



**PENENTUAN KLASSTER OPTIMAL PADA ALGORITMA K-MEDOIDS
DALAM PENGELOMPOKAN INDEKS PEMBANGUNAN MANUSIA
KABUPATEN/KOTA DI JAWA TENGAH**

**Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh Gelar Sarjana
Statistika**

SKRIPSI

**Oleh
RAHAYU AHADIATI PUTRI
B2A219049**

**PROGRAM STUDI S1 STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS MUHAMMADIYAH SEMARANG**

2021

PERSETUJUAN PEMBIMBING

Skripsi dengan Judul “**Penentuan Klaster Optimal pada Algoritma K-Medoids dalam Pengelompokan Indeks Pembangunan Manusia Kabupaten/Kota di Jawa Tengah**” yang disusun oleh:

Nama : Rahayu Ahadiati Putri

NIM : B2A219049

Program Studi : S1 Statistika

Telah disetujui oleh Dosen Pembimbing pada tanggal 10 Februari 2021.

Pembimbing Utama

Pembimbing Pendamping


Indah Manfaati Nur, M.Si.
NIK. 28.6.1026.221


Dr. Rochdi Wasono, M.Si
NIK. 28.6.1026.119

PENGESAHAN KELULUSAN

Skripsi dengan Judul “**Penentuan Kluster Optimal pada Algoritma K-Medoids dalam Pengelompokan Indeks Pembangunan Manusia Kabupaten/Kota di Jawa Tengah**” yang disusun oleh:

Nama : Rahayu Ahadiati Putri

NIM : B2A219049

Program Studi : S1 Statistika

Telah diajukan dalam Sidang Panitia Ujian Skripsi Program Sarjana, Universitas Muhammadiyah Semarang pada tanggal 10 Februari 2021.

Panitia Ujian
Ketua Tim Penguji


Tiani Wahyu Utami, M.Si.
NIK. 28.6.1026.341

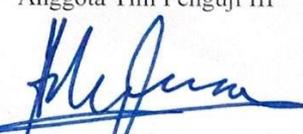
Anggota Tim Penguji I


Prizka Rismawati Arum, M. Stat
NIK. CP. 1026.071

Anggota Tim Penguji II


Indah Manfaati Nur, M.Si
NIK. 28.6.1026.221

Anggota Tim Penguji III


Dr. Rochdi Wasono, M.Si.
NIK. 28.6.1026.119

Mengetahui,
Ketua Program Studi


Indah Manfaati Nur, M.Si.
NIK. 28.6.1026.221

SURAT PERNYATAAN
PUBLIKASI KARYA ILMIAH

Nama : Rahayu Ahadiati Putri
NIM : B2A219049
Program Studi : FMIPA/S1 Statistika
Judul Skripsi : Penentuan Klaster Optimal pada Algoritma K-Medoids dalam Pengelompokan Indeks Pembangunan Manusia Kabupaten/Kota di Jawa Tengah

Dengan ini menyatakan bahwa saya menyetujui untuk:

1. Memberikan hak bebas royalti kepada perpustakaan UNIMUS atas penulisan Karya Ilmiah saya, demi pengembangan ilmu pengetahuan.
2. Memberikan hak menyimpan, mengalih mediakan/mengalih formatkan, mengelola dalam bentuk pangkalan data (*database*), mendistribusikannya, serta menampilkannya dalam bentuk *softcopy* untuk kepentingan akademis kepada perpustakaan UNIMUS tanpa perlu meminta ijin dari saya, selama tetap mencantumkan nama saya sebagai penulis/pencipta.
3. Bersedia dan menjamin untuk menanggung secara pribadi tanpa melibatkan pihak perpustakaan UNIMUS dari semua bentuk tuntutan hukum yang timbul atas pelanggaran hak cipta dalam Karya Ilmiah ini.

Demikian pernyataan ini saya buat dengan sesungguhnya dan semoga dapat digunakan sebagaimana mestinya.

Semarang, 10 Februari 2021

Yang Menyatakan,



Rahayu Ahadiati Putri
B2A219049

PENENTUAN KLASTER OPTIMAL PADA ALGORITMA K-MEDOIDS DALAM PENGELOMPOKAN INDEKS PEMBANGUNAN MANUSIA KABUPATEN/KOTA DI JAWA TENGAH

Rahayu Ahadiati Putri¹, Indah Manfaati Nur², Rochdi Wasono³

¹²³Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Muhammadiyah Semarang
Jl. Kedungmundu, Kec. Tembalang, Kota Semarang, Jawa Tengah 50273
E-Mail: rahayuahadiatip@gmail.com

ABSTRAK

Indeks Pembangunan Manusia (IPM) adalah indeks pencapaian kemampuan dasar pembangunan manusia yang dibangun melalui pendekatan tiga dimensi dasar yaitu umur panjang dan sehat, pengetahuan, dan kehidupan yang layak. Salah satu provinsi di Indonesia yang memiliki perkembangan yang cukup baik dalam pembangunan kependudukan adalah Provinsi Jawa Tengah. Untuk mengatasi hal ini pemerintah memerlukan gambaran umum kondisi sosial ekonomi kabupaten/kota di Jawa Tengah dengan pengelompokan di setiap bidang IPM. Salah satu metode yang dapat digunakan untuk pengelompokan variabel atau obyek adalah analisis cluster. Data yang digunakan dalam penelitian ini adalah data sekunder yaitu data Indeks Pembangunan Manusia (IPM) yang diperoleh dari publikasi BPS Provinsi Jawa Tengah 2019. Hasil penelitian gambaran umum Indikator Indeks Pembangunan Manusia (IPM) di Provinsi Jawa Tengah menunjukkan bahwa Rata-rata presentase Angka Partisipasi Sekolah (APS) tiap kabupaten/kota (X_1) di Jawa Tengah pada tahun 2019 sebesar 96,23 persen. Rata-rata Presentase Tamat SMP tiap kabupaten/kota (X_2) di Jawa Tengah pada tahun 2019 sebesar 19,38 persen. Rata-rata presentase Rumah Tangga dengan Akses Air Bersih tiap kabupaten/kota (X_3) di Jawa Tengah pada tahun 2019 sebesar 56,56 persen. Rata-rata Tingkat Partisipasi Angkatan Kerja tiap kabupaten/kota (X_4) di Jawa Tengah pada tahun 2019 sebesar 68,83 persen. Rata-rata Tingkat Pengangguran Terbuka (TPT) tiap kabupaten/kota (X_5) di Jawa Tengah pada tahun 2019 sebesar 4,44 persen. Sedangkan, rata-rata Produk Domestik Regional Bruto (PDRB) tiap kabupaten/kota (X_6) di Jawa Tengah pada tahun 2019 sebesar 5,48 persen. Hasil penentuan kluster optimal pada algoritma K-Medoids dalam pengelompokan Indeks Pembangunan Manusia kabupaten/kota di Jawa Tengah menunjukkan bahwa Metode K-medoids mampu memberikan nilai *Shilhouette* lebih dari 0,5 yang berarti struktur pengklasteran baik dengan 2 kluster. Kluster pertama, terdapat 27 kota/kabupaten dan Kluster kedua, terdapat 8 kota/kabupaten.

