

Implementasi Algoritma ST-DBSCAN dan K-MEANS Untuk Pengelompokan Indeks Pembangunan Manusia Kabupaten/Kota Pulau Jawa Tahun 2014-2016 Berbasis Web Di Badan Pusat Statistik

M. Ivan Putra Eriansya¹⁾, Muhammad Syafrullah²⁾

Program Studi Anda, Fakultas Teknologi Informasi, Universitas Budi Luhur
Jl. Raya Ciledug, Petukangan Utara, Kebayoran Lama, Jakarta Selatan 12260
Telp. (021) 5853753, Fax. (021) 5866369
E-mail : putra.ivan90@gmail.com

ABSTRAK

Pembangunan manusia sebagai ukuran kinerja pembangunan secara keseluruhan diukur oleh suatu nilai tunggal, yaitu Indeks Pembangunan Manusia (IPM). IPM merupakan indikator penting untuk mengukur keberhasilan pembangunan kualitas manusia di suatu daerah (BPS 2016). Bila dibandingkan dengan wilayah lain di Indonesia, Pulau Jawa memiliki peningkatan IPM terbesar pada periode 2014-2016. Dengan memiliki peningkatan IPM terbesar, Pulau Jawa menarik untuk diteliti persebarannya. Salah satu cara untuk mendapatkan pola persebaran kondisi IPM di Pulau Jawa adalah dengan menggunakan teknik pengelompokan atau clustering. Pada penelitian ini, metode yang akan digunakan untuk pengelompokan data nilai IPM kabupaten/kota di Pulau Jawa adalah ST-DBSCAN dan K-Means. Berdasarkan penelitian ini, hasil pengelompokan dari metode K-Means yang menggunakan parameter $k=4$ menghasilkan nilai Silhouette Coefficient yang lebih tinggi, yaitu sebesar 0.3444 bila dibandingkan dengan metode ST-DBSCAN yang menggunakan parameter nilai $Eps1=0.7$, $Eps2=3$, $MinPts=7$ dan $\Delta\epsilon = 2$, yaitu sebesar 0.1521. Waktu proses metode K-Means juga lebih cepat dengan nilai 0.4 detik bila dibandingkan dengan metode ST-DBSCAN yang memiliki waktu proses 3.4 detik.

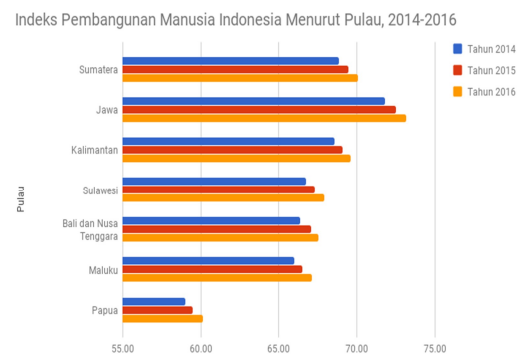
Kata kunci : Data Mining, Clustering, ST-DBSCAN, K-Means, Silhouette Coefficient

1. PENDAHULUAN

Pembangunan manusia sebagai ukuran kinerja pembangunan secara keseluruhan diukur oleh satu nilai tunggal, yaitu Indeks Pembangunan Manusia (IPM). IPM merupakan indikator penting untuk mengukur keberhasilan pembangunan kualitas manusia di suatu daerah [1]. Konsep IPM pertama kali diperkenalkan oleh United Nations Development Programme (UNDP) pada tahun 1990. Konsep ini diukur melalui tiga dimensi dasar manusia, yaitu umur panjang dan sehat, pengetahuan, dan standar hidup layak. Dimensi umur panjang dan sehat diwakili oleh indikator harapan hidup saat lahir. Dimensi pengetahuan diwakili oleh indikator harapan lama sekolah dan rata-rata sekolah. Dimensi standar hidup layak diwakili oleh pengeluaran per kapita.

Berdasarkan gambar 1, rata-rata IPM di Pulau Jawa memiliki peningkatan yang paling tinggi diantara pulau-pulau lain. Pada tahun 2014, IPM Pulau Jawa sebesar 71.8. Di tahun 2015, IPM Pulau

Jawa sebesar 72.46. Kemudian pada tahun 2016, IPM Pulau Jawa sebesar 73.12.



