MODIFIKASI K-MEANS BERBASIS *ORDERED WEIGHTED AVERAGING* (OWA) UNTUK KASUS KLASTERING

Millatul Ulya
Jurusan Teknologi Industri Pertanian, Fakultas Pertanian, Universitas Trunojoyo
Korespondensi : Jl. Raya Telang PO BOX Kamal-Bangkalan, Email: milatululya@trunojoyo.ac.id

ABSTRACT

K-means clustering method based on Ordered Weighted Averaging (OWA) was developed by Cheng et al (2009) to resolve problem in classification using integrating k-means clustering and OWA. K-means clustering is a method of clustering and OWA is an aggregation operator. OWA was able to reduce the complexity of experimental data and help in representing sophisticated relationships between the criteria. Based on the original function of k-means and OWA algorithm used, it is predicted that OWA-based k-means clustering (Cheng et al, 2009) works by modifying some of its stages. In this study, it will be done by modification of OWA-based k-means clustering (Cheng et al, 2009) and validated it in the clustering of iris data set. This research aims to apply OWA-based k-means clustering in clustering iris data sets for validation and measure accuracy rate of OWA-based k-means clustering in the iris data sets. Result showed that accuracy of OWA-based k-means clustering in clustering iris data sets is 96.67%, which was better than k-means clustering method of 89.33%.

Keywords : clustering, *k-means*, OWA.

PENDAHULUAN

Data mining adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar (Santosa, 2007). Salah satu tugas dalam *data mining* adalah klastering. Tujuan utama dari klastering adalah pengelompokan sejumlah data/obyek ke dalam klaster sehingga dalam setiap klaster akan berisi data yang semirip mungkin (Santosa, 2007).

Metode klastering yang umum digunakan adalah *k-means clustering*, yang termasuk metode *partition clustering,* yakni memilah-milah data/obyek ke dalam klaster-klaster yang ada. Menurut Jain (2009), Metode *k-means* telah mengalami banyak pengembangan, antara lain: 1) *Fuzzy-c-Means*, 2) *X-means*, 3) *k-medoid*, 4) *Kernel k-means*. Menurut Agusta (2008), terdapat pula metode *k-harmonic means* dan *k-modes*. Variasi metode k-means tersebut umumnya berhubungan dengan tiga hal yang telah disebutkan oleh Agusta (2007). Namun saat ini telah dihasilkan pengembangan *k-means clustering* berbasis OWA oleh Cheng dkk (2009) yang melakukan klastering nilai agregat, yang merupakan kumpulan dari nilai multi atribut yang ada. Metode ini berbeda dengan variasi metode *k-means* yang telah ada sebelumnya. Cheng dkk (2009) lebih fokus pada cara untuk mengurangi kompleksitas data set eksperimental dan keterkaitan antara berbagai kriteria yang ada, yang dapat diatasi dengan cara menggabungkan *k-means* dengan OWA.

Ditinjau dari sisi perkembangan metode *Ordered Weighted Averaging* (OWA), ada beberapa penelitian terdahulu yang telah menerapkan metode OWA ini pada kasus pengenalan pola (*pattern recognition*). Yager (1988) yang pertama kali memperkenalkan tentang OWA, menyatakan bahwa OWA dapat diaplikasikan untuk menyelesaikan berbagai problem, termasuk problem klasifikasi. Klasifikasi termasuk metode *supervised learning*, metode yang diterapkan menggunakan latihan (ada proses *training*) dan tanpa ada guru (*teacher*). Guru di sini adalah label (output/variabel respon) dari data. Label tersebut yang menandai kemana data akan dikelompokkan (Santosa, 2007). OWA dapat diterapkan untuk memberikan bobot yang berbeda untuk setiap atribut. OWA operator yang tepat dapat digunakan untuk

mewakili hubungan antar kriteria yang diagregasikan (Yager, 1988). Pernyataan ini didukung oleh penelitian Grandhi (2003), yang telah mengaplikasikan OWA untuk klasifikasi, pada kasus pendeteksian ranjau darat (*landmine detection*), dengan mengintegrasikan OWA dan *Feed Forward Neural Networks* (FOWA). Penelitian lainnya oleh Cheng dkk (2009) yang menggabungkan antara OWA dengan metode *k-means clustering* yang digunakan untuk memecahkan masalah klasifikasi untuk mengelompokkan *Key Performance Indicator* (KPI) menjadi dua kelas yaitu bagus dan normal pada perusahaan di Taiwan.

K-means sebenarnya merupakan metode klastering, namun dalam penelitian Cheng dkk (2009), *k-means* digunakan untuk menyelesaikan problem klasifikasi dengan melibatkan proses *training* dan adanya *teacher*. Artinya, Cheng dkk (2009) masih mengaplikasikan *k-means* berbasis OWA tersebut pada data yang memiliki label (output/variabel respon), dan hasil penelitian Cheng dkk (2009) menyatakan bahwa metode ini cukup valid untuk menyelesaikan kasus klasifikasi tersebut. Namun, mengingat bahwa *k-means* itu termasuk metode klastering, maka seharusnya metode *k-means* berbasis OWA oleh Cheng dkk (2009) juga dapat diaplikasikan untuk menyelesaikan kasus klastering.

OWA yang digunakan dalam penelitian Cheng dkk (2009) ini menggunakan persamaan OWA yang dikembangkan oleh Fuller and Majlender (2001), dimana persamaan-persamaan tersebut tidak memperhatikan apakah data yang kita analisis memiliki label (output/variabel respon) atau tidak. karena persamaan ini hanya memerlukan dua input parameter saja yaitu jumlah variabel dari data dan nilai *orness* (α) atau parameter situasi yang digunakan. Oleh karena itu, diduga bahwa *k-means* berbasis OWA ini juga dapat digunakan dalam menyelesaikan kasus klastering data, seperti pada data *iris*.

