

BAB II

LANDASAN TEORI

2.1 Tinjauan Pustaka

Ardiansyah, menurut temuan Muhammad (2014) Analisis kluster data *time series* merupakan teknik *data mining* yang bertujuan untuk mengkategorikan hal-hal yang berhubungan dengan data *time series*. Pemanfaatan data *time series* di beberapa bidang, seperti teknik, bisnis, dan ekonomi. Mengklasifikasikan data deret waktu menggunakan jarak berdasarkan autokorelasi, invarian kompleksitas, periodogram, dan *Dynamic Time Warping* (DTW). Ketika *cluster* hierarkis dengan keterkaitan rata-rata dalam jarak *Dynamic Time Warping* dianalisis, jarak terbaik diperoleh dengan menggunakan korelasi *cophenetic*. Menggunakan pendekatan K-means untuk meningkatkan hasil belajar *cluster* hirarki, jumlah kelompok yang dibentuk oleh *cluster* hirarkis. Teknik *K-means* menegaskan pengelompokan dengan menggunakan koefisien siluet. Nilai korelasi sebesar 0.9775802 ini menunjukkan bahwa *Dynamic Time Warping* merupakan jarak terbaik. Validasi pengelompokan dengan metode *K-means* menunjukkan bahwa jarak *Dynamic Time Warping* menghasilkan pengelompokan dengan koefisien siluet 0,257 yang dianggap sangat baik.

Penelitian serupa sebelumnya dilakukan oleh Pratiwi Eka Puspita dengan judul “*A Practical Evaluation of Dynamic Time Warping in Financial Time series Clustering*” oleh Zulkarnain (2020). Membatasi risiko investasi, pemilihan portofolio saham yang luas merupakan tantangan besar bagi investor. *Clustering* dapat digunakan untuk mengatasi masalah ini sebagai pekerjaan penambangan data. Metode ini mengelompokkan ekuitas dengan jarak yang lebih dekat. Teknik untuk mengevaluasi perbedaan sangat penting untuk menghasilkan pengelompokan berkualitas tinggi. Selama periode 2000-2020 terdiri dari 25 saham yang terlacak oleh Indeks Saham Syariah Indonesia (ISSI) dan 30 saham yang terlacak oleh Jakarta Islamic Index (JII). Dengan menggunakan indeks *Silhouette* dan waktu lari, efektivitas pendekatan jarak dalam pengelompokan dinilai. *Cluster* berbasis DTW memberikan hasil yang lebih besar yang secara statistik tidak berbeda dari *cluster*

berbasis *Euclidean* berdasarkan nilai rata-rata indeks *Silhouette*. Di sisi lain, jenis dataset memberikan kontribusi yang signifikan. ISSI lebih rendah dari indeks siluet untuk dataset JII (homogen). Oleh karena itu, ia memilih untuk mengelompokkan data menggunakan pendekatan kesamaan dengan waktu pemrosesan yang lebih cepat jika kualitas metode kesamaan sebanding.

Penelitian yang dilakukan oleh Shitian Shen, Min Chi (2017) dengan judul “*Clustering Student Sequential Trajectories Using Dynamic Time Warping*”. Penelitian ini membahas pengelompokkan siswa secara langsung berdasarkan lintasan interaktif sekuensial sistem-siswa dari waktu menuju waktu. Tujuannya riset berikut yakni guna membangun kerangka kerja pengelompokan temporal umum menangkap karakteristik berbeda dari pola perilaku sekuensial siswa, melacak apakah pengalaman belajar siswa adalah tidak menguntungkan, dan dapat mengidentifikasi individu seperti itu sedini mungkin mungkin sehingga pembelajaran dipersonalisasi dapat ditawarkan.

Pusat ide kerangka kerja didasarkan pada *Dynamic Time Warping* (DTW), menghitung jarak antara dua urutan temporal bahkan dengan panjang berbeda. Dalam makalah ini, penulis jelajahi DTW asli dan normalisasi penulis usulkan DTW untuk menghasilkan matriks jarak dan menerapkan Hirarki *Clustering* ke matriks jarak dihasilkan. Untuk sepenuhnya mengevaluasi kekuatan kerangka kerja pengelompokan sekuensial temporal, penulis menghitung matriks jarak pada tiga jenis granularitas dalam urutan yang meningkat dari: masalah, level, dan sesi di tiga set data pelatihan. Seperti yang diharapkan, hasil menunjukkan bahwa pengelompokan urutan temporal momen-ke-momen pada masalah granularity lebih efektif daripada level dan *session granularity*. Selain itu, DTW normalisasi kami usulkan lebih banyak efektif daripada DTW asli dan *Euclidean* dasar jarak.

Menurut penelitian Muhammad, Ardiansyah (2014), analisis kluster data deret waktu merupakan cara untuk mengurutkan sesuatu berdasarkan data deret waktu. Data deret waktu digunakan di berbagai bidang, seperti teknik, bisnis, dan ekonomi. Data deret waktu dimasukkan ke dalam kelompok menggunakan jarak berdasarkan autokorelasi, invarian kompleksitas, periodogram, dan *Dynamic Time Warping* (DTW). Dengan menggunakan korelasi *cophenetic*, Anda dapat

menemukan jarak terbaik antara kluster hierarkis yang rata-rata keterkaitannya berada dalam jarak *Dynamic Time Warping*. Menggunakan metode K-means untuk meningkatkan hasil belajar, jumlah kelompok yang dibentuk oleh hirarki *cluster*, dan jumlah hirarki *cluster*. Menggunakan koefisien siluet, metode K-means berfokus pada pengelompokan. Nilai korelasi sebesar 0.9775802 menunjukkan bahwa jarak terbaik adalah *Dynamic Time Warping*. Menggunakan metode K-means untuk memvalidasi pengelompokan menunjukkan bahwa jarak *Dynamic Time Warping* membuat pengelompokan dengan koefisien siluet 0,257, yang sangat baik.

“Analisis *Time series* Kluster Dalam Pengelompokan Provinsi di Indonesia Berdasarkan Nilai PDRB” adalah judul penelitian Luh Putu Widya Adnyani dan Pardomuan Robinson Sihombing yang akan diterbitkan pada tahun 2021. Penulis menjelaskan bahwa *cluster time series* merupakan cara untuk mengelompokkan data deret waktu yang dapat digunakan untuk mengelompokkan berbagai provinsi di Indonesia berdasarkan nilai PDRBnya. Jarak waktu dinamis (DTW) dan jarak *Euclidean* digunakan untuk menganalisis *cluster*. Kedua jarak tersebut dibandingkan untuk melihat mana yang memberikan hasil pengelompokan terbaik, terutama dalam hal menggabungkan provinsi-provinsi di Indonesia berdasarkan nilai PDRB-nya. Berdasarkan penelitian, dibuat enam kelompok, dan koefisien *Silhouette* untuk metode DTW adalah 0,68, sedangkan koefisien *Silhouette* untuk metode Eucliden adalah 0,605. Sehingga dapat dikatakan bahwa kedua cara pembuatan kelompok tersebut menghasilkan klasifikasi kelompok yang baik.

