

Metode K-Means Berbasis *Ordered Weighted Averaging* (OWA) pada Data Potensi Desa untuk Penentuan Status Desa

Aries Alfian Prasetyo ^{1*}

Suprapedi ²

Siska Narulita ³

Bayu Praharsena ⁴

^{1,4} Teknik Listrik Industri, Politeknik Negeri Madura

² Sekolah Pascasarjana, Program Doktor Ilmu Lingkungan, Universitas Brawijaya

³ Sistem dan Teknologi Informasi, Institut Teknologi dan Bisnis Semarang

INFO ARTIKEL

Histori artikel:

Diterima : 4 Agustus 2022
Revisi : 12 Agustus 2022
Disetujui : 19 Agustus 2022
Publikasi : 23 Agustus 2022

Kata kunci:

K-Means
Ordered Weighted Average
Potensi Desa

ABSTRACT

The Village Development Index, which was built from the 2014 Village Potential Data Collection (Podes), is to assess the level of village development, divided into 3 classifications, namely Independent, Developing and Disadvantaged Villages, which has 5 dimensions. The author aims to assess the level of village development through village status based on IPD data, determining village status using the clustering technique with the K-Means Method based on *Ordered Weighted Averaging* (OWA), OWA can reduce data complexity by combining multi-attribute values into aggregate values in the form of a single value with using expert judgment to determine the value of orness (α). The results of this study grouping IPD data in 2014 into 3 village statuses with the K-Means algorithm based on OWA shows the K-Means clustering method with OWA has a Davies Index of 1.42 better than the K-means method with euclidean which has a value of 1.65. . The final result of the study obtained the number of villages for each cluster, namely the Disadvantaged Village cluster as many as 98, the Developing Village cluster as many as 132, and the Independent Village cluster as many as 100 villages.

ABSTRAK

Indeks Pembangunan Desa yang dibangun dari Pendataan Potensi Desa (Podes) tahun 2014. untuk menilai tingkat perkembangan desa, dibagi menjadi 3 klasifikasi yaitu Desa Mandiri, Berkembang, dan Tertinggal, memiliki 5 dimensi. Penulis bertujuan untuk menilai tingkat perkembangan desa melalui status desa berdasar data IPD, penentuan status desa menggunakan teknik *clustering* dengan metode K-Means berbasis *Ordered Weighted Averaging* (OWA), OWA dapat mengurangi kompleksitas data dengan memadukan nilai multi atribut ke nilai agregat berupa nilai tunggal dengan menggunakan *expert judgement* untuk menentukan nilai orness (α). Hasil dari penelitian ini pengelompokan data IPD tahun 2014 kedalam 3 status desa dengan algoritma K-Means berbasis OWA menunjukkan metode *clustering* K-Means dengan OWA memiliki Index Davies 1,42 lebih baik

dari metode *K-Means* dengan *euclidean* yang memiliki nilai 1,65. Hasil akhir penelitian diperoleh jumlah desa untuk setiap *cluster*, yaitu *cluster* Desa Tertinggal sebanyak 98, *cluster* Desa Berkembang sebanyak 32, dan *cluster* Desa Mandiri sebanyak 100 desa.

PENDAHULUAN

Kementerian Desa, Pembangunan Daerah Tertinggal, dan Transmigrasi, BAPENNAS (2014) bersama BPS (Badan Pusat Statistik) mengeluarkan buku tentang Indeks Pembangunan Desa yang di dalamnya terdapat 3 status pengelompokan yang digunakan dalam pembangunan desa. Buku tersebut disertai dengan data indikator pembangunan desa tahun 2014 (IPD 2014) yang terdiri dari 74.093 desa (*instance*). Buku tersebut memuat data potensi desa yang memiliki 5 dimensi, 12 variabel dan 42 indikator sebagai parameter dasar pembangunan sebuah desa. Indeks Pembangunan Desa (IPD) dibangun berdasarkan data hasil Pendataan Potensi Desa (Podes) tahun 2014. Terdapat 5 (lima) dimensi indeks pembangunan desa (IPD) yaitu: 1) Pelayanan Dasar, 2) Kondisi Infrastruktur, 3) Aksesibilitas/Transportasi, 4) Pelayanan Publik, 5) Penyelenggaraan Pemerintahan. Kemendesa PDTT (2015).

K-Means merupakan salah satu metode data *clustering* non hirarki yang berusaha mempartisi data ke dalam *cluster* atau kelompok sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu *cluster* yang sama dan data yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam kelompok yang lain. Yager RR (2004) menyatakan tujuan dari *clustering* ini adalah untuk meminimalisasi *objective function* yang diset dalam proses *clustering*, yang ada umumnya berusaha meminimalisasi variasi di dalam suatu *cluster* dan memaksimalkan variasi antar *cluster*. Cheng CH, JW Wang, dan MC Wu (2009), *problem clustering* memiliki data set eksperimental yang kompleks. *Ordered Weighted Averaging* (OWA) dapat mengurangi kompleksitas data dengan memadukan nilai-nilai multi *attribut* ke nilai-nilai agregat yang berupa nilai tunggal. Nilai tunggal tersebut didapatkan karena persamaan ini hanya memerlukan dua *input* parameter saja yaitu jumlah variabel dari data dan nilai *orness* (α) atau parameter situasi yang digunakan. Masing-masing dimensi dalam Indeks Potensi Desa memiliki nilai yang berbeda, dimana ciri-ciri tersebut yang menentukan pada klaster mana yang sesuai untuk satu sampel bunga. L. Xu, C. Jiang (2014), kasus ini mirip dengan kasus pengambilan keputusan multikriteria (*Multicriteria Decision Making/MCDM*), dimana keputusan mengklasterkan bunga berdasarkan ciri-cirinya didasarkan pada multikriteria. *Ordered Weighted Averaging* (OWA) sangat berguna untuk proses MCDM yang sering kali memerlukan keterkaitan antar kriteria yang ada (Chang J. R. dan C.H. Cheng, 2006).