Kata Kunci: K-medoids, Indeks Pembangunan Manusia, Jawa Tengah

ABSTRACT

The Human Development Index (HDI) is an index of attaining the basic capabilities of human development which is built through a three-dimensional approach, namely long and healthy life, knowledge, and a decent life. One of the provinces in Indonesia that has a fairly good development in population development is Central Java Province. To overcome this, the government needs a general description of the socio-economic conditions of districts / cities in Central Java with groupings in each HDI sector. One method that can be used to group variables or objects is cluster analysis. The data used in this study are secondary data namely the Human Development Index (HDI) data obtained from the publication of the Central Java Province BPS 2019. The results of the research on an overview of the Human Development Index (HDI) Indicator in Central Java Province show that average percentage of School Participation Rate (APS) for each district / city (X_1) in Central Java in 2019 amounted to 96.23 percent. The average percentage of junior high school completion for each district / city (X_2) in Central Java in 2019 was 19.38 percent. The average percentage of households with access to clean water in each district / city (X_3) in Central Java in 2019 is 56.56 percent. The average labor force participation rate for each district / city (X_4) in Central Java in 2019 was 68.83 percent. The average Open Unemployment Rate (TPT) for each district / city (X_5) in Central Java in 2019 was 4.44 percent. Meanwhile, the average Gross Regional Domestic Product (PDRB) of each district / city (X_6) in Central Java in 2019 was 5.48 percent and the results of determining the optimal cluster in the K-Medoids algorithm in grouping the human development index of districts / cities in Java Middle shows that the K-

medoids Method is able to provide a Shilhouette value of more than 0.5 which means the clustering structure is good with 2 clusters. The first cluster, there are 27 cities / regencies aten and the second cluster, there are 8 cities / regencies..

Keywords: K-medoids, Human Development Index, Central Java

PENDAHULUAN

Pembangunan nasional Indonesia menempatkan rakyat sebagai titik sentral pembangunan. Untuk dapat ikut berpartisipasi dalam proses pembangunan, tentunya dibutuhkan masyarakat yang unggul dari segi kuantitas dan kualitas (Talaku, Leleury, & Talluta, 2017). Pembangunan manusia menjadi sentral dari paradigma pembangunan dewasa ini. Pembangunan manusia sendiri merupakan sebuah proses pembangunan yang bertujuan untuk menjadikan masyarakat memiliki lebih banyak pilihan, khususnya dalam hal pendidikan, kesehatan dan pendapatan. Salah satu tolak ukur pembangunan manusia suatu daerah adalah besaran angka Indeks Pembangunan Manusia (IPM) (BPS, 2019).

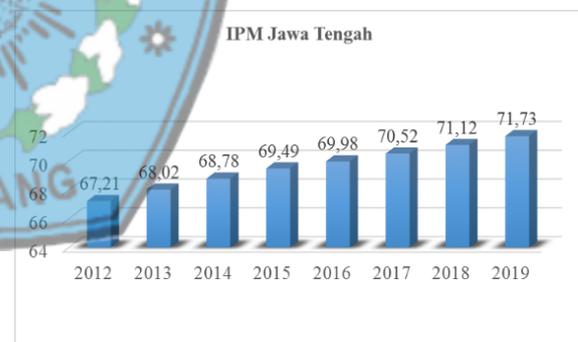
Indeks Pembangunan Manusia (IPM) adalah indeks pencapaian kemampuan dasar pembangunan manusia yang dibangun melalui pendekatan tiga dimensi dasar yaitu umur panjang dan sehat, pengetahuan, dan kehidupan yang layak. Untuk mengukur dimensi kesehatan, digunakan angka harapan hidup waktu lahir. Selanjutnya untuk mengukur dimensi pengetahuan digunakan gabungan indikator harapan lama sekolah dan rata-rata lama sekolah. Dimensi hidup layak digunakan indikator kemampuan daya beli masyarakat terhadap sejumlah kebutuhan pokok yang dilihat dari rata-rata besarnya pengeluaran per kapita disesuaikan (BPS, 2019).

Menurut Badan Pusat Statistik (BPS) klasifikasi Indeks Pembangunan Manusia (IPM) dibagi menjadi 4 kategori atau golongan, yaitu; Indeks Pembangunan Manusia (IPM) rendah jika < 60 , sedang $60 \leq IPM < 70$, tinggi $70 \leq IPM < 80$, dan ≥ 80 sangat tinggi (BPS, 2014). Karena pembangunan di Indonesia tidak merata maka Indeks Pembangunan Manusia (IPM) di wilayah-wilayah terutama kabupaten/kota sangatlah beragam (Darsyah, 2014).

Sejak tahun 2010, pembangunan manusia di Indonesia setiap tahunnya mengalami peningkatan yang cukup signifikan. Rata-rata pertumbuhan IPM di Indonesia adalah 0,89 persen per tahun. Dalam kurun waktu sembilan tahun terjadi kenaikan IPM

sebesar 5,39 poin. Perkembangan ini menunjukkan semakin membaiknya pembangunan manusia secara umum di Indonesia. Pada tahun 2019, IPM Indonesia tumbuh 0,74 persen atau meningkat sebesar 0,53 poin dibanding tahun sebelumnya, sehingga mencapai 71,92. Meskipun mengalami pertumbuhan, tetapi capaian ini lebih lambat dibandingkan pertumbuhan tahun sebelumnya yang mencapai 0,82 persen. Dengan capaian ini, status pembangunan manusia di Indonesia masih berada pada level tinggi (berada pada kisaran antara $70 \leq IPM < 80$).

