatu informasi, apakah anda harus melihat gambar, skema, ilustrasi, grafik, dan sejenisnya? Kedua, apakah bahasa tubuh dari pembicara sangat membantu anda dalam menangkap informasi? Ketiga, apakah seni, keindahan, dan estetika penting bagi Anda? Keempat, apakah visualisasi informasi dalam pikiran membantu

anda mengingat informasi secara lebih baik? Jika jawaban siswa atas keempat pertanyaan ini adalah ya, maka siswa tersebut memiliki preferensi gaya belajar visual.

2. Gaya belajar Aural

Siswa dengan gaya belajar Aural mengandalkan indera telinga atau pendengaran dalam menangkap informasi dan memahami informasi tersebut. Siswa sangat memperhatikan intonasi, lafal, dan kecepatan bicara dari guru pada saat menjelaskan, bertanya, atau menjawab pertanyaan. Keberhasilan dalam menangkap informasi dan memahami informasi sangat bergantung pada pengemasan informasi tersebut. Jika informasi disajikan dalam bentuk rekaman, presentasi, cerita, atau dibacakan dengan keras, maka siswa dengan preferensi gaya aural akan lebih senang, lebih nyaman, dan lebih mudah.

Untuk menguji seorang siswa memiliki kecenderungan gaya belajar aural, berikut ini beberapa pertanyaan yang biasa diujikan. Pertama, untuk memperoleh dan mengingat suatu informasi, apakah anda harus mendengarkan penjelasan, rekaman, presentasi, atau diskusi? Kedua, apakah kejelasan intonasi, lafal, dan kecepatan bicara dari pembicara sangat membantu anda dalam menangkap informasi? Ketiga, apakah situasi yang tenang, tidak ramai, tidak gaduh sangat penting bagi anda ketika mendengarkan penjelasan, presentasi, lagu, atau rekaman? Keempat, apakah mengulang-ulang (menghafal) informasi dalam hati atau pikiran membantu anda mengingat informasi? Jika jawaban siswa

atas keempat pertanyaan ini adalah ya, maka siswa tersebut memiliki preferensi gaya belajar aural.

3. Gaya Belajar Baca/Tulis

Siswa dengan gaya belajar baca/tulis lebih suka memperoleh informasi dalam bentuk teks grafis dan bukan gambar, yang memuat kata-kata, kalimat, paragraf, atau wacana. Siswa dengan gaya ini lebih nyaman dalam aktivitas belajar dengan input dan output dalam bentuk teks. Dengan demikian, kegiatan membaca buku (teks, pelajaran, ilmiah), koran (artikel, opini, berita, iklan, tajuk rencana, biografi), majalah (petunjuk mengoperasikan sesuatu, resep masakan, informasi hiburan), novel, esai, brosur, leaflet, surat, poster, serta menerjemahkan dengan kamus, menulis kembali, meringkas, mencatat, menulis pokok-pokok informasi, menulis kata-kata kunci, dan membuat parafrase, merupakan kunci keberhasilan dalam memperoleh dan memahami informasi. Untuk menguji seorang siswa memiliki kecenderungan gaya belajar baca/tulis, berikut ini beberapa pertanyaan yang biasa diujikan. Pertama, untuk memperoleh dan mengingat suatu informasi, apakah anda harus membaca suatu teks? Kedua, apakah kejelasan tulisan, termasuk tanda baca, pilihan kata (diksi), kelogisan kalimat, keefektifan kalimat, dan paragraf yang variatif, sangat membantu anda dalam menangkap informasi? Ketiga, apakah menuliskan kembali informasi dalam diagram ke dalam beberapa kalimat membantu anda dalam mengingat informasi tersebut? Keempat, apakah membuat catatan kecil, menandai kalimat-kalimat tertentu dalam buku, membantu anda mengingat

informasi? Jika jawaban siswa atas keempat pertanyaan ini adalah ya, maka siswa tersebut memiliki preferensi gaya belajar baca/tulis.