Muhammad Irfan Rizkia, Teguh Ammar Taqqiyuddin, dan Jessica Jesslyn Cerelia pada tahun 2021 melakukan penelitian bertajuk “K-Medoids *Clustering* Dengan *Dynamic Time Warping* Distance Dalam Pengelompokan Provinsi di Indonesia Berdasarkan Kasus Aktif COVID-19”. Dalam penelitiannya, Indonesia merupakan salah satu negara dengan kasus COVID-19 teraktif. Banyak provinsi yang dapat mempersulit pemerintah dan gugus tugas penanganan COVID-19 nasional untuk menangani kasus COVID-19 yang sudah terjadi. Karena setiap provinsi di Indonesia berbeda, penting untuk melihat perkembangan kasus aktif di masing-masing provinsi. Jadi, provinsi yang berbeda dimasukkan ke dalam

kelompok berdasarkan kesamaan yang mereka miliki sehingga perubahan jumlah kasus aktif dapat dipahami. Hal ini dapat memudahkan pemerintah dalam menangani kasus COVID19 yang masih terjadi dengan melihat kesamaan dari masing-masing klaster. Pengelompokan kasus aktif dilakukan dengan menganalisis klaster dengan metode K-Medoids *Clustering* dan mencari jarak antar *time series* dengan *Dynamic Time Warping* (DTW), karena data merupakan *time series* dari tanggal 21 Maret 2020 sampai dengan 4 Juli 2021 Pada tahap pengelompokan, jumlah kelompok terbaik adalah kelompok dengan koefisien *pseudo-F* terbesar. Ada tiga kelompok, dan perubahan jumlah kasus aktif tidak sama di setiap kelompok. Golongan 1 dan 2 wajib diwaspadai karena jumlah kasus aktif di dalamnya banyak berubah.

Penelitian Raditya Novidianto dan Andrea Tri Rian Dani berjudul “Analisis Klaster Kasus Aktif COVID-19 Menurut Provinsi di Indonesia Menggunakan Data *Time series*” (2021). Studinya menyebutkan bahwa Corona Virus Disease 2019 (COVID-19) adalah epidemi global utama yang menyerang lebih dari 240 negara. Pada 11 Maret 2020, WHO menyatakan bahwa COVID-19 adalah pandemi. Kini, wabah COVID-19 menyebar dengan lebih serius, terutama di Indonesia. Sejak berita tersebut keluar pada awal Maret hingga Juli, 80.094 orang dinyatakan terinfeksi COVID-19, 3.797 orang meninggal dunia, dan 39.094 orang dinyatakan sembuh. Salah satu masalah yang dihadapi pemerintah dalam menghadapi wabah COVID-19 adalah keadaan darurat dan strategi yang dipilih pemerintah daerah.

Setiap hari, langkah-langkah harus diambil untuk menangani masalah COVID-19 yang sedang berlangsung. Namun, setiap lokasi memiliki kepribadiannya masing-masing, sehingga wawasan harus digunakan setiap hari. Maka, tujuan dari penelitian ini adalah untuk menggunakan data *time series* dari setiap provinsi di Indonesia untuk melakukan analisis deskriptif kasus aktif COVID-19. Kemudian, data kasus aktif COVID-19 di Indonesia digunakan untuk mengelompokkan kasus tersebut. Metode pengelompokan menggunakan algoritma keterkaitan tunggal, lengkap, dan rata-rata, yang semuanya merupakan jenis pengelompokan hierarki aglomeratif. *Euclidean Distance* dan *Dynamic Time Warping* digunakan untuk mengetahui seberapa mirip dua hal (DTW).

Menggunakan ukuran seberapa baik sesuatu, yang disebut koefisien koneksi *cophenetic*, hasil penelitian menunjukkan bahwa jarak *Euclidean* adalah jalan keluar terbaik dari ketiganya yang digunakan untuk mengukur seberapa mirip dua hal. Dendrogram yang dibuat dari hasil pengelompokan menunjukkan bahwa ketiga metode tersebut mengelompokkan orang-orang yang memiliki kesamaan sifat. Artinya, data yang ditemukan kelompok ini dapat membantu pemerintah pusat dan daerah membuat rencana untuk menghentikan penyebaran virus COVID-19.

Penelitian yang dilakukan oleh Masagus Reyhan Fadhillah, Jadi Suprijadi, Titi Purwandari (2021) dengan judul “Pengelompokkan Emisi Karbon Pulau Sumatera dengan Jarak *Dynamic Time Warping* (DTW) “. Dalam penelitian tersebut dilakukan analisis pada pencemaran emisi gas karbon sebagai suatu masalah yang seringkali dijumpai pada negara Indonesia. Emisi gas karbon paling banyak dihasilkan dikarenakan kebakaran hutan yang terjadi di wilayah Indonesia. Pulau Sumatera merupakan pulau yang dekat dengan negara tetangganya misalnya Malaysia, Singapura, maka dalam penelitian difokuskan untuk meneliti di pulau Sumatera. Berdasarkan perbedaan pola curah hujan akibat dari banyaknya pengeluaran emisi karbon dari area yang terbakar, membuat analisis peramalan curah hujan untuk tahun-tahun berikutnya menjadi kompleks sehingga langkah yang bisa untuk mengantisipasi banyaknya pengeluaran jumlah berat karbon adalah dengan mengelompokkan area yang rawan dengan kejadian kebakaran hutan. Pengelompokkan spatial (wilayah) yang mempunyai jumlah berat emisi karbon berbeda dengan karakteristik masing-masing temporal (deret waktu) tertentu melalui menganalisis *clustering* mempergunakan metode *K-medoids* dan menghitung jarak antarderet waktu mempergunakan *Dynamic Time Warping* (DTW). Total k yang diambil di tahapan *clustering* mengacu pada nilainya koefisien *pseudo f* paling besar. Ketepatan pengelompokkan ini diukur menggunakan *Silhouette Coefficient* dimana nilai koefisien yang besar menunjukkan kualitas pengelompokkan yang baik.

Jonatha Sousa Pimentel, Paulo Canas Rodrigues (2022) dalam penelitiannya yang berjudul “*Clustering the World Currency Exchange Rates Using Hierarchical Methods Based on Dynamic Time Warping*”. Penulis meneliti analisis nilai tukar

mata uang sangat penting untuk menganalisis kesehatan ekonomi suatu negara. Dalam penelitian ini, penulis mengumpulkan dan menganalisis data historis nilai tukar semua mata uang yang tersedia, dengan mempertimbangkan dolar AS sebagai referensi. Secara khusus, penulis tertarik untuk mengelompokkan deret waktu harian yang dikumpulkan dengan menggunakan ukuran kesamaan berdasarkan *Dynamic Time Warping* (DTW). Secara total, pengamatan 150 mata uang, antara 3 Januari 2005 dan 30 April 2020, dianalisis. Hasil penelitian menunjukkan bahwa penggunaan *Dynamic Time Warping* sebagai ukuran jarak menghasilkan peningkatan interpretasi dendrogram, jika dibandingkan dengan ukuran kesamaan standar seperti jarak *Euclidean*.