Problem klastering *iris* (bunga) memiliki data set eksperimental yang kompleks. Yager (1988) menyatakan OWA *operator* dapat mengurangi kompleksitas data dengan memadukan nilai-nilai multi attribut ke nilai-nilai agregat yang berupa nilai tunggal. Masing-masing bunga memiliki sifat atau ciri-ciri yang berbeda, dimana ciri-ciri tersebut yang menentukan pada klaster mana yang sesuai untuk satu sampel bunga. Kasus ini mirip dengan kasus pengambilan keputusan multikriteria (*Multicriteria Decision Making*/MCDM), dimana keputusan mengklasterkan bunga berdasarkan ciri-cirinya didasarkan pada multikriteria. Menurut Yager (2004), *Ordered Weighted Averaging* (OWA) sangat berguna untuk proses MCDM yang sering kali memerlukan keterkaitan antar kriteria yang ada.

Berdasarkan fakta-fakta tersebut, baik dari sisi metode *k-means*, persamaan OWA yang digunakan, keunggulan OWA untuk menyelesaikan pengambilan keputusan multikriteria, karakteristik data set iris, maka metode *k-means* berbasis OWA yang telah dikembangkan oleh Cheng *et al.* (2009) diduga dapat diaplikasikan pada kasus klastering data set *iris*, namun ada beberapa tahapan dalam penelitian Cheng *et al.* (2009) yang harus dimodifikasi, karena tidak adanya *teacher* (label/output/variabel respon) pada data set *iris* dan tidak ada proses *training* dalam proses pembelajaran dari data tersebut. Tujuan dari penelitian ini adalah untuk mengaplikasikan *k-means* berbasis OWA pada klastering data set *iris* (pengelompokan jenis bunga menjadi 3, yaitu *Setosa*, *Virginica* dan *Versicolor*) dan mengukur tingkat akurasinya untuk proses validasi.

METODE

Mendefinisikan Indeks, Parameter dan Variabel dalam Penelitian

Pada bagian berikut ini akan didefinisikan indeks, parameter dan variabel dari metode yang akan digunakan dalam penelitian ini.

Indeks

i	data ke- (i = 1, 2, 3, …, m)
j	variable ke- (j = 1, 2, 3, …, n)
k	jumlah klaster
m	jumlah data
n	jumlah variable
r	klaster ke- (r = 1, 2, …, k)

Parameter

w_j	bobot variable ke- (j = 1, 2, 3, …, n)

α parameter situasi ($0 \leq \alpha \leq 1$)
a_i nilai agregat data ke- (i = 1, 2, 3, …, m)
C_r pusat klaster ke- dengan r = 1, 2,.., k.

Variabel

x_{ij} data ke-i pada variabel ke-j (i = 1, 2, 3, …, m ; dan j = 1, 2, 3, …, n)
μ_j mean dari variabel ke- (j = 1, 2, ..., n)
μ_r mean dari klaster ke- (r = 1, 2, ..., k)

Modifikasi Metode *k-means* Berbasis OWA

Critical Review Metode k-means Berbasis OWA oleh Cheng dkk (2009)

Metode *k-means* berbasis OWA telah dikembangkan sebelumnya oleh Cheng dkk (2009) yang digunakan untuk menyelesaikan kasus klasifikasi. Menurut Yager (1988), OWA memang dapat digunakan dalam pengenalan pola khususnya masalah klasifikasi. Oleh karena itu Cheng dkk (2009) mengembangkan metode baru untuk menyelesaikan kasus klasifikasi dengan cara menggabungkan metode OWA dan *k-means*. Namun sebenarnya *k-means* adalah metode untuk menyelesaikan kasus klastering, bukan metode klasifikasi. Penggunaan *k-means* untuk menyelesaikan kasus klasifikasi oleh Cheng dkk (2009) dimungkinkan karena adanya OWA yang digunakan. Metode OWA menghendaki adanya tahapan *ordering attributes* (pengurutan atribut), dimana atribut pada data yang dianalisis harus diurutkan menurut tingkat kepentingannya. Variabel yang penting adalah variabel yang jika dihilangkan akan mempengaruhi nilai variabel responnya. Sehingga, proses pengurutan atribut ini hanya dapat dilakukan pada data yang memiliki variabel respon, dalam artian bahwa kasus yang cocok adalah klasifikasi. Karena data pada kasus klasifikasi memiliki variabel respon.

Jika dilihat dari sisi fungsi yang sebenarnya dari *k-means*, yaitu sebagai metode klastering, maka seharusnya metode *k-means* berbasis OWA ini dapat juga digunakan untuk menyelesaikan kasus klastering. Namun tahapan pengurutan atribut harus dimodifikasi agar dapat dilakukan meski data yang dianalisis tidak memiliki variabel respon seperti pada kasus klastering.

Di samping itu, algoritma yang digunakan Cheng dkk (2009) untuk memperoleh bobot OWA adalah algoritma OWA oleh Fuller & Majlender (2001), yang tidak memperhatikan apakah data yang dianalisis memiliki variabel respon atau tidak. Karena pada algoritma ini hanya perlu memasukkan *n* (jumlah atribut) dan nilai α (orness) yang digunakan. Dengan kata lain, persamaan 2.8 – 2.10 (Fuller & Majlender, 2001) tersebut dapat digunakan untuk kasus klasifikasi atau pun klastering.

