

Reduksi Dimensi untuk Meningkatkan Performa Metode Fuzzy Klustering pada Big Data

Joko Eliyanto¹, Suparman²

^{1,2}Magister Pendidikan Matematika, Universitas Ahmad Dahlan, Indonesia

E-mail: joko.eliyanto@gmail.com

Abstract. Big data adalah data yang memiliki volume yang besar, jenis data yang beragam serta produksi data yang sangat cepat. Berbagai metode matematika dan statistika dikembangkan untuk menganalisis big data ini. Jumlah data yang begitu besar dan kompleks menjadi salah satu tantangan dalam analisis big data ini. Reduksi dimensi adalah pengurangan dimensi suatu dataset dengan pertimbangan bahwa informasi-informasi penting tetap dipertahankan. Reduksi dimensi dapat digunakan untuk penyederhanaan komputasi dan visualisasi big data. Dengan dimensi yang lebih rendah hasil analisis pada data hasil reduksi masih menghasilkan kesimpulan yang relevan. PCA adalah metode klasterisasi didasarkan pada komponen utama yang diperoleh dari kombinasi linear data dengan vector eigen terbaiknya. Sedangkan *Core and reduct* adalah metode reduksi dimensi dengan memilih atribut atau variable yang merupakan inti dari suatu dataset. Fuzzy klustering merupakan metode klustering yang baik untuk digunakan pada data-data yang belum memiliki kelas yang jelas. Fuzzy klustering telah diterapkan pada berbagai bidang penting seperti ekonomi, kesehatan dan teknologi. Performa metode fuzzy klustering dapat ditingkatkan menggunakan dimensi reduksi, yaitu menghasilkan kluster dengan akurasi yang sama namun dengan kecepatan yang lebih tinggi dan biaya komputasi yang rendah.

Kata kunci: Clustering, Fuzzy Clustering, Fuzzy C-Means. Reduksi Dimensi,

1. Pendahuluan

Segala sesuatu yang kita lakukan saat ini, selalu meninggalkan jejak digital. Mulai dari foto-foto yang diunggah di sosial media, teks yang kita kirim melalui pesan singkat, halaman-halaman web yang pernah kita kunjungi, ataupun produk-produk yang pernah kita temukan di toko online atau suatu *market place*. Bahkan ketika kita menggunakan aplikasi peta pada gawai kita masing-masing, kita meninggalkan jejak digital berupa koordinat posisi dan waktu ketika kita berada pada posisi tersebut. Tentu data-data di atas merupakan suatu data yang sangat besar. Kemampuan untuk menggali informasi dari data yang diperoleh tersebut adalah hal yang sangat berguna di segala bidang. Big data adalah data yang memiliki volume yang besar, jenis data yang beragam serta produksi data yang sangat cepat [1]. Berbagai metode matematika dan statistika dikembangkan untuk menganalisis big data ini. Jumlah data yang begitu besar dan kompleks menjadi salah satu tantangan dalam analisis big data ini.

Reduksi dimensi adalah pengurangan dimensi suatu dataset dengan pertimbangan tertentu [2]. Pertimbangan yang dimaksud adalah bahwa informasi-informasi penting pada dataset tetap dipertahankan meskipun melalui proses reduksi dimensi. Semakin kecil jumlah dimensi dataset akan mempermudah dan mempercepat proses-proses komputasi pada dataset tersebut. Dataset dimensi dua atau tiga juga memungkinkan untuk dilakukan visualisasi.

Klustering adalah pengelompokan data didasarkan pada kemiripan antar data di dalam suatu dataset. Setelah terkluster, data-data pada kluster yang sama akan memiliki tingkat kemiripan yang

tinggi, sedangkan data-data pada kluster yang berbeda akan memiliki tingkat kemiripan yang rendah [3]. Pengelompokan pada sebuah himpunan menjadi sangat penting untuk dilakukan guna mengambil informasi pada dataset tersebut atau memahami karakteristiknya. Klastering adalah proses pengelompokan dengan pembelajaran tidak tertuntun (*unsupervised learning*) karena data-data dikelompokkan berdasar kemiripan yang telah terkandung pada data tersebut. Metode yang digunakan tidak memerlukan uji terlebih dahulu pada himpunan dataset latihan (*training data set*) untuk kemudian diimplementasikan pada dataset lain yang serupa.