Sehingga berdasarkan uraian tersebut, dimana data IPD dalam bentuk *unsupervised* (tidak berlabel) dan memiliki kriteria masing-masing proses penentuan status desa menjadi 3, yaitu Maju, Berkembang, dan Tertinggal, dilakukan dengan menggunakan teknik *clustering* dengan Metode *K-Means* berbasis *Ordered Weighted Averaging* (OWA), dengan menggunakan OWA kompleksitas data dapat dikurangi dengan memadukan nilai-nilai multi atribut ke nilai-nilai agregat yang berupa nilai tunggal dengan menggunakan *expert judgement* untuk menentukan nilai *orness* (α) atau parameter situasi dan mengukur peningkatan rata-rata jarak ke *centroid*. Penelitian ini bertujuan untuk mengaplikasikan *K-Means* dengan bobot pada atribut berbasis *Ordered Weighted Averaging* OWA pada *clustering* data potensi desa (pengelompokan desa menjadi 3, yaitu Mandiri, Berkembang, dan Tertinggal) dan mengukur peningkatan rata-rata jarak ke *centroid*.

TINJAUAN PUSTAKA

Algoritma K-Means

Dr. Manju (2014) beserta rekannya Bhawana, pernah melakukan penelitian tentang perbandingan algoritma K-Means dan hierarki *clustering* untuk mencari kelemahan dan mengukur kekuatan dari masing-masing algoritma. Dalam penelitian ini, digunakan data literatur *review* dari penelitian sejenis yang sebelumnya diungkapkan bahwa algoritma K-Means memberikan *executing time*, performa, dan *clustering criteria* yang lebih baik dari *clustering* berbasis hierarki pada *dataset* yang besar.

Manish Verma, Mauliy Srivastava (2012), meneliti tentang perbandingan algoritma *clustering* dalam *data mining* dengan karakteristik *dataset* yang besar. Pada penelitian ini membandingkan enam algoritma *clustering* yaitu K-Means, hierarki, EM *algorithm Density Based*, *Optics*, dan DBSCAN dengan analisa menggunakan WEKA. Hasil pengujian menunjukkan algoritma K-Means kinerja lebih baik menggunakan *dataset* yang besar, algoritma K-Means lebih cepat dibanding algoritma *clustering* lain dan juga menghasilkan *cluster* yang berkualitas

Ordered Weighted Averaging (OWA)

Metode *clustering* 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: *Fuzzy C-Means*, *X-Means*, *K-Medoid*, dan *Kernel K-Means*. Menurut Augusta (2008), terdapat pula metode *K-Harmonic Means* dan *K-Modes*. Variasi metode K-Means tersebut umumnya berhubungan dengan tiga hal yang telah disebutkan oleh Augusta. Namun saat ini telah dihasilkan pengembangan K-Means *clustering* berbasis OWA oleh Cheng yang melakukan *clustering* nilai agregat, yang merupakan kumpulan dari nilai multi atribut yang ada. Metode ini berbeda dengan variasi metode K-Means yang telah ada sebelumnya. Cheng lebih fokus pada cara untuk mengurangi kompleksitas *dataset* eksperimental dan keterkaitan antara berbagai kriteria yang ada, yang dapat diatasi dengan cara menggabungkan K-Means dengan OWA.

Adapun tujuan dari *clustering* ini adalah untuk meminimalisasi *objective function* yang diset dalam proses *clustering*, yang ada umumnya berusaha meminimalisasi variasi di dalam suatu *cluster* dan memaksimalkan variasi antar *cluster*. Dalam proses pengelompokan data ke dalam suatu *cluster*, nilai dari suatu atribut sangat menentukan. Karena konsep K-Means yang mengelompokan data sesuai dengan karakteristik, maka pemilihan *centroid* sebagai titik pusat jarak tergantung pada nilai di atribut tersebut. Jika nilai pada atribut memiliki jarak nilai terlalu kecil dengan nilai lain maka hasil *cluster* juga akan berubah ubah, sesuai pilihan pusat *centroid*. *Problem clustering* memiliki *dataset* eksperimental yang kompleks.

Indeks Potensi Desa

Indeks Pembangunan Desa (IPD) dibangun berdasarkan data hasil Pendataan Potensi Desa (Podes) tahun 2014. Terdapat 5 (lima) dimensi indeks pembangunan desa (IPD), yaitu Pelayanan Dasar, Kondisi Infrastruktur, Aksesibilitas/Transportasi, Pelayanan Publik, dan Penyelenggaraan Pemerintahan (Kemendesa PDPT, 2015).

IPD disusun untuk menunjukkan tingkat perkembangan pembangunan di suatu desa. Nilainya mempunyai rentang 0 s/d 100. Untuk memudahkan interpretasi, maka dilakukan pengelompokan desa menjadi 3 kategori yaitu desa mandiri, berkembang, dan tertinggal. Desa mandiri adalah desa yang mempunyai ketersediaan dan akses terhadap pelayanan dasar yang mencukupi, infrastruktur yang memadai, aksesibilitas/transportasi yang tidak sulit, pelayanan publik yang bagus, dan

penyelenggaraan pemerintahan yang sudah sangat baik. Pada teknisnya desa mandiri merupakan desa dengan nilai IPD lebih dari 75. Klasifikasi selanjutnya adalah desa berkembang, yaitu desa mempunyai ketersediaan dan akses terhadap pelayanan dasar, infrastruktur, aksesibilitas/transportasi, pelayanan publik, dan penyelenggaraan pemerintahan yang cukup memadai. Pada teknisnya desa mandiri merupakan desa dengan nilai IPD lebih dari 50 namun kurang dari sama dengan 75. Klasifikasi terakhir adalah desa tertinggal yaitu desa mempunyai ketersediaan dan akses terhadap pelayanan dasar, infrastruktur, aksesibilitas/transportasi, pelayanan publik, dan penyelenggaraan pemerintahan yang masih minim. Pada teknisnya desa mandiri merupakan desa dengan nilai IPD kurang dari sama dengan 50.

Davies Bouldin Index

Davies Bouldin Index (DBI) merupakan cara validasi *cluster* yang dibuat oleh D. L. Davies, salah satu metode evaluasi internal yang mengukur evaluasi *cluster* pada suatu metode pengelompokan yang didasarkan pada nilai kohesi dan separasi. Dalam suatu pengelompokan, kohesi didefinisikan sebagai jumlah dari kedekatan data terhadap *centroid* dari *cluster* yang diikuti. Sedangkan separasi didasarkan pada jarak antar *centroid* dari *clusternya*. DBI adalah fungsi rasio dari jumlah distribusi di dalam *cluster* untuk pemisahan antar *cluster* (S. Ding, 2014). Pengukuran menggunakan DBI bertujuan untuk memaksimalkan jarak *inter-cluster*. Dalam penelitian ini, DBI digunakan untuk melakukan validasi data pada setiap *cluster*.