4. Gaya Belajar Kinestetik

Siswa dengan gaya belajar kinestetik lebih suka memperoleh informasi melalui aktivitas praktik yang melibatkan fisik dan mengalami langsung dalam situasi kelas atau di luar kelas. Pengalaman merupakan hal yang penting bagi siswa kinestetik. Aktivitas seperti melakukan percobaan, membuat sesuatu, mendemonstrasikan suatu gerakan, bermain drama atau role playing, merupakan aktivitas yang mereka minati. Untuk menguji seorang siswa memiliki kecenderungan gaya belajar kinestetik, berikut ini beberapa pertanyaan yang biasa diujikan.

Pertama, untuk memperoleh dan mengingat suatu informasi, apakah anda harus melakukan aktivitas fisik, seperti mempraktikkan, mendemonstrasikan, atau memeragakan?

Kedua, apakah pengalaman langsung dengan objek dan bahan sangat membantu anda dalam menangkap informasi?

Ketiga, apakah sulit bagi anda untuk duduk diam untuk beberapa saat dan fokus memahami suatu informasi?

Keempat, apakah aktivitas praktik lebih menarik daripada aktivitas berpikir dan mengingat? Jika jawaban siswa atas keempat pertanyaan ini adalah ya, maka siswa tersebut memiliki preferensi gaya belajar kinestetik.

2.3 Data Mining

Data Mining sebagai proses untuk mendapatkan informasi yang berguna dari gudang basis data yang besar. Data mining juga dapat diartikan sebagai pengekstrakan informasi baru yang diambil dari bongkahan data besar yang membantu dalam pengambilan keputusan. Istilah data mining kadang disebut juga knowledge discovery. Data Mining merupakan proses ataupun kegiatan untuk mengumpulkan data yang berukuran besar kemudian mengekstraksi data tersebut menjadi informasi-informasi yang nantinya dapat digunakan (Mulyani, Rihadisha, Greshanayathi, Saputri, & Wulansari, 2020).

Langkah-langkah proses pelaksanaan data mining dalam tiga aktivitas adalah :

1. Eksplorasi Data, terdiri dari aktivitas pembersihan data, transformasi data, pengurangan dimensi, pemilihan ciri, dan lain-lain.
2. Membuat model dan Pengujian Validitas Model, merupakan pemilihan terhadap model-model yang sudah dikembangkan yang cocok dengan kasus yang dihadapi. Dengan kata lain, dilakukan pemilihan model secara kompetitif.
3. Penerapan model dengan data baru untuk menghasilkan perkiraan dari kasus yang ada. Tahapan ini merupakan tahapan yang menentukan apakah model yang telah dibangun dapat menjawab permasalahan yang dihadapi.

2.3.1 Komponen-Komponen Data Mining

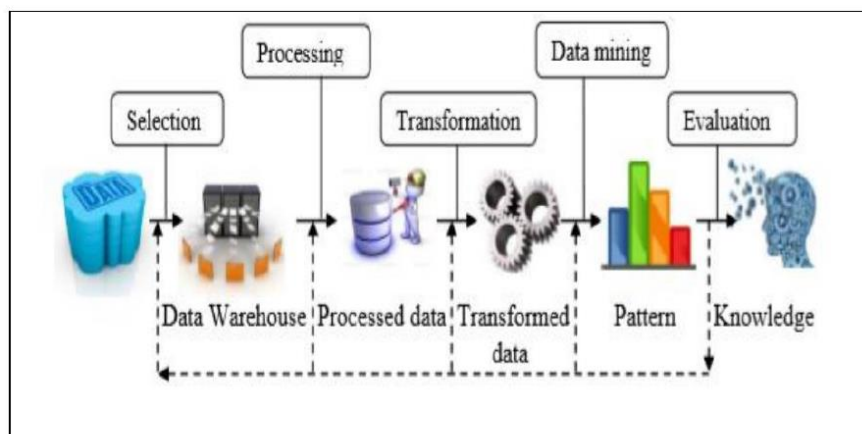
Komponen-komponen utama dari proses klasifikasi adalah (Mulyani et al., 2020) :

1. Kelas, merupakan variabel tidak bebas yang merupakan label dari hasil klasifikasi.

2. Prediktor, merupakan variabel bebas suatu model berdasarkan dari karakteristik atribut data yang diklasifikasikan, misalnya merokok, minuman beralkohol, tekanan darah, status perkawinan, dan sebagainya.
3. Set Data Pelatihan, merupakan sekumpulan data lengkap yang berisi kelas dan prediktor yang dilatih agar model dapat mengelompokkan ke dalam kelas yang tepat.
4. Set Data Uji, berisi data-data baru yang akan dikelompokkan oleh model guna mengetahui akurasi dari model yang telah dibuat

