

KLUSTERING DENGAN K-MEANS BERBASIS LVQ DAN K-MEANS BERBASIS OWA

Dian Eka Ratnawati¹, Indriati²

^{1,2}Program Studi Informatika/Ilmu Komputer, Fakultas Ilmu Komputer
Universitas Brawijaya
Email: ¹dian_ilkom@ub.ac.id, ²indriati.tif@ub.ac.id

(Naskah masuk: 29 Januari 2015, diterima untuk diterbitkan: 17 Februari 2015)

Abstrak

Pada penelitian ini dilakukan perbandingan hasil klustering pada data *car evaluation* dengan menggunakan K-Means berbasis LVQ (Learning Vector Quantization) dan K-Means berbasis OWA (*Ordered Weighted Averaging*). Pada kedua metode ini sama-sama mempergunakan K-Means tetapi yang sudah mengalami modifikasi.

Hasil dari penelitian sebelumnya secara terpisah yang membandingkan metode K-Means modifikasi tersebut dengan K-Means konvensional menunjukkan bahwa kedua metode modifikasi tersebut sama-sama lebih baik daripada K-Means konvensional. Tetapi belum pernah ada penelitian yang membandingkan akurasi hasil klustering kedua metode modifikasi tersebut. Sehingga pada penelitian ini dilakukan klustering dengan menggunakan kedua metode tersebut untuk data *car evaluation*, karena dari penelitian sebelumnya kedua metode tersebut cukup handal dalam melakukan klustering. Hasil dari uji coba menunjukkan rata-rata hasil akurasi dimulai yang tertinggi adalah K-Means berbasis LVQ (86,50%), K-Means berbasis OWA (86,16%) kemudian K-Means konvensional (56,50%). Tetapi dengan urutan atribut yang benar dan pemilihan nilai alpha yang tepat yakni 0,8, K-Means berbasis OWA bisa menghasilkan akurasi yang lebih tinggi yakni 93,33%.

Kata kunci: K-Means berbasis LVQ, K-Means, K-Means berbasis OWA, bobot

Abstract

In this paper do a comparison with the results of clustering using K-Means based LVQ (Learning Vector Quantization) and K-Means based OWA (Ordered Weighted Averaging). In both of these methods used K-Means but which has been modified.

Results from previous studies have shown that both methods are better than conventional K-Means. But there has never been a study comparing the accuracy of clustering results of the two methods. So in this study conducted clustering using both methods for data car evaluation, because of previous studies both methods are reliable enough to perform clustering. In the researchs before it, both method are prefer than conventional K-Means, but there are no researchs which compare them. So, in the research, we will compare it by using same data that is car evaluation. In order to know what it is method is the best. The result of research are that in the average, K-Means LVQ (86.50%) is more accuracy than K-Means – OWA (86,16%) and conventional K-Means (56,50). But if the order of selection attributes and alpha values is correct, K Means based OWA can generate higher accuracy that is 93.33 % using the alpha value of 0.8

Keywords: K-Means based LVQ, K-Means, K-Means based OWA, weight

1. PENDAHULUAN

K-Means konvensional selain mempunyai kelebihan yakni sangat cepat dalam proses klustering, mempunyai beberapa kelemahan antara lain: memungkinkan suatu kluster tidak mempunyai anggota dan hasil klusteringnya selalu berubah-ubah. Hal tersebut yang melatarbelakangi munculnya berbagai metode K-Means modifikasi. Dua diantaranya dari K-Means modifikasi tersebut adalah K-Means berbasis LVQ dan K-Means berbasis OWA.

Untuk K-Means berbasis LVQ yang akan dilakukan adalah : Pertama : memodifikasi dari K-Means konvensional, yakni melakukan proses K-Means secara *iterative* sampai nilai *threshold* tertentu, jika pada *threshold* tersebut belum didapatkan 1 kluster dengan anggota yang sama, maka baru dilakukan proses LVQ (*Learning Vector Quantization*). Dengan menggunakan data *breast-cancer-wisconsin* hasil klustering dengan K-Means berbasis LVQ mempunyai tingkat akurasi (92,67%), yang jauh di atas K-Means konvensional dengan tingkat akurasi : 60,2%. Hal ini menunjukkan dengan data tersebut, K-Means modifikasi

mempunyai akurasi yang lebih tinggi daripada K-Means konvensional.

