# BAB II LANDASAN TEORI

## 

## 2.1 Sampah Organik

Menurut Undang-Undang No. 18 Tahun 2008 sampah adalah sisa kegiatan sehari-hari manusia dan/atau proses alam yang berbentuk padat. Sesuai dengan ketentuan Undang-undang yang menekankan bahwa perlu adanya pengelolaan secara sistematis, meyeluruh dan berkesinambungan yang meliputi pengurangan dan penanganan sampah.

Sampah rumah tangga meliputi sampah organik, anorganik, dan sampah B3 (Bahan Beracundan Berbahaya). Sampah berpotensi menyebabkan gangguan lingkungan yang dapat berupa pecemaran air, tanah, dan udara. Pencemaran selanjutnya dapat menimbulkan gangguan kesehatan dan sosial ekonomi.( Renna Eliana Warjoto et al.,2018).

Salah satu cara untuk menangani sampah organik rumah tangga adalah melalui proses pembuatan kompos. Komposting merupakan proses penguraian materi-materi organik dengan bantuan mikroorganisme. Pada umumnya, komposting alami berlangsung cukup lama, yakni sekitar 3-4 bulan. Oleh sebab itu, akselerasi proses pembuatan kompos dapat dilakukan melalui penggunaan bioreaktor yang berisi campuran bahan organik dan mikroorganisme pengurai.( Meda Canti et al.,2018).

Sampah organik dapat diolah menjadi pupuk dengan menggunakan proses fermentasi. Pupuk organik yang dibuat dengan menggunakan proses fermentasi disebut dengan kompos. Kompos merupakan istilah untuk pupuk organik yang dibuat oleh manusia melalui proses pembusukkan sisa-sisa buangan makhluk hidup (tanaman maupun hewan). Secarakeseluruhan, proses pengomposan ini bisa disebut juga dengan dekomposi ( Retno Wulan Sekarsari et al.,2020).

## *2.2 Data Mining*

*Data mining* adalah proses pencarian pola yang penting dari data yang ada didalam  *database* yang besar (Rante Rerung, 2018) Data *mining*  merupakan istilah *pattern recognition* merupakan algoritma untuk pengolahan data guna menemukan pola data menjadi pengetahuan baru . Data yang diolah dengan teknik data  *mining* akan menghasilkan suatu pengetahuan bersumber dari data lama, hasilnya untuk menentukan keputusan bisnis. (Sembiring & Sari, 2019).

Data *mining* dibagi menjadi 2 kategori utama, yaitu:

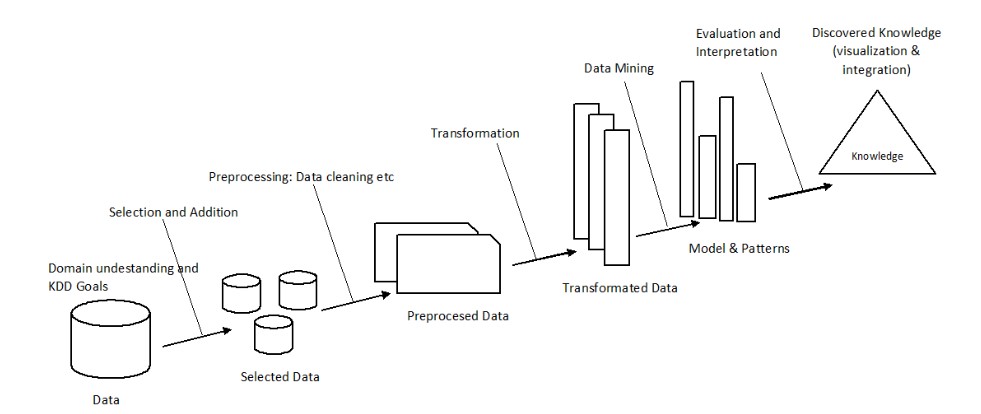
1. *Descriptive mining*, tahapan untuk menemukan karakteristik penting dari data dalam satu basis data.

2.  *Predictive*, tahapan untuk mempelajari bagaimana pattern design dari data set yang berupa variabel.

Teknik pengolahan data dengan bantuan algoritma data *mining*. pengolahan dilakukan dengan membangun sebuah *pattern design*, kemudian model tersebut membentuk pola data yang lain yang tidak berada dalam basis data. Tujuannya untuk memperolah *pattern design* untuk menjadikan data sebagai grafik untuk menjadi pengetahuan. (Gustian et al., 2018).

Data *mining* juga disebut dengan *Knowledge Discovery in Database* (KDD) yaitu sebuah proses yang secara otomatis saat pencarian data didalam *space memory* yang amat besar dari data untuk mengetahui pola dengan menggunakan teknik seperti klasifikasi hubungan (*association*) atau pengelompokan (*clustering*). (Rizki Muliono & Zulfikar Sembiring, 2019).