METODE PENELITIAN

Dataset yang digunakan merupakan data kuantitatif berupa dokumentasi yang dikumpulkan oleh Badan Pusat Statistik (BPS) berdasarkan tingkat perkembangan desa (*village specific*) di Indonesia dengan menjadikan desa sebagai unit analisis. Data Potensi Desa dibangun dari data Sensus Potensi Desa Tahun 2014 dan digunakan sebagai rujukan indikator-indikator utama penyusun indek; serta data administrasi pemerintahan yang digunakan sebagai rujukan standar untuk jumlah desa terintegrasi di Indonesia. Dalam merumuskan konsep ataupun metode pengukurannya, Data Potensi Desa mempertimbangkan kekhasan persoalan desa-desa di Indonesia.

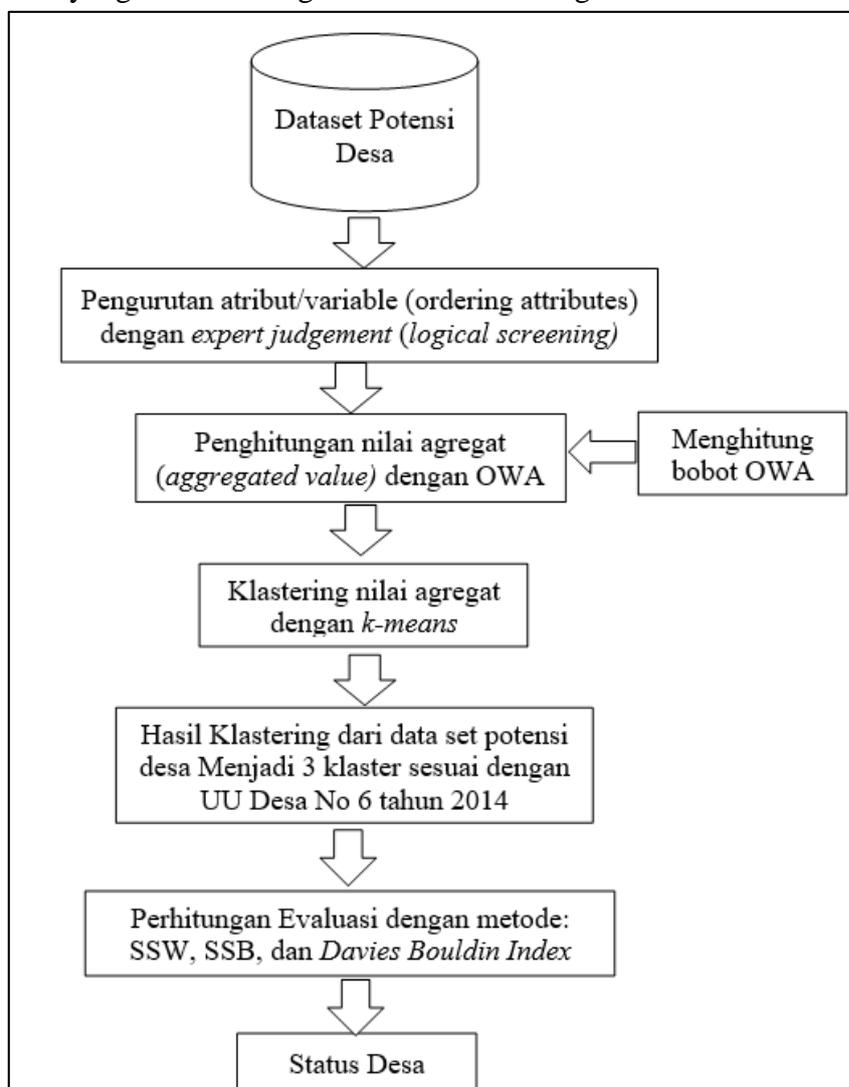
Pada Data Potensi Desa terdapat 5 dimensi yang terdiri dari pelayanan dasar, kondisi infrastruktur, aksesibilitas/transportasi, pelayanan publik, dan penyelenggaraan pemerintahan. Dari 5 dimensi tersebut terbagi lagi menjadi 12 variabel yaitu pelayanan pendidikan, pelayanan kesehatan, infrastruktur ekonomi, infrastruktur energi, infrastruktur kesehatan dan sanitasi, infrastruktur komunikasi dan informasi, sarana transportasi, aksesibilitas transportasi, kesehatan masyarakat, olah raga, kemandirian, dan kualitas sumber data manusia.

Data yang digunakan dalam penelitian ini merupakan data sekunder. Data sekunder ini merupakan data Indikator Pembangunan Desa yang diperoleh dari Bappenas dan BPS yang bekerja sama dengan Kementerian Desa PDTT, Pembangunan Daerah Tertinggal dan Transmigrasi Republik Indonesia yang berjumlah 74.093 data. *Dataset* Indikator Pembangunan Desa yang digunakan adalah 42 Indikator Potensi Desa yang berada pada 5 dimensi yaitu Pelayanan Dasar, Kondisi Infrastruktur, Aksesibilitas Transportasi, Pelayanan Publik dan Penyelenggaraan Pemerintahan. Atribut-atribut tersebut adalah:

1. Dimensi Pelayanan Dasar
2. Dimensi Kondisi Infrastruktur
3. Dimensi Aksesibilitas Transportasi
4. Dimensi Pelayanan Publik

5. Dimensi Penyelenggaraan Pemerintahan

Dari 42 atribut yang terdiri dari 5 dimensi tersebut akan digunakan untuk menentukan desa tersebut berada dalam status Desa Tertinggal, Desa Berkembang, Desa Mandiri (Chang J. R., 2006). Dalam penelitian ini metode yang diusulkan digambarkan dalam diagram berikut:



Gambar 1. Metode Penelitian

HASIL DAN PEMBAHASAN

Data yang digunakan dalam penelitian ini merupakan data sekunder. Data sekunder ini merupakan data Indikator Pembangunan Desa yang diperoleh dari Bappenas dan BPS yang bekerja sama dengan Kementerian Desa PDTT, Pembangunan Daerah Tertinggal dan Transmigrasi Republik Indonesia yang berjumlah 74.093 data. *Dataset* Indikator Pembangunan Desa yang digunakan adalah 5 dimensi yaitu Pelayanan Dasar, Kondisi Infrastruktur, Aksesibilitas Transportasi, Pelayanan Publik dan Penyelenggaraan Pemerintahan. Sampel yang digunakan adalah salah satu kabupaten di Provinsi Jawa Tengah yaitu Kabupaten Majalengka dengan 330 baris dengan 5 dimensi data potensi desa.