2.3.2 Proses Tahapan Data Mining

Data mining merupakan salah satu dari *rangkaian Knowledge Discovery in Database (KDD)*, KDD merupakan berhubungan dengan teknik integrase dan penemuan ilmiah, interpretasi, dan visualisasi dari pola-pola sejumlah data. Serangkaian proses tersebut memiliki tahapan yang dapat dilihat pada Gambar 2.1 (Hadianto, Novitasari, & Rahmawati, 2019):



Gambar 2.2 Tahapan *Data Mining*

1. Pembersihan Data (*Data Cleaning*)

Pembersihan data merupakan proses menghilangkan *noise* dan data yang tidak konsisten atau data tidak relevan.

2. Integrasi Data (*Data Integration*)

Integrasi data merupakan penggabungan data dari berbagai *database* baru.

3. Seleksi Data (*Data Selection*)

Data yang ada pada *database* sering kali semuanya dipakai, oleh karena itu hanya data yang sesuai untuk dianalisis yang akan diambil dari *database*.

4. Transformasi Data (*Data Transformtion*)

Data diubah atau digabung kedalam format yang sesuai untuk diproses dalam *data mining*.

5. Proses Mining

Merupakan suatu proses utama saat metode diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data. Beberapa metode yang dapat digunakan berdasarkan pengelompokan data mining.

6. Evaluasi Pola (*Pattern Evaluation*)

Untuk mengidentifikasi pola-pola menarik kedalam Knowledge Based yang ditentukan.

7. *Knowledge*, yaitu sebuah hasil yang dicapai berupa pengetahuan atau sebuah informasi.

2.3.3 Fungsi Data Mining

Menurut (Hadianto et al., 2019) fungsi data mining dapat dikelompokkan ke dalam enam kelompok yaitu:

1. Klasifikasi (*classification*): menggeneralisasi struktur yang diketahui untuk diaplikasikan pada data-data baru. Misalkan, klasifikasi penyakit ke dalam sejumlah jenis, klasifikasi *email* ke dalam spam atau bukan.
2. Klasterisasi (*clustering*): mengelompokkan data, yang tidak diketahui label kelasnya, ke dalam sejumlah kelompok tertentu sesuai dengan ukuran kemiripannya.
3. Regresi (*regression*): menemukan suatu fungsi yang memodelkan data dengan galat (kesalahan prediksi) seminimal mungkin.
4. Deteksi anomali (*anomaly detection*): mengidentifikasi data yang tidak umum, bisa berupa *outlier*, perubahan atau deviasi yang mungkin sangat penting dan perlu investigasi lebih lanjut.
5. Pembelajaran aturan asosiasi (*association rule mining*) atau pemodelan kebergantungan (*dependency modeling*): mencari relasi antar variabel..
6. Perangkuman (*summarization*): menyediakan representasi data yang lebih sederhana, meliputi visualisasi dan pembuatan laporan.

2.4 Clustering

Clustering merupakan upaya untuk mengelompokkan record, observasi, atau mengelompokkan ke dalam kelas yang memiliki kesamaan objek (Kusuma V. M., 2017). Pengklasteran berbeda dengan klasifikasi yang tidak adanya variable target dalam pengklasteran. Pengklasteran tidak digunakan untuk melakukan

klasifikasi, mengestimasi, atau memprediksi nilai dari target. Pengklasteran digunakan untuk melakukan pembagian terhadap keseluruhan data menjadi kelompok – kelompok yang memiliki kemiripan.

2.5 K-Means

Berikut ini langkah-langkah yang terdapat pada algoritma K-Means (Ediyanto et al., 2013) :

1. Tentukan k sebagai jumlah *cluster* yang dibentuk

Untuk menentukan banyaknya *cluster* k dilakukan dengan beberapa pertimbangan seperti pertimbangan teoritis dan konseptual yang mungkin diusulkan untuk menentukan berapa banyak *cluster*. Penelitian ini akan menggunakan metode *elbow criterion* dimana metode ini sangat praktis untuk memilih jumlah *cluster* k yang akan digunakan untuk pengelompokan data pada algoritma K-Means. (Madhulatha, 2012). Metode *elbow* ini, dapat dihasilkan dari perbandingan hasil SSE (*Sum of Squared Error*) dengan rumus SSE seperti dibawah ini (Irwanto, et. al, 2012):

$$SSE = \sum_{k=1}^K \sum_{x_i \in S_k} \|X_i - C_k\|^2 \dots\dots\dots (2.1)$$

Dimana : X_i menyatakan *norma euclid* (L2) dan C_k adalah pusat kluster S_k yang dihitung berdasarkan rata-rata jarak titik-titik kluster ke pusat kluster.