Berikut merupakan Tahapan *Knowledge Discovery in Database* dapat dilihat pada Gambar 2.2(Inna Alvi Nikmatun et al., 2019):



Gambar 2.1 Tahapan *Knowledge Discovery in Database*

Berikut adalah penjelasan dari setiap tahapan KDD :

1. *Domain Understanding and KDD Goals*. Tujuan ditentukan dari sudut pandang user dan digunakan untuk mengembangkan dan pemahaman tentang domain aplikasi dan pengetahuan sebelumnya.
2. *Selection and Additions.* Tahap kedua berfokus pada penentuan data target dan subset dari data sampel atau variabel.
3. *Preprocessing: Data Cleaning etc*. Pembersihan dan preprocessingdata merupakan operasi dasar untuk menyelesaikan data yang konsisten tanpa *noisy*.
4. *Transformation Transformasi* data dari satu bentuk ke bentuk lainnya sehingga data di implementasikan dengan mudah.
5. *Data Mining (Chosing the Suitable Data Mining Task)* Memillih metode data *mining* yang sesuai berdasarkan tujuan tertentu yang telah didefinisikan pada tahap pertama, contoh dari metode data *mining* adalah *classification, regression, clustering* dan *summarization.*
6. *DataMining (Chosing the Suitable Data Mining Algorithm)* Memilih algoritma yang tepat untuk pencarian pola-pola data, algoritma yang dipilih berdasarkan kecocokan kriteria dengan metode data *mining*.
7. *Data Mining (Imploying Data Mining Algorithm)* Pada tahap ini algoritma yang telah dipilih diimplementasikan.
8. *Evaluation and Interpretation*. Tahap ini berfokus pada interpretasi dan evaluasi yang mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan hipotesa yang ada sebelumnya.
9. *Discovered Knowledge*. Penggunaan pengetahuan yang ditemukan dari proses KDD, dimana memutuskan apa yang akan dilakukan dengan pengetahuan dihasilkan.

Hasil dari pengolahan data dengan metode data *mining* ini dapat digunakan untuk mengambil keputusan di masa depan. Data *mining* ini juga dikenal dengan istilah *pattern recognition.* Data *mining* merupakan metode pengolahan data berskala besar oleh karena itu data *mining* ini memiliki peranan penting dalam bidang industri, keuangan, cuaca, ilmu dan teknologi. Secara umum kajian data *mining* membahas metode-metode seperti, *clustering, klasifikasi, regresi, seleksi variable*, dan market basket analisis.(Nurul Rofiqo et al., 2018).

**2.3 *Clustering***

*Clustering* ialah teknik data mining yang digunakan untuk menganalisis dan mengkaji data untuk menyelesaikan permasalahan dalam pengelompokkan data membagi dari suatu dataset ke dalam subset. Teknik *clustering* tujuannya untuk kasus pendistribusian (orang,objek, peristiwa, dan lainnya) ke dalam suatu kelompok, hingga derajat keterkaiatan antara member cluster yang sama ialah lemah dan kuat antara anggota *cluster* dengan yang lainnya. Teknik *cluster* mempunyai dua cara dalam pengelompokkan yaitu *hierarchical clustering* dan *non-hierarchical clustering*. *hierarchical clustering*  ialah metode mengelompokkan data yang kerjanya dengan cara mengelompokkan dua atau lebih data yang mempunyai kesamaan atau kemiripan, kemudian diteruskan ke objek lain yang mempunyai kedekatan dua, proses berlangsung sampai suatu cluster membentuk semacam tree dimana tingkatan atau hirarki yang jelas antar objek dari yang paling tidak mirip hingga yang paling mirip. Namun secara logika seluruh objek pada akhirnya akan membentuk sebuah *cluster*. (Putrama Alkhairi.,2018).

Analisis Pengelompokan / *Clustering* merupakan proses membagi data dalam suatu himpunan ke dalam beberapa kelompok yang kesamaan datanya dalam suatu kelompok lebih besar daripada kesamaan data tersebut dengan data dalam kelompok lain. Potensi *clustering* adalah dapat digunakan untuk mengetahui struktur dalam data yang dapat dipakai lebih lanjut dalam berbagai aplikasi secara luas seperti klasifikasi, pengolahan gambar, dan pengenalan pola.(Nurul Rafiqo et al.,2018).

*Data Clustering* merupakan salah satu metode Data Mining yang bersifat (*unsupervised)* atau tanpa adanya suatu arahan. Ada dua jenis data *clustering* yang sering digunakan yaitu, *hierarchical clustering* dan *nonhierarchical clustering*.*K-Means* adalah contoh metode *nonhierarchical clustering* yang memecah data ke dalam satu atau beberapa cluster yang memiliki karakteristik yang sama dan tidak. Metode ini berupaya meminimalisir variasi datadi dalam *cluster* dan memper banyak variasi antar *cluster.*(Nugroho Dwi Saksono et al., 2018).