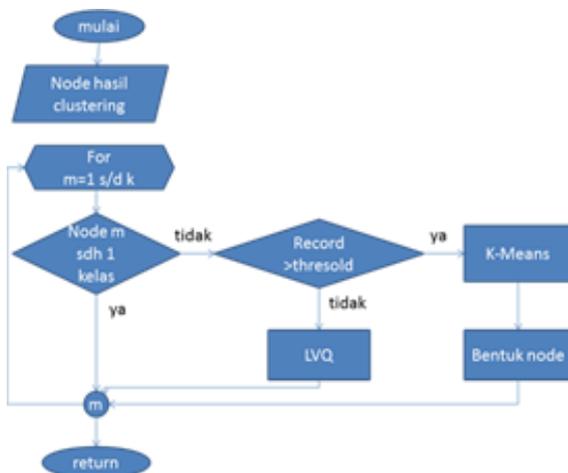
Sedangkan K-Means berbasis OWA (*Ordered Weighted Averaging*) diperkenalkan oleh Cheng dkk (2009) yang melakukan klustering nilai agregat, yang merupakan kumpulan dari nilai multi atribut yang ada. Metode ini berbeda dengan metode K-Means yang telah ada sebelumnya. Cheng dkk (2009) lebih fokus pada cara untuk mengurangi kompleksitas *data set* eksperimental dan keterkaitan antara berbagai kriteria yang ada, yang dapat diatasi dengan cara menggabungkan K-Means dengan OWA. Dari penelitian yang dilakukan oleh Ulya yang dipergunakan untuk klustering iris menunjukkan metode ini lebih bagus akurasinya dibandingkan dengan K-Means konvensional, yakni K-Means berbasis OWA akurasinya : 96.67 sedangkan K-Means konvensional 89.33.

Dengan melihat keunggulan kedua metode tersebut jika dibandingkan dengan K-Means konvensional, maka pada penelitian ini akan dibandingkan akurasi dari kedua metode modifikasi tersebut yakni K-Means berbasis LVQ dan K-Means berbasis OWA.

2. TINJAUAN PUSTAKA

2.1. K-Means berbasis LVQ

Flowchart untuk K-Means berbasis LVQ bisa dilihat pada Gambar 1.



Gambar 1. Flowchart K-Means berbasis LVQ

Seperti pada Gambar 1, tahapan untuk melakukan klustering dengan K-Means berbasis LVQ adalah :

1. Mengelompokkan data dengan metode *K-Means*
2. Dilakukan pengecekan label terhadap semua kelompok yang dihasilkan dari proses K-Means sebelumnya. Yakni apakah anggota pada satu kluster tersebut sudah berlabel sama atau belum. Jika sudah berlabel sama (satu kelas semua), maka pemrosesan terhadap kluster tersebut berhenti, jika tidak lakukan pengecekan ke langkah 3

3. Bandingkan isi kluster dengan *threshold* (isi minimal anggota suatu kluster). Jika isi kluster > *threshold* maka dilakukan K-Means lagi, sedang jika < *threshold* dilakukan LVQ.

2.2 K- Means berbasis OWA

Langkah yang dilakukan untuk K-Means berbasis OWA adalah (Cheng, 2007) :

1. Pemilihan Fitur . Pada dataset *car evaluation* ini ada 6 fitur yang dipilih.
2. Menghitung bobot dengan menggunakan Operator OWA . Algoritma untuk pembobotan dengan OWA mengacu pada Cheng, 2007.
3. Penghitungan nilai *agregat* berdasar operator OWA dari bobot OWA
4. Kluster nilai *agregat* . *K-Means* digunakan untuk menghasilkan kluster yang didasarkan pada nilai-nilai agregat.
5. Prediksi klustering. Dari langkah 4 diperoleh bobot optimal, kemudian dari bobot itu akan dilakukan prediksi klustering terhadap data uji, dengan persamaan (1) :

$$E_Disti = \sqrt{(Ci_{center} - aij)^2} \quad (1)$$

2.3. Pembobotan dengan OWA

Algoritma untuk metode OWA seperti terlihat pada Gambar 2

```

For different  $\alpha$  and  $n$ , we can get different OWA weight.
OWA ( $n, \alpha$ )
/*  $n$  is the number of attributes;  $\alpha$  is the situation
parameter */
If  $\alpha < 0.5$ 
Then  $\alpha = 1 - \alpha$ 
If  $\alpha > 0.5$ 
Then  $w_1 [((n-1)\alpha + 1 - n w_1)]^n =$ 
 $[(n-1)\alpha]^{n-1} [((n-1)\alpha - n)w_1 + 1]$  //Calculate  $w_1$ 
 $w_n = [((n-1)\alpha - n)w_1 + 1] / [(n-1)\alpha + 1 - n w_1]$ 
//Calculate  $w_n$ 
For  $i = 2$  to  $n-1$  do
 $w_i = \sqrt[n]{w_1^{n-i} w_n^{i-1}}$  //Calculate  $w_i$ 
    
```