Gambar 1: Indeks Pembangunan Manusia Menurut Pulau, 2014-2016

Berdasarkan data tersebut, didapatkan bahwa IPM di Pulau Jawa dari tahun 2014-2016 terjadi peningkatan sebesar 1.32. Hal ini membuat Pulau Jawa memiliki peningkatan IPM terbesar bila

dibandingkan dengan pulau-pulau lain. Dengan memiliki peningkatan IPM terbesar, Pulau Jawa menarik untuk diteliti persebarannya.

Pada penelitian ini, metode yang akan digunakan untuk pengelompokan data nilai IPM kabupaten/kota di Pulau Jawa adalah ST-DBSCAN dan K-Means. ST-DBSCAN adalah algoritma pengelompokan berbasis kepadatan (*density*) yang memiliki kemampuan untuk mencari pengelompokan berdasarkan data spasial, data temporal dan data non-spasial dari suatu objek. K-means adalah algoritma pengelompokan partisi yang digunakan untuk membagi objek data yang diberikan ke dalam k-cluster yang berbeda melalui metode pengulangan [2].

Berdasarkan penelusuran pustaka, Budiman (2016) melakukan perbandingan pengelompokan metode K-means dan DBSCAN pada data rumah kost mahasiswa di kelurahan Tembalang, Semarang [3]. Maitry & Vaghela (2014) melakukan analisis perbandingan metode pengelompokan berbasis kepadatan seperti DBSCAN, FDBSCAN, ODBSCAN, VDBSCAN, dan ST-DBSCAN untuk pengelompokan data spasial [4]. Belum ditemukan penggunaan metode K-means untuk pengelompokan data nilai IPM di kabupaten/kota. Namun, Penelitian terkait penggunaan algoritma ST-DBSCAN untuk mencari pola persebaran nilai IPM sudah beberapa kali dilakukan. Penelitian Afifah (2014) yang berjudul Clustering Data Indeks Pembangunan Manusia (IPM) Pulau Jawa Menggunakan Algoritme ST-DBSCAN Dan Bahasa Pemrograman R menghasilkan 9 cluster dan 47 titik noise [5]. Penelitian Trisnaningtyas (2014) yang berjudul Pengelompokan Data Indeks Pembangunan Manusia Di Pulau Jawa Dengan Algoritme ST-DBSCAN Dan Bahasa Pemrograman Python menghasilkan 4 cluster dan 18 titik noise [6]. Kedua penelitian tersebut memiliki beberapa kesamaan, seperti pengelompokan hanya dilakukan di Pulau Jawa dan penggunaan data nilai IPM hanya menggunakan satu jenis data temporal, yaitu tahun 2012.

2. LANDASAN TEORI

2.1. Indeks Pembangunan Manusia

Indeks Pembangunan Manusia (IPM) adalah suatu pengukuran untuk mengetahui pencapaian pembangunan manusia berbasis sejumlah komponen dasar kualitas hidup. Sebagai ukuran kualitas hidup, IPM dibangun melalui pendekatan tiga dimensi. Ketiga dimensi itu adalah umur panjang dan sehat, pengetahuan, dan standar hidup layak. Dimensi umur

Harapan Hidup (AHH) saat lahir. Dimensi pengetahuan diwakili oleh indikator Harapan Lama Sekolah (HLS) dan Rata-rata Lama Sekolah (RLS). Sementara itu, dimensi standar hidup layak hidup diwakili oleh pengeluaran per kapita.

Konsep IPM pertama kali diperkenalkan oleh *United Nations Development Programme* (UNDP) pada tahun 1990. Sampai dengan tahun 2017, UNDP telah beberapa kali melakukan revisi metode penghitungan IPM. Revisi yang cukup besar dilakukan pada tahun 2010. UNDP menyebut revisi itu dengan era baru pembangunan manusia. UNDP memperkenalkan dua indikator baru yang sekaligus menggantikan dua indikator metode lama. Indikator harapan lama sekolah menggantikan indikator melek huruf, sementara Pendapatan Nasional Bruto (PNB) per kapita menggantikan Produk Domestik Bruto (PDB) per kapita.