Salah satu provinsi di Indonesia yang memiliki perkembangan yang cukup baik dalam pembangunan kependudukan adalah Provinsi Jawa Tengah. Hal ini ditunjukkan bahwa selama beberapa tahun terakhir pencapaian angka Indeks Pembangunan Manusia (IPM) Jawa Tengah terus meningkat. Berikut pertumbuhan IPM di Jawa Tengah pada tahun 2012-2019, dapat dilihat pada Gambar 1:



Gambar 1. IPM Jawa Tengah

Berdasarkan Gambar 1. dapat diketahui bahwa IPM di Jawa Tengah mengalami kenaikan setiap tahunnya. Pada tahun 2019 IPM di Jawa Tengah mengalami peningkatan sebesar 0,86 persen dibandingkan IPM Jawa Tengah pada tahun 2018. Namun, persentase indeks pembangunan manusia provinsi Jawa Tengah masih dibawah rata-rata IPM Indonesia yaitu sebesar 71,92 dan tinggi rendahnya IPM kabupaten/kota di Provinsi Jawa Tengah hanya ditunjukkan melalui indeks komposit, tetapi tidak ditunjukkan indikator yang dominan terhadap tinggi rendahnya peringkat IPM, maka pengelompokan wilayah kabupaten/kota di

Jawa Tengah juga perlu dilakukan sebagai bahan perencanaan dan evaluasi sasaran pemerintah (Purnamasari, Yasin, & Wuryandari, 2014). Untuk mengatasi hal ini pemerintah memerlukan gambaran umum kondisi sosial ekonomi kabupaten/kota di Jawa Tengah dengan pengelompokan di setiap bidang IPM. Salah satu metode yang dapat digunakan untuk pengelompokan variabel atau obyek adalah analisis cluster (Putri & Widodo, 2017).

Cluster adalah suatu *unsupervised learning* atau yang bersifat tanpa arahan, dimana sekelompok data langsung dikelompokkan berdasarkan tingkat kemiripannya tanpa dilakukan supervisi. Cluster juga dapat mengelompokkan data yang berdasarkan tingkat kemiripannya dan juga berdasarkan tingkat akurasinya (Han & Kamber, 1999). Tujuan utama dari penelitian ini adalah penggunaan data mining. Data mining adalah sekumpulan metode yang meliputi kegiatan pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar. Sehingga dalam penelitian ini fokus pada metode data mining dengan kasus pengelompokan analisis (clustering) kelompok yang diinginkan (Santosa, 2007).

Penelitian tentang pengelompokan Kabupaten/Kota berdasarkan IPM pernah dilakukan oleh (Talakua et al., 2017) untuk mengelompokkan Kabupaten/Kota di Provinsi Maluku Berdasarkan Indikator Indeks Pembangunan Manusia Tahun 2014 dengan metode *K-Means*. Hasil penelitian menunjukkan bahwa pengelompokan kabupaten/kota berdasarkan angka Indeks Pembangunan Manusia, Angka Harapan Hidup, Angka Melek Huruf, angka Rata-rata Lama Sekolah, dan angka Pengeluaran per Kapita kurang terjadi pada tahun 2014 menghasilkan 3 cluster atau kelompok.

Metode *K-Means* mampu mengelompokkan dokumen dalam jumlah yang besar dengan waktu komputasi yang cepat. Prinsip dari metode ini adalah mempartisi suatu koleksi dokumen menjadi beberapa cluster dan menentukan *centroid* (titik pusat awal) secara acak. Tetapi, metode *K-Means* ini juga memiliki kelemahan yaitu sensitif terhadap outlier (Wu & Kumar, 2009).

Selain *K-Means*, terdapat metode lain untuk mengelompokkan yaitu *K-Medoids*. Prinsip dari metode ini hampir sama dengan *K-Means* hanya saja *centroid* yang digunakan tidak mengambil nilai rata-rata dari objek dalam sebuah cluster sebagai titik acuan, namun menggunakan medoid

sebagai titik acuan yang merupakan objek dalam sebuah cluster yang paling terpusat. Kelebihan dari metode ini adalah termasuk metode yang fleksibel karena hampir dapat bekerja pada setiap jenis data matriks dan mampu mengelompokkan dokumen dalam jumlah yang besar dengan waktu komputasi yang cepat dan mampu mengatasi outlier (Ikasakti, 2017). Berdasarkan uraian di atas Penulis bermaksud mengadakan penelitian dengan judul “Penentuan Klaster Optimal pada Algoritma *K-Medoids* dalam Pengelompokan Indeks Pembangunan Manusia Kabupaten/Kota di Jawa Tengah”.

TINJAUAN PUSTAKA

Analisis Deskriptif

Statistika deskriptif adalah metode – metode yang berkaitan dengan pengumpulan dan penyajian suatu data sehingga memberikan informasi yang berguna (Walpole, 1995). Statistik deskriptif berfungsi untuk mendeskripsikan atau memberi gambaran terhadap objek yang diteliti melalui data sampel atau populasi, statistika deskriptif juga memberikan informasi hanya mengenai data yang dipunyai dan sama sekali tidak menarik inferensia atau kesimpulan (Sugiyono, 2007).

Analisis Cluster

Analisis Cluster adalah teknik pengelompokan objek berdasarkan karakteristik yang dimiliki oleh objek tersebut. Analisis cluster mengelompokkan objek yang paling mirip kedalam satu kelompok yang sama sehingga, karakteristik objek dalam satu cluster bersifat homogen sedangkan karakteristik objek antar cluster yang lain bersifat heterogen (Hair, et al. 2006).

Analisis Cluster terbagi menjadi dua, yaitu metode hierarki dan non hierarki. Berikut penjelasan untuk masing-masing jenis :

1. *Hierarchical Methods* (Metode Hierarki)

Metode ini biasa digunakan untuk individu yang tidak terlalu banyak, dan jumlah kelompok yang hendak dibentuk belum diketahui. Pengelompokan ini disajikan dalam bentuk dendogram, yang mirip dengan struktur diagram pohon atau tree diagram (Usman dan Sobari, 2013). Metode hierarki terbagi menjadi dua, yaitu *Agglomerative* atau Metode Penggabungan dan *Divisive* atau Metode Pembagian.

Metode *agglomerative* dimulai dengan setiap objek dalam satu kelompok yang terpisah.