Penelitian yang dilakukan oleh Young Seon Jeong (2010) dengan judul "*Weighted Dynamic Time Warping for time series classification*". Dalam penelitiannya membahas *Dynamic Time Warping* (DTW), yang menemukan jalur minimum dengan menyediakan keberpihakan *non-linear* antara dua deret waktu, telah banyak digunakan sebagai ukuran jarak untuk klasifikasi deret waktu dan kekelompokan. Namun, DTW tidak memperhitungkan kepentingan relatif mengenai perbedaan fase antara titik referensi dan titik pengujian. Ini dapat menyebabkan kesalahan klasifikasi terutama dalam aplikasi di mana kesamaan bentuk antara dua urutan merupakan pertimbangan utama untuk pengenalan yang akurat.

Oleh karena itu, pada penelitiannya penulis mengusulkan ukuran jarak baru, yang disebut *weight dynamic time wrapping* (WDTW), yang merupakan DTW berbasis penalti. Pendekatannya menghukum poin dengan perbedaan fase lebih tinggi antara titik referensi dan titik pengujian untuk mencegah distorsi jarak minimum yang disebabkan oleh *outlier*. Alasan mendasari ukuran jarak yang diusulkan ditunjukkan dengan beberapa contoh ilustrasi. Baru fungsi bobot, yang disebut fungsi bobot logistik yang dimodifikasi (MLWF), juga diusulkan untuk secara sistematis menetapkan bobot sebagai fungsi dari perbedaan fasa antara titik referensi dan titik pengujian. Oleh menerapkan bobot yang berbeda ke titik yang berdekatan, algoritma yang diusulkan dapat meningkatkan deteksi kesamaan antara dua deret waktu. Penulis menunjukkan bahwa beberapa ukuran jarak populer

seperti DTW dan Jarak *Euclidean* adalah kasus khusus dari ukuran *weight dynamic time wrapping* WDTW yang penulis usulkan. Memperluas ide yang diusulkan kevarian DTW lainnya seperti turunan *derivative dynamic time wrapping* (DDTW) dan mengusulkan pembobotan versi DDTW. Penulis pada penelitian ini telah membandingkan kinerja prosedur yang diusulkan dengan populer lainnya pendekatan menggunakan kumpulan data publik yang tersedia melalui Arsip Penambangan Data Seri Waktu UCR untuk keduanya klasifikasi deret waktu dan masalah pengelompokan. Hasil percobaan menunjukkan bahwa usulan pendekatan dapat mencapai akurasi yang lebih baik untuk klasifikasi deret waktu dan masalah pengelompokan.

Penelitian oleh Santosh, Shubhra (2018) yang judulnya “*Optimal cluster analysis using hybrid K-Means and Ant Lion Optimizer*”. Dalam penelitiannya membahas analisis *clustering* menggunakan *K-Means Clustering*. *K-Means* adalah metode analisis *cluster* populer yang bertujuan untuk mempartisi sejumlah titik data menjadi *K cluster*. Ini telah berhasil diterapkan pada sejumlah masalah. Namun, efisiensi *K-Means* tergantung pada inisialisasi pusat *cluster*. Teknik kecerdasan *swarm* yang berbeda diterapkan pada masalah pengelompokan untuk meningkatkan kinerja. Dalam karya ini pendekatan *clustering hybrid* berdasarkan *K-means* dan *Ant Lion Optimization* telah dipertimbangkan untuk analisis *cluster* yang optimal. *Ant Lion Optimization* (ALO) adalah model optimasi global stokastik. Kinerja algoritma yang diusulkan dibandingkan dengan kinerja metode pengelompokan *K-Means*, *KMeans-PSO*, *KMeans-FA*, *DBSCAN* dan Revisi *DBSCAN* berdasarkan metrik kinerja yang berbeda. Eksperimen dilakukan pada delapan set data, yang analisis statistiknya dilakukan. Hasil yang diperoleh menunjukkan bahwa metode *hybrid K-Means* dan *Ant Lion Optimization* berkinerja lebih baik daripada tiga algoritma lainnya dalam hal jumlah jarak *intracluster* dan *F-measure*.

Penelitian terdahulu yang serupa yaitu penelitian Aniket, Pavant, Prashant (2021) dengan judul “*Performance Evaluation of the Data Clustering Techniques and Cluster Validity Indices for Efficient Toolpath Development for Incremental Sheet Forming*”. Tujuan dari penelitian ini adalah untuk membandingkan teknik *clustering* data dan indeks validitas *cluster* untuk ekstraksi ciri geometris

menggunakan *point cloud*. Di sini, titik awan dihasilkan dengan mengiris permukaan desain berbantuan komputer (CAD), dan data pada setiap irisan digunakan sebagai input ke algoritme pengelompokan. Teknik *clustering* digunakan untuk mendeteksi beberapa kontur tertutup pada dataset planer. Dalam makalah ini, empat teknik pengelompokan yang paling populer, yaitu, partisi berbasis (*K-means*), berbasis kepadatan (DBSCAN), pengelompokan hierarki tautan tunggal, varian dari pengelompokan hierarki, dan teknik pengelompokan berbasis grafik (spektral). dibandingkan dengan menggunakan empat dataset yang berbeda. Pengelompokan hierarki tautan tunggal lebih disukai daripada varian lain karena mendeteksi *cluster* berbentuk secara efisien dan efektif. Perbandingan tersebut didasarkan pada kemampuan keberhasilan pendeteksian kontur tertutup pada dataset planer, waktu yang dibutuhkan, dan input yang dibutuhkan untuk algoritma. Dari investigasi, ditemukan bahwa DBSCAN merupakan teknik yang paling cocok untuk pengembangan *feature-based toolpaths* (FBTs). Selain itu, untuk penilaian kualitas solusi pengelompokan dan untuk menentukan indeks validitas superlatif, teknik seperti *Calinski-Harabasz*, *Ball-Hall*, *Davies-Bouldin*, *Dunn*, *Det Ratio*, *Silhouette*, *Trace WiB*, dan *Log Det Ratio*. Solusi dari teknik *clustering* digunakan untuk membandingkan indeks validitas. Berdasarkan analisis komparatif, disimpulkan bahwa indeks *Ball-Hall*, *Dunn*, *Det Ratio*, dan *Log Det Ratio* merupakan kriteria validitas terbaik untuk kontur tertutup berbentuk arbitrer. Hasil keseluruhan dari penelitian ini akan membantu dalam membangun algoritme untuk strategi pengembangan jalur alat berbasis fitur untuk berbagai proses manufaktur menggunakan ilmu data dan teknik pembelajaran mesin. Jurnal penelitian dapat dilihat pada Tabel 2.1