Proses klastering telah diterapkan di berbagai bidang. Pada bidang kesehatan misalnya, klastering digunakan untuk mengelompokkan orang-orang yang memiliki resiko terpapar suatu wabah penyakit tertentu, sehingga kelompok tersebut memerlukan perlakuan khusus untuk menanggulangi resiko penyakit tersebut. Pada bidang bisnis, terlebih pada era saat ini, klastering digunakan untuk segmentasi pasar. Perusahaan sangat memerlukan informasi tentang calon konsumennya. Kemampuan untuk mensegmentasi konsumen yang baik akan sangat membantu perusahaan untuk menghasilkan produk yang tepat untuk konsumen yang tepat bahkan pada saat yang tepat sehingga perusahaan mampu memaksimalkan keuntungan. Pada bidang *image processing*, sebuah citra akan dikonversikan menjadi gradasi, baik warna ataupun gradasi hitam putih. Pengelompokan warna-warna yang mirip digunakan untuk menentukan batas-batas suatu objek pada gambar tersebut [3].

Klastering dibagi menjadi dua yaitu *hard clustering* dan *fuzzy clustering*. Pada *hard clustering* algoritma klastering ini akan membentuk sejumlah k-kelompok, sehingga setiap objek hanya akan menjadi anggota sebuah kelompok. Sedangkan pada *fuzzy clustering* setiap objek diasumsikan memiliki kemungkinan untuk masuk ke dalam beberapa kelompok dengan derajat kemungkinan yang berbeda-beda untuk tiap kelompoknya [3]. Untuk data yang belum memiliki pengelompokan yang jelas, *fuzzy clustering* lebih tepat untuk diaplikasikan [4].

Klastering *fuzzy c means* (FCM) adalah metode dari *fuzzy clustering* yang populer untuk dipakai. Pada metode FCM data yang diubah ke dalam bentuk fuzzy adalah jarak antara objek dengan pusat kluster yang diberikan, Fungsi objektif dari FCM adalah hasil kali dari derajat keanggotaan data pada suatu kluster dengan kuadrat dari jarak titik data ke pusat kluster. Dengan meminimumkan fungsi ini maka akan diperoleh kluster yang anggota-anggota di dalamnya sangat mirip dan perbedaan antar kluster tinggi.

Beberapa penelitian terkait *fuzzy clustering* terkhusus FCM telah dilakukan untuk meningkatkan kualitas hasil klastering. Abas Majdi dan Morteza Beiki menerapkan algoritma optimisasi evolusioner yaitu *Genetic Algorithm* (GA) dan *particle swarm optimization* (PSO) untuk merancang dan mengoptimasikan kalsterisasi *fuzzy-c-means* (FCM) untuk selanjutnya digunakan untuk memprediksi modulus dari deformasi massa batuan. Metode baru ini menghasilkan hasil klastering yang lebih akurat jika dibandingkan dengan FCM yang sudah ada [5]. Klastering FCM memiliki kelemahan utama yang dapat terjebak pada beberapa optimum lokal. Untuk mengatasi kekurangan ini, Adil Baykasoglu dkk menggunakan algoritma metaheuristik generasi baru. *Weighted Superposition Attraction Algorithm* (WSA) adalah metode baru berbasis kecerdasan segerombolan yang menarik inspirasi dari prinsip superposisi fisika dalam kombinasi dengan gerakan agen yang tertarik. Karena kemampuan konvergensi dan kepraktisannya yang tinggi, algoritma WSA digunakan untuk meningkatkan kinerja klastering FCM. Hasilnya menunjukkan peningkatan signifikan atas algoritma FCM yang biasa [6]. Dengan memodifikasi fungsi objektif pada klastering FCM, Chundru Raja Ramesh dkk berhasil meningkatkan akurasi FCM hingga 35% [7].