*Clustering* pada suatu data adalah suatu tahapan untuk menggolongkan himpunan data yang atribut kelasnya belum dideskripsikan, secara konsep *clustering* adalah untuk memaksimalkan dan meminimalkan kemiripan intra antar kelassebagai contoh, ada suatu himpunan obyek, proses pertama dapat di klasterisasi menjadi beberapa himpunan kelas selanjutnya menjadi sebuah himpunan beraturan sehingga dapat diturunkan berdasarkan kelompok klasifikasi tertentu. *Cluster* juga dapat diartikan sebagai kelompok. Maka analisa *clustering* pada dasarnya akanmenghasilkan sejumlah *cluster* (kelompok). Sebelum kita melakukan analisa perlu diterapkan pemahaman bahwa suatu himpunan dari data tertentu sebenarnya sudah memiliki kemiripan di antara anggotanya tersebut.Oleh karena itu, setiap anggota-anggota yang memiliki kemiripan karasteristik dikelompokkan ke dalam satu atau lebih dari suatu kelompok. Tujuan dari pada *clustering* data adalah untuk meminimalisasikan fungsi tujuan yang ditetapkan didalam proses *clustering*, dan umumnya selalu meminimalisasi variasi suatu *cluster* dan memaksimalisasi kan variasi antar *cluster.*(Rizki Muliono & Zulfikar Sembiring, 2019)*.*

Contoh-contoh penerapan analisa *cluster* dapat dijumpai saat ini. Berikut ini beberapa diantaranya (Herlawati & Rahmadya Trias Handayanto, 2020):

1. *Segmentasi* Pasar adalah pengklasteran data yang membagi pelanggan menjadi grup-grup tertentu yang akan mempermudah bagian penjualan (*marketing*) dalam memasarkan produk-produknya, seperti rumah, kendaraan, dan sebagainya.
2. Pengklasteran Dokumen. Dokumen-dokumen yang memiliki kemiripan yang sama, misalnya politik, ekonomi, dan bidang lainnya dikumpulkan dalam satu grup. Manfaat yang diperoleh adalah kemudahan dalam mencari, mengorganisir, dan mensuplai data-data yang akan dimanfaatkan oleh pengguna pada bidang yang sesuai.
3. Pengklasifikasian Penyakit. Penyakit tertentu dapat didetiksi dari gejala-gejala yang menyertainya. Oleh karena itu pengklasifikasian penyakit berdasarkan gejala sangat membantu para praktisi kesehatan dalam aktivitas kesehariannya, sehingga perlakuan yang tepat dapat diterapkan untuk tiap kasus penyakit tertentu.
4. Pengklasifikasian dalam Biologi. Biologi sangat membutuhkan proses klasifikasi, misalnya dalam bioinformatika untuk mencari gen-gen terbaik berdasarkan kelas-kelas yang terbentuk.

## 2.4 *K-Means Clustering*

## *K-Means Clustering* adalah algoritms yang termasuk dalam klasifikasi terhadap objek berdasarkan atribut. Permatisian ini dilakukan untuk mencari nilai jarak minimum antara data dan nilai *centroid* yang telah di set baik secara *random* dengan cara menentukan nilai *centroid* berdasarkan K yang berupa *object* yang berurutan . *Centroid* merupakan hasil rata-rata dari sebuah bentuk objek dari seluruh titik dalam objek. (Falentino Sembiring & Sudin Saepudin, 2020).

Data  *Clustering*  merupakan salah satu metode Data Mining yang bersifat (unsupervised) atau tanpa adanya suatu arahan. Ada dua jenis data  *clustering* yang sering digunakan,  *hierarchical clustering*  dan *nonhierarchical clustering. K-Means*  adalah contoh metode *nonhierarchical clustering* yang memecah data ke dalam satu atau beberapa cluster yang memiliki karakteristik yang sama dan tidak. Metode ini berupaya meminimalisir.(Nugroho Dwi Saksono.,2018)

Metode *K-means clustering* merupakan metode *clustering* yang dikenalkan oleh Lloyd, S, 1982.Metode K-means adalah metode yang terkenal cepat dan simpel. *K-means clustering* merupakan salah satu metode data *clusteringnon-hirarki* yang mengelompokan data dalam bentuk satu atau lebih *cluster* atau kelompok. Data-data yang memiliki karakteristik yang sama dikelompokan dalam satu *cluster* kelompok dan data yang memiliki karakteristik yang berbeda dikelompokan dengan *cluste*r kelompok yang lain sehingga data yang berada dalam satu *cluster* kelompok memiliki tingkat variasi yang kecil.(Ade Bastian et al., 2018).

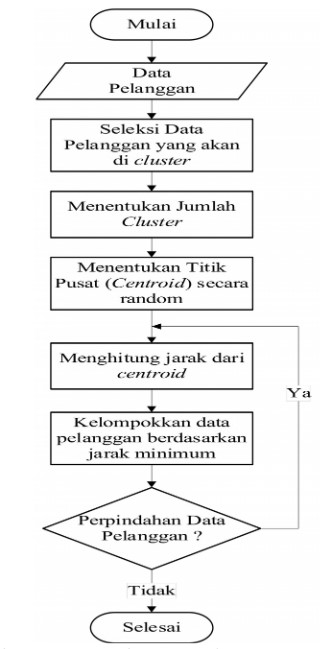
penerapan algoritma *K-means* dihasilkan nilai *centroid* dari data yang diperoleh dengan syarat bahwa pengelompokan yang ingin ditentukan adalah 2, Penentuan *cluster* dibagi atas dua bagian yakni *cluster* tingkat distribusi tinggi (C1) dan *cluster* tingkat distribusi rendah (C2). Maka nilai centroid juga terdapat 2 titik. Penentuan titik *cluster* ini dilakukan dengan mengambil nilai terbesar (maksimum) untuk *cluster* tingkat distribusi tinggi (C1) dan nilai terkecil (minimum) untuk *cluster* tingkat distribusi rendah (C2). Berikut ini adalah Ilustrasi nilai titik *centroid* dapat dilihat pada Tabel 2.1 (Cici Astria et al., 2019) :