2.2. Analisis Cluster

Analisis *Cluster* atau clustering adalah suatu proses membagi sekumpulan data menjadi beberapa kelompok atau *cluster*. Semua objek yang berada di satu cluster memiliki kemiripan (*similar*) antara satu dengan lainnya dan berbeda (*dissimilar*) dengan objek cluster lain. *Clustering* disebut juga *unsupervised learning*. Berbeda dengan *classification* yang menggunakan pendekatan *learning by examples*, *clustering* menggunakan pendekatan observasi untuk melakukan proses pengelompokan sehingga tidak ditentukan oleh label-label kelas yang diberikan. Proses pengelompokan tersebut tidak dilakukan oleh manusia, melainkan dengan menggunakan clustering algorithm. Karena itu, clustering dapat digunakan untuk mencari tahu kelompok yang sebelumnya tidak diketahui di data.

Clustering telah banyak digunakan di banyak aplikasi. Pada *business intelligence*, *clustering* dapat digunakan untuk mengatur sejumlah besar pelanggan ke dalam kelompok. Di pengenalan gambar, *clustering* digunakan untuk menemukan cluster di sistem pengenalan tulisan tangan. *Clustering* juga telah ditemukan di banyak aplikasi *Web search*, seperti mencari relevansi dari kata kunci hingga mengelompokan berbagai berita yang memiliki kemiripan.

Sebagai fungsi *data mining*, analisis *cluster* dapat digunakan sebagai sarana independen (*standalone tool*) untuk mendapatkan pengetahuan dari distribusi data, meneliti berbagai karakteristik setiap cluster, dan memfokuskan pada sekelompok untuk analisis lebih lanjut. *Clustering* juga dapat digunakan sebagai langkah preprocessing untuk

algoritma lain, seperti penentuan label untuk proses klasifikasi.

2.2. Metode Clustering Berbasis Kepadatan (Density-based Method): ST-DBSCAN

Algoritma *Spatial-Temporal Density Based Spatial Clustering of Application with Noise* (ST-DBSCAN) adalah suatu algoritma untuk pengelompokan yang dibangun dengan memodifikasi algoritma DBSCAN. Modifikasi yang dimiliki algoritma ST-DBSCAN adalah, (i) algoritma ST-DBSCAN dapat mengelompokkan data *spatiotemporal* sesuai dengan atribut non-spasial, spasial dan temporal, (ii) DBSCAN tidak mendeteksi titik noise ketika kepadatan bervariasi tetapi algoritma ST-DBSCAN mampu mengatasi masalah ini dengan menetapkan faktor kepadatan untuk setiap penggerombolan, (iii) Untuk mengatasi konflik pada perbatasan objek dilakukan dengan membandingkan nilai rata-rata penggerombolan yang akan datang dengan nilai baru.

Algoritma ST-DBSCAN membutuhkan empat parameter, yaitu Eps_1 , Eps_2 , $MinPts$, dan $\Delta\epsilon$. Eps_1 adalah jarak untuk atribut spasial. Eps_2 adalah jarak untuk atribut temporal. Perhitungan jarak dengan menggunakan *Euclidean*, *Manhattan*, atau *Minkowski distance* dapat digunakan untuk Eps_1 dan Eps_2 . $MinPts$ adalah jumlah minimum objek yang berada dalam jarak Eps_1 dan Eps_2 sedangkan $\Delta\epsilon$ digunakan untuk mencegah ditemukannya kombinasi cluster akibat perbedaan nilai yang cukup besar dari atribut non-spasial dari objek lainnya. Berikut tahapan dari algoritma ST-DBSCAN:

- 1) Algoritma dimulai dengan objek pertama pada data *spatiotemporal* dan mengambil semua titik yang bersifat *directly density-reachable*.
- 2) Jika jumlah objek yang *directly density-reachable* lebih kecil dari nilai $MinPts$, objek dinilai sebagai noise.
- 3) Noise dapat diubah kemudian jika titik bukan *directly density-reachable*, tetapi merupakan *density-reachable*.
- 4) Jika jumlah objek pada *density-reachable* lebih besar atau sama dengan nilai $MinPts$, objek tersebut adalah *core object* dan sebuah cluster baru dibangun. Semua *directly density-reachable* dari *core object* ditandai dengan cluster yang sama dari *core object*.
- 5) Algoritma secara iteratif memasukan *directly density-reachable* pada *core object* sebelumnya ke dalam *stack*. *Stack* dibutuhkan untuk menemukan objek-objek

density-reachable dari objek-objek *directly density-reachable*.