Untuk bekerja pada data yang sangat besar, algoritma untuk metode klastering biasanya akan bekerja dengan waktu yang cukup lama. Oleh karena itu reduksi dimensi dapat digunakan untuk mengatasi hal tersebut. Metode reduksi dimensi yang biasa digunakan adalah PCA. Misalkan data yang akan direduksi terdiri dari tupel atau vektor data yang dijelaskan oleh n variabel atau dimensi. *Principal component analysis* (PCA) mencari vektor ortogonal berdimensi k yang paling baik digunakan untuk mewakili data, di mana $k \leq n$. Data asli tersebut dengan demikian diproyeksikan ke ruang yang jauh lebih kecil, menghasilkan reduksi dimensi. PCA yang diterapkan bersama dengan *rapid centroid estimation* terbukti dapat meningkatkan hasil klastering k means [8]. PCA dan analisis kluster diterapkan secara bersamaan telah diterapkan dalam beberapa bidang [9], [10], [11], [12], [13]. Berdasar hasil tersebut, PCA dapat diuji untuk meningkatkan klastering fuzzy pada big data. Namun

demikian, pencarian nilai eigen dan eigen vektor pada PCA menghasilkan kesulitan tersendiri pada dimensi yang sangat besar, sehingga pada n yang sangat besar PCA mungkin akan kurang berjalan dengan baik. Metode *Core and Reduct* adalah metode reduksi dimensi yang efektif termasuk juga merumuskan inti dari variabel-variabel pada dataset [14]. *Core and Reduct* adalah salah satu manfaat dari teori rough set [15]. Penelitian tentang aplikasi metode ini juga masih sangat sedikit. Sehingga reduksi dimensi menggunakan PCA dan *Core and Reduct* dapat digunakan untuk meningkatkan performa FCM.

2. Metode

2.1. Pengumpulan literatur

Studi pustaka dilakukan dengan diawali mengumpulkan literatur-literatur akademik baik buku teks maupun jurnal-jurnal internasional yang terbaru. Penelitian ini membahas pokok-pokok teori tentang big data, klasterisasi fuzzy, reduksi dimensi, PCA, *core and reduct*. Selain itu, perkembangan terbaru tentang beberapa metode di atas baik tentang kualitas metode atau implementasinya dalam berbagai bidang, di bahas berdasarkan jurnal-jurnal yang terkait.

2.2. Pengkajian literature

Penulis membaca dan memahami teori-teori tentang big data, klasterisasi fuzzy, reduksi dimensi, PCA dan *core and reduct* dan materi lain berdasar literatur – literatur yang telah dikumpulkan. Penulis mencatat hal-hal pokok seperti gagasan utama teori, bagaimana prosedur penerapannya, kelebihan dan kekurangannya serta perkembangan teori tersebut hingga masa sekarang tentang klastering *fuzzy*. Kekurangan dan kelebihan metode ini menjadi pusat perhatian pada penelitian ini. Hasil – hasil penelitian dari jurnal-jurnal yang dikumpulkan digunakan sebagai pendukung kelebihan atau solusi untuk menutupi kekurangan yang ada. Metode baru yang pernah berhasil di bahasan lain namun memiliki esensi yang sama ditawarkan untuk menutupi kekurangan metode klastering *fuzzy*.

2.3. Penyusunan hasil kajian

Hasil kajian disusun pada Tabel 1 sebagai berikut:

Tabel 1. Penyusunan Hasil Kajian

Pendahuluan	Meliputi latar belakang masalah pada big data, klastering <i>fuzzy</i> dan pengembangan metode FCM menggunakan dimensi reduksi. Hasil-hasil penelitian terbaru diungkapkan sekilas guna mendukung bahasan pokok penelitian. Ide dasar solusi dari masalah diungkapkan pada akhir bagian pendahuluan.
Metode	Meliputi langkah-langkah kajian pustaka.
Hasil dan Pembahasan	Menguraikan gagasan utama teori big data, FCM, reduksi dimensi dengan PCA dan <i>Core and Reduct</i> . Kelebihan dan kekurangan metode di bahas dan kemudian solusi yang ditawarkan berdasarkan hasil-hasil penelitian yang sudah ada sebelumnya.
Kesimpulan	Berisi kesimpulan dari hasil kajian