Berdasarkan kondisi tersebut, maka metode *k-means* OWA oleh Cheng dkk (2009) diduga dapat diaplikasikan untuk menyelesaikan kasus klastering. Namun, mengingat ada perbedaan dalam metode pembelajaran dalam klastering (un*supervised learning*) dan klasifikasi (*supervised learning*), maka harus ada modifikasi dari prosedur dalam *k-means* OWA yang telah dikembangkan oleh Cheng dkk (2009) tersebut.

Modifikasi yang Dikembangkan

Pada prosedur *k-means* OWA (Cheng dkk, 2009), ada tahap *feature selection, feature reduction* & *ordering attributes* dengan cara meranking atribut menggunakan *stepwise regression*. Tahap *feature selection* & *feature reduction* adalah tahapan kunci yang membedakan antara metode klasifikasi dan klastering. Poin penting dalam *f*eature *selection* adalah proses pemilihan fitur/variabel data pada *data mining* untuk mendapatkan data yang optimal dalam sistem klasifikasi (Jain & Cadrasekaran, 1982). Proses *feature selection* dengan *stepwise regression* dilakukan dengan cara mencari variabel mana yang korelasinya tinggi terhadap variabel respon. Jika korelasi variabel rendah terhadap variabel respon, maka variabel tersebut tidak digunakan lagi dalam tahapan selanjutnya. Artinya, ada tahap *feature reduction* (pengurangan fitur) juga dalam metode ini.

Kondisi di atas hanya dapat dilakukan jika data yang dianalisis memiliki variabel respon (label kelompok). Sehingga proses tersebut hanya cocok digunakan untuk kasus klasifikasi. Jika kasusnya klastering,

data yang dianalisis tidak memiliki variabel respon, maka proses *feature selection* dan *feature reduction* tidak dapat dilakukan. Namun, adanya OWA yang digunakan dalam metode ini, menuntut bahwa fitur atau variabel dalam data yang dianalisis harus urut sesuai dengan tingkat kepentingan atau bobot dalam pengelompokan data tersebut. OWA (*Ordered Weighted Averaging*) adalah bobot rata-rata ter-urut, yang harus dipasangkan dengan variabel yang urut pula, mulai dari yang penting sampai yang tidak penting. Oleh karena itu, proses *ordering attributes* atau pengurutan atribut/variabel/fitur (Cheng dkk, 2009) tetap perlu dilakukan namun harus dimodifikasi agar dapat digunakan untuk kasus klastering. Tidak menggunakan pendekatan statistik seperti korelasi, namun menggunakan *expert judgment* yang diperoleh dari jurnal-jurnal penelitian sebelumnya yang berkaitan dengan variabel dari data yang kita analisis.

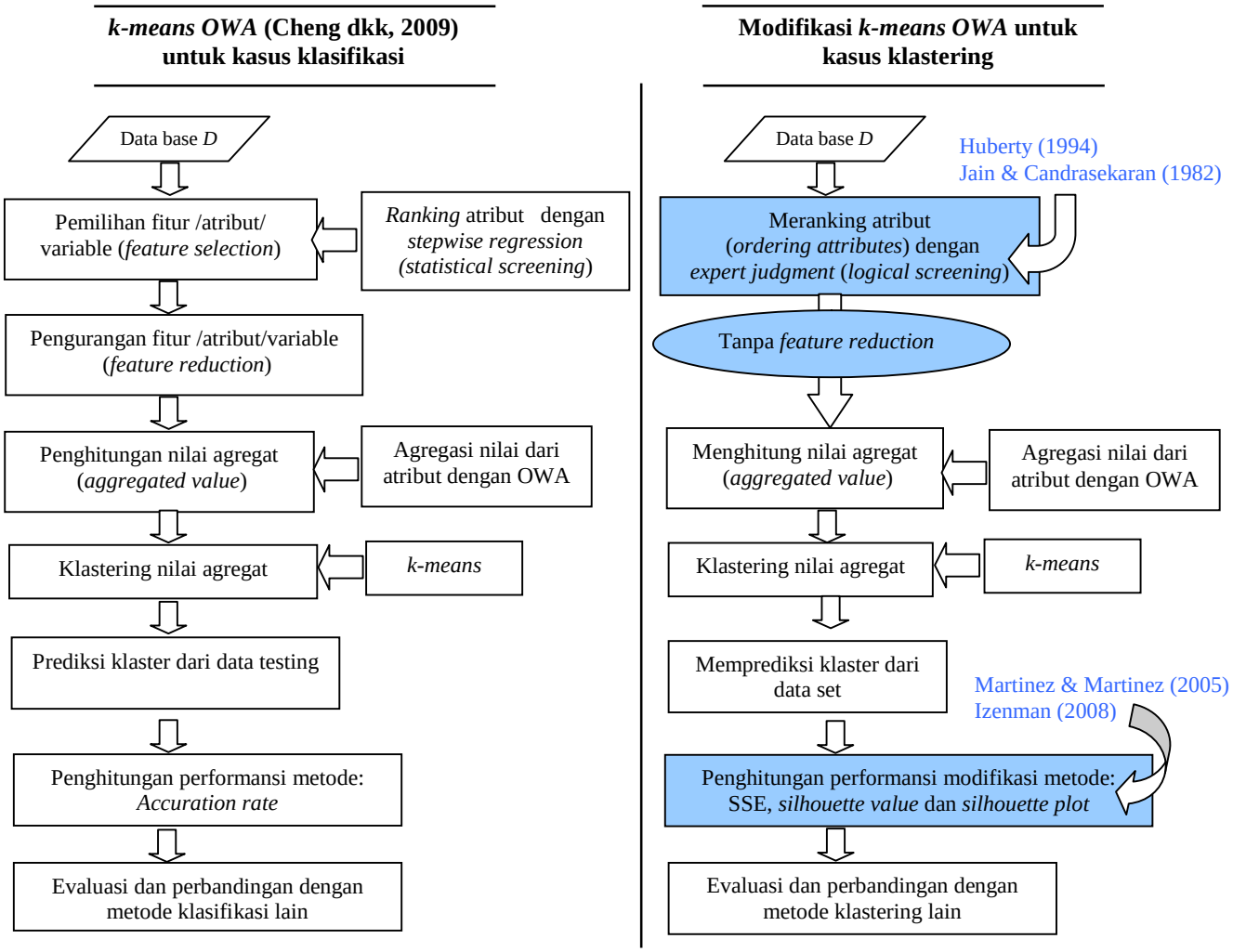
Huberty (1994) menyarankan tiga cara dalam menyaring (*screening*) variabel, termasuk menggunakan *logical screening*. Pendekatannya menggunakan teori-teori, keandalan dan *practical grounds* untuk menyaring variabel. Huberty (1994) juga menyarankan untuk menggunakan pengetahuan spesifik yang subyektif untuk menemukan variabel-variabel yang mungkin secara teoritis berhubungan dengan kelompok/klaster. Cara yang kedua menurut Huberty (1994) adalah *statistical screening*, yang dilakukan dengan cara mencari variabel yang berkorelasi tinggi terhadap variabel respon. Namun cara kedua ini tidak dapat dilakukan, karena data yang dianalisis tidak memiliki variabel respon.