- 6) Jika *stack* tidak kosong, satu persatu objek yang di dalam *stack* akan dicari *density-reachable*. Jika jumlah objek *density-reachable* tersebut lebih besar dari $MinPts$, lakukan perulangan untuk setiap objek-objek *density-reachable* tersebut.
- 7) Jika objek tidak ditandai sebagai noise atau tidak dalam sebuah cluster, dan perbedaan antara nilai rata-rata non-spasial dari sebuah cluster dan nilai IPM baru dari objek lebih kecil dari $\Delta\epsilon$, objek tersebut ditempatkan ke dalam cluster tersebut.
- 8) Jika dua cluster C_1 dan C_2 saling berdekatan, sebuah objek p mungkin termasuk dalam kedua cluster tersebut. Namun, algoritma ini akan menyatakan objek p sebagai bagian dari cluster yang pertama kali menemukan.

2.3. Metode Clustering Berbasis Partisi (Partitioning Method): K-Means

Algoritma K-Means pertama kali diusulkan oleh Macqueen pada tahun 1967 yang merupakan algoritma yang tidak rumit dan termasuk ke dalam *non-supervised learning* [2]. K-means merupakan salah satu metode *clustering* berbasis partisi yang mengelompokkan data dalam bentuk satu atau lebih *cluster* yang masing-masing berbeda. Data yang memiliki kemiripan akan dikelompokkan ke dalam satu *cluster*. Algoritma ini membutuhkan parameter berupa jumlah cluster yang dihasilkan. Langkah-langkah penerapan clustering dengan menggunakan algoritma K-means adalah sebagai berikut:

- 1) Masukkan parameter k .
- 2) Inisiasi titik pusat atau centroid dari semua cluster secara acak.
- 3) Untuk setiap objek data di D , tentukan centroid yang terdekat dari objek dengan menggunakan pengukuran jarak seperti *Euclidean Distance* dan nyatakan objek tersebut sebagai bagian dari cluster.
- 4) Untuk setiap *cluster* dari 1 sampai k , hitung kembali titik centroid dari *cluster* dengan cara menghitung nilai rata-rata dari setiap objek yang berada di dalam *cluster* tersebut.
- 5) Ulangi ke langkah ke-3 sampai algoritma mengalami konvergensi atau ketika centroid dari setiap cluster tidak ada yang berubah.

2.4 Euclidean Distance Matrices

Euclidean Distance Matrices (EDM) adalah metode umum yang digunakan untuk mengukur jarak dua objek. Metode pengukuran ini sering ditemukan pada psychometrics, crystallography, machine learning, wireless sensor network dan banyak lagi [7]. EDM dari dua n-dimensional vector didefinisikan:

$$d(i, j) = \sqrt{(x_{i1} - x_{j1})^2 + \dots + (x_{in} - x_{jn})^2}$$

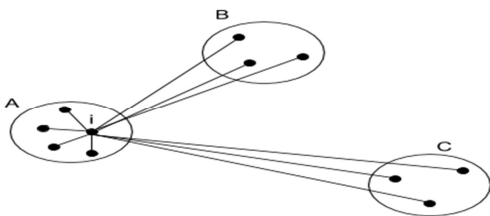
Berdasarkan hasil eksperimen Sinwar & Kaushik (2014) menunjukkan bahwa EDM menghasilkan performa yang lebih baik dibandingkan dengan metode lainnya seperti *Manhattan Distance* dalam jumlah iterasi yang dilakukan selama proses perhitungan [8].

2.4 Silhouette Coefficient

Silhouette Coefficient digunakan untuk mengukur kualitas dari cluster yang terbentuk. Nilai Silhouette Coefficient suatu objek *i* adalah:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}$$

Dengan *a(i)* merupakan jarak rata-rata antara objek *i* dengan seluruh objek yang berada dalam satu cluster yang sama (intra-cluster). *b(i)* adalah jarak rata-rata antara objek *i* dengan seluruh objek yang berada pada cluster terdekat [9]. Sebagai ilustrasi pada Gambar 2.6, titik *i* berada dalam cluster A, maka cluster B adalah cluster terdekatnya, dengan *a(i)* adalah nilai rata-rata jarak titik *i* dengan titik lain di dalam cluster A dan *b(i)* adalah rata-rata jarak titik *i* ke semua titik dalam cluster B.