Table 2.1 Ilustrasi Nilai titik *Centroid*

|  |  |
| --- | --- |
| Atribut | Nilai |
| *Cluster* Tinggi | 37722.47 |
| *Cluster* Rendah | 60.2433 |

Menurut (Usep Tatang Suryadi., 2019) terdapat 3 komponen didalam Algoritma *K-Means* yaitu :

1. Jumlah *Cluster K*

Dalam metode ini jumlah k harus ditentukan dulu, setelah jumlah k didapatkan dengan melalui pendekatan metode hirarki dapat melakukan pengambilan *cluster* awal. Aturan khusus dalam menentukan jumlah *cluster* k bahkan tidak ada, namun ada juga jumlah *cluster* yang diinginkan sesuai dengan kebutuhan subjektif seseorang.

1. *Cluster* Awal

Ada banyak pendapat saat melakukan pengambilan *cluster* awal untuk metode *K-Means* misalnya pemilihan terhadap *interval* dari jumlah setiap *observasi*, melalui pendekatan salah satu metode hirarki dan ada juga dengan melalui pemilihan *cluster* secara acak dari sekumpulan *observasi*. Maka dengan adanya beberapa cara pengambilan *cluster* awal tersebut dapat memungkinkan solusi terbaik yang dihasilkan.

1. Ukuran Jarak

Pada tahapan ini ukuran jarak juga penting dalam menempatkan *observasi* ke dalam *cluster* berdasarkan nilai *centroid* terdekat. *Euclidian Distance* adalah jarak yang digunakan untuk mengukur jarak dalam metode *K-Means*.

### 2.4.1 Tahapan Metode *K-Means*

Di dalam metode  *K-Means* memiliki beberapa tahapan yang harus dilakukan saat melakukan perhitungan untuk mencari data *cluster*  hingga iterasi menjadi sama dengan hitungan yang sebelumnya, berikut adalah tahapan *K-Means* (Falentino Sembiring & Sudin Saepudin, 2020):

1. Menentukan K sebgai Jumlah *cluster* **(*C1, C2 dan C3* )**yang ingin dibentuk.
2. Tentukan nilai untuk pusat *cluster* (*centroid*).
3. Hitung jarak untuk setiap data input terhadap masing-masing *centroid* menggunakan Rumus *Euclidean Distance* hingga ditemukan jarak yang paling dekat dari setiap data dengan *centroid.* Dibawah ini adalah persamaan *Euclidean Distance* :

……………...………….……….…………(2.1)

Keterangan :

d (x,y) : jarak antara data pada posisi titik x dan y

x : posisi titik data pertama (pusat*cluster*)

y : posisi titik data kedua (data dari N)

n : jumlah atribut data

Jarak terpendek antara *centroid* dengan objek menentukan titik cluster antar objek. Adapun rumus iterasinya didefinisikan sebagai berikut :

…………………………….…………………………(2.2)

Keterangan :

= *Centroid*, *Cluster* pertama pada variable ke-j

= jumlah data yang menjadi *cluster* ke-i

i,k = indeks dari *cluster*

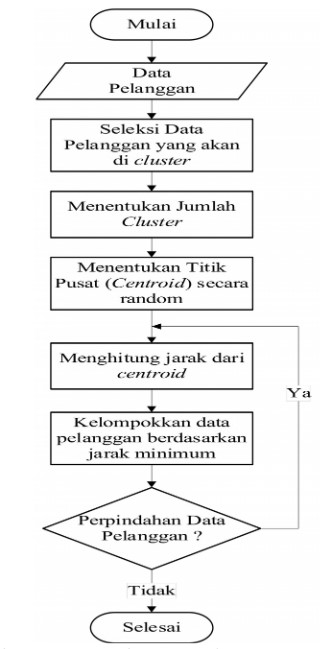
j = indeks dari variable

= nilai data ke-k yang ada didalam *cluster* ini untuk variable ke-j.

1. Kemudian lakukan iterasi selama partisi data masih bergerak (tidak ada lagi objek yang bergerak ke partisi lain), bila masih maka ke point 3
2. Bila grup data sekarang sama dengan grup data sebelumnya, maka hentikan iterasi. Data telah di partisi sesuai nilai *centroid* akhir.