Berdasarkan kondisi di atas, maka *k-means* berbasis OWA (Cheng dkk, 2009) harus dimodifikasi prosedurnya, agar dapat diaplikasikan untuk menyelesaikan kasus klastering. Secara visual, modifikasi metode *k-means* berbasis OWA (Cheng dkk, 2009) dapat digambarkan pada Gambar 1. Modifikasi prosedur *k-means* OWA (Cheng dkk, 2009) tersebut antara lain:

1. Prosedur *feature reduction*/pengurangan variabel (Cheng dkk, 2009) dihilangkan. Menurut Breiman (2001) dalam Izenman (2008), Dugaan atau pikiran yang membuat suatu variabel dianggap 'penting' tidak dapat dipahami dengan baik, tapi ada suatu interpretasi bahwa variabel yang penting adalah variabel yang jika dihilangkan akan berdampak serius pada akurasi prediksi kita. Namun, kita tidak dapat mengetahui bagaimana dampak prediksi kita, karena data yang kita analisis tidak memiliki variabel respon dan justru kita yang akan memberikan label atau variabel respon pada data yang kita analisis. Dalam kasus seperti ini, kita tidak boleh mengurangi variabel karena setiap variabel mengandung informasi yang penting buat kita dan dapat digunakan sebagai dasar dalam pengklasteran yang akan dilakukan pada data tersebut.
2. Proses *ordering attributes*/pengurutan atribut pada Cheng dkk (2009) menggunakan *stepwise regression* atau pendekatan statistik diubah menggunakan pendekatan *logical screening* (Huberty, 1994)
3. Pengukuran performansi metode oleh Cheng dkk (2009) menggunakan *accuration rate* (tingkat akurasi) diubah menjadi pengukuran *Sum of Squares Error* (SSE) dan *Silhouette value* (nilai siluet) serta plot siluet (Martinez & Martinez (2005); Izenman (2008)).

Validasi Modifikasi Metode *k-means* Berbasis OWA

Validasi metode dilakukan untuk menjamin bahwa modifikasi metode *k-means* berbasis OWA yang telah dilakukan dapat diaplikasikan pada data riil dan memberikan performansi yang baik atau cukup valid. Proses validasi ini dilakukan dengan cara mengaplikasikan modifikasi *k-means* OWA untuk klastering data set *iris* yang merupakan data set yang umum dijadikan *sample data* untuk menguji apakah metode baru yang dikembangkan oleh peneliti cukup valid atau tidak. Ukuran performansi untuk menilai metode klastering yang digunakan dalam validasi metode ini adalah *accuration rate* (tingkat akurasi). Metode ini dinilai valid atau lebih baik daripada yang lain, jika nilai tingkat akurasi metode *k-means* OWA lebih besar dari metode klastering yang lain. Tahapan proses validasi dapat diringkas dalam diagram alir berikut ini.



Gambar 1. Modifikasi Metode *k-means* Berbasis OWA untuk Kasus Klastering

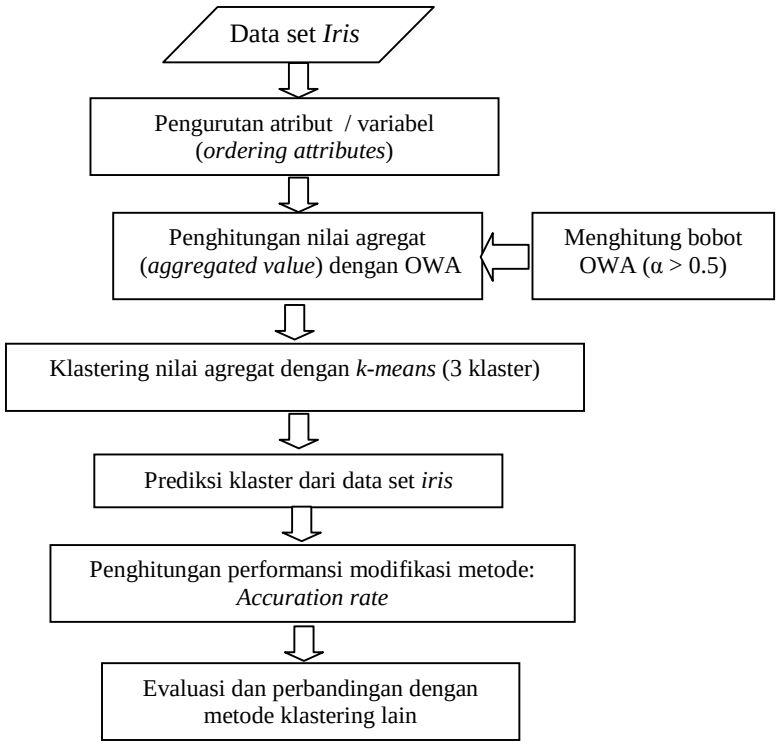
Tahapan validasi modifikasi metode *k-means* berbasis OWA secara rinci dapat dijelaskan sebagai berikut:

1. Data set *Iris*

Data set *iris* ini memiliki label kelompok atau variabel respon, sehingga lazim digunakan oleh para peneliti untuk memvalidasi metode-metode baru yang mereka kembangkan. Data set *Iris* didownload dari UCI *Machine Learning Repository*. Data set *Iris* adalah data tentang klasifikasi jenis bunga menjadi 3 (tiga) kelas, yaitu Setosa, Virginica dan Versicolor berdasarkan 4 atribut, yaitu *sepal length*, *sepal width*, *petal length* dan *petal width* (Frank &Asuncion, 2010).

2. Pengurutan atribut / variabel (*ordering attributes*)

Proses ini dilakukan untuk memperoleh urutan atribut mulai dari yang paling penting sampai yang tidak penting. Atribut yang paling penting nanti akan dikalikan dengan bobot OWA yang pertama (w_1), atribut terpenting kedua dikalikan w_2, dan seterusnya. Proses pengurutan atribut ini dilakukan dengan cara mencoba semua kombinasi urutan variabel hingga menemukan urutan yang menghasilkan SSE yang paling rendah dan nilai siluet paling tinggi. Untuk data set *iris*, karena ada 4 variabel dalam data ini, maka diperoleh 24 kombinasi urutan variabel yang mungkin untuk diklasterkan dengan modifikasi metode *k-means* berbasis OWA.



Gambar 2. Tahapan Validasi Modifikasi Metode *k-means* Berbasis OWA

3. Menghitung bobot OWA

Penentuan bobot OWA dihitung menggunakan persamaan yang dikembangkan oleh Fuller dan Majlender (2001), yaitu mengubah persamaan OWA Yager's ke persamaan polinomial dengan menggunakan *Lagrange multipliers*. Menurut pendekatan mereka, vektor bobot yang terkait dapat diperoleh dengan persamaan (2.8) – (2.10). Menurut Marichal (1999), nilai *orness* (α) adalah ukuran toleransi yang digunakan. Nilai parameter situasi (α) yang digunakan adalah α > 0,5.

Nilai bobot OWA untuk tiap data set akan berbeda jika jumlah variabel masing-masing data set tersebut berbeda. Penghitungan bobot OWA ini dilakukan untuk α > 0.5, dan nanti semua α tersebut akan diproses pada langkah-langkah selanjutnya sampai diperoleh hasil klastering dan diuji performansinya, sehingga dapat diketahui α berapa yang menghasilkan performansi paling bagus.

Nilai α atau *orness* adalah ukuran toleransi dari pembuat keputusan. toleransi (*tolerant*) pembuat keputusan dapat diterima jika hanya beberapa kriteria dipenuhi, hal ini dapat disamakan dengan *disjunctive behavior* (*orness* atau $C\mu > 0.5$). Sedangkan *intolerant* pembuat keputusan menginginkan bahwa sebagian besar kriteria terpenuhi bersama-sama, sama dengan *conjunctive behavior* (*orness* atau $C\mu < 0.5$). Untuk nilai *orness* atau $C\mu = 0.5$ berhubungan dengan keputusan yang adil (Marichal, 1999).

4. Menghitung nilai agregat dengan operator OWA

Setiap data set memiliki sejumlah atribut (n), jumlah atribut pada data set *iris* adalah 4. Dari langkah 2 dan 3, kita peroleh urutan atribut dan bobot OWA. Untuk menghitung nilai agregat, kita kalikan nilai-nilai urutan atribut dengan bobot OWA yang sesuai, itu dapat dinyatakan dalam persamaan berikut.

$$\text{nilai agregat}(a_i) =$$

$$w_1 x_{i1} + w_2 x_{i2} + w_3 x_{i3} + \ldots + w_i x_{ij} + w_n x_{in}$$

dimana w_j adalah OWA dari variabel input ke-*j*, x_{ij} elemen terbesar ke-*j* dalam variabel input, dan a_i adalah nilai agregat ke-i dari *m* data.

5. Mengklasterkan nilai agregat dengan *k-means*

Langkah ini, mengklasterkan nilai-nilai yang telah diagregasikan dengan *k-means*. Klastering data set *iris* dilakukan menjadi 3 klaster, sesuai dengan data aslinya.

6. Menghitung *accuration rate* (tingkat akurasi)

Hasil klastering pada semua nilai α yang kita usulkan dibandingkan dengan label asli kelompok yang ada pada data set *iris* dengan cara menghitung *accuration rate*nya. *Acuration rate* merupakan persentase jumlah data yang label prediksinya sama dengan label aslinya dibandingkan dengan jumlah semua data yang ada. Label prediksi adalah label dari hasil klastering oleh metode yang kita usulkan.

$$\textit{Accuration rate}\ (\%) = \frac{\textit{jumlah prediksi yang benar}}{\textit{jumlah semua data}} x100\% \quad (3.2)$$

7. Evaluasi dan pembandingan dengan metode lain

Langkah ini bertujuan untuk memvalidasi metode yang kita usulkan. *accuration rate* tertinggi pada modifikasi *k-means* berbasis OWA ini dievaluasi dan dibandingkan dengan *accuration rate* metode klastering lain yang telah ada sebelumnya, untuk mengetahui apakah metode ini sudah cukup valid untuk digunakan atau tidak dan pada α berapa yang menghasilkan tingkat akurasi tertinggi. Metode pembanding yang digunakan antara lain: *k-means* menggunakan jarak *Cityblock* dan *Euclidean* serta *hierarchical clustering* dengan jarak *single linkage* dan *complete linkage*.