Gambar 2: Ilustrasi Silhouette Coefficient

Hasil perhitungan nilai *Silhouette Coefficient* dapat bervariasi antara -1 hingga 1. Semakin nilai *s(i)* mendekati 1, hasil clustering pada objek *i* sudah berada dalam cluster yang tepat. Sebaliknya, jika nilai *s* mendekati -1 maka hasil clustering pada objek *i* terjadi overlapping sehingga objek tersebut lebih tepat dimasukkan ke dalam cluster lain. Untuk menghitung semua nilai dari *Silhouette Coefficient* setiap objek, dapat menggunakan pustaka yang sudah

tersedia di Scikit-learn. Hasil keluarannya ialah nilai rata-rata *Silhouette Coefficient* dari semua objek data.

3. RANCANGAN SISTEM DAN APLIKASI

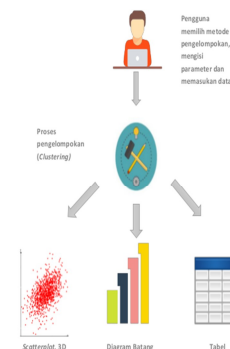
3.1 Analisa Masalah

Ketersediaan data IPM tiap wilayah di Pulau Jawa perlu dilakukan proses statistik untuk mendapatkan data dan informasi tentang pencapaian pembangunan antar wilayah. Ada banyak sekali keuntungan yang bisa di dapat dengan melakukan pengolahan data IPM di Pulau Jawa seperti mengetahui pola persebarannya, penyusunan kebijakan-kebijakan yang berdampak pada peningkatan pembangunan dan memantau capaian pembangunan manusia antar wilayah.

Oleh karena itu, diperlukan penelitian untuk mengimplementasi metode pengelompokan dan menentukan metode yang tepat untuk mengembangkan aplikasi untuk melakukan pengelompokan pada data IPM. Hasil penelitian tersebut diharapkan dapat membantu BPS untuk menganalisis data IPM sehingga pihak tersebut dapat menyediakan informasi yang berkualitas kepada pemerintah dalam upaya penyusunan kebijakan-kebijakan untuk meningkatkan kesejahteraan tiap daerah.

3.2 Penyelesaian Masalah

Dari analisis masalah yang telah diuraikan di atas, diperlukan penelitian untuk mengimplemntasi metode pengelompokan dan mencari tahu metode yang paling optimal untuk melakukan proses data mining pada data IPM di Pulau Jawa. Hasil penelitian tersebut juga dapat dijadikan rekomendasi untuk mengembangkan aplikasi yang dapat melakukan proses pengelompokan pada data IPM Pulau Jawa sehingga BPS dapat memberikan informasi yang berkualitas untuk pemerintah dalam upaya untuk mewujudkan kesejahteraan masyarakat khususnya di Pulau Jawa. Teknik data mining yang digunakan adalah clustering. Penerapan proses clustering telah diputuskan oleh peneliti karena data IPM yang tersedia tidak memiliki label pengelompokan.



Gambar 3: Rancangan Aplikasi Pengelompokan

aplikasi terdapat pada Gambar 3.

3.3 Praproses Data

Data yang digunakan pada penelitian ini adalah data IPM tiap kabupaten/kota Pulau Jawa pada tahun 2014-2016 yang sudah tersedia pada publikasi online di website resmi Badan Pusat Statistika dan data *shapefile* (.shp) wilayah administratif Indonesia yang dapat diunduh dari website Ina-Geospasial (<http://tanahair.indonesia.go.id>). Website tersebut merupakan portal resmi Geospasial Indonesia yang diluncurkan oleh Badan Informasi Geospasial (BIG). Data *shapefile* digunakan untuk memperoleh data centroid (titik tengah) setiap wilayah kabupaten/kota di Pulau Jawa. Selain itu, data tersebut digunakan sebagai visualisasi hasil clustering.