Tabel 1. Data Kabupaten Majalengka (Jawa Tengah)

Nama Desa	Pelayanan Dasar	Kondisi Infrastruktur	Transportasi dan Akses	Pelayanan Publik	Pemerintahan
AMPEL	57,39	63,87	53,11	68,37	79,22
ANDIR	65,35	47,32	85,76	74,82	75,57

ANGGRAWATI	71,92	46,19	74,65	62,55	92,70
ARGALINGGA	73,34	45,73	81,93	62,23	78,56
ARGAMUKTI	43,04	38,85	81,93	41,97	64,43
ARGASARI	61,45	43,70	77,44	62,23	92,92
BABAJURANG	57,67	49,25	63,21	49,65	79,22
BABAKANSARI	63,41	48,65	73,24	40,96	74,27
BAGJASARI	60,21	37,09	75,70	54,25	76,42
BALAGEDOG	64,09	43,34	76,05	49,65	76,22

Sumber: Data yang Diolah (2022)

Cluster Data Desa Menggunakan Algoritma K-Means Berbasis OWA Menghitung Bobot OWA

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. C. Shah and A. Jivani (2013), untuk menghitung bobot OWA dapat dihitung dengan persamaan 1 dan 2.

$$orness(W) = (1 - 1 \sum_{i=1}^n ((n - i) w_i)) \quad (1)$$

Menentukan nilai parameter situasi (α) yang digunakan adalah $\alpha > 0,5$.

$$orness(0,5) = (1 - 0,5)/(5 - 1) = 0,125$$

$$orness(0,6) = (1 - 0,6)/(5 - 1) = 0,1$$

$$orness(0,7) = (1 - 0,7)/(5 - 1) = 0,075$$

$$orness(0,8) = (1 - 0,8)/(5 - 1) = 0,05$$

$$orness(0,9) = (1 - 0,9)/(5 - 1) = 0,025$$

Sehingga: $f(a_1, a_2, \dots, a_n) = \sum_{i=1}^n w_i b_i$

b_i : elemen terbesar ke-i dari *aggregated object* { a_1, \dots, a_n } (2)

Tabel 2. Bobot OWA dengan *Expert Judgment*

	A1	A2	A3	A4	BI
Weighted for average 1	0,125	0,125	0,125	0,125	0,5
Weighted for average 2	0,1	0,1	0,1	0,1	0,6
Weighted for average 3	0,075	0,075	0,075	0,075	0,7
Weighted for average 4	0,05	0,05	0,05	0,05	0,8
Weighted for average 5	0,025	0,025	0,025	0,025	0,9

Sumber: Data yang Diolah (2022)

Menghitung Nilai Agregat dengan Operator OWA

Setiap *dataset* memiliki sejumlah atribut (n), jumlah atribut pada *dataset* Potensi Desa adalah 5. 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:

Menghitung nilai agregat (a_i)

$$w_1x_{i1}w_2x_{i2}w_3x_{i3} \cdots w_ix_{ij}w_nx_{in} \quad (3)$$

Hasil perhitungan bobot OWA pada tabel 2 kemudian dihitung dengan menggunakan rumus (3) pada tabel 1 yaitu tabel data Kabupaten Majalengka.

$$0,5 = (57,39 \times 0,125) + (63,87 \times 0,125) + (0,125 \times 0,125) + (68,37 + 0,125) + (79,22 \times 0,5) = 63,33$$

$$0,6 = (57,39 \times 0,1) + (63,87 \times 0,1) + (0,125 \times 0,1) + (68,37 + 0,1) + (79,22 \times 0,6) = 66,51$$

$$0,7 = (57,39 \times 0,075) + (63,87 \times 0,075) + (0,125 \times 0,075) + (68,37 + 0,075) + (79,22 \times 0,7) = 69,68$$

$$0,8 = (57,39 \times 0,05) + (63,87 \times 0,05) + (0,125 \times 0,05) + (68,37 + 0,05) + (79,22 \times 0,8) = 72,86$$

$$0,9 = (57,39 \times 0,025) + (63,87 \times 0,025) + (0,125 \times 0,025) + (68,37 + 0,025) + (79,22 \times 0,9) = 76,04$$

Tabel 3. Hasil *Ordering Attribute*

DESA	0,5	0,6	0,7	0,8	0,9
	Pelayanan Dasar	Kondisi Infrastruktur	Transportasi dan Akses	Pelayanan Publik	Pemerintahan
AMPEL	63,33	66,51	69,68	72,86	76,04
ANDIR	71,94	72,67	73,39	74,12	74,84
ANGGRAWATI	78,26	81,15	84,04	86,92	89,81
ARGALINGGA	72,19	73,46	74,74	76,01	77,29
ARGAMUKTI	57,94	59,24	60,54	61,84	63,13
ARGASARI	77,06	80,23	83,41	86,58	89,75
BABAJURANG	67,08	69,51	71,94	74,36	76,79
BABAKAN	73,87	74,21	74,55	74,89	75,23
BABAKANANYAR	69,76	69,81	69,86	69,90	69,95
BABAKANKAREO	72,87	75,46	78,06	80,65	83,25
BABAKANMANJETI	75,18	76,80	78,42	80,03	81,65
BABAKANSARI	65,42	67,19	68,96	70,73	72,50
BAGJASARI	66,62	68,58	70,54	72,50	74,46
BALAGEDOG	67,25	69,04	70,84	72,63	74,43

Sumber: Data yang Diolah (2022)

Mengklusterkan Nilai Agregat dengan K-Means Berbasis OWA

Selanjutnya langkah ini, adalah mengklusterkan nilai-nilai yang telah diagregasikan dengan K-Means dan OWA. Setelah data agregat OWA didapatkan langkah selanjutnya adalah menghitung nilai hasil agregat OWA ke dalam Algoritma K-Means. Langkah yang pertama:

Tabel 4. Nilai Atribut dengan OWA

No	Nama Desa	Pelayanan Dasar	Kondisi Infrastruktur	Transportasi dan Akses	Pelayanan Publik	Pemerintahan
1	AMPEL	63,33	66,51	69,68	72,86	76,04
2	ANDIR	71,94	72,67	73,39	74,12	74,84
3	ANGGRAWATI	78,26	81,15	84,04	86,92	89,81