Gambar 2. Algoritma untuk pembobotan dengan OWA. Sumber : Cheng, 2007

3. Dataset

Penelitian ini akan menggunakan *data car evaluation*. Data ini diambil dari website dengan alamat <http://archive.ics.uci.edu/ml/dataset/car+evaluation>. Jumlah data tersebut 1728. Dari data tersebut diambil 200 record secara acak sebagai data latih dan 60 record sebagai data uji. Data tersebut memiliki 8 atribut, yaitu :

1. id : nomor data
2. buying: dapat bernilai vhigh, high, med, low.
3. maint: dapat bernilai vhigh, high, med, low.

4. doors: dapat bernilai 2, 3, 4, 5 more.
5. persons: dapat bernilai 2, 4, more.
6. lug_boot: dapat bernilai small, med, big.
7. safety: dapat bernilai low, med, high.
8. class : kelas data, dapat bernilai unacc, acc, good, vgood

Atribut nomor 1 dan no 8 tidak dipakai pada penelitian.

4. Hasil Uji Coba dan Analisa

Pada penelitian ini dilakukan dua skenario ujicoba terhadap data *car evaluation* dengan jumlah data uji 60. Skenario pertama, uji coba dilakukan dengan menggunakan metode K-Means berbasis OWA. Skenario kedua akan membandingkan hasil uji coba antara K-Means berbasis LVQ, K-Means berbasis OWA dan K-Means konvensional dengan menggunakan parameter terbaik pada masing-masing metode.

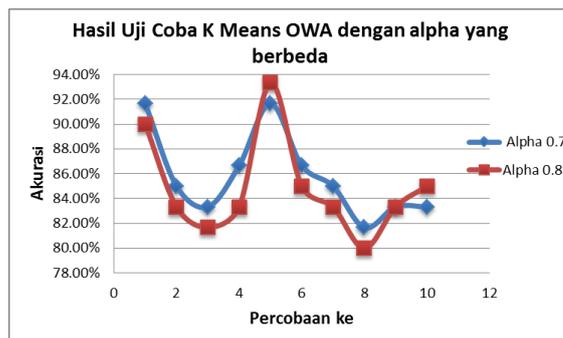
4.1. Hasil Uji coba dengan menggunakan K-Means berbasis OWA

Ada 2 hal yang ingin diketahui dari uji coba pertama ini : Pertama untuk mengetahui apakah urutan atribut berpengaruh terhadap hasil akurasi. Jika berpengaruh berarti bobot dari setiap atribut seharusnya tidak sama. Kedua untuk mengetahui apakah nilai alpha yang merupakan salah satu parameter dari OWA berpengaruh terhadap hasil akurasi.

Untuk point yang kedua, maka pada uji coba dipergunakan nilai alpha 0.7 dan 0.8. Hal ini dengan pertimbangan nilai alpha 0.7 dan 0.8 adalah paling baik dari semua nilai alpha antara 0 dan 1 yang sebelumnya telah dilakukan uji coba terlebih dahulu. Hasil ujicoba bisa dilihat pada Tabel.1 dan Gambar 3

Tabel 1. Hasil Ujicoba K-Means OWA terhadap nilai alpha

Uji ke	Kombinasi	alpha 0.7	alpha 0.8
1	150432	91.66%	90.00%
2	104532	85.00%	83.33%
3	14325	83.33%	81.66%
4	150324	86.66%	83.33%
5	150423	91.66%	93.33%
6	150324	86.66%	85.00%
7	12453	85.00%	83.33%
8	103452	81.66%	80.00%
9	15423	83.33%	83.33%
10	154230	83.33%	85.00%
Rata2		85.83%	84.83%



Gambar 3. Hasil Uji coba K-Means OWA dengan nilai alpha yang berbeda

Seperti terlihat pada Tabel 1 dan Gambar 3 menunjukkan bahwa nilai alpha dan urutan atribut mempengaruhi hasil akurasi. Secara rata-rata hasil akurasi dengan menggunakan parameter nilai alpha 0,7 lebih baik (85,83%) daripada dengan menggunakan nilai alpha 0,8 yang mempunyai hasil akurasi 84,83% meski tidak berbeda terlalu jauh.