3. Hasil dan Pembahasan

3.1. Big Data

Definisi khusus terkait big data tidak ditentukan sebagaimana teori-teori lain. Hal ini disebabkan karena teori ini masih terus menerus dikembangkan. Bahkan di saat yang sama, big data telah, sedang dan akan terus digunakan pada berbagai bidang, terutama bisnis dan teknologi. Pengkajian mendalam tentang definisi big data telah dilakukan oleh Ylijoki dan Poran pada tahun 2016 [16]. Mereka mengkaji 17 definisi big data pada 62 makalah yang relevan. Hasilnya adalah sebagai berikut: Big data adalah data yang memiliki tiga karakteristik utama yaitu volume yang besar, variasi data yang beragam dan kecepatan data itu diproduksi. Definisi ini sering disebut dengan definisi 3V. Big data dihasilkan oleh kita setiap saat. Big data muncul setelah adanya perkembangan teknologi yang pesat, membuat data tersedia begitu banyak dan cepat. Bahkan, di tahun 2020 diprediksi 1,7 mb data akan diproduksi tiap detik oleh setiap manusia di dunia [1]. Volume yang sangat besar, data yang beragam dan kecepatan data yang tinggi merupakan tantangan utama untuk mengambil informasi penting dari data-data yang ada untuk mendukung atau bahkan sebagai dasar pengambilan keputusan.

3.2. Klustering

Klustering data atau klustering, juga disebut kluster analisis, segmentasi analisis, taksonomi analisis atau klasifikasi tidak tertuntun adalah metode pengelompokan objek dalam suatu kelompok atau kluster sedemikian sehingga objek-objek di dalam kluster tersebut sangat mirip dan objek-objek pada kluster yang berbeda memiliki perbedaan yang signifikan. Klustering diterapkan di berbagai bidang, pada bidang bisnis klustering digunakan untuk segmentasi pasar, di bidang kesehatan psikologi klustering digunakan untuk mengelompokkan orang-orang agar menerima pelayanan kesehatan yang tepat. Pada bidang pengolahan citra, klustering digunakan untuk menentukan batas-batas objek pada suatu gambar/citra [2].

3.3. Fuzzy Klustering

Diberikan himpunan $X = \{x_1, x_2, \dots, x_n\}$, himpunan fuzzy dari himpunan X adalah himpunan dari tingkat keanggotaan setiap objek antara 0 dan 1. Secara formal, himpunan fuzzy, S , dapat dimodelkan sebagai fungsi berikut: $F_S : X \rightarrow [0,1]$.

Secara formal, diberikan objek o_1, o_2, \dots, o_n fuzzy klustering untuk sejumlah k kluster fuzzy C_1, \dots, C_k dapat direpresentasikan dengan matriks partisi $M = [w_{ij}] (1 \leq i \leq n, 1 \leq j \leq k)$, di mana w_{ij} adalah derajat keanggotaan o_i pada kluster fuzzy C_j . Matriks partisi tersebut harus memenuhi kriteria berikut:

- Untuk setiap objek o_i dan kluster C_j $0 \leq w_{ij} \leq 1$. (Syarat kluster fuzzy)
- Untuk setiap objek o_i , $\sum_{j=1}^k w_{ij} = 1$

Kelebihan menggunakan fuzzy klustering [4]:

- Komputasi yang mudah dan sederhana
- Dapat menangani ketidakpastian di berbagai tahap.

- Untuk pengambilan keputusan tingkat yang lebih tinggi dalam klasifikasi pola, pendekatan fuzzy sangat berguna karena memberikan tingkat informasi keanggotaan.
- Pendekatan ini mendukung penalaran manusia dalam bentuk alami sampai batas tertentu, dengan menetapkan nilai keanggotaan parsial pada data.
- Ketika mengelompokkan objek berdimensi tinggi, akurasi dan efisiensi algoritma pengelompokan fuzzy telah sangat baik dibandingkan dengan pengelompokan tradisional karena konsep keanggotaan fuzzy tersebut.

Kerugian menggunakan fuzzy klustering:

Model dari sistem fuzzy cukup terbatas untuk dikembangkan secara luas. Butuh penelitian lanjutan untuk pengembangan sistem fuzzy.

3.4. Algoritma Fuzzy C Means

Algoritma c-means adalah metode klustering yang membolehkan sebuah data menjadi anggota dua kluster atau lebih. Misalkan $X = \{x_1, x_2, \dots, x_M\}$ adalah himpunan data numeris pada semesta \square^n . Misalkan c adalah bilangan bulat $1 < c < M$. Sejumlah c subhimpunan fuzzy $\{w_i : X \rightarrow [0,1]\}$ adalah c -partisi dari X dengan ketentuan sebagai berikut:

$$0 \leq u_{ij} \leq 1 \quad \forall i, j \quad (1)$$

$$\sum_{i=1}^c w_{ij} = 1 \quad (2)$$

$$0 < \sum_{j=1}^M w_{ij} < n, \quad \forall i \quad (3)$$

di mana $w_{ij} = w_i(x_j)$, $1 \leq i \leq c$ dan $1 \leq j \leq M$. Misalkan cM adalah nilai w_{ij} memenuhi kondisi di atas disusun sebagai matriks ukuran $c \times M$ yaitu $W = [w_{ij}]$. Kemudian himpunan dari semua matriks tersebut adalah *nondegenerate fuzzy c-partitions* dari X :

$$M_{f,c,M} = \{W \in \square^{cM} : w_{ij} \text{ memenuhi kondisi (1),(2) dan (3) } \forall i \text{ dan } j\}.$$

Jika semua w_{ij} bernilai 0 atau 1, maka subhimpunan *hard c-partitions* dari X :

$$M_{cM} = \{W \in M_{f,c,M} : w_{ij} = 0 \text{ atau } 1, \forall i \text{ dan } j\}.$$

Setiap elemen w_{ij} dapat diperlakukan sebagai tingkat keanggotaan x_j pada kluster ke- i dari X . Diasumsikan c diketahui. Fungsi objektif untuk klustering di dalam himpunan X adalah fungsi *within-group sum squared errors* (WGSS), yang disebut sebagai J_1 :

$$J_1(W, \bar{v}; X) = \sum_{i=1}^c \sum_{j=1}^M w_{ij} (\|x_j - v_i\|_1)^2,$$

dengan $\bar{v} = (v_1, v_2, \dots, v_c)$ adalah vektor pusat kluster, $v_i \in \mathbb{R}^N$ untuk $1 \leq i \leq c$ dan $W \in M_{cM}$.

Tujuannya adalah untuk mencari partisi optimal W^* dari X sedemikian sehingga pasangan (W^*, \bar{v}^*) adalah *local minimize* pada J_1 .

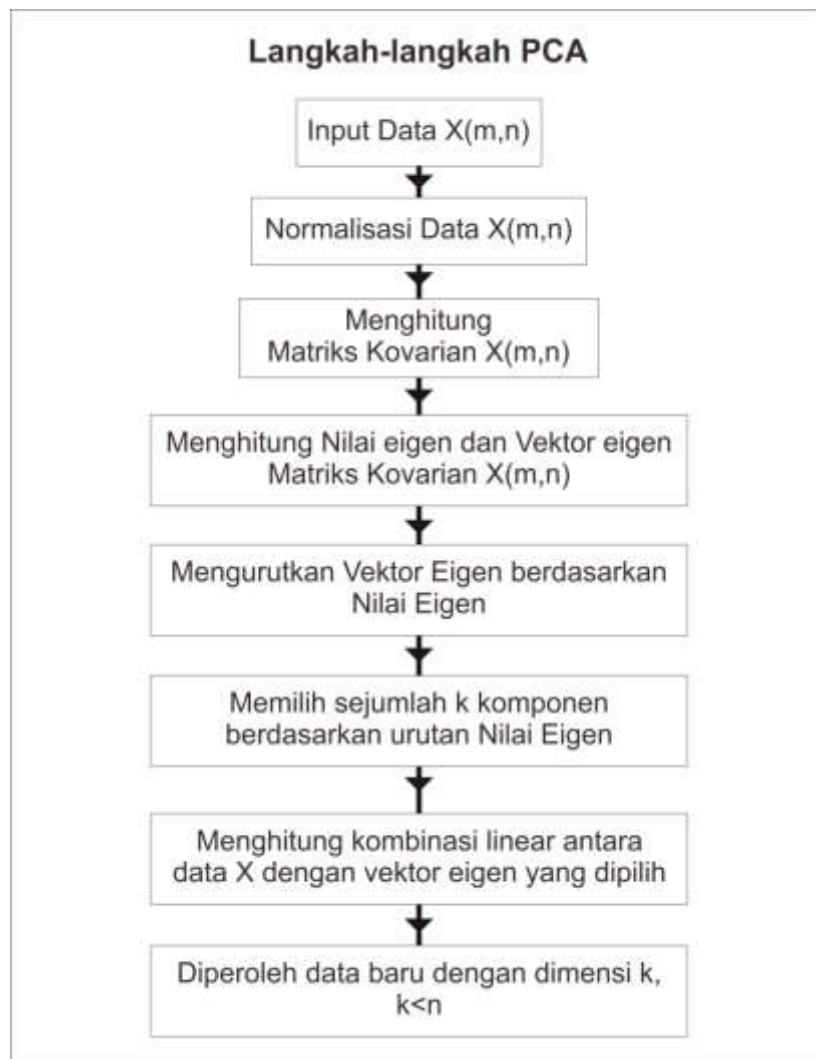
Penerapan algoritma FCM sudah diteliti dan diaplikasikan pada banyak bidang. Salah satunya adalah analisis tindak kekerasan menggunakan pendekatan FCM. Hasil penelitian ini dapat digunakan untuk mengevaluasi sistem keamanan di negara-negara bagian Amerika Serikat [11]. Pada paper [12], konsep fuzzy klustering juga membantu untuk mendeteksi penyakit kanker payudara. Pada bidang teknik, prediksi secara fuzzy dari status kekuatan lithium baterai ion dapat dilakukan menggunakan FCM [13].