4	ARGALINGGA	72,19	73,46	74,74	76,01	77,29
5	ARGAMUKTI	57,94	59,24	60,54	61,84	63,13

Sumber: Data yang Diolah (2022)

Kemudian data hasil agresi dengan OWA dihitung dengan *instance* yang telah dipilih sehingga akan menghasilkan *cluster* awal:

1. Perhitungan jarak *centroid* pada *cluster* 1 awal:

$$D(X1,Y1) = \text{SQRT}(((63,33-57,94)^2) + ((66,51-59,24)^2) + ((69,6860,54)^2) + ((72,86-61,84)^2) + ((76,04-63,13)^2)) = 29,39$$

2. Perhitungan jarak *centroid* pada *cluster* 2 awal:

$$D(X1,Y2) = \text{SQRT}(((63,33-62,83)^2) + ((66,51-63,71)^2) + ((69,8-64,59)^2) + ((72,86-65,47)^2) + ((76,04-66,35)^2)) = 20,50$$

3. Perhitungan jarak *centroid* pada *cluster* 3 awal:

$$D(X1,Y2) = \text{SQRT}(((63,33-74,65)^2) + ((66,51-75,52)^2) + ((69,8-76,39)^2) + ((72,86-77,26)^2) + ((76,04-78,13)^2)) = 20,50$$

Dari hasil perhitungan jarak *centroid* awal pada *cluster* 1, *cluster* 2, dan *cluster* 3 selanjutnya digunakan untuk mengelompokkan setiap *instance* data dari hasil perhitungan *centeroid* terdekat.

Tabel 5. *Cluster* Awal dari Iterasi 1

No	Nama Desa	C1	C2	C3	C1	C2	C3
1	AMPEL	29,39	20,50	6,86			*
2	ANDIR	28,80	19,68	6,72			*
4	ARGALINGGA	31,75	22,72	3,91			*
11	BABAKANMANJETI	39,99	31,00	5,11			*
-	-	-	-	-	-	-	-
328	WARINGIN	42,00	32,94	6,60			*
329	WERAGATI	52,61	43,62	17,43			*
330	WERASARI	45,19	36,28	10,55			*

Sumber: Data yang Diolah (2022)

Setelah ditemukan *cluster* selanjutnya data cluster akan digunakan untuk menentukan *centroid* baru dengan cara mengambil nilai rata-rata dari setiap indikator yang sesuai dengan *cluster*. Sehingga diperoleh *centroid* baru.

Tabel 6. *Centroid* Baru Iterasi 1 *Euclidean*

<i>Centroid</i> Baru	Pelayanan Dasar	Kondisi Infrastruktur	Transportasi dan Akses	Pelayanan Publik	Pemerintahan
C1	60,01	59,96	59,92	59,87	59,82
C2	65,11	65,41	65,71	66,01	66,31
C3	73,50	75,06	76,63	78,19	79,76

Sumber: Data yang Diolah (2022)

Dari tabel *centroid* baru, setiap data pada tabel hasil perhitungan OWA, diukur jaraknya terhadap *centroid* baru dengan *euclidean distance*.

1. Perhitungan jarak *centroid* pada *cluster* 1 *centroid* baru:

$$D(X1,Y1) = \text{SQRT}(((63,33-60,01)^2) + ((66,51-59,96)^2) + ((69,68-59,92)^2) + ((72,86-59,87)^2) + ((76,04-59,82)^2)) = 31,31$$

2. Perhitungan jarak *centroid* pada *cluster 2 centroid* baru:
 $D(X1,Y2) = \text{SQRT}(((63,33-65,11)^2) + ((66,51-65,41)^2) + ((69,8-65,71)^2) + ((72,86-66,01)^2) + ((76,04-66,31)^2)) = 18,44$
3. Perhitungan jarak *centroid* pada *cluster 3 centroid* baru:
 $D(X1,Y2) = \text{SQRT}(((63,33-73,50)^2) + ((66,51-75,06)^2) + ((69,8-76,63)^2) + ((72,86-78,19)^2) + ((76,04-79,76)^2)) = 6,70$

Dari hasil perhitungan jarak *centroid* awal pada *cluster 1*, *cluster 2*, dan *cluster 3* selanjutnya digunakan untuk mengelompokkan setiap *instance* data dari hasil perhitungan *centeroid* terdekat.

Tabel 7. *Cluster* Baru dari Iterasi 1 Perhitungan *Euclidean*

No	Nama Desa	C1	C2	C3	C1	C2	C3
1	AMPEL	31,31	18,44	6,70			*
2	ANDIR	30,23	17,23	7,71			*
11	BABAKANMANJETI	41,70	28,72	4,00			*
12	BABAKANSARI	21,03	8,64	17,16		*	
-	-	-	-	-	-	-	-
328	WARINGIN	43,54	30,54	6,19			*
329	WERAGATI	54,31	41,32	16,63			*
330	WERASARI	47,03	34,08	9,35			*

Sumber: Data yang Diolah (2022)

Mengulangi ke langkah 3 jika posisi *centroid* baru tidak konvergen. Pengecekan konvergensi dilakukan dengan membandingkan matriks *group assignment* pada iterasi sebelumnya dengan matriks *group assignment* pada iterasi yang sedang berjalan. Jika hasilnya sama maka algoritma K-Means *cluster* dinyatakan sudah konvergen dan proses berhenti. Tetapi jika belum konvergen maka perlu mengulangi iterasi ke langkah 3.