Tahapan praproses data yang dilakukan sebagai berikut:

- 1) **Pemilihan Data**
 Pada proses ini dilakukan pemilihan data IPM yang hanya terdapat di Pulau Jawa dengan rentang tahun 2014-2016 dan data *shapefile* wilayah administratif pulau jawa. Penelitian ini menggunakan perangkat lunak QuantimGIS untuk memperoleh centroid di setiap kabupaten/kota Pulau Jawa Ketika sudah mendapatkan data centroid , data tersebut diubah menjadi format *comma delimited* (.csv).
- 2) **Reduksi Data**
 Reduksi data bertujuan untuk mengurangi jumlah atribut yang tidak dibutuhkan pada tahap *data mining* . Reduksi yang dilakukan dengan cara memilih empat atribut dari delapan atribut. Atribut yang dipilih adalah Nama Provinsi, IPM tahun 2014, IPM tahun 2015 dan IPM tahun 2016. Untuk data centroid, hanya diambil data koordinat (longitude dan latitude) tiap kabupaten/kota di Pulau Jawa.
- 3) **Integrasi Data**
 Integrasi data dilakukan dengan menggabungkan atribut-atribut yang sudah ditentukan. Koordinat centroid yang sudah diubah menjad *comma delimited* (.csv)

digabungkan dengan data IPM di Pulau Jawa dengan rentang tahun 2014-2016.

4) Transformasi Data

Transofrmasi data dilakukan dengan menggabungkan atribut tahun menjadi satu sehingga setiap satu record mewakili satu tahun dan untuk setiap tiga record mewakili satu kabupaten/kota. Berikut contoh nilai atribut yang sudah dilakukan tahap praproses ditunjukkan pada Tabel 1.

Tabel 1: Contoh nilai atribut dari data hasil praproses

kabupaten_kota	longitude_centroid	latitude_centroid	tahun	ipm
Bandung	107.52634	-7.010638	2014	69.06
Bandung	107.52634	-7.010638	2015	70.05
Bandung	107.52634	-7.010638	2016	70.69
Bandung Barat	107.491977	-6.865221	2014	64.27
Bandung Barat	107.491977	-6.865221	2015	65.23
Bandung Barat	107.491977	-6.865221	2016	65.81

3.4. Penentuan Nilai Parameter

Penentuan parameter pada metode pengelompokan sangat berpengaruh terhadap cluster yang dihasilkan. Parameter yang terdapat pada metode ST-DBSCAN adalah Eps1 (jarak atribut spasial), Eps2 (jarak atribut temporal), MinPts (jumlah minimum objek yang berada dalam jarak Eps1 dan Eps2) sedangkan $\Delta\epsilon$ adalah batas selisih nilai non-spasial. Parameter yang terdapat pada metode K-Means adalah k (jumlah cluster).

Pada penelitian ini, nilai k yang digunakan untuk parameter K-Means adalah 4. Pemberian nilai tersebut mengikuti jumlah pengelompokan wilayah berdasarkan nilai IPM menurut UNDP, yaitu 4 kelompok. Pada metode ST-DBSCAN, nilai Eps2 yang digunakan adalah 3. Nilai tersebut digunakan agar proses pengelompokan melibatkan semua tahun sehingga tiap pengelompokan tidak terpisah berdasarkan tahun. Penggunaan nilai MinPts = 7 mengikuti saran dari Birant dan Kut (2007), yang menyatakan bahwa nilai yang cukup optimal pada MinPts adalah $\ln(n)$ dimana nilai n adalah jumlah dari total keseluruhan dataset yang digunakan sedangkan pemilihan $\Delta\epsilon = 2$ mengikuti penelitian yang sudah dilakukan oleh Afifah (2014) [5] dan Trisnaningtyas

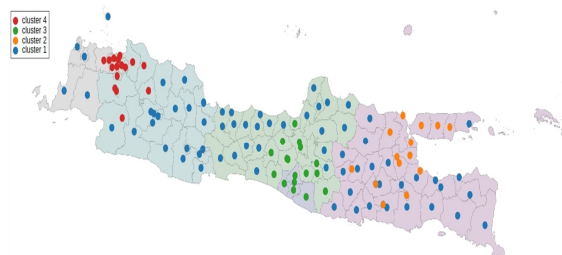
(2014) [6]. Dengan menggunakan paramter tersebut, maka hal tersebut menjadi pertimbangan pemilihan parameter Eps1. Agar hasil pengelompokan memiliki jumlah yang sama dengan hasil pengelompokan metode K-Means, maka nilai yang digunakan untuk Eps1 adalah 0.7.

4. HASIL DAN PEMBAHASAN

4.1. Hasil Pengelompokan ST-DBSCAN

Pada penelitian ini, proses algoritma ST-DBSCAN akan menggunakan data spasial (lokasi titik tengah atau centroid pada kabupaten/kota), data temporal (tahun 2014 sampai 2016) dan data non spasial (data IPM di tahun 2014-2016). Dengan menggunakan parameter Eps1=0.7, Eps2=2, MinPts=7 dan $\Delta\epsilon=2$, diperoleh 4 cluster.