3.5. Reduksi dimensi

Reduksi dimensi adalah pengurangan dimensi suatu dataset pada dataset dengan pertimbangan tertentu [2]. Pertimbangan yang dimaksud adalah bahwa informasi-informasi penting pada dataset tetap dipertahankan meskipun melalui proses reduksi dimensi. Semakin kecil jumlah dimensi pada dataset akan mempermudah dan mempercepat proses-proses komputasi pada dataset serta memungkinkan untuk melakukan visualisasi, misal pada data dua dimensi dan tiga dimensi. Untuk mereduksi data dengan jumlah besar akan dikaji dua metode. Metode pertama adalah PCA dan kedua adalah *Core and Reduct*. Metode PCA dipilih karena terbukti dapat meningkatkan hasil dari klustering data [8] sedangkan metode *Core and Reduct* dipilih karena metode ini memiliki efektifitas yang baik [14].

3.6. Principal Component Analysis

Algoritma PCA adalah algoritma untuk ekstraksi variabel, artinya variabel-variabel yang 'tidak penting' akan dibuang sehingga dimensi dapat direduksi. Variabel-variabel yang dibuang adalah variabel-variabel yang berkorelasi sehingga hasil dari PCA adalah variabel-variabel yang independen. Ini mengubah satu dataset variabel yang saling terkait menjadi yang tidak berkorelasi yang disebut komponen utama. Jumlah komponen utama lebih kecil dari jumlah variabel dataset awal. Komponen utama ini sebenarnya adalah vektor eigen yang diperoleh dengan mendekomposisi matriks kovarian data. Sebelum mendekomposisi nilai eigen / vektor eigen dari matriks kovarians, perlu untuk menormalkan variabel-variabel dengan mengurangi rata-rata dari masing-masing dimensi data. Setelah itu, matriks kovarians dari titik data akan dihitung dan kemudian vektor eigen dan nilai eigen yang sesuai akan dihitung, terpecahkan. Selanjutnya, vektor eigen menurut nilai eigennya diurutkan dalam urutan menurun. Kemudian dipilih sejumlah k komponen didasarkan pada urutan nilai eigen sehingga diperoleh dimensi baru sejumlah k yang lebih kecil dari dimensi awal. Akhirnya, PCA akan mengubah titik data dimensi asli dalam dimensi baru yang lebih kecil. Titik kelemahan utama dari PCA adalah pada penghitungan vektor eigen dan nilai eigen. Langkah-langkah pada PCA dapat dilihat pada gambar 1. Pada data yang berdimensi sangat besar, penghitungan nilai eigen rentan terhadap kesalahan hitung computer atau kesalahan yang dihasilkan algoritma untuk menghitungnya.



Gambar 1. Langkah-langkah Principal Component Analysis.

Core and Reduct

Selain PCA, metode reduksi dimensi yang lain adalah *Core dan Reduct*. *Core dan Reduct* adalah dua konsep terpenting dari teori *rough set*. *Reduct* adalah pengurangan subhimpunan dari himpunan asli yang mempertahankan akurasi dari himpunan set asli. *Reduct* sering digunakan dalam proses pemilihan variabel untuk mengurangi variabel yang tidak perlu terhadap aplikasi pengambilan keputusan. Beberapa hal yang penting dari *Reduct* adalah

- Reduksi dari tabel keputusan adalah seperangkat atribut kondisi yang cukup untuk mendefinisikan atribut keputusan.
- Setiap pengurangan memungkinkan kita untuk mengurangi variabel kondisi.
- *Reduct* tidak mengandung atribut redundan terhadap hasil klasifikasi.
- Mengurangi biaya perhitungan untuk *rule generation*.