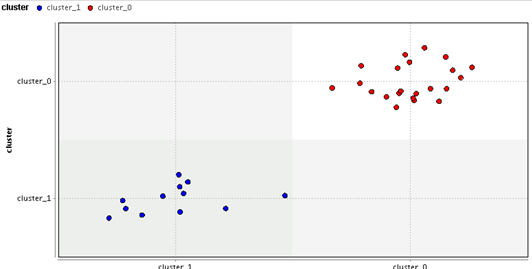
Atau apabila perubahan nilai *threshold* yang ditentukn atau apabila perubahan nilai pada objective function yang digunakan di atas nilai *threshold* yang ditentukan.*Distance space* digunakan untuk menghitung jarak antara data dan *centroid.* Adapun persamaan yang dapat digunakan salah satunya yaitu *Euclidean Distance Space. Euclidean distance space* sering digunakan dalam perhitungan jarak, hal ini dikarenakan hasil yang diperoleh merupakan jarak terpendek antara dua titik yang diperhitungkan (Dwi Retno Lestari, 2019).

Berikut ini adalah ilustrasi penerapan metode *K-Means* dengan Diagram Alir Segmentasi data pelanggan menggunakan Algoritma *K-Means*(Nisa Hanum Harani et al., 2020):



Gambar 2.2Diagram Alir Segmentasi Data Menggunakan Algoritma *K-Means*

Berikut ini adalah ilustrasi Tampilan Hasil *Clustering* menggunakan metode *K-Means* menunjukan bagaimana tahapan titik-titik tanpa *cluster* kemudian dibagi menjadi dua *cluster* yang telah di kelompokkan (Abdi Rahim Damanik, 2020).



Gambar 2.3Ilustrasi Tampilan Hasil *Clustering*

Berikut ini adalah contoh ilustrasi penggunaan metode *k-means* oleh (Abdi Rahim Damanik, 2020) tentang pengelompokaandata tingkat pengangguran dan data tenaga kerja Penelitian ini menggunakan dua data yaitu data tingkat pengangguran dan data tenaga kerja yang akan dikelompokan dan dilakukan perbandingan antara data persentasi pengangguran dan data persentasi tenaga kerja. Berikut adalah data penelitian yang digunakan.

**Tabel 1. Data Tingkat Pengangguran**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Provinsi | 2015 | 2016 | 2017 | 2018 |
| **Aceh** | 8,83 | 7,85 | 6,98 | 6,455 |
| **SumateraUtara** | 6,55 | 6,165 | 6,005 | 5,575 |
| **SumateraBarat** | 6,44 | 5,45 | 5,69 | 5,55 |
| **Riau** | 7,275 | 6,685 | 5,99 | 5,96 |
| **Jambi** | 3,535 | 4,33 | 3,77 | 3,755 |
| **SumateraSelatan** | 5,55 | 4,125 | 4,095 | 4,125 |
| **Bengkulu** | 4,06 | 3,57 | 3,275 | 3,105 |
| **Lampung** | 4,29 | 4,58 | 4,38 | 4,195 |
| **Kep.BangkaBelitung** | 4,82 | 4,385 | 4,12 | 3,63 |
| **Kep.Riau** | 7,625 | 8,36 | 6,8 | 6,775 |
| **DkiJakarta** | 7,795 | 5,945 | 6,25 | 5,79 |
| **JawaBarat** | 8,56 | 8,73 | 8,355 | 8,165 |
| **Jawa Tengah** | 5,15 | 4,415 | 4,36 | 4,37 |
| **Di Yogyakarta** | 4,07 | 2,765 | 2,93 | 3,205 |
| **JawaTimur** | 4,39 | 4,175 | 4,05 | 3,92 |
| **Banten** | 9,065 | 8,435 | 8,515 | 8,145 |
| **Bali** | 1,68 | 2,005 | 1,38 | 1,115 |
| **NusaTenggara Barat** | 5,335 | 3,8 | 3,59 | 3,55 |
| **NusaTenggara Timur** | 3,475 | 3,42 | 3,24 | 2,995 |
| **KalimantanBarat** | 4,965 | 4,405 | 4,29 | 4,205 |
| **KalimantanTengah** | 3,84 | 4,245 | 3,68 | 3,595 |
| **KalimantanSelatan** | 4,875 | 4,54 | 4,15 | 4,18 |
| **KalimantanTimur** | 7,335 | 8,405 | 7,73 | 6,75 |
| **KalimantanUtara** | 5,735 | 4,575 | 5,355 | 4,95 |
| **SulawesiUtara** | 8,86 | 7 | 6,65 | 6,475 |
| **Sulawesi Tengah** | 3,545 | 3,375 | 3,39 | 3,31 |
| **SulawesiSelatan** | 5,88 | 4,955 | 5,19 | 5,365 |
| **SulawesiTenggara** | 4,585 | 3,25 | 3,22 | 3,025 |
| **Gorontalo** | 3,855 | 3,32 | 3,965 | 3,825 |
| **SulawesiBarat** | 2,58 | 3,025 | 3,095 | 2,805 |
| **Maluku** | 8,325 | 7,015 | 8,53 | 7,325 |
| **Maluku Utara** | 5,805 | 3,72 | 5,075 | 4,71 |
| **PapuaBarat** | 6,345 | 6,595 | 7,005 | 5,985 |
| **Papua** | 3,855 | 3,16 | 3,79 | 3,055 |

Tabel 1 merupakan data tingkat pengangguran mulai dari tahun 2015-2018, dan berikut adalah data tenaga kerja yang digunakan.