Pembandingan Modifikasi Metode *k-means* Berbasis OWA dengan Metode Klastering Lainnya

Pengukuran Accuration Rate

Nilai *accuration rate* merupakan perbandingan antara jumlah anggota klaster yang sesuai dengan label pada data aslinya dibandingkan dengan jumlah data keseluruhan.

$$\text{Accuration rate} = \frac{\textit{Jumlah data yang sesuai label aslinya}}{\textit{Jumlah data keseluruhan}} x100\%$$

Pembandingan Performansi k-means Berbasis OWA dengan Metode Klastering yang Lain

Data set *iris*, selain diklasterkan dengan metode *k-means* berbasis OWA juga diklasterkan dengan metode klastering yang lain, yaitu metode *k-means* (tanpa OWA) dengan jarak *Euclidean* dan *Cityblock* dan klastering hirarki (*hierarchical clustering*) *single linkage* dan *complete linkage*. dan dihitung *silhouette value* dan SSEnya. Kedua kriteria tersebut kemudian dibandingkan antar metode klastering yang digunakan, untuk mengetahui apakah performansi metode *k-means* berbasis OWA yang digunakan lebih baik atau tidak dari metode klastering lain yang telah ada.

Teknik Analisis Data

Semua tahapan penelitian ini dilakukan menggunakan bantuan software Matlab 7.01, dengan cara membuat program sederhana (*Matlab code*) yang dapat digunakan secara cepat untuk memperoleh hasil klastering pada masing-masing metode yang digunakan. Program tersebut kemudian diaplikasikan pada Matlab command window dengan memasukkan input: data yang dianalisis (x), jumlah klaster yang diinginkan (*k*), dan jumlah iterasi yang digunakan (I). Sehingga proses analisis data pada penelitian ini diharapkan lebih efisien waktu dan hasilnya lebih akurat.

HASIL DAN PEMBAHASAN

Validasi Modifikasi Metode *k-means* Berbasis OWA Menggunakan Data Set *Iris*

Data Set Iris

Data set *iris*, merupakan data pengelompokan 150 bunga *iris* menjadi 3 kelompok, yaitu Setosa, Versicolor dan Virginica berdasarkan empat variabel, yaitu: 1) sepal length, 2) sepal width, 3) petal length dan 4) petal width (panjang dan lebar dari kelopak dan mahkota bunganya). Data set *iris* ini merupakan data yang umum digunakan di bidang *data mining* untuk membantu proses validasi berbagai metode baru, karena data ini telah memiliki label data atau variabel respon. Hasil pengelompokan dengan metode baru dibandingkan dengan label data yang

ada, lalu dihitung berapa obyek/data yang label hasil prediksi tidak sesuai dengan label aslinya. Data set *iris* secara lengkap dapat dilihat pada Lampiran 2.

Sebagai tahap awal, perlu dihitung terlebih dahulu bobot OWA untuk data *iris*. Penentuan bobot OWA dihitung dengan persamaan (7), (8) dan (9) yang dikembangkan oleh Fuller & Majlender (2001). Persamaan ini memerlukan input parameter yaitu nilai n (jumlah variabel) dan nilai *orness* (α atau parameter situasi). Di dalam metode OWA, nilai α merupakan toleransi yang dapat diterima oleh pengambil keputusan. Nilai n dalam data set *iris* adalah 4 dan nilai *orness* yang digunakan adalah α > 0,5. Bobot OWA dengan jumlah variabel 4 dapat dilihat pada Tabel 1. berikut.

Tabel 1. Bobot OWA untuk n = 4

	α = 0.5	α = 0.6	α = 0.7	α = 0.8	α = 0.9	α = 0.10
w_1	0,2500	0,3474	0,4614	0,5965	0,7641	1,0000
w_2	0,2500	0,2722	0,2756	0,2520	0,1821	0,0000
w_3	0,2500	0,2133	0,1647	0,1065	0,0434	0,0000
w_4	0,2500	0,1671	0,0984	0,0450	0,0103	0,0000

Tabel 2. Perhitungan Nilai Agregat (α = 0.6 ; urutan variabel : SL, SW, PL, PW)

No.	Sepal Length * w_1	Sepal Width * w_2	Petal Length * w_3	Petal Width * w_4	Nilai Agregat
1	5.1 * 0.3474	3.5 * 0.2722	1.4 * 0.2133	0.2 * 0.1671	3.05648
2	4.9 * 0.3474	3.0 * 0.2722	1.4 * 0.2133	0.2 * 0.1671	2.85090
3	4.7 * 0.3474	3.2 * 0.2722	1.3 * 0.2133	0.2 * 0.1671	2.81453
4	4.6 * 0.3474	3.1 * 0.2722	1.5 * 0.2133	0.2 * 0.1671	2.79523
…	…	…	…	…	
147	6.3 * 0.3474	2.5 * 0.2722	5.0 * 0.2133	1.9 * 0.1671	4.25311
148	6.5 * 0.3474	3.0 * 0.2722	5.2 * 0.2133	2.0 * 0.1671	4.51806
149	6.2 * 0.3474	3.4 * 0.2722	5.4 * 0.2133	2.3 * 0.1671	4.61551
150	5.9 * 0.3474	3.0 * 0.2722	5.1 * 0.2133	1.8 * 0.1671	4.25487