Tabel 2.1 Jurnal Penelitian

No	Penulis	Tahun	Judul	Metode	Hasil
1	Muhamad, Ardiansyah	2014	“Penggunaan Jarak <i>Dynamic Time Warping</i> (DTW) pada Analisis <i>Cluster</i>	<i>Autocorrelation-based</i> , <i>Complexity in variant</i> , <i>Periodogram-</i>	Nilai korelasi <i>cophenetic</i> memperlihatkan bahwa DTW merupakan jalan

			Data Deret Waktu”	<i>based, Dynamic Time Warping, K Means Clustering</i>	terbaik dengan nilai <i>cophonetic</i> sebesar 0,9775805, algoritma <i>K Means</i> menunjukkan jarak DTW terbagi kelompok yang cukup baik dengan nilai <i>Silhouette</i> sebesar 0,257.
2	Pratiwi Eka Puspita, Zulkarnain	2020	“ <i>A Practical Evaluation of Dynamic Time Warping in Financial Time series Clustering</i> ”	<i>Dynamic Time Warping (DTW) dan K-Means Clustering</i>	Indeks <i>Silhouette</i> untuk dataset JII (homogen) lebih baik daripada ISSI (heterogen)
3	Shitian Shen, Min Chi	2017	“ <i>Clustering Student Sequential Trajectories Using Dynamic Time Warping</i> ”	<i>Dynamic Time Warping (DTW) dan Hirarki Clustering</i>	pengelompokan urutan temporal momen-ke-momen pada masalah granularity lebih efektif daripada level dan session

					granularity. Selain itu, metode perhitungan jarak DTW lebih efektif daripada <i>Euclidean Distance</i>
4	Iftitah Ayundari, Sutikno	2019	”Penentuan Zona Musim di Mojokerto Menurut Karakteristik Curah Hujan dengan Metode <i>Time series Based Clustering</i> ”	<i>Clustering Time series Based, Dynamic Time Warping (DTW)</i> dan <i>autocorrelation function (ACF)</i> .	Strategi yang dipilih dengan standar deviasi terendah adalah pengelompokan dengan <i>Dynamic Time Warping (DTW)</i> .
5	Luh Putu Widya Adnyani, Pardomuan Robinson Sihombing	2021	“Analisis <i>Cluster Time series</i> dalam Pengelompokan Provinsi di Indonesia Berdasarkan Nilai PDRB”	<i>Clustering Time series Based, Dynamic Time Warping (DTW)</i>	Akibat terbentuknya enam <i>cluster</i> , koefisien <i>Silhouette</i> untuk teknik DTW adalah 0,68. Koefisien siluet dari teknik <i>Euclidean</i> adalah 0,605%. Kedua teknik

					klasifikasi tersebut mengelompok dengan kategori klasifikasi efektif.
6	Irfan Rizkia, Teguh Ammar Taqqiyuddin, Jessica Jesslyn Cerelia	2021	“ <i>K-Medoids Clustering</i> dengan Jarak <i>Dynamic Time Warping</i> dalam Mengelompokkan Provinsi di Indonesia Berdasarkan Kasus Aktif COVID-19”	<i>Clustering Time series Based, Dynamic Time Warping (DTW), K-Meloids Clustering</i>	Ada tiga grup yang dibuat, dengan fluktuasi kasus aktif menunjukkan berbagai fitur di antara grup. Grup 1 dan 2 memerlukan perawatan khusus karena variasi yang besar dalam kasus aktif.
7	Raditya Novidia nto, Andrea Tri Rian Dani	2021	“Analisis Kluster Kasus Aktif COVID-19 Menurut Provinsi di Indonesia Berdasarkan Data Deret Waktu”	<i>Dynamic Time Warping (DTW), Euclidean Distance, agglomerative hierarchical clustering</i>	Jarak <i>Euclidean</i> adalah pengukuran kesamaan terbaik di antara ketiga teknik yang digunakan.
8	Jonatha Sousa	2022	“ <i>Clustering the World Currency</i> ”	<i>Dynamic Time Warping</i>	Menggunakan pembengkokan

	Pimente l, Paulo Canas Rodrigo es		<i>Exchange Rates Using Hierarchical Methods Based on Dynamic Time Warping”</i>	(DTW), <i>Euclidean Distance</i>	waktu dinamis sebagai metrik jarak meningkatkan interpretasi dendrogram dibandingkan dengan ukuran kesamaan konvensional seperti jarak DTW <i>Euclidean.</i>
9	Young Seon Jeong	2021	<i>“Weighted Dynamic Time Warping for time series classification. Dalam penelitiannya membahas Dynamic Time Warping (DTW)”</i>	<i>weight dynamic time wrapping (WDTW), Dynamic Time series (DTW), derivative dynamic time wrapping (DDTW)</i>	Metode yang disarankan meningkatkan ketepatan klasifikasi deret waktu dan masalah pengelompokan.
10	Santosh, Shubhra	2018	<i>“Optimal cluster analysis using hybrid K-Means and Ant Lion Optimizer”</i>	<i>hybrid K-Means Clustering dan Ant Lion Optimization</i>	Dalam hal total jarak <i>intracluster</i> dan <i>F-measure</i> , pendekatan hibrid <i>K-Means</i> dan <i>Ant Lion</i>

					<i>Optimization</i> mengungguli tiga algoritma lainnya
11	Aniket, Pavant, Prashant	2021	<i>“Performance Evaluation of the Data Clustering Techniques and Cluster Validity Indices for Efficient Toolpath Development for Incremental Sheet Forming”</i>	<i>K-Means Clustering</i> dan DBSCAN	Untuk kontur tertutup bentuk arbitrer, indeks <i>Ball-Hall</i> , <i>Dunn</i> , <i>Det Ratio</i> , dan <i>Log Det Ratio</i> adalah kriteria validitas terbaik.

2.2 Dasar Teori

2.2.1 Revenue atau Pendapatan

Pendapatan perusahaan adalah salah satu faktor yang paling penting dalam menghasilkan laporan laba rugi. Banyak orang bingung dengan istilah pendapatan. Karena pendapatan dapat diartikan sebagai pendapatan dan keuntungan, pendapatan dapat dianggap sebagai keuntungan, sedangkan pendapatan mengacu pada pendapatan atau keuntungan. Kamus Besar Bahasa Indonesia menyatakan bahwa pendapatan adalah hasil kerja (usaha atau sebagainya). Dalam bahasa manajemen, pendapatan mengacu pada uang yang dihasilkan oleh individu, bisnis, dan organisasi lain dalam bentuk upah, gaji, sewa, bunga, komisi, biaya, dan keuntungan. Penjualan produk bisnis menghasilkan uangnya. Arus masuk aset, pertumbuhan aset lain, atau pembayaran liabilitas yang timbul dari penyerahan, penyediaan layanan, atau aktivitas lain yang merupakan bisnis utama atau inti entitas.