**Tabel 2. Data Tenaga Kerja**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Provinsi | 2015 | 2016 | 2017 | 2018 |
| **Aceh** | 39,5 | 42,83 | 40,42 | 40,77 |
| **SumateraUtara** | 39,86 | 40,02 | 42 | 43,28 |
| **SumateraBarat** | 35,59 | 38,2 | 35,61 | 35,48 |
| **Riau** | 51,36 | 46,37 | 47,3 | 45,3 |
| **Jambi** | 41,35 | 37,91 | 41,47 | 44,13 |
| **SumateraSelatan** | 38,69 | 39,29 | 37,6 | 38,76 |
| **Bengkulu** | 31,19 | 34,08 | 33,61 | 31,83 |
| **Lampung** | 29,17 | 29,9 | 29,71 | 29,2 |
| **Kep,BangkaBelitung** | 45,87 | 44,43 | 51,62 | 50,12 |
| **Kep,Riau** | 71,9 | 66,2 | 68,51 | 67,14 |
| **DkiJakarta** | 72,96 | 68,71 | 71,55 | 69,82 |
| **JawaBarat** | 49,61 | 51,36 | 50,22 | 51,78 |
| **Jawa Tengah** | 38,28 | 37,83 | 39,71 | 39,06 |
| **Di Yogyakarta** | 48,79 | 45,09 | 43,45 | 45,59 |
| **JawaTimur** | 36,81 | 37,88 | 38,65 | 38,08 |
| **Banten** | 61,46 | 61,51 | 58,07 | 56,92 |
| **Bali** | 47,2 | 45,68 | 50,2 | 50,37 |
| **NusaTenggara Barat** | 26,8 | 26,89 | 26,38 | 28,22 |
| **NusaTenggara Timur** | 21,4 | 25,2 | 24,14 | 23,96 |
| **KalimantanBarat** | 34,44 | 38,19 | 38,57 | 40,52 |
| **KalimantanTengah** | 45,56 | 41,35 | 45,67 | 48,61 |
| **KalimantanSelatan** | 41,48 | 37,81 | 38,98 | 40,71 |
| **KalimantanTimur** | 60,34 | 55,25 | 59,96 | 59,91 |
| **KalimantanUtara** | 56,82 | 56,98 | 57,76 | 57,83 |
| **SulawesiUtara** | 40,45 | 38,68 | 44,54 | 44,67 |
| **Sulawesi Tengah** | 33,37 | 33,55 | 31,94 | 33,36 |
| **SulawesiSelatan** | 36,95 | 35,11 | 37,63 | 36,99 |
| **Sulawesi Tenggara** | 30,48 | 31,47 | 36,29 | 37,01 |
| **Gorontalo** | 37,14 | 39,69 | 39,49 | 38,74 |
| **SulawesiBarat** | 26,32 | 29,85 | 31,58 | 25,53 |
| **Maluku** | 31,98 | 35,76 | 37,18 | 35,87 |
| **Maluku Utara** | 31,4 | 35,97 | 38,05 | 33,24 |
| **PapuaBarat** | 39,45 | 40,57 | 40,9 | 46,16 |
| **Papua** | 18,51 | 21,63 | 21,81 | 22,37 |

Selanjutnya masuk dalam tahap perhitungan menggunakan metode *k-means*:

1. Penetuan Pusat Cluster Awal

Penentuan titik *cluster* ini dilakukan dengan mengambil nilai terbesar (maksimum)

untuk (C1), nilai terkecil untuk (C2) sebagai berikut:

**Tabel 3. Cluster Awal Data Pengangguran**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Cluster** | **2015** | **2016** | **2-17** | **2018** |
| **Tinggi** | 9,065 | 8,73 | 8,53 | 8,165 |
| **Rendah** | 1,68 | 2,005 | 1,38 | 1,115 |

1. Perhitungan Jarak Cluster

Untuk menghitung jarak antara data dengan pusat *cluster* menggunakan persamaan

(1)(2) berikut adalah proses menghitung jarak antara data dengan pusat *cluster*:

Iterasi 1:

D(Aceh,1) = (7,85−8,73)2+

(6,98−8,53)2+(6,455−8,165)2  =2,481174

D(Aceh,2) =(7,85−8,73)2+

(6,98−8,53)2+(6,455−8,165)2 =12,04832

D(SumUt,1) =

(6,005−1,38)2+(5,575−1,115)2 = 5,09786

D(SumUt,,2) =

(6,005−1,38)2+(5,575−1,115)2 = 9.072195

Perhitungandilakukanhinggaprovinsiterakhirbaikuntukclustertinggimaupunrendah.Berikutadalahhasildari perhitunganjarakclusteruntukiterasi 1.