Kelompok dibentuk dengan mengelompokkan objek ke dalam kelompok yang semakin membesar. Proses ini dilanjutkan sampai semua objek menjadi anggota dari suatu kelompok tunggal (*a single cluster*). Sedangkan metode divisive dimulai dari semua objek dikelompokkan menjadi kelompok tunggal. Kemudian kelompok tersebut dibagi atau dipisah, sampai setiap objek berada dalam kelompok yang terpisah (Sangga, 2018). Beberapa metode pengklasteran hierarki antara lain.

- a. *Single linkage* (jarak terdekat atau tautan tunggal), pengklasteran ini memberikan hasil bila kelompok-kelompok digabungkan menurut jarak antara anggota-anggota yang terdekat di antara dua kelompok (Prasetyo, 2012).
- b. *Complete linkage* (jarak terjauh atau tautan lengkap), pengklasteran ini terjadi bila kelompok-kelompok digabungkan menurut jarak antara anggota-anggota yang terjauh di antara dua kelompok (Prasetyo, 2012).
- c. *Average linkage* (jarak rata-rata atau tautan rata-rata), pengklasteran ini menggabungkan objek menurut jarak rata-rata pasangan-pasangan anggota masing-masing pada himpunan antara dua kelompok (Prasetyo, 2012).
- d. *Ward's method*, metode ini menggunakan perhitungan yang lengkap dan memaksimalkan homogenitas di dalam satu kelompok (Dillon dan Goldstein, 1984).

2. *Non Hierarchical Methods* (Metode Non Hierarki)

Metode Non Hierarki Analisis klaster dengan metode non hierarki merupakan metode klaster yang menentukan jumlah klaster secara manual.. Prosedur pada metode non hierarki dimulai dengan memilih sejumlah nilai cluster awal sesuai dengan jumlah yang diinginkan, teknik analisis klaster nonhierarki dirancang untuk mengelompokkan item bukan variabel menjadi kumpulan K klaster. Metode non hierarki ini meliputi metode *sequential threshold*, *parallel threshold*, dan *optimizing partitioning* (Gudono, 2011).

- a. *Metode Sequential Threshold*, pada metode Sequential Threshold dengan pemilihan satu cluster dan menempatkan semua obyek yang berada pada jarak terdekat ke dalam cluster tersebut.
- b. *Metode Parallel Threshold*, secara prinsip sama dengan metode sequential threshold, hanya saja pada metode parallel threshold

dilakukan pemilihan terhadap beberapa obyek awal cluster sekaligus dan kemudian melakukan penggabungan obyek ke dalamnya secara bersamaan.

- c. *Metode Optimization*, Metode Optimization hampir mirip dengan metode Sequential Threshold dan metode Parallel Threshold yang membedakan adalah metode optimization ini memungkinkan untuk menempatkan kembali obyek-obyek ke dalam cluster yang lebih dekat atau dengan melakukan optimasi pada penempatan obyek yang ditukar untuk cluster lainnya dengan pertimbangan kriteria optimasi. Ada dua masalah utama pendekatan non hierarki. Pertama, banyaknya cluster harus ditentukan terlebih dahulu. Kedua, pemilihan pusat cluster tidak menentu (pasti). Seterusnya, hasil pengclustering tergantung pada bagaimana pusat cluster dipilih. Banyak program yang dimulai dengan memilih kasus pertama k (k = jumlah cluster) sebagai pusat cluster awal. Jadi, hasil pengclustering tergantung pada observasi data. Kekurangan metode ini dapat dilakukan dengan cepat dan sangat bermanfaat kalau jumlah observasi besar (Simamora, 2005).

K-Medoids

K-Medoids atau *Partitioning Around Medoids* (PAM) adalah algoritma clustering yang mirip dengan *K-Means*. Algoritma *K-Medoids* ini diusulkan oleh Leonard Kaufmann dan Peter J. Rousseeuw pada tahun 1987 (Setiyawati, 2017). *K-Medoids* merupakan suatu algoritma yang digunakan untuk menemukan medoids didalam sebuah kelompok (*cluster*) yang merupakan titik pusat dari suatu kelompok (*cluster*).

Algoritma *K-Medoids* lebih baik dibandingkan dengan *K-Means* karena pada *K-Medoids* kita menemukan k sebagai objek yang representatif untuk meminimalkan jumlah ketidaksamaan objek data, sedangkan pada *K-Means* menggunakan jumlah jarak *euclidean distances* untuk objek data (Marlina D, 2008). Analisis cluster yang menggunakan prosedur partisi salah satunya adalah *Medoids*. *Medoids* adalah objek yang dianggap mewakili cluster sekaligus sebagai pusat cluster. *K-Medoids Clustering* hadir untuk mengatasi kelemahan *K-Means Clustering* yang sensitif terhadap *outlier* karena suatu objek dengan suatu nilai yang besar mungkin secara substansial menyimpang dari distribusi data (Han dan Kamber, 2006).

Ukuran Jarak

Pada metode cluster membutuhkan suatu ukuran seberapa mirip objek (jarak) yang akan di kelompokkan. Adapun Jarak yang biasa digunakan dalam analisis pengklasteran diantaranya ada *Euclidian* (Charles, 2015) sebagai berikut:

1. Jarak Euclidean

Jarak *Euclidean* adalah akar dari jumlah kuadrat perbedaan/deviasi di dalam nilai untuk setiap variabel (Supranto, 2004). Jarak *Euclidean* juga biasa disebut sebagai metode perhitungan jarak yang didasarkan pada ruang berdimensi terbatas bernilai riil (Kumari dan Bhagat, 2013). Adapun persamaan untuk menghitung jarak Euclidean sebagai berikut:

$$d_{ij} = \sqrt{\sum_{k=1}^n (X_{ik} - X_{jk})^2}$$

Keterangan :

d_{ij} : Tingkat perbedaan jarak antara objek i dan objek j

n : Jumlah variabel cluster

X_{ik} : Dari data variabel ke i pada variabel ke k

X_{jk} : Dari data variabel ke j pada variabel ke k

Validasi Cluster

Menurut Dalton, et al. (2009), salah satu pengukuran validasi internal adalah *Silhouette Coefisien*. *Silhouette Coefisien* mengevaluasi objek secara visual baik di dalam cluster maupun yang berada diluar, berdasarkan rentang nilai *silhouettenya*. Nilai *silhouette* objek mendefinisikan kedekatannya dengan kelompok sendiri dengan kelompok lain. Rumus rentang *silhouette* sebagai berikut:

$$SC = \frac{1}{N} \sum_{i=1}^N S(xi)$$

dengan

$$S(xi) = \frac{b - a}{\max[b, a]}$$

Keterangan :

a : rata – rata kemiripan antara objek ke- i dengan objek lainnya dalam satu cluster

b : minimum kemiripan antara objek ke- i dengan objek lain di masing – masing cluster