Semakin besar pendapatan perusahaan, semakin besar kapasitasnya untuk mendukung semua biaya dan operasinya. Pendapatan terdiri dari semua penerimaan

kas dan nonkas dari penjualan produk atau jasa selama periode waktu tertentu (pendapatan pendapatan) (Solihin, 2010: 621). Selain itu, pendapatan dapat dianggap sebagai uang dari operasi utama perusahaan atau penjualan produk untuk jasa diikuti dengan dikurangi pengeluaran untuk mendapatkan laba kotor (Munawir, 2002: 26). Terdapat tiga kategori pendapatan yaitu,

- a. Penghasilan berupa uang, termasuk penghasilan tetap berupa uang yang sering diterima sebagai upah atau kompensasi.
- b. Penerimaan berupa barang meliputi semua penghasilan tetap dan rutin yang selalu merupakan balas jasa dan diterima dalam bentuk barang dan jasa.
- c. Penghasilan yang bukan penghasilan terdiri dari semua penerimaan redistributif yang biasanya mempengaruhi keuangan keluarga..

2.2.2 Data Mining

Menggunakan metode dari bidang statistik, matematika, AI, dan pembelajaran mesin, "penambangan data" mengekstrak wawasan yang bermakna dari kumpulan data yang sangat besar. Penemuan pengetahuan adalah fokus utama dari bidang akademik yang dikenal sebagai "penambangan data/*data mining*". Penemuan Pengetahuan dalam Basis Data adalah istilah umum untuk ini (KDD). Tujuan dari penemuan pengetahuan dalam data, atau KDD, adalah untuk menemukan hubungan yang bermakna antara potongan-potongan informasi yang tampaknya tidak terkait dalam kumpulan data yang sangat besar.

Penambangan data dan Penemuan Pengetahuan dalam Basis Data (KDD) keduanya terkait dengan proses menemukan hubungan dan pola yang sebelumnya tidak diketahui dalam kumpulan data masif, sebagaimana dinyatakan dalam sebuah buku oleh Fayyad (Kusrini, 2009). Ada hubungan antara nama-nama ini, meskipun memiliki arti yang berbeda. Penambangan data adalah langkah dalam proses Penemuan Pengetahuan dalam Basis Data (KDD). Pendekatan KDD dijelaskan sebagai berikut oleh Narwati (2010),

1. Memisahkan informasi relevan dari data tidak relevan

Pemilihan data dari kumpulan data operasional perlu dilakukan sebelum melanjutkan ke proses penambangan informasi KDD. Teknik

penambangan data menggunakan file untuk menyimpan data secara independen dari basis data operasional.

2. Metode Pembersihan dan Pretreatment

Pembersihan adalah prasyarat untuk menambang di KDD. Proses pembersihan data terdiri dari menghilangkan duplikasi, memverifikasi ketidakkonsistenan, dan memperbaiki kesalahan tipografi. Selain itu, KDD "memperkaya" datanya dengan informasi tambahan dari dunia luar jika menurutnya sangat relevan atau berharga untuk tujuannya.

3. Perubahan

Encoding adalah proses membuat penyesuaian diperlukan untuk data telah dipilih sehingga dapat ditambang secara efektif. KDD mengandalkan solusi pengkodean kreatif sangat kontekstual dengan jenis atau pola data ditambang.

4. Pertambahan untuk Informasi

Istilah "penambangan data" mengacu pada praktik menggunakan berbagai metode dan alat untuk menyaring sejumlah besar data untuk mencari pola atau wawasan. Ada banyak variasi dalam metode, strategi, dan algoritma penambangan data yang digunakan. Sasaran dan seluruh proses KDD diperhitungkan saat memutuskan metode atau algoritme terbaik untuk digunakan.

5. Analisis dan penilaian

Hasil penambangan data yang mengungkapkan pola dalam data harus disajikan dengan cara yang dapat dipahami oleh siapa saja. Fase interpretasi KDD mengacu pada proses ini. Cari tahu apakah data baru tersebut bertentangan dengan apa yang sudah Anda ketahui atau hipotesiskan.

2.2.3 Analisis *Clustering*

Analisis klaster adalah cara terbaik untuk mengelompokkan hal-hal yang mirip ke dalam kelompok. *Cluster* yang dihasilkan sangat mirip di dalam dan memiliki banyak perbedaan di luar (Hair, dkk., 2010: 486). Metode pengelompokan dapat dikelompokkan menjadi dua kelompok utama: yang menggunakan hierarki

dan yang tidak. *Clustering* adalah jenis penambangan data yang melibatkan penempatan data atau objek ke dalam kelas atau *cluster* berdasarkan kesamaan yang mereka miliki.

Pengelompokan yang baik mengelompokkan hal-hal yang memiliki banyak kesamaan dengan hal-hal lain dalam kelompok atau klaster yang sama tetapi tidak banyak kesamaan dengan hal-hal yang ada dalam kelompok atau klaster lain. Seperti yang telah kita lihat, sekarang ada dua cara untuk mengelompokkan berbagai hal. Mereka disebut pengelompokan hierarkis dan pengelompokan partisi. Metode *Hierarchical Clustering* meliputi *Complete Linkage Clustering*, *Single Linkage Clustering*, dan *Average Linkage Clustering*. Metode partisi termasuk *KMeans*, *K-Medoids*, dan *Fuzzy K-Means*. Strategi pengelompokan hirarkis terdiri dari dua jenis teknik: teknik *agglomerative* dan teknik *divisive*.

Dalam pendekatan *agglomerative*, proses mengelompokkan sesuatu dimulai dengan masing-masing item itu sendiri. Jadi, jumlah *cluster* sama dengan jumlah benda. Bagian-bagian yang paling mirip berkumpul terlebih dahulu untuk membuat *cluster*, dan seterusnya sampai *cluster* dibuat. Strategi memecah belah, di sisi lain, adalah kebalikan dari strategi *agglomerative*. Metode ini dimulai dengan menempatkan semua objek pengamatan ke dalam satu kelompok. Kemudian, benda-benda yang sangat berbeda satu sama lain akan dimasukkan ke dalam kelompok baru, demikian seterusnya, hingga jumlah klaster sama dengan jumlah benda yang dimasukkan ke dalam kelompok (Johnson & Wichern, 2007: 680).

Secara umum ada dua metode dalam analisis *cluster* yaitu,

1. Metode Non-Hierarki

Pendekatan non-hierarkis adalah metode pengelompokan dimana *cluster* yang akan dihasilkan ditentukan, sehingga item-item dipilah ke dalam *k grup* yang telah ditentukan.