Tabel 3. Jumlah Data yang Label Prediksi Tidak Sesuai dengan Label Aslinya pada Klastering Data *Iris* Menggunakan *k-means* Berbasis OWA

Kombinasi urutan	1234	1243	1342	1324	1423	1432	2341	2314	2413	2431	3412	3421	3241
α = 0.5	22	22	22	22	22	22	22	22	22	22	22	22	22
α = 0.6	24	25	22	22	23	22	23	25	29	25	19	18	19
α = 0.7	33	37	24	24	26	24	28	31	42	30	15	8	18
α = 0.8	47	43	24	26	31	25	44	45	71	52	13	10	13
α = 0.9	50	45	36	36	34	34	69	73	87	87	5	9	12
α = 1.0	50	55	48	42	56	42	66	78	79	71	16	7	7
Kombinasi urutan	**3214**	**4123**	**4132**	**2143**	**2134**	**1423**	**1432**	**4213**	**4231**	**3124**	**3142**	**4321**	**4312**
α = 0.5	22	22	22	22	22	22	22	22	22	22	22	22	22
α = 0.6	21	21	20	29	27	23	22	22	19	22	21	19	19
α = 0.7	20	21	19	45	42	26	24	25	19	20	20	18	18
α = 0.8	18	19	17	74	66	31	25	16	24	17	17	6	7
α = 0.9	15	6	6	81	77	34	34	10	9	15	11	6	6
α = 1.0	16	50	6	73	78	42	48	50	6	7	7	6	6

Keterangan:
1 = Sepal Length
2 = Sepal Width
3 = Petal Length
4 = Petal Width

Tahap selanjutnya adalah *ordering attributes*, yaitu mengurutkan 4 variabel pada data *iris* mulai dari yang penting sampai yang tidak penting dalam menentukan proses pengelompokan. Karena problem ini adalah problem klastering, maka pengurutan variabel dapat dilakukan mencoba semua urutan yang mungkin dari 4 variabel yang ada, sehingga akan diperoleh 24 urutan yang mungkin dipilih. Hasil urutan variabel tersebut kemudian akan dikalikan dengan bobot OWA mulai dari w_1, w_2, w_3, dan w_4 yang kemudian diagregasikan menjadi satu nilai. Perhitungan nilai agregat ini dilakukan pada 24 urutan yang ada dan masing-masing dihitung pada semua nilai *orness* (α). Di bawah ini ditampilkan contoh perhitungan nilai agregat pada data *iris* dengan nilai α = 0.6 dan urutan variabel: 1) Sepal Length/SL, 2) Sepal Width/SW, 3) Petal Length/PL dan 4) Petal Width/PW.

Selanjutnya, nilai agregat pada kolom terakhir diklasterkan menggunakan metode *k-means*. Label klaster dari proses klastering tersebut adalah label hasil prediksi, yang kemudian akan dibandingkan dengan label asli (variabel respon) dari data set *iris*. Untuk contoh di atas (Tabel 2), hasil klastering tersebut menghasilkan 126 data yang label prediksinya sesuai dengan label aslinya. Jadi, ada 24 data yang label prediksinya tidak sesuai dengan label aslinya, sehingga tingkat akurasinya sebesar 84%. Modifikasi metode *k-means* berbasis OWA diaplikasikan pada seluruh kombinasi urutan variabel dan pada semua nilai *orness* (α) untuk dicari pada nilai *orness* (α) berapa dan bagaimana urutan variabel yang menghasilkan tingkat akurasi paling tinggi. Aplikasi *k-means* berbasis OWA pada keseluruhan kombinasi urutan variabel data *iris* dapat dirangkum dalam Tabel 3.

Berdasarkan Tabel 3 terlihat bahwa pada urutan variabel: 1) petal length, 2) petal width, 3) sepal length dan 4) sepal width dan pada nilai *orness* (α) = 0.9 menghasilkan 5 data yang tidak sesuai labelnya, artinya ada 145 data yang label prediksinya sesuai dengan label aslinya. Sehingga tingkat akurasinya adalah 96,67%, paling tinggi dibandingkan urutan variabel dan nilai *orness* (α) yang lain.

Perbandingan Tingkat Akurasi yang Diperoleh dengan Metode Klastering yang Lain

Untuk mengetahui sampai sejauh mana validitas metode *k-means* berbasis OWA ini, maka perlu dilakukan pembandingan dengan metode lain. Dengan data set yang sama, yaitu data *iris* diklasterkan menggunakan metode klastering lain seperti *k-means* (tanpa OWA) dan *hierarchical clustering* dan label prediksinya akan dibandingkan dengan label asli dari data *iris* tersebut. Perbandingan berbagai metode pada klastering data *iris* dapat dilihat pada Tabel 4. berikut.

Tabel 4. Perbandingan dengan Metode Lain pada Klastering Data *Iris*

Metode klastering	Jumlah data dengan label prediksi tidak sama dengan label asli	Jumlah data dengan label prediksi sama dengan label asli	Tingkat akurasi
K-means berbasis OWA (urutan variabel: Petal Length, Petal Width, Sepal Length dan Sepal Width serta nilai α = 0.9)	5	145	**96.67**
K-means (dengan menggunakan jarak *cityblock*)	17	133	88.67
K-means (dengan menggunakan jarak *Euclidean*)	16	134	89.33
Hierarchical clustering (*Single linkage*)	48	102	68
Hierarchical clustering (*Complete linkage*)	24	126	84

Berdasarkan Tabel 4. tersebut, maka metode *k-means* berbasis OWA merupakan metode yang lebih baik dibandingkan dengan *k-means* (tanpa OWA) dan *hierarchical clustering* karena menghasilkan tingkat akurasi yang paling tinggi, yaitu 96,67% dalam pengklasteran data set *iris*. Dengan demikian modifikasi *k-means* berbasis OWA ini memungkinkan untuk diterapkan dalam klastering data set yang lain, namun pengukuran performansinya dilakukan dengan menghitung nilai SSE dan *silhouette value*nya karena data set yang riil untuk kasus klastering tidak memiliki label kelompok. Jika diterapkan pada data riil maka perlu dilakukan perbandingan dengan metode klastering yang lain dan dibandingkan performansinya untuk mengetahui apakah metode modifikasi *k-means* berbasis OWA ini lebih baik atau tidak.