Hasil *Clustering* dengan K-Means Berbasis OWA

Dari hasil pengujian yang dihasilkan dengan menggunakan data potensi desa tahun 2014 menggunakan algoritma K-Means berbasis *Ordered Weighted Average* (OWA) untuk 3 *cluster* (Mandiri, Tertinggal, dan Maju) untuk 5 dimensi yaitu Pelayanan Dasar, Kondisi Infrastruktur, Aksesibilitas Transportasi, Pelayanan Publik, dan Penyelenggaraan Pemerintahan

Tabel. 8 *Centroid* Setiap *Cluster* dari Metode K-Means dan OWA

Atribut	Cluster 1	Cluster 2	Cluster 3
Pelayanan Dasar	60,01	65,11	73,50
Kondisi Infrastruktur	59,96	65,41	75,06
Transportasi dan Akses	59,92	65,71	76,63
Pelayanan Publik	59,87	66,01	78,19
Pemerintahan	59,82	66,31	79,76

Sumber: Data yang Diolah (2022)

Dari penggunaan K-Means pada data potensi desa dengan menghasilkan 3 *cluster* pada 5 dimensi, dengan *centroid* yang dihasilkan, diperoleh hasil *clustering* sebagai berikut:

Tabel 9. Hasil *Clustering* Menggunakan *K-Means* Berbasis OWA

No	Desa	Pelayanan Dasar	Kondisi Infrastruktur	Transportasi dan Akses	Pelayanan Publik	Pemerintahan	Cluster
310	TAJUR	65,80	65,35	64,91	64,47	64,03	cluster_1
323	WADOWETAN	52,91	52,65	52,38	52,12	51,85	cluster_1
72	CIKONENG	70,90	70,03	69,17	68,30	67,44	cluster_2
75	CIMANGGUHILIR	70,22	71,88	73,55	75,22	76,89	cluster_2
-	-	-	-	-	-	-	-
189	MAJASARI	77,57	81,32	85,08	88,84	92,59	cluster_3
190	MAJASUKA	80,82	83,24	85,66	88,08	90,50	cluster_3
184	LIGUNG LOR	83,25	85,19	87,12	89,05	90,99	cluster_3

Sumber: Data yang Diolah (2022)

Dari penggunaan algoritma *K-Means* berbasis OWA untuk mengelompokan Indek Potensi Desa tahun 2014 yang berjumlah 74.093 desa dengan studi kasus di Kabupaten Majalengka Provinsi Jawa Tengah dari 330 desa, diperoleh jumlah desa dari masing-masing *cluster* sebagai berikut:

1. *Cluster 1*: 98 desa tertinggal
2. *Cluster 2*: 132 desa berkembang
3. *Cluster 3*: 100 desa mandiri

Evaluasi dan Perbandingan dengan Metode Lain

Hasil perhitungan Data Potensi Desa dengan metode *K-Means* yang menghasilkan 3 *cluster* (Tertinggal, Maju, Mandiri) sesuai dengan UU Desa No. 6 tahun 2014, kemudian diuji hasil *clustering*nya menggunakan DBI. Selanjutnya data hasil perhitungan *K-Means* dengan OWA yang menghasilkan 3 *cluster* (Tertinggal, Maju, Mandiri) dan diuji hasil *clustering* dengan dihitung DBI nya. Kedua kriteria tersebut kemudian dibandingkan antar metode *clustering* yang digunakan, untuk mengetahui apakah performansi metode *K-Means* berbasis OWA yang digunakan lebih baik atau tidak dari metode *clustering K-Means* biasa.

Perhitungan Centroid *K-Means* dengan *Euclidean*

Dari hasil pengujian yang dihasilkan dengan menggunakan data potensi desa tahun 2014 menggunakan algoritma *K-Means* untuk 3 *cluster* (Mandiri, Tertinggal, dan Maju) untuk 5 dimensi yaitu Pelayanan Dasar, Kondisi Infrastruktur, Aksesibilitas Transportasi, Pelayanan Publik, dan Penyelenggaraan Pemerintahan.

Tabel 10. Centroid Setiap *Cluster* dari Metode *K-Means Euclidean*

Atribut	Cluster 1	Cluster 2	Cluster 3
Pelayanan Dasar	57,08	64,37	70,64
Kondisi Infrastruktur	39,77	50,40	53,21
Transportasi dan Akses	82,22	68,56	86,95
Pelayanan Publik	49,17	54,33	60,65
Pemerintahan	64,78	73,47	76,85

Sumber: Data yang Diolah (2022)

Dari penggunaan *K-Means* pada data potensi desa dengan menghasilkan 3 *cluster* pada 5 dimensi, dengan *centroid* yang dihasilkan, diperoleh hasil *clustering* sebagai berikut:

Tabel 11. Hasil *Clustering* Menggunakan *K-Means*

	Desa	Pelayanan Dasar	Kondisi Infrastruktur	Transportasi dan Akses	Pelayanan Publik	Pemerintahan	Cluster
310	TAJUR	77,23	50,49	88,53	55,78	63,58	cluster_1
314	TANJUNGSARI	64,78	48,56	90,98	62,23	57,16	cluster_1
317	TEGALSARI	63,40	50,37	88,51	56,10	77,27	cluster_1
286	SUKADANA	70,74	43,37	76,05	55,78	83,27	cluster_2
-	-	-	-	-	-	-	-
184	LIGUNG LOR	79,87	66,49	85,76	62,23	92,92	cluster_3
185	LOJI	78,47	48,75	93,39	74,82	77,29	cluster_3
188	MAJA UTARA	78,92	75,18	89,91	74,82	83,92	cluster_3

Sumber: Data yang Diolah (2022)

Dari penggunaan algoritma *K-Means* dengan metode perhitungan jarak *euclidean distance* untuk mengelompokan Indek Potensi Desa tahun 2014 yang berjumlah 74.093 desa dengan studi kasus di Kabupaten Majalengka Provinsi Jawa Tengah dari 330 desa, diperoleh jumlah desa dari masing-masing *cluster* sebagai berikut:

1. *Cluster* 1: 106 desa tertinggal
2. *Cluster* 2: 106 desa berkembang
3. *Cluster* 3: 118 desa mandiri

Perhitungan dengan *K-Means* dan OWA

Dari hasil pengujian yang dihasilkan dengan menggunakan data potensi desa tahun 2014 menggunakan algoritma *K-Means* berbasis *Ordered Weighted Average* (OWA) untuk 3 *cluster* (Mandiri, Tertinggal, dan Maju) untuk 5 dimensi yaitu Pelayanan Dasar, Kondisi Infrastruktur, Aksesibilitas Transportasi, Pelayanan Publik, dan Penyelenggaraan Pemerintahan

Tabel 12. *Centroid* Setiap *Cluster* dari Metode *K-Means* dan OWA

Atribut	Cluster 1	Cluster 2	Cluster 3
Pelayanan Dasar	60,01	65,11	73,50
Kondisi Infrastruktur	59,96	65,41	75,06
Transportasi dan Akses	59,92	65,71	76,63
Pelayanan Publik	59,87	66,01	78,19
Pemerintahan	59,82	66,31	79,76

Sumber: Data yang Diolah (2022)