2. Metode Hirarki

Teknik hirarkis dapat digunakan jika jumlah kelompok yang dibutuhkan tidak diketahui, dan sering digunakan untuk objek dengan jumlah observasi yang sedikit (Simamora, 2005).

Teknik hierarki menggunakan strategi koreksi jarak berikut (Supranto,

2004),

a. Metode pautan tunggal (*Single Linkage*)

Item individual yang dengan jelas menggabungkan jarak terdekatnya membuat *cluster*. Pada setiap langkah setelah terbentuknya *cluster* baru (UV), jarak antara UV dan *cluster* lainnya, seperti W, bertambah.:

$$d_{(uv)w} = \min \{d_{uw}, d_{vw}\} \quad (2.1)$$

Keterangan:

d_{uw} = jarak antara *cluster* U dan W

d_{vw} = jarak antara *cluster* V dan W

$d_{(uv)w}$ = jarak antara *cluster* (UV) dan W

b. Metode pautan lengkap (*Complete Linkage*)

Metode *complete linkage* dilakukan dengan cara yang sama seperti metode *single linkage*, dengan perbedaan pada setiap langkah jarak antar i ditentukan oleh jarak antara dua item, dengan objek yang paling jauh dari setiap *cluster* berfungsi sebagai faktor penentu. Setelah menggabungkan *cluster* U dan V menjadi *cluster* (UV), jarak antara UV dan klaster lainnya, seperti W, dihitung.:

$$d_{(uv)w} = \max \{d_{uw}, d_{vw}\} \quad (2.2)$$

Keterangan:

d_{uw} = jarak antara *cluster* U dan W

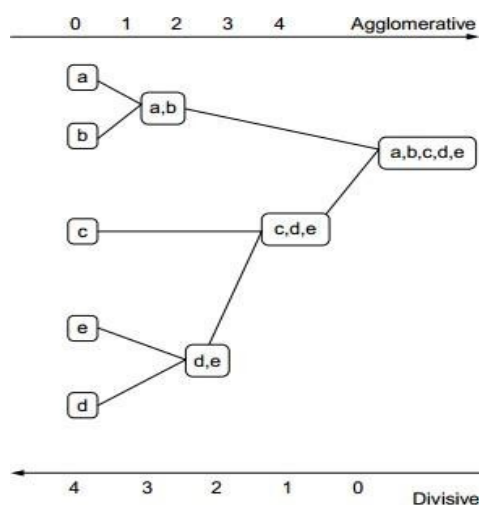
d_{vw} = jarak antara *cluster* V dan W

$d_{(uv)w}$ = jarak antara *cluster* (UV) dan W

c. Metode pautan rata-rata (*Average Linkage*)

Metode pautan rata-rata dijalankan hampir sama dengan metode sebelumnya, dengan pengecualian bahwa dik (Johnson dan Wichern, 1996:594) adalah jarak antara item I dalam gugus (UV) dan objek k dalam gugus W. $N(UV)$ dan N_W masing-masing mewakili jumlah item dalam kelompok (UV) dan (W).

$$d_{(uv)w} = \frac{\sum_i \sum_k d_{ik}}{N_{(UV)} N_W} \quad (2.3)$$



Gambar 2.1 Contoh Struktur Pohon Metode *Clustering* Hirarki
(Sumber : Kauffman & Rousseeuw (1970) dalam Everitt, dkk. (2011:72))

2.2.4 Uji Validitas *Cluster*

Setelah menggunakan teknik pengelompokan data untuk mendapatkan hasil atau solusi untuk setiap *cluster*, penting untuk memeriksa validitas setiap *cluster*. Uji validitas kluster diperlukan untuk memastikan bahwa hasil analisis kluster benar atau baik. Indeks *Silhouettes* adalah salah satu cara agar hasil metode pengelompokan hierarkis dapat diperiksa. Koefisien korelasi kofenetik adalah korelasi antara elemen matriks pertidaksamaan asli (matriks jarak *Euclidean*) dengan elemen matriks dendrogram (matriks kofenetik) (Silva & Dias, 2013:589-590).

Indeks *Sillhoutte* adalah salah satu cara untuk membandingkan subset, *cluster*, atau seluruh *cluster* data. Menggabungkan nilai kohesi dan separasi, ini adalah cara paling umum untuk memeriksa apakah kluster sudah benar (Eko

Prasetyo, 2014). Untuk mengetahui SI dari titik data ke- i , diperlukan dua variabel, a_i dan b_i . Setelah a dan b diketahui, nilai Indeks *Silhouette* dapat dihitung.

2.2.5 K-Means Clustering

K-Means adalah algoritma pengelompokan iteratif yang membagi kumpulan data menjadi K kategori yang telah ditentukan. Metode *K-Means* mudah digunakan, cukup cepat, fleksibel, dan digunakan secara luas. *K-Means* secara historis menjadi salah satu pendekatan data mining yang paling penting (Wu dan Kumar, 2009). Lloyd (1957, 1982), Forgey (1965), Friedman dan Rubin (1967), dan McEwen mengembangkan pendekatan K-means (1967). (V. Kumar, 2009). Lloyd menciptakan konsep *clustering* pada tahun 1957, namun baru dipublikasikan pada tahun 1982. Forgey juga mengembangkan teknik pada tahun 1965 yang sering dikenal dengan pendekatan Lloyd-Fordy. K-Means adalah strategi pengelompokan data non-hierarkis yang mencoba mengklasifikasikan data yang ada ke dalam beberapa kelompok. Pendekatan ini mengatur data ke dalam *cluster* atau kelompok sedemikian rupa sehingga data dengan kualitas yang identik (*high intra-class similarity*) dikelompokkan menjadi satu, sedangkan data dengan kualitas yang berbeda (Hukum kesamaan antar kelas) dikelompokkan secara terpisah. Pengelompokan data ini bertujuan untuk mengurangi fungsi tujuan dalam suatu kelompok dan meningkatkan keragaman lintas kelompok (Eko Prasetyo, 2012: 178).

Dasar algoritma *K-Means* adalah sebagai berikut,

1. Tentukan nilai k yang mewakili jumlah *cluster* yang diinginkan.
2. Inisialisasi k sebagai centroid berkemampuan acak.
3. Hitung jarak antara setiap titik data dan setiap centroid menggunakan persamaan *Euclidean Distance* berikut:

$$d(P, Q) = \sqrt{\sum_{j=1}^p (x_j(P) - x_j(Q))^2}$$

4. Kelompokkan setiap potongan data menurut jarak terpendek antara itu dan pusat massanya.

5. Tentukan lokasi *centroid* baru (k).
6. Kembali ke langkah 3 jika lokasi *centroid* baru berbeda dengan lokasi *centroid* sebelumnya.