$S(xi)$: lebar *Silhouette* cluster ke – i

Menurut Lewis (2010), jika nilai *silhouette coefisien* (sci) < 0 maka objek berada dalam kelompok yang salah sedangkan jika nilai

silhouette coefisien (sci) > 0 maka objek berada dalam kelompok yang benar dan jika nilai *silhouette coefisien* (sci) = 0 maka objek berada diantara dua cluster sehingga objek tersebut belum dapat diputuskan masuk di dalam cluster yang mana. Ketika nilai rata-rata rentang *silhouette coefisien* objek secara keseluruhan (SC) > 0,51 menunjukkan bahwa struktur hasil pengelompokan sudah dianggap baik, sedangkan rata-rata rentang *silhouette coefisien* objek secara keseluruhan (SC) > 0,71 menunjukkan bahwa struktur hasil pengelompokan sangat baik.

METODE PENELITIAN

Sumber Data

Data yang digunakan dalam penelitian ini adalah data sekunder yaitu data Indikator Indeks Pembangunan Manusia (IPM) yang diperoleh dari publikasi BPS Provinsi Jawa Tengah 2019.

Variabel Penelitian

Variabel yang digunakan dalam penelitian ini adalah indikator Indeks Pembangunan Manusia (IPM) Provinsi Jawa Tengah. Definisi operasional dari 6 variabel tersebut disajikan dalam Tabel 3.1 berikut ini.:

Tabel 1. Variabel Penelitian

Indikator	Variabel	Keterangan	Skala
Pendidikan	X ₁	Angka Partisipasi Sekolah	Rasio
	X ₂	Presentase Tamat SMP	Rasio
Kesehatan	X ₃	Rumah Tangga dan Akses Air Bersih	Rasio
Kependudukan	X ₄	Tingkat Partisipasi Angkatan Kerja	Rasio
	X ₅	Tingkat Pengangguran Terbuka	Rasio
	X ₆	PDRB	Rasio

Analisis Data

Pengelompokan Indeks Pembangunan Manusia (IPM) Kabupaten/Kota di Jawa Tengah. Langkah – langkah penelitian sebagai berikut:

1. Pengumpulan data BPS tahun 2020 dengan pengampilan data dengan indikator Indeks Pembangunan Manusia (IPM), Pendidikan, Kesehatan, Kependudukan Tahun 2020.
2. Menginput data yang sudah ada.
3. Menentukan jumlah cluster dalam objek cluster.
4. Menentukan centroid sebagai awal *medoids*.
5. Menghitung kemiripan dan Menempatkan objek *non medoids* dengan *medoids*

menggunakan jarak *Euclidian*. Berikut adalah persamaan *Euclidian*;

$$d_{x_i y_i} = \sqrt{(x_i - E_x)^2 + (y_i - E_y)^2}$$

6. Menghitung Nilai S, apabila S kurang 0 maka mengulang kembali dimulai dari tahap 5 sampai mendapat S lebih besar dari 0.
7. Menampilkan output dari pengklasteran.
8. Melakukan validasi cluster untuk mengevaluasi cluster yang didapat sudah baik atau tidak.

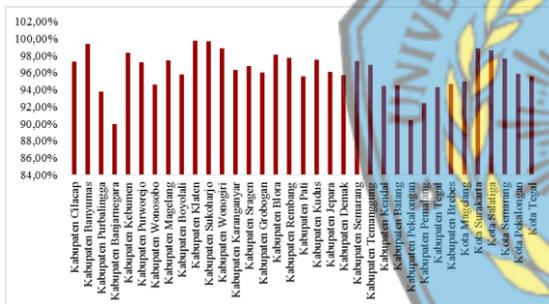
HASIL DAN PEMBAHASAN

Analisis Deskriptif

Analisis deskriptif digunakan untuk mengetahui gambaran umum tentang variabel penelitian.

Angka Partisipasi Sekolah (APS)

Indikator yang digunakan secara matematis oleh BPS dalam mengukur tingkat partisipasi disebut Angka Partisipasi Sekolah.

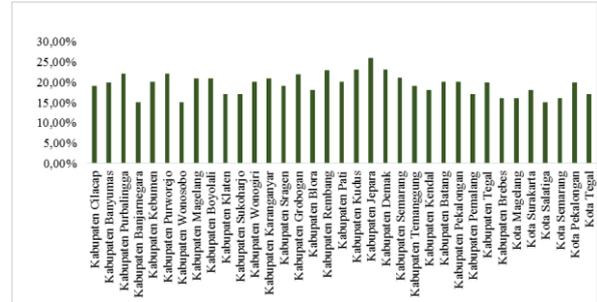


Gambar 2. Angka Partisipasi Sekolah

Gambar 2. menunjukkan bahwa rata-rata presentase Angka Partisipasi Sekolah (APS) tiap kabupaten/kota (X_1) di Jawa Tengah pada tahun 2019 sebesar 96,23 persen. Angka Partisipasi Sekolah (APS) terendah yaitu sebesar 89,98 persen adalah Kabupaten Banjarnegara, sedangkan Angka Partisipasi Sekolah (APS) tertinggi adalah Kabupaten Klaten yaitu sebesar 99,77 persen.

Rata-rata Presentase Tamat SMP

SMP (Sekolah Menengah Pertama) merupakan pendidikan formal pada jenjang pendidikan dasar.