Dari penggunaan *K-Means* pada data potensi desa dengan menghasilkan 3 *cluster* pada 5 dimensi, dengan *centroid* yang dihasilkan pada tabel 12, diperoleh hasil *clustering* sebagai berikut:

Tabel 13. Hasil *Clustering* Menggunakan *K-Means* Berbasis OWA

No	Desa	Pelayanan Dasar	Kondisi Infrastruktur	Transportasi dan Akses	Pelayanan Publik	Pemerintahan	Cluster
310	TAJUR	65,80	65,35	64,91	64,47	64,03	cluster_1
314	TANJUNGSARI	61,90	60,95	60,00	59,05	58,10	cluster_1
319	TEJAMULYA	59,98	60,22	60,47	60,72	60,96	cluster_1
310	TAJUR	65,80	65,35	64,91	64,47	64,03	cluster_1
66	CIKARACAK	65,54	67,80	70,06	72,32	74,58	cluster_2
67	CIKEUSAL	67,84	68,49	69,14	69,78	70,43	cluster_2

65	CIKALONG	73,72	72,98	72,23	71,49	70,74	cluster_2
-	-	-	-	-	-	-	-
184	LIGUNG LOR	83,25	85,19	87,12	89,05	90,99	cluster_3
187	MAJA SELATAN	82,34	82,40	82,46	82,53	82,59	cluster_3
190	MAJASUKA	80,82	83,24	85,66	88,08	90,50	cluster_3

Sumber: Data yang Diolah (2022)

Dari penggunaan algoritma *K-Means* berbasis OWA untuk mengelompokan Indek Potensi Desa tahun 2014 yang berjumlah 74.093 desa dengan studi kasus di Kabupaten Majalengka Provinsi Jawa Tengah dari 330 desa, diperoleh jumlah desa dari masing-masing *cluster* sebagai berikut:

1. *Cluster* 1: 98 desa tertinggal
2. *Cluster* 2: 132 desa berkembang
3. *Cluster* 3: 100 desa mandiri

Pengujian *Davies Bouldin Index*

Sum of square within cluster atau lebih di kenal dengan SSW merupakan metrik pada sebuah *cluster*. SSW lebih dikenal dengan jarak rata-rata antara instan data dengan *centroid* atau *medoid* dalam *cluster* yang sama. m_i adalah jumlah data yang ada pada *cluster* i . Hasil perhitungan Data Potensi Desa dengan metode *K-Means* yang menghasilkan 3 *cluster* (Tertinggal, Maju, Mandiri) kemudian diuji hasil *clustering*nya menggunakan DBI. Selanjutnya data hasil perhitungan *K-Means* dengan OWA yang menghasilkan 3 *cluster* (Tertinggal, Maju, Mandiri) dan diuji hasil *clustering* dengan dihitung DBI nya dapat dilihat pada:

Tabel 14. Hasil SSW

Algoritma K-Means	
<i>Cluster</i>	Nilai SSW <i>Euclidean</i>
<i>Cluster</i> 1	18,97
<i>Cluster</i> 2	18,45
<i>Cluster</i> 3	17,90
Algoritma K-Means + OWA	
<i>Cluster</i>	Nilai SSW OWA
<i>Cluster</i> 1	9,05
<i>Cluster</i> 2	14,33
<i>Cluster</i> 3	9,17

Sumber: Data yang Diolah (2022)

Sum of square between cluster atau lebih dikenal dengan SSB merupakan metrik separasi antar dua *cluster* dengan mengukur jarak (ketidakmiripan) antar *centroid* satu dengan yang lain seperti persamaan sebagai berikut:

Tabel 15. Metrik nilai SSB pada Algoritma *K-Means* dengan Metode *Euclidean distance*

SSB	<i>Cluster</i> 1	<i>Cluster</i> 2	<i>Cluster</i> 3
<i>Cluster</i> 1	0,00	21,33	25,78
<i>Cluster</i> 2	21,33	0,00	20,90

Cluster 3	25,78	25,78	0,00
------------------	-------	-------	------

Sumber: Data yang Diolah (2022)

Tabel 16. Metrik Nilai SSB pada Algoritma K-Means dengan Metode *Ordered Weighted Average*

SSB	Cluster	Cluster	Cluster
	1	2	3
Cluster 1	0,00	13,00	37,71
Cluster 2	13,00	0,00	24,74
Cluster 3	37,71	37,71	0,00

Sumber: Data yang Diolah (2022)

Adapun metrik Rasio (R_{ij}) dari hasil penelitian *clustering* data Indeks Potensi Desa tahun 2014 dengan metode pengukuran jarak dapat ditampilkan:

Tabel 17. Metrik Nilai Rasio pada Algoritma K-Means dengan Metode Euclidean

R	Cluster	Cluster	Cluster	Max
	1	2	3	
Cluster 1	0	1,75	1,43	1,75
Cluster 2	1,75	0	1,74	1,75
Cluster 3	1,43	1,41	0	1,43

Sumber: Data yang Diolah (2022)

Tabel 18. Metrik Nilai Rasio pada Algoritma K-Means dengan OWA

R	Cluster	Cluster	Cluster	Max
	1	2	3	
Cluster 1	0	1,80	0,48	1,80
Cluster 2	1,80	0	0,95	1,80
Cluster 3	0,48	0,62	0	0,62

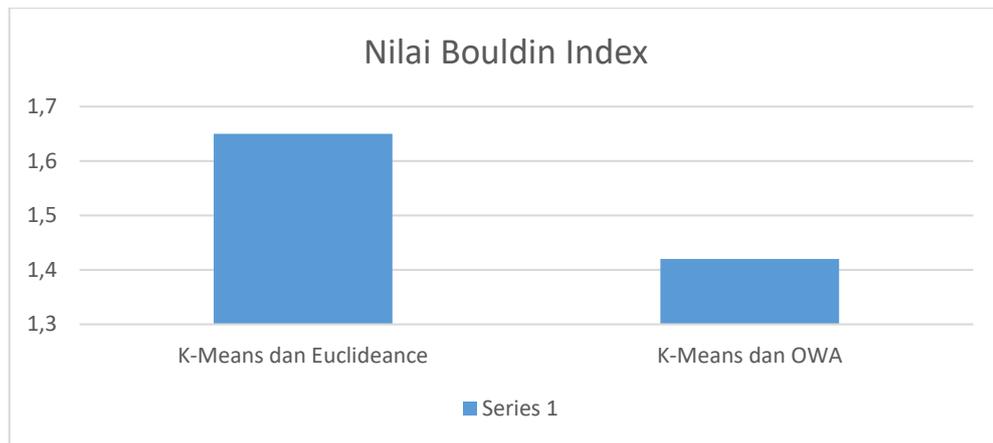
Sumber: Data yang Diolah (2022)

Dari perhitungan SSW, SSB, dan Rasio (R_{ij}) dapat diambil hasil akhir dari penelitian ini yaitu memperoleh nilai Indeks *Davies* dari komparasi 2 metode *clustering* berbasis *measure distance* menggunakan data Indeks Potensi Desa tahun 2014 dapat dilihat pada tabel 19.