Ada beberapa keuntungan menggunakan *k-means clustering*, termasuk fakta bahwa *clustering* selalu konvergen atau mampu melakukan *clustering*, tidak memerlukan operasi matematika yang kompleks, sederhana, dan beban komputasi relatif ringan, memungkinkan *clustering* dilakukan dengan cepat terlepas dari jumlah data dan jumlah kelompok yang diinginkan. Namun, ada kelemahan *k-means*, seperti sensitivitas metode terhadap *outlier* data, karena objek dengan nilai besar dapat menyimpang dari hasil grup turunan.

2.2.6 *Dynamic Time Warping (DTW)*

Dengan menggunakan pendekatan *Dynamic Time Warping (DTW)*, dimungkinkan untuk membandingkan beberapa bentuk data deret waktu dengan panjang variabel. DTW mengidentifikasi semua jalur layak dan memilih salah satu menghasilkan jarak terkecil antara dua deret waktu, menggunakan matriks jarak di mana setiap kolom mencerminkan jarak kumulatif dari nilai minimum tiga tetangga terdekat (Ayundari & Sutikno, 2019).

Karena sifat data yang dinamis, algoritma pengelompokan tidak dapat lagi menentukan kedekatan objek menggunakan jarak *Euclidean*. Jarak *Euclidean* cocok untuk data atau fungsi statis. Selama pengembangan matriks kedekatan jarak antar objek dalam data deret waktu, juga diperlukan untuk menghitung jarak antara deret waktu dinamis, memerlukan teknik pengelompokan yang berbeda dari yang diperlukan untuk data statis (Adnyani & Sihombing, 2021). Dengan demikian, *Dynamic Time Warping* adalah kemungkinan teknik pengukuran alternatif (DTW). Satuan pengukuran ini memungkinkan perbandingan berbagai panjang periode waktu.

Dua kumpulan data deret waktu digunakan untuk membuat matriks jarak m-dimensi: vektor A dengan m baris dan vektor B dengan n baris. Masuknya matriks jarak E, atau elemen I j, memuat semua kemungkinan jarak antara dua deret waktu,

salah satunya akan dipilih berdasarkan nilai kumulatif terkecil dari persamaan 1 (Niennattrakul & Ratanamahatana, 2007).

$$e_{ij} = d_{ij} + \min \{e_{(i-1)(j-1)}, e_{(i-1)j}, e_{i(j-1)}\} \quad (1)$$

Dimana

$$d_{ij} = (a_i - b_j)^2$$

e_{ij} = elemen (i,j)

Matriks jarak E yang berisi rute-rute potensial akan digunakan untuk menentukan rute ideal yang meminimalkan total jarak. Dengan menggunakan rumus pada Persamaan 2, jarak DTW minimal antara dua data deret waktu A dan B dihitung.

$$D_{DTW}(A, B) = \min_w \left[\sqrt{\sum_{l=1}^L e_{wl}} \right] \quad (2)$$

Dengan

$$w = 1, 2, \dots, W$$

W : keseluruhan jalur *warping* yang mungkin

w : elemen dari W yang ada pada matriks E

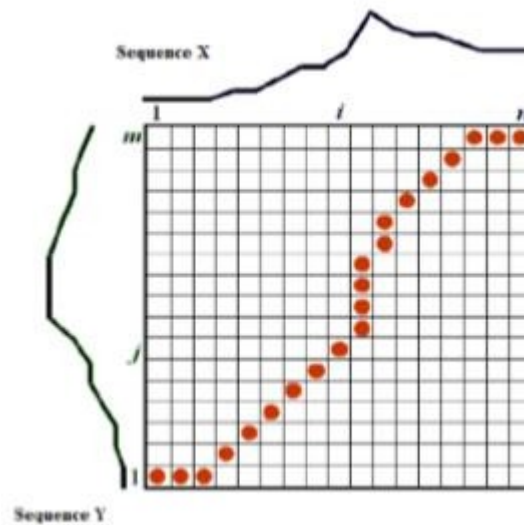
L : *length* atau panjang dari jalur *warping*

l : elemen dari L yang ada pada matriks E

Tujuan dari pendekatan DTW adalah untuk menyelaraskan dua urutan vektor dengan memutar sumbu waktu berulang kali sampai ditemukan kecocokan yang sempurna. Metode ini menyelaraskan dua sinyal dengan melakukan pemetaan sumbu linier parsial. Asumsikan ada dua urutan vektor dalam ruang n-dimensi:

$$x = [x_1, x_2, \dots, x_n] \text{ dan } y = [y_1, y_2, \dots, y_n]$$

Kedua urutan disejajarkan di sisi kotak, dengan satu di kanan dan satu lagi di kiri. Kedua urutan dimulai di sudut kiri bawah kisi.



Gambar 2.2 Contoh Jarak *Grid* Global

Menurut dua urutan tersebut, pengukuran jarak diposisikan dan dibandingkan di dalam setiap sel. Jarak antara dua lokasi ditentukan dengan menggunakan rumus jarak *Euclidean*.

$$Dist(x, y) = |x - y| = [(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_n - y_n)^2]^{1/2}$$

Rute melintasi grid yang meminimalkan jarak keseluruhan antara dua urutan ini, sering dikenal sebagai jarak globalnya, adalah kecocokan atau penyelarasan yang optimal. Jarak global ditentukan dengan mengidentifikasi dan melintasi semua jalur layak melintasi grid, masing-masing menghasilkan jarak global. Jarak global minimal sama dengan total jarak *Euclidean* antara setiap elemen pada rute, dibagi dengan jumlah fungsi pembobotan. Untuk setiap urutan dengan panjang yang cukup, jumlah jalur alternatif melintasi grid akan cukup tinggi. Rumus rekursif digunakan untuk menghitung jarak global.

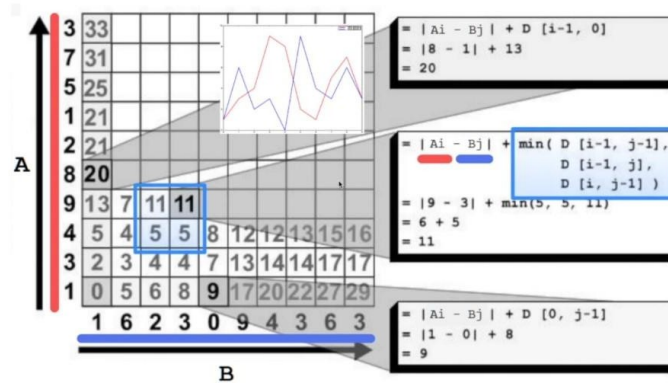
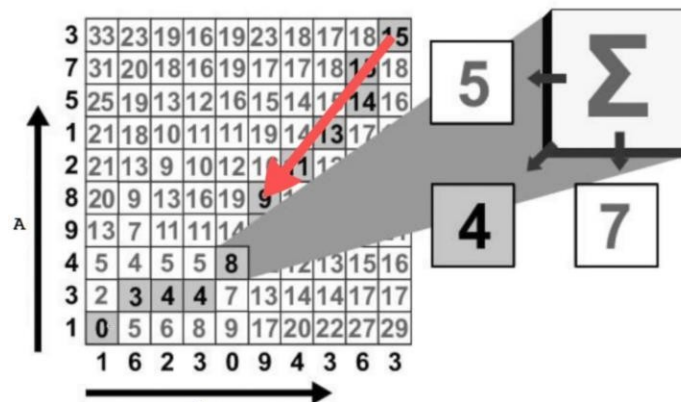
$$GD_{xy} = LD_{xy} + \min(GD_{x-1 y-1}, GD_{x-1 y}, GD_{x y-1})$$