2. Tentukan k centroid (titik pusat cluster) awal secara random Penentuan centroid awal dilakukan secara random/acak dari objek-objek yang tersedia sebanyak k cluster, kemudian untuk menghitung centroid cluster ke-i berikutnya, digunakan rumus sebagai berikut :

$$v = \frac{\sum_{i=1}^n x_i}{n}; i = 1, 2, 3, \dots, n \dots\dots\dots (2.2)$$

Dimana; v : centroid pada cluster
 xi : objek ke-i
 n : banyaknya objek/jumlah objek yang menjadi anggota cluster

3. Hitung jarak setiap objek ke masing-masing centroid dari masing-masing cluster. Untuk menghitung jarak antara objek dengan centroid dapat menggunakan Euclidian Distance

$$d(x, y) = \|x - y\| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}; i = 1, 2, 3, \dots, n \dots\dots\dots (2.3)$$

Dimana; xi : objek x ke-i
 yi: daya y ke-i
 n : banyaknya objek

4. Alokasikan masing-masing objek ke dalam centroid yang paling dekat. Untuk melakukan pengalokasian objek kedalam masing-masing cluster pada saat iterasi secara umum dapat dilakukan dengan cara hard kmeans dimana secara tegas setiap objek dinyatakan sebagai anggota cluster dengan mengukur jarak kedekatan sifatnya terhadap titik pusat cluster tersebut.
5. Lakukan iterasi, kemudian tentukan posisi centroid baru dengan menggunakan persamaan (2.2)
6. Ulangi langkah 3 jika posisi centroid baru tidak sama

2.6 X-Means

Algoritma X-means merupakan salah satu algoritma clustering, algoritma x-means dimulai dengan asumsi memiliki sejumlah minimum cluster, dan kemudian secara dinamis meningkatkannya. X-means menggunakan kriteria pemisahan tertentu untuk mengontrol proses pengelompokan kluster. Algoritma xmeans adalah variasi dari k-means clustering yang memurnikan tugas cluster dengan berulang kali mencoba pembagian, dan menjaga hasil perpecahan terbaik, sampai beberapa kriteria tercapai.

Pelleg dan Moore (2000) mendefinisikan algoritma X-means yang diadaptasi dari K-means, x-means beraksi setelah masing-masing menjalankan Kmeans, membuat keputusan tentang bagian mana dari arus pusat-pusat harus memisahkan diri untuk mendapatkan data yang lebih baik. Pembelahan keputusan dilakukan dengan menghitung kriteria BIC. Pendekatan baru ini mengusulkan solusi efisien untuk satu kelemahan utama K-means yaitu pencarian dari jumlah cluster K. Selain itu, X-means memiliki biaya komputasi yang rendah. Tapi hasil dari properti non-konvergensi dari K-means algoritma. Langkah-langkah algoritma x-means adalah sebagai berikut:

1. Inisialisasi $K = K_{min}$
2. Jalankan algoritma K-means
3. Untuk $k = 1, \dots, K$: Ganti setiap centroid μ_k dengan dua centroid $\mu(1)$ dan $\mu(2)$.

Dua centroid baru untuk inisialisasi masing-masing algoritma K-means diperoleh dengan mempengaruhi centroid asli dalam dua arah berlawanan

sepanjang vektor yang dipilih secara acak dengan jumlah yang proporsional dengan ukuran kluster yang bersangkutan.