Dari Tabel 1 dan Gambar 2 terlihat urutan kombinasi atribut yang mempunyai akurasi tertinggi adalah 150423 dengan akurasi 91.66% (alpha 0.7) dan 93.33% (alpha 0.8). Rata-rata akurasi untuk urutan 150423 adalah 92.50%. Arti urutan 150423 berarti susunan atributnya adalah *maint*(1), *safety*(5), *buying*(0), *lug_boot*(4), *doors*(2) dan *persons*(3). Susunan tersebut menunjukkan urutan dari bobot, semakin ke kanan maka bobotnya akan semakin tinggi. Sehingga karena atribut *maint* berada di paling kiri atau posisi pertama, berarti atribut *maint* mempunyai bobot paling rendah dari kelima atribut yang ada, sedangkan atribut *persons* karena berada di posisi terakhir berarti mempunyai bobot paling tinggi.

Urutan berikutnya adalah 150432 dengan akurasi 91.66% (alpha 0.7) dan 90.00% (alpha 0.8). Rata-rata akurasi untuk urutan 150423 adalah 90.83%.

Dari hasil uji coba di atas bisa disimpulkan bahwa untuk data *car evaluation* setiap atribut mempunyai bobot yang berbeda. Pembobotan yang berbeda ini menunjukkan bahwa ada atribut yang sangat berpengaruh terhadap penentuan kluster, sementara ada atribut lain yang kurang berpengaruh. Hanya dengan uji coba bisa diketahui atribut mana yang paling dominan perannya. Dari data tersebut nilai alpha yang paling bagus adalah 0.7 dan 0.8, meski secara rata-rata nilai alpha 0.7 sedikit lebih unggul daripada alpha 0.8.

4.2. Hasil Ujicoba K-Means LVQ, K-Means OWA dan K-Means konvensional

Untuk skenario uji coba yang kedua ini ingin mengetahui perbandingan ke 3 metode dari segi akurasi dan waktu komputasi. Pada penelitian ini dilakukan 10 kali ujicoba. Hasil uji coba bisa

dilihat pada Tabel 2 dan Gambar 4. Parameter yang dipergunakan untuk uji coba K-Means berbasis LVQ adalah diambil parameter yang menghasilkan akurasi tinggi, dimana :

Parameter K-Means

- Jumlah kluster : 4,
- *threshold* :25

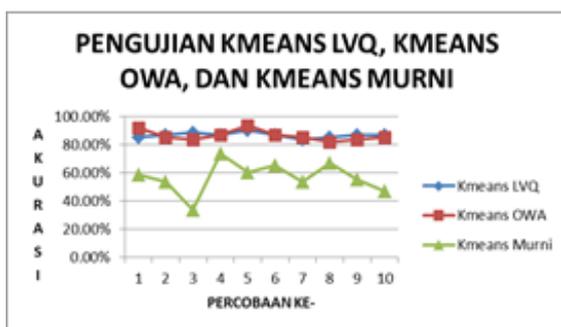
Parameter untuk atribut LVQ :

- *error minimum* : 0.000001
- *Max epoch* : 100
- *Learning rate* : 0.05
- Faktor pengurang : 0,01

Selanjutnya hasil terbaik dari setiap metode dapat dilihat seperti pada Tabel 2 dan Gambar 3.

Tabel 2 Hasil uji coba penelitian dengan K-Means berbasis OWA, K-Means berbasis LVQ dan K-Means konvensional

Uji ke	K-Means LVQ	K-Means OWA	K-Means murni
1	85.00%	91.66%	58.33%
2	86.66%	85.00%	53.33%
3	88.33%	83.33%	33.33%
4	86.66%	86.66%	73.33%
5	90.00%	93.33%	60.00%
6	86.66%	86.66%	65.00%
7	83.33%	85.00%	53.33%
8	85.00%	81.66%	66.66%
9	86.66%	83.33%	55.00%
10	86.66%	85.00%	46.66%
Rata 2	86.50%	86.16%	56.50%



Gambar 4. Hasil uji coba penelitian dengan K-Means berbasis OWA, K-Means berbasis LVQ dan K-Means konvensional

Dari hasil uji coba yang ada pada Tabel 2 dan Gambar 4 menunjukkan bahwa hasil akurasi dari K-Means berbasis OWA dan K-Means berbasis LVQ masih lebih unggul dibandingkan dengan K-Means konvensional. Dengan data *car evaluation*, K-Means konvensional mempunyai rata-rata akurasi dibawah 60%, yakni 56,50%, sedangkan K-Means berbasis OWA mempunyai rata-rata akurasi 86,16% dan K-

Means berbasis LVQ mempunyai rata-rata akurasi 86.50%. Rata-rata akurasi tertinggi ada pada metode K-Means berbasis LVQ meski selisihnya hanya 0,34%.