Hasil clustering akan ditampilkan berupa centroid setiap kabupaten/kota yang sudah dikelompokkan tiap cluster. Cluster terbentuk berdasarkan kedekatan jarak antara satu centroid kabupaten/kota dengan yang lain. Setiap anggota clustering memiliki persamaan yaitu berkumpul dan memiliki jarak radius 0.7 derajat antara centroid. Yang membedakan antara wilayah satu dengan yang lain adalah jarak pada titik paling luar (*border*). Adanya perbedaan ini diakibatkan oleh selisih nilai rata-rata atribut non-spatial dalam cluster dan nilai non-spatial objek yang akan diseleksi lebih besar dari nilai batas $\Delta\epsilon$.



Gambar 4: Hasil Clustering ST-DBSCAN

Berdasarkan Gambar 4, dapat diketahui bahwa hasil clustering dari ST-DBSCAN cenderung memadat di suatu wilayah tertentu, kecuali cluster 2 yang menyebar di Pulau Jawa. Tiap cluster yang terbentuk memiliki karakteristik wilayah yang berbeda-beda berdasarkan aspek spasial, aspek temporal dan aspek non-spatial.

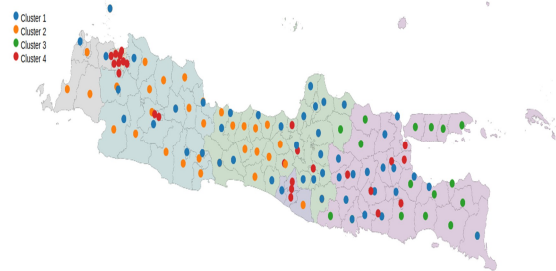
Tabel 2: Hasil Silhouette-Coeffecien dan waktu clustering metode ST-DBSCAN

Silhouette Coeffecien	Waktu
0.15	3.480 detik

Untuk mengukur tingkat performansi dari metode ST-DBSCAN, dilakukan pengukuran *Silhouette Coeffecien* dengan bantuan *library* Scikit-learn dan menghitung waktu proses clustering . Hasil yang didapat ialah, metode ST-DBSCAN memiliki Silhouette Coeffecien sebesar 0.15213 dan waktu proses 3.480 detik.

4.2. Hasil Pengelompokan K-Means

Dari data IPM di Pulau Jawa dengan rentang tahun 2014-2016 dilakukan proses *clustering* menggunakan K-means dengan menggunakan yang sama dimasukkan oleh algoritma ST-DBSCAN, yaitu data spasial (lokasi centroid pada kabupaten/kota), data temporal (tahun 2014 sampai 2016) dan data non spasial (data IPM di tahun 2014-2016). Dengan menggunakan parameter K =4, diperoleh 4 cluster.



Gambar 4: Hasil Clustering ST-DBSCAN

Bila dibandingkan dengan ST-DBSCAN, tiap cluster yang terbentuk dari metode K-Means menyebar ke seluruh Pulau Jawa seperti pada Gambar 4.13. Bila ditinjau lebih lanjut, hasil pengelompokan K-Means cenderung mengikuti persebaran berdasarkan nilai IPM. Nilai IPM memiliki variasi yang lebih besar dari ciri lain seperti longitude, latitude, dan tahun, yaitu dengan rentang 56.98 sampai 85.32 sehingga nilai tersebut mempunyai pengaruh besar terhadap pembentukan cluster pada metode K-means. Namun, antara K-means dan ST-DBSCAN memiliki kesamaan, yaitu terdapat satu cluster yang menyebar ke seluruh Pulau Jawa.

Pengukuran nilai *Silhouette Coeffecien* dan waktu proses pengelompokan juga dilakukan pada metode K-Means. Berdasarkan percobaan tersebut, didapatkan hasil pengelompokan dengan menggunakan metode K-Means dengan menggunakan parameter k=4 memiliki nilai *Silhouette Coeffecien* sebesar 0.3445 dan waktu

proses 0.495 milidetik. Bila dibandingkan dengan metode ST-DBSCAN, metode K-Means memiliki keunggulan pada nilai *Silhouette Coefficient* yang lebih tinggi dan waktu proses yang lebih cepat.