4. Jalankan algoritma K-means dengan $K = 2$ di atas k cluster Ganti atau pertahankan setiap centroid berdasarkan kriteria pemilihan model. Algoritma melakukan tes pemilihan model BIC untuk menentukan apakah dua cluster baru adalah model yang lebih baik daripada kluster tunggal asli di masing-masing kasus. $BIC_{\{k\}} = -2 \cdot \log L + k \log N$ yang N adalah jumlah bobservasi dan k adalah jumlah cluster dan $\log L$ adalah log-likelihood.
5. Jika kondisi konvergensi tidak puas, lanjutkan ke Langkah (2). Jika sudah, Berhenti.

2.7 K-Medoids

K-Medoids atau *Partitioning Around Medoids* (PAM) adalah algoritma clustering yang mirip dengan K-Means. Perbedaan dari kedua algoritma ini yaitu algoritma K-Medoids atau PAM menggunakan objek sebagai perwakilan (medoid) sebagai pusat cluster untuk setiap cluster, sedangkan K-Means menggunakan nilai rata-rata (mean) sebagai pusat cluster. Algoritma K-Medoids memiliki kelebihan untuk mengatasi kelemahan pada pada algoritma K-Means yang sensitive terhadap *noise* dan *outlier*, dimana objek dengan nilai yang besar yang memungkinkan menyimpang pada dari distribusi data. Kelebihan lainnya yaitu hasil proses clustering tidak bergantung pada urutan masuk dataset. Langkah-langkah algoritma K-Medoids:

1. Inisialisasi pusat cluster sebanyak k (jumlah cluster)
2. Alokasikan setiap data (objek) ke cluster terdekat menggunakan persamaan ukuran jarak *Euclidian Distance* dengan persamaan:

$$d(x,y)=\|x-y\|=\sqrt{\sum_{i=1}^n(x_i-y_i)^2}; 1,2,3,\dots,n$$
3. Pilih secara acak objek pada masing-masing cluster sebagai kandidat medoid baru.
4. Hitung jarak setiap objek yang berada pada masing-masing cluster dengan kandidat medoid baru.
5. Hitung total simpangan (S) dengan menghitung nilai total distance baru – total distance lama. Jika $S < 0$, maka tukar objek dengan data cluster untuk membentuk sekumpulan k objek baru sebagai medoid.

Ulangi langkah 3 sampai 5 hingga tidak terjadi perubahan medoid, sehingga didapatkan cluster beserta anggota cluster masing-masing

2.8 Indeks *Davis-Bouldin*

Metode *Davies Bouldin-Index* merupakan suatu metode untuk mengevaluasi klaster yang ditemukan oleh David L. Davies dan Donald W. Bouldin. Evaluasi menggunakan metode *Davies Bouldin Index* dapat dilihat dari kohesi untuk mengukur berapa besar kedekatan atau jarak antara data-data yang berada pada klaster yang sama. Selain itu dilihat juga dari separasi untuk mengukur perbedaan jarak data-data yang berbeda dengan klaster lain (Saikhu & Gita, 2013).

Terdapat Langkah-langkah perhitungan *Davies Bouldin-Index* adalah sebagai berikut:

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

Pada perhitungan *Sum of Squere Within Cluster* dapat mengetahui nilai kohesi dalam sebuah klaster ke I. Berikut Persamaan 2.4 untuk memperoleh nilai *Sum of Squere Within Cluster*.

$$SSW_i = \frac{i}{m_i} \sum_{j=i}^{m_i} d(x_j, c_j) \dots\dots\dots (2.4)$$

Dimana:

m_i = Jumlah data dalam klaster ke-i

c_i = Centroid klaster ke-i

$d(x_j, c_j)$ = Jarak setiap data ke centroid i yang dihitung menggunakan *euclidean*

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

Pada perhitungan *Sum of Squere Between Cluster* dapat mengetahui nilai separasi antar klaster. Berikut Persamaan 2.5 untuk memperoleh nilai *Sum of Squere Between Cluster*.

$$SSB_{i,j} = d(x_i, x_j) \dots\dots\dots (2.5)$$

Dimana:

$d(x_i, x_j)$ = Jarak antara data ke-i dengan data ke-j pada klaster yang lain.