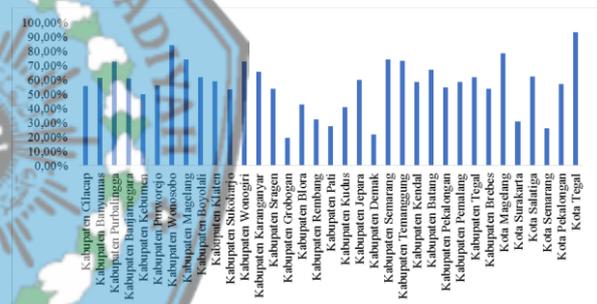


Gambar 3. Presentase Tamat SMP

Rata-rata Presentase Tamat SMP tiap kabupaten/kota (X_2) di Jawa Tengah pada tahun 2019 sebesar 19,38 persen. Peringkat terendah diduduki oleh Kota Salatiga dengan presentase 15,04 persen, sedangkan peringkat tertinggi diduduki oleh Kabupaten Jepara yaitu sebesar 26,00 persen.

Rumah Tangga dan Akses Air Bersih

Kebutuhan akan akses air bersih yang mudah dan murah merupakan hak azasi bagi setiap warga negara.

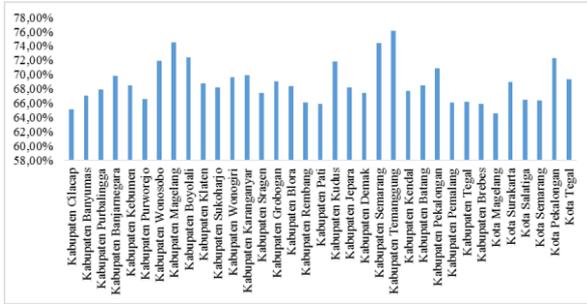


Gambar 4. Rumah Tangga dan Akses Air Bersih

Rata-rata presentase Rumah Tangga dan Akses Air Bersih tiap kabupaten/kota (X_3) di Jawa Tengah pada tahun 2019 sebesar 56,56 persen. Kabupaten Grobogan menempati peringkat terendah dengan presentase sebesar 19,47 persen. Sedangkan Kota Tegal mempunyai presentase tertinggi yaitu sebesar 93,45 persen.

Tingkat Partisipasi Angkatan Kerja

Tingkat partisipasi angkatan kerja sendiri merupakan suatu indikator ketenagakerjaan yang memberikan gambaran tentang penduduk yang aktif secara ekonomi dalam kegiatan sehari-hari merujuk pada suatu waktu dalam periode survei.

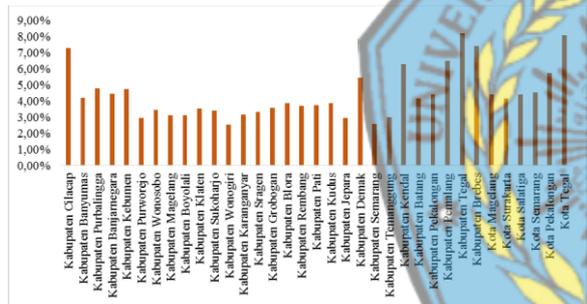


Gambar 5. Tingkat Partisipasi Angkatan Kerja

Rata-rata Tingkat Partisipasi Angkatan Kerja tiap kabupaten/kota (X_4) di Jawa Tengah pada tahun 2019 sebesar 68,83 persen. Kota Magelang menempati peringkat terendah dengan angka sebesar 64,62 persen, sedangkan Kabupaten Temanggung menduduki peringkat tertinggi yaitu sebesar 76,09 persen.

Tingkat Pengangguran Terbuka (TPT)

Pengangguran adalah seseorang yang tergolong dalam angkatan kerja dan secara aktif mencari pekerjaan tetapi belum memperolehnya.

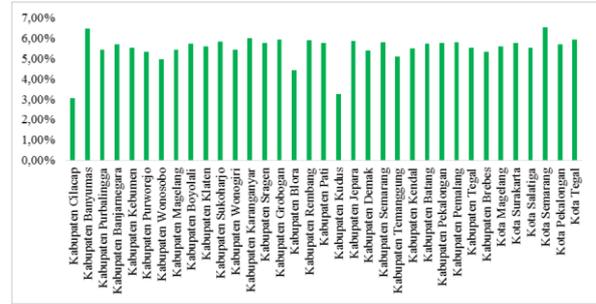


Gambar 6. Tingkat Pengangguran Terbuka (TPT)

Rata-rata Tingkat Pengangguran Terbuka (TPT) tiap kabupaten/kota (X_5) di Jawa Tengah pada tahun 2019 sebesar 4,44 persen. Peringkat terendah diduduki oleh Kabupaten Wonogiri dengan presentase 2,54 persen, sedangkan peringkat tertinggi diduduki oleh Kabupaten Tegal yaitu sebesar 8,21 persen.

Produk Domestik Regional Bruto (PDRB)

Produk Domestik Regional Bruto pada dasarnya merupakan jumlah nilai tambah yang dihasilkan oleh seluruh unit usaha.

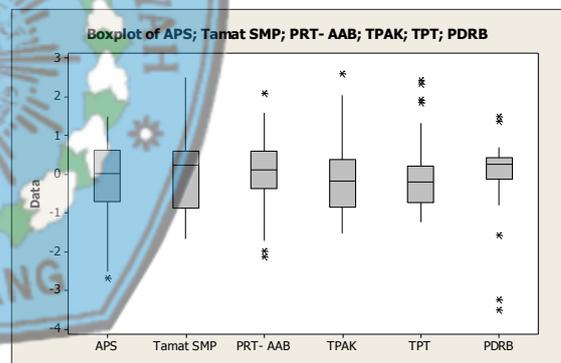


Gambar 7. Produk Domestik Regional Bruto (PDRB)

Rata-rata Produk Domestik Regional Bruto (PDRB) tiap kabupaten/kota (X_6) di Jawa Tengah pada tahun 2019 sebesar 5,48 persen. Kabupaten Cilacap menempati peringkat terendah dengan angka sebesar 3,05 persen, sedangkan Kota Semarang menduduki peringkat tertinggi yaitu sebesar 6,52 persen.

Pemeriksaan Outlier

Pengujian dilakukan untuk mendeteksi outlier yang terdapat pada dalam data. Pengujian outlier dalam penelitian ini menggunakan boxplot yang telah diolah dengan aplikasi Minitab.



Gambar 8. Boxplot Setiap Variabel

Berdasarkan Gambar 4.7 dapat dilihat bahwa variabel-variabel yang di gunakan mengandung outlier atau pencilan yaitu Angka Partisipasi Sekolah, Rumah Tangga dan Akses Air Bersih, Tingkat Partisipasi Angkatan Kerja, Tingkat Pengangguran Terbuka, dan PDRB.