Tabel 19. Daftar Nilai *Davies Bouldin Index*

Metode	Algoritma K-Means	Algoritma K-Means + OWA
	1,65	1,42

Sumber: Data yang Diolah (2022)



Gambar 2. Perbandingan Nilai *Bouldin Index*

Sumber: Data yang Diolah (2022)

Hasil *clustering* terhadap data yang telah divalidasi menggunakan DBI maka suatu *cluster* akan dianggap memiliki skema *clustering* yang optimal jika yang memiliki *Index Davies* minimal. Dari tabel 18 dapat diketahui bahwa nilai yang paling optimal dari perbandingan Algoritma *K-Means* dan *K-Means* menggunakan *Ordered Weighted Average* (OWA) adalah Algoritma *Clustering* menggunakan *K-Means* berbasis OWA dengan nilai 1,42.

SIMPULAN DAN SARAN

Simpulan

Pengelompokan data Indeks Potensi Desa tahun 2014 ke dalam 3 status desa dengan algoritma *K-Means* berbasis *Ordered Weighted Average* (OWA) metode perhitungan jarak menunjukkan bahwa metode *clustering K-Means* dengan *Ordered Weighted Average* (OWA) metode lebih baik daripada metode *K-Means* dengan *euclidean* yang memiliki nilai 1,65 dan hasil akhir penelitian ini diperoleh jumlah desa untuk setiap *cluster* yaitu *cluster* Desa Tertinggal sebanyak 98 desa, *cluster* Desa Berkembang sebanyak 132 desa, dan *cluster* Desa Mandiri sebanyak 100 desa. Pembobotan atribut *dataset* menggunakan metode *Ordered Weighted Average* (OWA) terbukti dapat meningkatkan kualitas *cluster* yang dihasilkan oleh algoritma *K-Means*. Peningkatan kualitas model dapat dilihat dari peningkatan hasil pencarian rata-rata jarak data ke *centroid* dan *Davies Bouldin Index* (DBI) mengalami peningkatan, dimana nilainya semakin kecil.

Saran

Dari pengelompokan data Potensi Desa tahun 2014 ke dalam 3 kelompok menggunakan algoritma *K-Means* dengan berbasis *Ordered Weighted Averaging* terbukti dapat meningkatkan kualitas *cluster* yang dihasilkan oleh algoritma *K-Means*. Peningkatan kualitas model dapat dilihat dari peningkatan hasil pencarian rata-rata jarak data ke *centroid* dan *Davies Bouldin Index* (DBI) mengalami peningkatan, dimana nilainya semakin kecil, namun ada hal yang perlu diperhatikan supaya menjadi lebih baik kedepannya. Dimana 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 kuesioner 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 *dataset*.

DAFTAR PUSTAKA

- Kementerian Desa, Pembangunan Daerah Tertinggal, dan Transmigrasi. 2014. *Indeks Pembangunan Desa 2015, Tantangan Pemenuhan Stabdar Pelayanan Minimum Desa*. Jakarta: Kementerian Desa, Pembangunan Daerah Tertinggal.
- Yager RR. 2004. *Modeling Prioritized Multi Criteria Decision Making*. IEEE Transactions on System, Man, and Cybernetics - Part B: Cybernetics, 23(6), 2396–2403.
- Cheng CH, JW Wang, dan MC Wu. 2009. *OWA-Weighted Based Clustering Method for Classification Problem*. Expert Systems with Application, 26, 4988-4995.
- L. Xu, C. Jiang, J. Wang, J. Yuan, and Y. Ren. 2014. "Information Security in Big Data: Privacy and Data Mining". IEEE Access: The Journal for Rapid Open Access Publishing, vol. 1, pp. 1149-1176, 9 October 2014.
- 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.
- Dr. Manju Bhawana. 2014. *Comparative Study of K-Means and Hierarchical Clustering Techniques*.
- N. Claypo and S. Jaiyen. 2015. "Opinion Mining for Thai Restaurant Reviews using K-Means Clustering and MRF Feature Selection". Knowledge and Smart Technology (KST), Chonburi.
- M. Verma, M. Srivastava, N. Chack, A. K. Diswar, and N. Gupta. 2012. "A Comparative Study of Various Clustering Algorithms in Data Mining". International Journal of Engineering Research and Applications (IJERA), vol. 2, no. 3, pp. 1379-1384.
- Jain AK. 2009. *Data Clustering: 50 Years Beyond K-Means, Pattern Recognition Letters*.
- Agusta Y. 2008. *K-Means*. Artikel Internet. Diakses pada 4 Maret 2010, dari <http://yudiagusta.wordpress.com/k-means>.
- Ahn BS. 2006. *On The Properties of OWA*.
- Huberty. 1994. *Applied Discriminant Analysis*. New York: Willey Interscience.
- Grandhi R. 2003. *Integration of Ordered Weighted Averaging Operators with Feed-Forward Neural Networks for Optimal Feature Subset Selection And Pattern Classification*. Tesis Master. Florida: Universitas Florida.
- Kementerian Desa, Pembangunan Daerah Tertinggal, dan Transmigrasi. 2014. "Indeks Pembangunan Desa 2015, Tantangan Pemenuhan Stabdar Pelayanan Minimum Desa". Jakarta: Kementerian Desa, Pembangunan Daerah Tertinggal, pp 20-25.

S. Ding, F. Wu, Q. Jun, H. Jia, and F. Jin. 2001. “*Research on Data Stream Clustering Algorithms*”. *Artificial Intelligence Review*, vol. 43, no. 4, pp. 593-600.

C. Shah and A. Jivani. 2013. “*Comparison of Data Mining Clustering Algorithms*”. Nirma University International Conference on Engineering.