Tetapi dengan urutan atribut yang benar dan pemilihan nilai alpha yang tepat, K-Means berbasis OWA bisa menghasilkan akurasi yang lebih tinggi yakni 93.33%. Hal yang menyebabkan K-Means berbasis OWA bisa menghasilkan akurasi yang lebih tinggi daripada K-Means berbasis LVQ dan K-Means konvensional adalah karena K-Means berbasis OWA membedakan bobot antara atribut satu dengan atribut lainnya. Sedangkan untuk K-Means berbasis LVQ dan K-Means konvensional semua atribut dianggap mempunyai bobot yang sama.

Berarti untuk data *car evaluation* lebih cocok jika mempergunakan metode K-Means berbasis OWA, dimana setiap atribut memiliki bobot yang berbeda. Kelemahan dari metode ini adalah harus dicari semua kombinasi dari sejumlah atribut yang ada, semakin banyak atribut semakin banyak uji coba yang dilakukan. Jadi pertimbangan untuk mempergunakan metode mana yang lebih sesuai tergantung data yang dipakai, apakah memang ada perbedaan bobot terhadap data tersebut atau tidak.

Untuk waktu komputasi, urutan waktu dari yang paling cepat adalah K-Means konvensional, K-Means berbasis OWA, kemudian K-Means berbasis LVQ. K-Means konvensional paling cepat karena bobot dari setiap atribut dianggap sama jadi tidak perlu dilakukan pembobotan terlebih dahulu pada setiap atributnya, langsung dilakukan klustering dengan menggunakan K-Means. Berbeda untuk K-Means berbasis OWA, sebelum dilakukan klustering dengan menggunakan K-Means maka setiap atribut diberikan bobot yang berbeda. Untuk pembobotannya dengan menggunakan metode OWA. Sedangkan untuk K-Means berbasis LVQ waktu komputasinya paling lama. Hal ini karena proses untuk klustering dengan menggunakan K-Means nya sendiri dilakukan secara berulang-ulang hingga mencapai *threshold* tertentu. Setelah mencapai *threshold* masih dilakukan pengecekan ulang terhadap isi klusternya apakah sudah selesai atau perlu dilakukan klasifikasi dengan menggunakan metode LVQ.

5. Kesimpulan

Dari penelitian ini bisa diambil kesimpulan :

Dengan menggunakan dataset *car evaluation* rata-rata hasil akurasi dimulai yang tertinggi adalah K-Means LVQ(86.50%), K-Means berbasis OWA(86,16%) kemudian K-Means konvensional(56,50%). Tetapi dengan urutan atribut yang benar dan pemilihan nilai alpha yang tepat, K-Means berbasis OWA bisa menghasilkan akurasi yang lebih tinggi yakni 93.33% dengan menggunakan nilai alpha 0.8. Dari segi waktu

komputasi, K-Means konvensional mempunyai waktu yang paling cepat dibanding metode K-Means modifikasi.

6. Daftar Pustaka:

- AHN BS. 2006. On The Properties of OWA Operator Weights Functions with Constant Level of Orness, *IEEE Transactions on Fuzzy Systems* (4),511–515.
- CHENG, CHING-HUSE, JING-WEI LIU, MING-CHANG WU,2007. Owa Based Information Fusion Techniques For Classification Problem, Proceedings of the Sixth International Conference on Machine Learning and Cybernetics, Hong Kong, 19-22 August 2007, -4244-0973-X/07/\$25.00 ©2007 IEEE.
- CHENG CH, JW WANG, DAN MC WU. 2009. OWA-Weighted Based Klustering Method for Classification Problem, *Expert Systems with Application*4988-4995.
- FAYYAD, USAMA., dkk., 1996. From Data Mining to Knowledge Discovery in Databases. American Association for Artificial Intelligence.
- FAUSETT, L., 1994, *Fundamentals of NeuralNetwork, Architecture, Algorithms and Applications*, Prentice Hall, New Jersey
- KANTARDZIC, MEHMED. 2003. Data Mining : Concepts, Models, Methods and Algorithm. John Wiley & Sons. New York.
- RATNAWATI, DIAN, MARJI, LAILIL M. 2012. Pengembangan Metode Klasifikasi berdasarkan K-Means dan LVQ, Penelitian DIPA 2012.
- RATNAWATI, DIAN, MARJI, 2014. Klustering dengan K-Means dan K-Means modifikasi , proceeding KNSI 2014.
- ULYA, MILYATUL, 2011. Modifikasi K-Means Berbasis Ordered Weighted Averaging (Owa) Untuk Kasus Klastering, *AGROINTEK* Volume 5, No.2 Agustus 2011