Silhouette Coefficient

Silhouette Coefficient digunakan untuk mengukur validasi hasil clustering dan mengidentifikasi derajat kepemilikan setiap data yang ada di dalam cluster. Hasil perhitungan Silhouette Coefficient dapat dilihat pada Tabel 4.2:

Tabel 2. Hasil *Silhouette*

<i>k</i>	<i>Silhouette</i>
2	0.5871324
3	0.5234986
4	0.3823225
5	0.3682600
6	0.3008831
7	0.2680734
8	0.2868400
9	0.2787291
10	0.2783458

Nilai silhouette yang mendekati 1 atau sama dengan 1 disebut struktur data yang kuat atau jumlah cluster tersebut optimal. Berdasarkan Tabel 4.3 nilai silhouette terbesar adalah 0.5871324 yang terletak pada jumlah cluster sebanyak dua.

Pengelompokkan Kabupaten/Kota di Jawa Tengah dengan Metode K-Medoids

Algoritma K-Medoids mengelompokkan Kabupaten/Kota di Jawa Tengah menggunakan aplikasi *R Studio*. Berdasarkan *Silhouette Coefficient* didapatkan jumlah klaster optimal yang digunakan adalah 2. Hasil pengelompokan yang diperoleh dari aplikasi *R Studio* kemudian ditampilkan dalam bentuk peta menggunakan aplikasi GeoDa versi 1.12. Berikut hasil yang diperoleh:

Tabel 3. Hasil Perhitungan Centroid K-Medoids

Centroid	Klaster 1	Klaster 2
X ₁	96.33	95.54
X ₂	21.00	20.03
X ₃	65.76	27.62
X ₄	69.92	65.9
X ₅	3.15	3.74
X ₆	5.98	5.74



Gambar 9. Hasil Clustering dengan Metode K-Medoids

Distribusi peta hasil clustering menggunakan Algoritma *K-Medoids* dengan jumlah cluster 2 di ilustrasikan pada Gambar 4.8, dengan peta berwarna coklat muda menunjukkan kabupaten/kota yang masuk ke dalam cluster 1 dan berwarna coklat tua menunjukkan kabupaten/kota yang masuk ke dalam cluster 2. Hasil lebih lengkap ilustrasi peta diatas dapat dilihat di Tabel 4. di bawah ini:

Tabel 4. Keanggotaan Setiap Cluster dengan K-Medoids

Cluster 1	Cluster 2
Kabupaten Cilacap	Kabupaten Jepara
Kabupaten Banyumas	Kabupaten Semarang
Kabupaten Purbalingga	Kabupaten Temanggung
Kabupaten Banjarnegara	Kabupaten Kendal
Kabupaten Kebumen	Kabupaten Batang
Kabupaten Purworejo	Kabupaten Pekalongan
Kabupaten Wonosobo	Kabupaten Pemalang
Kabupaten Magelang	Kabupaten Tegal
Kabupaten Boyolali	Kabupaten Brebes
Kabupaten Klaten	Kabupaten Kota Magelang
Kabupaten Sukoharjo	Kabupaten Kota Salatiga
Kabupaten Wonogiri	Kabupaten Kota
Kabupaten Karanganyar	Kabupaten Kota Tegal
Kabupaten Sragen	

Berdasarkan klaster Kabupaten/Kota yang terbentuk diatas maka dapat diketahui bahwa:

1. Klaster pertama terdiri dari 27 kabupaten/kota yaitu Cilacap, Banyumas, Purbalingga, Banjarnegara, Kebumen, Purworejo, Wonosobo, Magelang, Boyolali, Klaten, Sukoharjo, Wonogiri, Karanganyar, Sragen, Jepara, Semarang, Temanggung, Kendal, Batang, Pekalongan, Pemalang, Tegal, Brebes, Kota Magelang, Kota Salatiga, Kota Pekalongan, Kota Tegal. Klaster pertama ini

memiliki karakteristik jumlah angka partisipasi sekolah rendah, jumlah presentase tamat smp rendah, jumlah rumah tangga dan akses air bersih rendah, jumlah tingkat partisipasi angkatan kerja rendah, jumlah tingkat pengangguran terbuka rendah, jumlah PDRB rendah.

2. Klaster kedua terdiri dari 8 kabupaten/kota yaitu Grobogan, Blora, Rembang, Pati, Kudus, Demak, Kota Surakarta, Kota Semarang. Klaster kedua ini memiliki karakteristik jumlah angka partisipasi sekolah tinggi, jumlah presentase tamat smp tinggi, jumlah rumah tangga dan akses air bersih tinggi, jumlah tingkat partisipasi angkatan kerja tinggi, jumlah tingkat pengangguran terbuka tinggi, jumlah PDRB tinggi

KESIMPULAN DAN SARAN

Kesimpulan

Hasil penerapan Penentuan Klaster Optimal Pada Algoritma K-Medoids Dalam Pengelompokan Indeks Pembangunan Manusia Kabupaten/Kota Di Jawa Tengah.

1. Rata-rata presentase Angka Partisipasi Sekolah (APS) tiap kabupaten/kota (X_1) di Jawa Tengah pada tahun 2019 sebesar 96,23 persen. Rata-rata Presentase Tamat SMP tiap kabupaten/kota (X_2) di Jawa Tengah pada tahun 2019 sebesar 19,38 persen. Rata-rata presentase Rumah Tangga dengan Akses Air Bersih tiap kabupaten/kota (X_3) di Jawa Tengah pada tahun 2019 sebesar 56,56 persen. Rata-rata Tingkat Partisipasi Angkatan Kerja tiap kabupaten/kota (X_4) di Jawa Tengah pada tahun 2019 sebesar 68,83 persen. Rata-rata Tingkat Pengangguran Terbuka (TPT) tiap kabupaten/kota (X_5) di Jawa Tengah pada tahun 2019 sebesar 4,44 persen. Sedangkan, rata-rata Produk Domestik Regional Bruto (PDRB) tiap kabupaten/kota (X_6) di Jawa Tengah pada tahun 2019 sebesar 5,48 persen.
2. Metode K-medoids mampu memberikan nilai *Shilhouette* lebih dari 0,5 yang berarti struktur pengklasteran baik dengan 2 klaster. Pengelompokan menggunakan Algoritma K-Medoids sebagai berikut:
 - a. Klaster pertama, terdapat 27 Kota/Kabupaten yaitu Cilacap, Banyumas, Kebumen Purbalingga, Banjarnegara, Purworejo, Wonosobo, Magelang, Boyolali, Klaten, Sukoharjo, Wonogiri,

Karanganyar, Sragen, Jepara, Semarang, Temanggung, Kendal, Batang, Pekalongan, Pemalang, Tegal, Brebes, Kota Magelang, Kota Salatiga, Kota Pekalongan, Kota Tegal. Klaster pertama ini memiliki karakteristik Angka Partisipasi Sekolah tinggi, Presentase Tamat SMP rendah, Rumah Tangga dan Akses Air Bersih tinggi, Tingkat Partisipasi Angkatan Kerja tinggi, Tingkat Pengangguran Terbuka rendah, PDRB rendah.