3. *Ratio*

Pada perhitungan *Ratio (Ri,j)* agar dapat mengetahui nilai perbandingan klaster ke-I dan klaster ke-j untuk menghitung nilai rasio yang dimiliki

oleh masing-masing klaster. Berikut Persamaan 2.6 untuk memperoleh nilai Ratio.

$$R_{i,j,\dots,n} = \frac{SSW_i + SSW_j + \dots + SSW_n}{SSB_{i,j} + \dots + SSB_{n,i,j}} \dots\dots\dots (2.6)$$

Dimana: S

SSW_i = Sum of Square Within Cluster pada centroid i

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

4. *Davies-Bouldin Index (DBI)*

Setelah menghitung nilai ratio selanjutnya melakukan perhitungan Davies-Bouldin Index (DBI) dimana nilai ratio tersebut digunakan dalam perhitungan ini. Berikut Persamaan 2.7 untuk memperoleh nilai Davies-Bouldin Index (DBI).

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} (R_{i,j,\dots,k}) \dots\dots\dots (2.7)$$

Dimana:

$R_{i,j}$ = Nilai ratio dari nilai SSW dan SSB

2.9 Metode *Elbow*

Metode Elbow merupakan metode visualisasi selisih *Within Sum of Square Error* (WSSE) yang membentuk sudut siku pada grafik. Evaluasi beberapa metode optimasi jumlah cluster pada algoritma K-Means telah dilakukan dengan hasil Metode Elbow memiliki keunggulan waktu komputasi lebih efisien dibandingkan dengan metode *Gap Statistic*, metode *Silhouette Coefficient*, metode *Canopy* jika diterapkan pada *dataset* yang kecil dengan hasil jumlah *cluster* yang dihasilkan sama, namun jika yang digunakan adalah dataset besar maka metode *Canopy* adalah

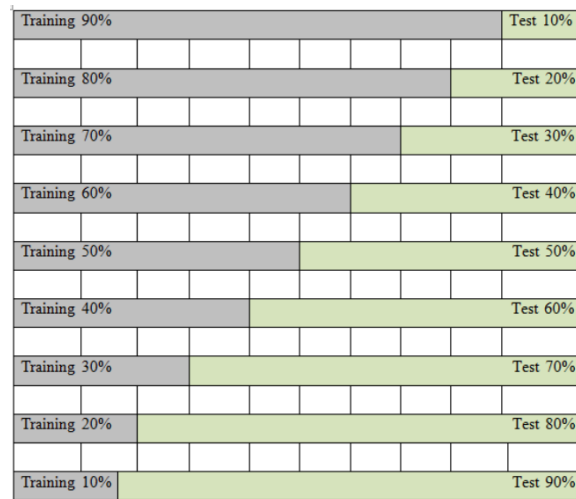
yang paling efisien (Yuan dan Yang, 2019). Berdasarkan penelitian-penelitian tersebut dapat dikatakan penentuan jumlah cluster terbaik menggunakan metode *Elbow* dapat meningkatkan efektifitas dan efisiensi algoritma K-Means untuk menghasilkan *cluster* yang baik.

Tujuan dari metode *elbow* adalah untuk memilih nilai k yang kecil dan masih memiliki nilai *withinss* yang rendah. Penentuan optimal jumlah *cluster* pada penelitian ini menggunakan salah satu metode analisis *cluster* yaitu metode *Elbow*, dengan memperhatikan nilai perbandingan (dari perhitungan SSE untuk setiap nilai *cluster*) antara jumlah *cluster* yang akan membentuk siku pada suatu titik, sehingga semakin besar jumlah *cluster* k maka nilai SSE akan semakin kecil. Rumus SSE sebagai berikut.

$$\sum_{k=1}^k \sum_{x_1 \in S_k} \|x_1 - C_k\|_2^2 \dots\dots\dots (2.8)$$

2.10 Split Validation

Split Validation adalah teknik validasi yang membagi data menjadi dua bagian secara acak, sebagian sebagai *data training* dan sebagian lainnya sebagai *data testing*. Dengan menggunakan *Split Validation* akan dilakukan percobaan *training* berdasarkan *split ratio* yang telah ditentukan sebelumnya, untuk kemudian sisa dari *split ratio data training* akan dianggap sebagai *data testing*. *Data training* adalah data yang akan dipakai dalam melakukan pembelajaran sedangkan *data testing* adalah data yang belum pernah dipakai sebagai pembelajaran dan akan berfungsi sebagai data pengujian kebenaran atau keakurasian hasil pembelajaran.



Gambar 2.3 Ilustrasi *Split Validation*