Tabel 3: Hasil *Silhouette-Coefficien* dan waktu clustering metode K-Means

<i>Silhouette Coefficien</i>	Waktu
0.3445	0.495 milidetik

5. KESIMPULAN

Melalui proses pengerjaan dan pengujian dalam penelitian ini, maka dapat disimpulkan beberapa hal:

- 1) Algoritma ST-DBSCAN dan K-Means dapat diimplementasikan untuk melakukan pengelompokan data IPM kabupaten/kota Pulau Jawa dengan rentang tahun 2014-2016.
- 2) Bila dilihat dari scatterplot, hasil pengelompokan dari metode ST-DBSCAN cenderung memadat di wilayah tertentu sedangkan metode K-Means menghasilkan pengelompokan yang menyebar.
- 3) Metode K-Means yang menggunakan parameter nilai $k=4$ masih memiliki nilai *Silhouette Coefficient* lebih tinggi dengan nilai rata-rata 0.344 dan waktu proses lebih cepat dibandingkan dengan metode ST-DBSCAN yang menggunakan parameter nilai $Eps1=0.7$, $Eps2=3$, $MinPts=7$ dan $\Delta\epsilon = 2$.

Beberapa saran yang dapat diberikan untuk pengembangan lebih lanjut dengan harapan dapat menghasilkan penelitian yang lebih baik lagi di kemudian hari, berikut saran yang dapat diberikan:

- 1) Pada penelitian selanjutnya, diharapkan dilakukan pengujian dengan metode pengelompokan lain.
- 2) Untuk pengembangan selanjutnya, diharapkan penambahan metode untuk mencari nilai parameter yang optimal secara otomatis sehingga pengguna tidak perlu memasukan nilai parameter.
- 3) Diharapkan ada peningkatan dataset yang mencakup seluruh Indoensia dan jangkauan tahun yang lebih lama sehingga dapat dilihat persebarannya secara menyeluruh.

DAFTAR PUSTAKA

- [1] Badan Pusat Statistik, 2016. Indeks Pembangunan Manusia Tahun 2016 [internet]. Tersedia di: <http://www.bps.go.id/publications/publikasi2016.php?key=indeks+pembangunan+manusia> [Diakses November 14, 2017].
- [2] Sehgal, G. dan Garg, Dr. Kanwal, 2014. Comparison of Various Clustering Algorithms. *International Journal of Computer Science and Information Technologies (IJCSIT)* , Vol. 5(3). hal. 3074-3076.
- [3] Budiman, S. A. D. et al., 2016. Perbandingan Metode K-Means Dan Metode DBSCAN Pada Pengelompokan Rumah Kost Mahasiswa Di Kelurahan Tembalang Semarang. *Jurnal Gaussian* , Vol. 5, No. 4. hal. 757-762.
- [4] Maitry, N. dan Vaghela, D., 2014. Survey on Different Density Based Algorithms on Spatial Dataset. *International Journal of Advance Research in Computer Science and Management Studies* , Vol 2, Issue 2. hal. 362-366.
- [5] Afifah, N., 2014. Clustering Data Indeks Pembangunan Manusia (IPM) Pulau Jawa Menggunakan Algoritme ST-DBSCAN Dan Bahasa Pemrograman R [SKRIPSI].
- [6] Trisnaningtyas, A., 2014. Pengelompokan Data Indeks Pembangunan Manusia Di Pulau Jawa Dengan Algoritme ST-DBSCAN Dan Bahasa Pemrograman Python [SKRIPSI]. Institut Pertanian Bogor.
- [7] Dokmanic, I., et al, 2015. Euclidean Distance Matrices: Essential Theory, Algorithms and Applications. *IEEE Signal Processing Magazine*. hal 1-17.
- [8] Sinwar, D. dan Kaushik, R., 2014. Study of Euclidean and Manhattan Distance Metrics using Simple K-Means Clustering. *International Journal For Research In Applied Science And Engineering Technology (IJRASET)* , Vol. 2 Issue V. hal 270-274.
- [9] Scikit-learn. Dokumentasi silhouette_score [internet]. Tersedia di http://scikit-learn.org/stable/modules/generated/sklearn.metrics.silhouette_score.html [Diakses Desember 25, 2017].