### Tabel4.HasilIterasi1DataPengangguran

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Provinsi | Tinggi | Rendah | Jarak | C1 | C2 |
| **Aceh** | 2,481174 | 12,04832 | 2,481174 | 1 |  |
| **SumateraUtara** | 5,09786 | 9,072195 | 5,09786 | 1 |  |
| **SumateraBarat** | 5,70551 | 8,530589 | 5,70551 | 1 |  |
| **Riau** | 4,324321 | 9,896088 | 4,324321 | 1 |  |
| **Jambi** | 9,594092 | 4,639865 | 4,639865 |  | 1 |
| **SumateraSelatan** | 8,339789 | 5,99188 | 5,99188 |  | 1 |
| **Bengkulu** | 10,24179 | 3,957872 | 3,957872 |  | 1 |
| **Lampung** | 8,544386 | 5,650586 | 5,650586 |  | 1 |
| **Kep,BangkaBelitung** | 8,770027 | 5,418194 | 5,418194 |  | 1 |
| **Kep,Riau** | 2,671236 | 11,71072 | 2,671236 | 1 |  |
| **DkiJakarta** | 4,495348 | 9,92418 | 4,495348 | 1 |  |
| **JawaBarat** | 0,534462 | 13,81713 | 0,534462 | 1 |  |
| **Jawa Tengah** | 8,107859 | 6,109372 | 6,109372 |  | 1 |
| **Di Yogyakarta** | 10,79319 | 3,613904 | 3,613904 |  | 1 |
| **JawaTimur** | 8,982988 | 5,200954 | 5,200954 |  | 1 |
| **Banten** | 0,296057 | 14,00758 | 0,296057 | 1 |  |
| **Bali** | 14,16294 | 0 | 0 |  | 1 |
| **NusaTenggara Barat** | 9,160766 | 5,233964 | 5,233964 |  | 1 |
| **NusaTenggara Timur** | 10,68444 | 3,495461 | 3,495461 |  | 1 |
| **KalimantanBarat** | 8,31714 | 5,879407 | 5,879407 |  | 1 |
| **KalimantanTengah** | 9,582445 | 4,596042 | 4,596042 |  | 1 |
| **KalimantanSelatan** | 8,377161 | 5,805289 | 5,805289 |  | 1 |
| **KalimantanTimur** | 2,395986 | 12,04221 | 2,395986 | 1 |  |
| **KalimantanUtara** | 6,983536 | 7,318181 | 6,983536 | 1 |  |
| **SulawesiUtara** | 3,070086 | 11,53278 | 3,070086 | 1 |  |
| **Sulawesi Tengah** | 10,44687 | 3,770046 | 3,770046 |  | 1 |
| **SulawesiSelatan** | 6,587143 | 7,676008 | 6,587143 | 1 |  |
| **SulawesiTenggara** | 10,23311 | 4,125864 | 4,125864 |  | 1 |
| **Gorontalo** | 9,802399 | 4,526166 | 4,526166 |  | 1 |
| **SulawesiBarat** | 11,52697 | 2,765452 | 2,765452 |  | 1 |
| **Maluku** | 2,04803 | 12,60725 | 2,04803 | 1 |  |
| **Maluku Utara** | 7,720217 | 6,821576 | 6,821576 |  | 1 |
| **PapuaBarat** | 4,362872 | 9,908978 | 4,362872 | 1 |  |
| **Papua** | 10,33193 | 3,954282 | 3,954282 |  | 1 |

Berdasarkan hasil pada tabel di atas maka didapatkan pengelompokkan sebagai berikut :

C1= Aceh, Sumatera Utara, SumateraBarat, Riau, Kep.Riau, DkiJakarta, JawaBarat, Banten, KalimantanTimur, KalimantanUtara, SulawesiUtara, SulawesiSelatan, Maluku, PapuaBarat.

C2 = Jambi, Sumatera Selatan, Bengkulu, Lampung, Kep,Bangka Belitung, JawaTengah, Di Yogyakarta, Jawa Timur, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Sulawesi Tengah, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku Utara, Papua.

1. Penetuan pusatclusterbaru

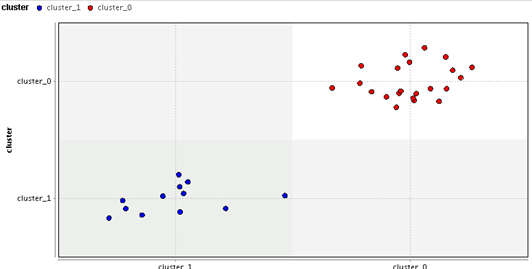
Setelah didapatkan hasil dari setiap *cluster* kemudian pusat *cluster* baru dihitungberdasarkan data member tiap – tiap *cluster* yang sudah didapatkan menggunakanpersamaan(2)yangsesuaidenganpusatmember*cluster* sebagaiberikut:

### Tabel5.PusatClusterIterasi2

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Cluster** | **2015** | **2016** | **2017** | **2018** |
| **Tinggi** | 7,472857 | 6,868929 | 6,788929 | 6,376071 |
| **Rendah** | 4,213 | 3,7305 | 3,69225 | 3,53375 |

Iterasi selanjutnya dilakukan dengan cara yang sama dengan iterasi 1 dan begitu juga untuk pengelompokan data tenaga kerja. Dari proses pengelompokan data tingkat pengangguran dengan 4 iterasi dan data tenaga kerja dengan 2 iterasi, diperopleh hasil pengelompokan sebagai berikut.