Reduksi dapat diperoleh dengan menggunakan algoritma generasi reduksi. Menggunakan matriks *discernibility*, reduct dari tabel keputusan dapat ditentukan. Inti dapat ditemukan sebagai himpunan semua entri tunggal dalam bentuk matriks *discernibility*. Reduksi adalah elemen minimal dalam matriks *discernibility*, yang memotong semua elemen dari matriks *discernibility* [14]. Menggunakan *Core and Reduct* sangat mungkin untuk melakukan reduksi dimensi pada dimensi data yang besar atau variabel yang sangat besar.

3.7. Reduksi dimensi untuk fuzzy klustering.

Data yang akan diklusterisasi, sebelumnya melalui proses reduksi dimensi. Dengan demikian proses komputasi pada proses klusterisasi dilakukan pada dimensi yang rendah. Hal ini juga sangat membantu untuk melakukan visualisasi data. Untuk mengetahui kualitas dari hasil klusterisasi maka akan dilakukan beberapa uji kualitas hasil klusterisasi. Gambar 2 menunjukkan skema reduksi dimensi pada fuzzy klustering



Gambar 2. Skema klasterisasi FCM dengan reduksi dimensi.

Perbandingan antara reduksi dimensi dengan PCA dan *Core and Reduct* disajikan dalam Tabel 2 sebagai berikut:

Table 2. Perbandingan antara PCA dan *Core and Reduct*

	PCA	<i>Core and Reduct</i>
Konsep Dasar	Memilih komponen dengan nilai eigen dan vektor eigen terbaik dari matriks kovarian	Mencari Indispensible Atribut pada Rough Set
Tipe Data	Bergantung pada tipe data, karena menghitung kovarian data	Tidak tergantung pada tipe data, karena hanya mempertimbangkan kesamaan anggota pada atribut atau record data
Jumlah penelitian	PCA telah banyak dikaji dan diterapkan pada proses klustering	Masih sedikit jumlah penelitian terutama pada bahasan klustering
Dimensi Besar	Tidak disarankan, karena kerawanan pada penghitungan nilai eigen	Memungkinkan untuk data berdimensi besar

4. Kesimpulan

Berdasarkan uraian di atas dapat disimpulkan bahwa dimensi reduksi dapat digunakan untuk meningkatkan hasil klustering dari fuzzy klustering. PCA dan *Core and Reduct* adalah metode dimensi reduksi yang cocok untuk data berdimensi cukup besar dan sangat besar. Untuk mengatasi

penghitungan nilai eigen dan vector eigen yang sangat rawan pada data berdimensi tinggi, maka *Core and Reduct* dapat diaplikasikan. Penggunaan metode reduksi dimensi pada fuzzy klustering diharapkan mampu menghasilkan klasterisasi dengan akurasi yang sama namun dengan kecepatan komputasi yang lebih tinggi serta biaya komputasi yang rendah.

Acknowledgment

Terimakasih diucapkan kepada Ibu dan Bapak yang telah, sedang dan selalu mendoakan serta mendukung penulis untuk berprestasi. Kepada bapak Sugiyarto, Ph.D Prodi Matematika UAD yang selalu membimbing penulis dalam belajar data mining dan big data. Kepada Prof. Dr Mustafa bin Mat Deris Universiti Tun Husein On Malaysia yang telah member masukan yang sangat membantu penulis dalam menyelesaikan makalah ini. Penulis mengucapkan terimakasih yang tidak terkira kepada beliau berdua dan semoga Allah senantiasa memberkahi hidup mereka. Terimakasih juga diucapkan kepada pimpinan Universitas Ahmad Dahlan dan Fakultas Pascasarjana UAD yang selalu memberikan dukungan kepada kami untuk selalu mengembangkan diri. Kepada ketua prodi Magister Pendidikan Matematika yang telah membimbing kami untuk menjadi mahasiswa S2 yang berkualitas dan berkaliber internasional dengan menghasilkan paper yang berkualitas. Kepada bapak ibu dosen prodi Magister Pendidikan Matematika yang selalu membagi ilmu kepada mahasiswa dengan tulus ikhlas.