Dimana,

GD = *Global Distance (overall distance)*

LD = *Local Distance (Euclidean distance)*

Gambar 2.3 dan 2.4 adalah ilustrasi dari algoritma DTW untuk menentukan jarak lokal dan optimal path :

Gambar 2.3 Ilustrasi Perhitungan *Local Distance*Gambar 2.4 Ilustrasi *Backtracking* Untuk *Optimal Path*

2.2.7 Analisis *Multidimensional Scalling* (MDS)

Analisis *Multidimensional Scalling* (MDS) adalah pendekatan multivariabel yang dapat digunakan untuk memperkirakan lokasi item lain berdasarkan evaluasi kesamaannya, serta ketergantungan atau saling ketergantungan antara variabel atau data (Johnson, 1992). Keterkaitan ini tidak dapat ditentukan dengan pengurangan atau pengelompokan variabel; melainkan ditentukan dengan membandingkan variabel setiap item menggunakan peta perseptual. MDS mengacu pada proses pembuatan peta yang menjelaskan lokasi satu item relatif terhadap objek lain berdasarkan kesamaannya. MDS juga merupakan metode yang dapat membantu peneliti dalam mengidentifikasi aspek-aspek utama yang mendasari penilaian responden terhadap sesuatu (pelanggan).

Menurut Hair, dkk (1998), penskalaan dua dimensi adalah teknik untuk menentukan parameter fundamental yang menjadi dasar evaluasi responden terhadap suatu item. Sebagai ilustrasi bagaimana penskalaan multidimensi diterapkan dalam dunia bisnis, pertimbangkan bagaimana ini digunakan untuk menguraikan metrik di balik peringkat bintang produk dan layanan oleh pelanggan. Penskalaan multidimensi, juga dikenal sebagai pemetaan perseptual, adalah teknik untuk menyampaikan atau memetakan evaluasi subjektif seseorang terhadap sejumlah entitas yang berbeda (perusahaan, produk, atau lainnya yang terkait dengan persepsi).

Menurut A. Walundungo (2014), *Multidimensional Scaling* (MDS) menggunakan nilai yang disebut kedekatan, yang selanjutnya dipisahkan menjadi persamaan dan perbedaan, untuk melakukan analisis data. MDS metrik dan MDS nonmetrik adalah dua kategori besar MDS yang dapat dipecah lebih lanjut bergantung pada sifat data.

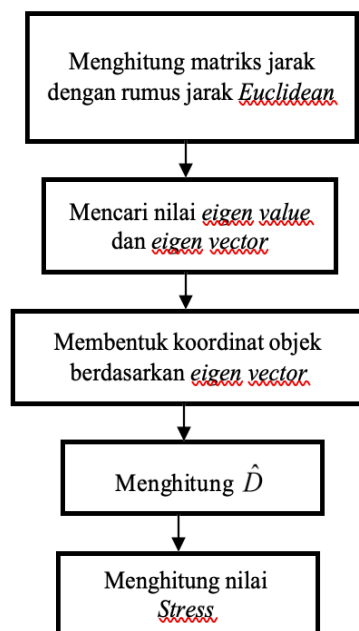
1. Pengukuran Skala dalam Beberapa Dimensi

Informasi jarak untuk penskalaan seringkali berbentuk data interval atau rasio. Metrik Penskalaan Multidimensi tidak peduli apakah data yang diberikan mewakili jarak secara akurat. Untuk mendapatkan kecocokan terbaik antara jarak keluaran dan jarak masukan, metode ini hanya memodifikasi geometri objek titik.

2. Skala Multidimensi Non-metrik

Informasi jarak digunakan untuk menjalankan operasi aritmatika pada nilai selisih, mengubah jarak menjadi nilai selisih urutan dalam transformasi monoton (sama) ke data dunia nyata yang mendasarinya. Dengan menyesuaikan jarak antar objek yang tidak sesuai dengan urutan nilai perbedaan sehingga terus memenuhi urutan nilai ketidaksamaan dan mendekati jarak aslinya, transformasi monoton mempertahankan urutan nilai ketidaksamaan.

Diagram alir berikut menggambarkan langkah-langkah yang terlibat dalam analisis penskalaan multidimensi, dapat dilihat pada Gambar 2.5



Gambar 2.5 Tahapan Analisis *Multidimensional Scaling* (MDS)

2.2.8 Prosedur Analisis *Multidimensional Scaling*

Pemetaan persepsi multidimensi (peta perseptual) adalah tujuan dari analisis skala multidimensi. Menurut Ginanjar (2008), analisis Penskalaan Multidimensi memerlukan beberapa langkah analisis yaitu,

1. Gunakan jarak *Euclidean* untuk menghitung matriks jarak. Hitung kedekatan antara item pada peta perseptual menggunakan jarak *Euclidean* antara objek pertama dan ke-j.
2. Cari nilai eigen dan vektor eigen.
3. Membentuk koordinat objek berdasarkan vektor eigen, dan kemudian menghitung \hat{D} , jarak *Euclidean* dari koordinat yang dibuat
4. Menghitung nilai stress.

2.2.9 Matlab

MATLAB ditemukan oleh matematikawan dan pemrogram komputer Cleve Moler. Ide untuk MATLAB didasarkan pada tesis PhD tahun 1960-an. Moler menjadi profesor matematika di University of New Mexico dan mulai mengembangkan MATLAB untuk murid-muridnya sebagai hobi. Dia

mengembangkan pemrograman aljabar linier awal MATLAB pada tahun 1967 dengan penasihat tesis satu kali, George Forsythe. Ini diikuti oleh kode Fortran untuk persamaan linier pada tahun 1971.

MathWorks menciptakan bahasa pemrograman multi-paradigma dan lingkungan komputasi numerik MATLAB (kependekan dari "*Matrix Laboratory*"). MATLAB mengizinkan manipulasi matriks, tampilan fungsi dan data, penerapan algoritme, pembuatan antarmuka pengguna, dan interaksi dengan program yang ditulis dalam bahasa lain.

Meskipun MATLAB terutama dirancang untuk perhitungan numerik, mesin simbolik MuPAD *toolbox* opsional memungkinkan kemampuan komputasi simbolik. Simulink adalah perangkat lunak tambahan yang menggabungkan desain berbasis model dan simulasi multi-domain grafis untuk sistem dinamis dan tertanam.