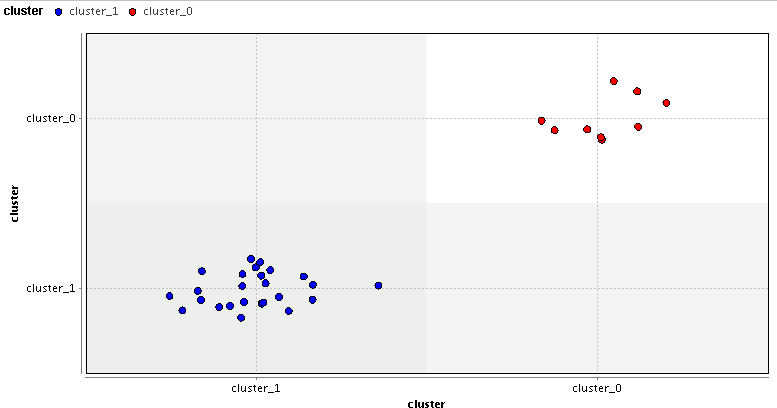
1. Hasil Pengelompokan data tingkat pengangguran diperoleh 12 provinsi pada cluster tinggi, dan 22 provinsi pada cluster rendah.

Gambar 1. Hasil Cluster Menggunakan Rapidminer Data TingkatPengangguran

Cluster tinggi: Aceh, Sumatera Utara, Sumatera Barat, Riau, Kep.Riau, Dki Jakarta, Jawa Barat, Banten, Kalimantan Timur, Sulawesi Utara, Maluku, Papua Barat.

Cluster rendah : Jambi, Sumatera Selatan, Bengkulu, Lampung, Kep,Bangka Belitung, Jawa Tengah, Di Yogyakarta, Jawa Timur, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku Utara, Papua.

1. Hasil Pengelompokan data tenaga kerja diperoleh 9 provinsi pada cluster tinggi, dan 25 provinsi pada cluster rendah.



### Gambar2. Hasil Cluster Menggunakan Rapid miner Data Tenaga Kerja

Cluster tinggi: Riau, Kep.Riau, Kep,Bangka Belitung, Dki Jakarta, Jawa Barat, Banten, Bali, Kalimantan Timur, Kalimantan Utara.

Cluster rendah : Aceh, Sumatera Utara, Sumatera Barat, Jambi, Sumatera Selatan, Bengkulu, Lampung, Jawa Tengah, Di Yogyakarta, Jawa Timur, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Papua Barat, Papua.

Berdasarkan hasil diatas, dapat dilihat bahwa jumlah provinsi dengan tingkat pengangguran lebih besar 57% daripada jumlah provinsi dengan tenaga kerja dengan persentasi 43%.

## Bahasa Pemrograman *Python*

*Python* adalah bahasa pemrograman *interpretatif multiguna* dengan filosofi perancangan yang berfokus pada tingkat keterbacaan kode. *Python* diklaim sebagai bahasa yang menggabungkan kapabilitas, kemampuan, dengan sintaksis kode yang sangat jelas, dan dilengkapi dengan fungsionalitas pustaka standar yang besar serta komprehensif. *Python* juga didukung oleh komunitas yang besar.

Python mendukung multi paradigma pemrograman, utamanya; namun tidak dibatasi; pada pemrograman berorientasi objek, pemrograman imperatif, dan pemrograman fungsional. Salah satu fitur yang tersedia pada python adalah sebagai bahasa pemrograman dinamis yang dilengkapi dengan manajemen memori otomatis. Seperti halnya pada bahasa pemrograman dinamis lainnya, python umumnya digunakan sebagai bahasa script meski pada praktiknya penggunaan bahasa ini lebih luas mencakup konteks pemanfaatan yang umumnya tidak dilakukan dengan menggunakan bahasa script. Python dapat digunakan untuk berbagai keperluan pengembangan perangkat lunak dan dapat berjalan di berbagai platform sistem operasi.

Saat ini kode python dapat dijalankan di berbagai platform sistem operasi, beberapa diantaranya adalah:

1. Linux/Unix
2. Windows
3. Mac OS X
4. Java Virtual Machine
5. Amiga
6. Palm
7. Symbian (untuk produk-produk Nokia)

*Python* didistribusikan dengan beberapa lisensi yang berbeda dari beberapa versi. Namun pada prinsipnya *Python* dapat diperoleh dan dipergunakan secara bebas, bahkan untuk kepentingan komersial. Lisensi *Python* tidak bertentangan baik menurut definisi *Open Source* maupun *General Public License* (GPL)( Akbar Nur Syahrudin dan Tedi Kurniawan,2018).