KESIMPULAN DAN SARAN

Metode *k-means* berbasis OWA yang dimodifikasi cukup valid untuk data *iris* dan dapat diaplikasikan pada klastering data set yang lain. Tingkat akurasi metode *k-means* berbasis OWA dalam klastering data set *iris* adalah 96.67% yang lebih tinggi dari metode klastering yang lain. Pada penelitian ini, proses *ordering attributes* (mengurutkan atribut) merupakan tahapan kunci yang membedakan antara kasus klastering dan klasifikasi yang memanfaatkan metode OWA. Pada penelitian ini, digunakan pendekatan *logical screening* dengan cara sintesis *expert judgment* dan data-data sekunder dari penelitian terdahulu untuk mengurutkan atribut pada data yang dianalisis. Untuk mengatasi kelemahan tersebut, masih memungkinkan untuk menggunakan pendekatan lain dalam tahapan *ordering attributes* ini, misalnya menggunakan teknik pembobotan terhadap atribut dengan pembobotan melalui penilaian pakar secara langsung menggunakan kuisioner dan perlu juga diteliti lebih lanjut tentang kemungkinan penggunaan operator agregasi yang lain selain *Ordered Weighted Averaging* (OWA) yang digabungkan dengan metode *k-means* untuk mengklasterkan suatu data set.

DAFTAR PUSTAKA

Agusta Y. 2007. K-means, Penerapan, Permasalahan dan Metode Terkait, *Jurnal Sistem dan Informatika* Vol, **3**

Agusta Y. 2008. *K-means*, Artikel Internet. http://yudiagusta,wordpress,com/ k-means/ [diakses tanggal 4 Maret 2010]

Ahn BS. 2006. On The Properties of OWA Operator Weights Functions with Constant Level of Orness, *IEEE Transactions on Fuzzy Systems* (**4**), 511–515.

Chang JR. dan CH Cheng. 2006. MCDM aggregation model by ME-OWA and MEOWGA operators, *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*. **14(4)**, 421–443.

Cheng CH, JW Wang, dan MC Wu. 2009. OWA-Weighted Based Clustering Method for Classification Problem, *Expert Systems with Application* **26**, 4988-4995.

Dunham MH. 2002. *Data Mining Introductory and Advanced Topics*, Pearson Education, Inc.

Frank A dan A Asuncion. (2010), *UCI Machine Learning Repository* [http://archive,ics,uci,edu/ml], Irvine, CA: University of California, School of Information and Computer Science,

Fuller R. dan P Majlender. 2001. An Analytic Approach for Obtaining Maximal Entropy OWA Operator Weights. *Fuzzy Sets and Systems*. **124**. 53–57.

Grandhi R. 2003. *Integration of Ordered Weighted Averaging Operators with Feed-Forward Neural Networks for Optimal Feature Subset Selection And Pattern Classification*, Tesis Master, Universitas Florida, Florida.

Han J. dan M Kamber. 2006. *Data mining: Concepts and techniques (2nd ed,)*, Elsevier Inc.

Huberty. 1994. *Applied Discriminant Analysis*. New York: Wiley Interscience.

Huh MH dan YB Lim. 2009. Weighting Variables in K-means Clustering, *Journal of Applied Statistics*. **36**: 1, 67 – 78.

Izenman A.J. 2008. *Modern Multivariate Statistical Techniques, Regression, Classification, and Manifold Learning*, Springer Science&Business Media, LLC, USA.

Jain AK. 2009, Data Clustering: 50 Years Beyond K-Means, *Pattern Recognition Letters*.

Jain AK dan B Chandrasekaran. 1982, Dimensionality and Sample Size Considerations in Pattern Recognition Practice. In *Handbook of Statistics*, Eds: P.R. Krishnalah & Kanal, L.N., Vol.2, halaman 835 – 855, North-Holland.

Liu X. 2004. Three methods for generating monotonic OWA operator weights with given orness leve. *Journal of Southeast University*. **20(3),** 369–373.

Marichal JL. 1999. *Aggregation Operators for Multicriteria Decision Aid,* [Desertasi, University of Liège, Belgium]

Martinez AL. dan AR Martinez. 2005, *Exploratory Data Analysis with MATLAB.* CRC Press Company, USA.

O'Hagan M. 1988. "Aggregating Template or Rule Antecedents in Real-Time Expert Systems with Fuzzy Set Logic", dalam *Proceedings of the 22nd Annual IEEE Asilomar Conference Signals, Systems, Computers*, hal. 681–689, Pacific Grove, CA.

Santosa B. 2007. *Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Jakarta: Graha Ilmu.

Yager RR. 1988. On Ordered Weighted Averaging Aggregation Operators in Multicriteria Decision Making. *IEEE Transactions on SMC*. **18**. 183–190.

Yager RR. 2004. Modeling prioritized multi criteria decision making, *IEEE Transactions on System, Man, and Cybernetics – Part B: Cybernetics*, **23(6)**. 2396–2403.