References

- [1] Marr B 2016 *Big data in practice: how 45 successful companies used big data analytics to deliver extraordinary results* (West Sussex: John Wiley & Sons)
- [2] Han J Kambel M Pei J 2006 *Data Mining Concepts and Techniques* (Waltham: Morgan Kaufman)
- [3] Gan G, Ma C, Wu J 2007 *Data clustering: theory, algorithms, and applications* (Philadelphia: SIAM)
- [4] Patel N U 2014 Comparative Study between Fuzzy Clustering and Hard Clustering. *International Journal of Futuristic Trends in Engineering and Technology* **1** 16-19
- [5] Majdi A, Beiki, M 2019 Applying evolutionary optimization algorithms for improving fuzzy C-mean clustering performance to predict the deformation modulus of rock mass. *International Journal of Rock Mechanics and Mining Sciences* **113** 172-182
- [6] Baykasoğlu A, Gölcük İ, Özsoydan F B 2018 Improving fuzzy c-means clustering via quantum-enhanced weighted superposition attraction algorithm. *Hacettepe Journal of Mathematics and Statistics* **48** 859-882
- [7] Hakyemez T C, Bozanta A, Coşkun M 2019 *K-Means vs. Fuzzy C-Means: A Comparative Analysis of Two Popular Clustering Techniques on the Featured Mobile Applications Benchmark* (Ankara: IMISC)
- [8] Nababan E B 2019 Improvement of K-Means Performance Using a Combination of Principal Component Analysis and Rapid Centroid Estimation *In Journal of Physics: Conference Series* **1230** 012003
- [9] Chin C S, Ji X, Woo W L, Kwee T J, Yang W 2019 Modified multiple generalized regression neural network models using fuzzy C-means with principal component analysis for noise prediction of offshore platform. *Neural Computing and Applications* **31(4)** 1127-1142
- [10] Moreira B R A, Viana R. S, Lisboa L A M, Lopes P R M, Figueiredo P A M, Ramos S B, May A 2019 Classifying Hybrids of Energy Cane for Production of Bioethanol and Cogeneration of Biomass-Based Electricity by Principal Component Analysis-Linked Fuzzy C-Means Clustering Algorithm. *Journal of Agricultural Science*, **11**
- [11] Wu X, Zhu J, Wu B, Zhao C, Sun J, Dai C 2019 Discrimination of Chinese Liquors Based on Electronic Nose and Fuzzy Discriminant Principal Component Analysis. *Foods* **8** 38

- [12] Hamed M A R 2019 Application of Surface Water Quality Classification Models Using Principal Components Analysis and Cluster Analysis *Available at SSRN 3364401*
- [13] Vijayarajan R, Muttan S 2014 Fuzzy C-Means Clustering Based Principal Component Averaging Fusion *International Journal of Fuzzy Systems* **16**
- [14] Vashist R, Garg M L 2011 Rule generation based on reduct and core: A rough set approach. *Int. J. Comput. Appl.*, **29** 0975-8887
- [15] Khan J, Li J P, Khan G A, Malik A, Parveen S, Shahid M 2019 A Survey on Rough Set Theory and Their Extension For Data Mining *Asian Journal For Convergence In Technology (AJCT)* **5**
- [16] Ylijoki O, Porras J 2016 Perspectives to definition of big data: a mapping study and discussion. *Journal of Innovation Management* **4** 69-91
- [17] Yamini M P C 2019 A violent crime analysis using fuzzy c-means clustering approach *ICTACT Journal on Soft Computing* **9** 1939-1944
- [18] Parvathavarthini S, Karthikeyani Visalakshi N, Shanthi S 2019 Breast Cancer Detection using Crow Search Optimization based Intuitionistic Fuzzy Clustering with Neighborhood Attraction. *Asian Pacific journal of cancer prevention: APJCP* **20** 157
- [19] Wang D, Yang F, Gan L, Li Y 2019 Fuzzy prediction of power lithium ion battery state of function based on the fuzzy c-means clustering algorithm. *World Electric Vehicle Journal* **10** 1.
- [20] Karami A 2019 Application of fuzzy clustering for text data dimensionality reduction. *arXiv preprint arXiv:1909.10881*.