- b. Klaster kedua, terdapat 8 Kota/Kabupaten yaitu Grobogan, Blora, Rembang, Pati, Kudus, Demak, Kota Surakarta, Kota Semarang. Klaster kedua ini memiliki karakteristik Angka Partisipasi Sekolah tinggi, Presentase Tamat SMP rendah, Rumah Tangga dan Akses Air Bersih rendah, Tingkat Partisipasi Angkatan Kerja rendah, Tingkat Pengangguran Terbuka rendah, PDRB rendah.

Saran

Berdasarkan kesimpulan di atas, maka pada penelitian selanjutnya untuk mengatasi masalah pada clustering dapat dilakukan pembobotan dalam perhitungan jarak, baik jarak *Euclidean* maupun jarak *Manhattan*.

DAFTAR PUSTAKA

- BPS. (2019). *Indeks pembangunan manusia 2019*.
- Charles E. Mongi, "Penggunaan Analisis Two Step Clustering Untuk Data Campuran", *JdC*, Vol. 4, No. 1, (2015), hal. 10-11.
- Dalton, et.al.2009. *Clustering Algorithms: On Learning, Validation, Performance, and Applications to Genomics* <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2766793/>. Diakses 15 Februari 2015.
- Darsyah, M. Y. (2014). *Klasifikasi Indeks Pembangunan Manusia (IPM) dengan Pendekatan K-Nearest Neighbor (K-NN)*. *Seminar Nasional Pendidikan, Sains Dan Teknologi*, 29–35.
- Dillon, W.R., dan Goldstein, M. 1984. *Multivariate Analysis Second Edition*. London: Heinemann Educational Books.
- Gudono. (2011). *Analisis Data Multivariat Edisi Pertama*. Yogyakarta: BPFE.
- Hair, et al.2006. *Multivariate data analysis*. edisi

- ke-6.pearson education.
- Han, J., & Kamber, M. (1999). *Data Mining: Concepts and Techniques* (2nd ed.). San Fransisco: Morga Kaufmann Publishers.
- Han, J. dan M. Kamber. 2006. *Data Mining Concepts and Techniques Second Edition*. San Francisco: Morgan Kaufmann.
- Ikasakti, N. A. (2017). *Efektivitas Algoritma K-Means, K-Medoids, dan Optimalisasi Jumlah Cluster untuk Modifikasi Algoritma K-Medoids pada Kumpulan Data Besar*.
- Lewis, Paul D. 2010. *R For Medicine and Biology*. Jones and Bartlett Publisher
- Marlina, D. *et al.* 'Implementasi Algoritma K-Medoids dan K-Means untuk Pengelompokan Wilayah Sebaran Cacat pada Anak', *Jurnal CoreIT: Jurnal Hasil Penelitian Ilmu Komputer dan Teknologi Informasi*, 4(2), p. 64. doi: 10.24014/coreit.v4i2.4498. 2018.
- Purnamasari, S. ., Yasin, H., & Wuryandari, T. (2014). *Pemilihan Cluster Optimum pada Fuzzy C-Means*. G (Vol. 3).
- Putri, R., & Widodo, E. (2017). *Analisis Klaster Hierarki Untuk Pengelompokan Kabupaten / Kota Di Jawa Tengah Berdasarkan Indikator Indeks Pembangunan Manusia (IPM) Tahun 2015*, 1(1), 229–239.
- Prasetyo, Lambang Agung. 2012. *Aplikasi Sistem Informasi Geografis pada Kerusakan Bangunan Akibat Erupsi Merapi Tahun 2010 di Kabupaten Sleman Daerah Istimewa Yogyakarta Berbasis WEB*. Skripsi Mahasiswa Statistika, FMIPA, UII.
- Sangga, V. A. P. (2018). *Perbandingan K-Means dengan K-Medoids dalam Pengelompokan Komoditas Peternakan di Jawa Tengah pada 2015*. *Journal of Visual Languages & Computing* (Vol. 11).
- Santosa, B. (2007). *Data Mining: Teknik Pemanfaatan Data untuk Keperluan Bisnis* (1st ed.). Yogyakarta: Graha I.
- Setiyawati, W. (2017). *Implementasi Algoritma Partitioning Around Medoids (PAM) untuk Pengelompokan Sekolah Menengah Atas di DIY berdasarkan Nilai Daya Serap Ujian Nasional*. Yogyakarta: (Skripsi) Jurusan Teknik Informatika Universitas Sanata Dharma.)
- Sugiyono. 2007. *Metode Penelitian Kuantitatif Kualitatif dan R&D*. Bandung: Alfabeta.
- Supranto, Johannes. 2004. *Analisis Multivariat : Arti dan Interpretasi*. Jakarta : PT. Rineka Cipta.
- Talakua, M. W., Leleury, Z. A., & Talluta, A. W. (2017). *Analisis Cluster dengan Menggunakan Metode K-Means untuk Pengelompokan Kabupaten/Kota di Provinsi Maluku berdasarkan Indikator Indeks Pembangunan Manusia Tahun 2014*. *Ilmu Matematika Dan Terapan*, 11, 119–128.
- Usman, Hardius dan Sobari, Nurdin. 2013. *Aplikasi Teknik Multivariat Untuk Riset Pemasaran*. Jakarta : PT. Raja Grafindo Persada.
- Walpole, R.E., dan Myers, R.H. 1995. *Ilmu Peluang dan Statistika untuk Insinyur dan Ilmuwan Edisi ke-4*. Bandung : Penerbit ITB.
- Wu, X., & Kumar, V. (2009). *The Top Ten Algorithms in Data Mining*. London: CRC Press