

## BAB II

### TINJUAN PUSTAKA

#### 2.1 *Coronavirus Disease 2019 (COVID-19)*

COVID-19 merupakan jenis penyakit yang diakibatkan dari suatu virus baru dimana sebelumnya belum pernah teridentifikasi pada manusia. *Coronavirus* adalah kelompok virus yang ditemukan pada hewan dan manusia yang disebabkan oleh infeksi *Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2)* dimana pertama kali ditemukan di Wuhan, China pada Desember 2019. Umumnya, penyakit ini menimbulkan gejala ringan seperti pilek, batuk, sakit tenggorokan, dan demam. Namun pada beberapa orang, gejalanya dapat lebih parah hingga menimbulkan radang paru-paru dan kesulitan bernapas. *Coronavirus* menyebar dengan cepat dan menyebar di banyak negara termasuk Indonesia (Sarmigi, 2020).

*Coronavirus* pertama kali ditemukan di pasar hewan dan makanan laut Kota Wuhan, dimana saat itu mulai banyak laporan mengenai orang-orang yang terinfeksi virus ini. Pasar tersebut menjual hewan-hewan liar, seperti ular, kelelawar, dan lainnya. Dugaan awalnya, *coronavirus* menyebar dari hewan ke manusia kemudian menginfeksi antar manusia. *Coronavirus* menyebar dari manusia ke manusia lain melalui tetesan pernapasan dari hidung atau mulut ketika seseorang batuk, bersin, atau menghembuskan napas. Tetesan ini dapat menempel ke benda-benda yang disentuh oleh orang yang terinfeksi dan bisa menularkan ke orang lain ketika menyentuh benda tersebut (WHO, 2020).

Kemenkes RI Nomor HK.01.07/MENKES/413/2020 tentang Pedoman Pencegahan. dan Pengendalian *Coronavirus Disease 2019* (COVID-19), menyatakan virus ini dapat mati dalam rentang waktu lima hingga tujuh hari, Masa inkubasi *coronavirus* paling pendek berlangsung selama dua sampai tiga hari sedangkan paling lama dapat mencapai 12 hari. Rentang waktu tersebut dibutuhkan oleh virus untuk menjangkiti inangnya dan menampakkan gejala-gejala awal, namun masih sulit untuk dideteksi. *Coronavirus* sensitif terhadap panas dengan suhu sekitar 56 derajat *celcius* selama 30 menit. Hingga saat ini, *coronavirus* belum dapat diobati dengan penanganan media apapun. Meskipun demikian, *coronavirus* yang masuk ke tubuh manusia sebenarnya dapat mati dalam rentang waktu lima hingga tujuh hari dengan sistem imun tubuh yang baik, sehingga virus tidak mudah menyebar ke seluruh anggota tubuh. Pencegahan yang bisa dilakukan untuk sekarang dan dinilai efektif yaitu dengan melakukan vaksinasi COVID-19.

## 2.2 **Vaksinasi COVID-19**

Vaksin merupakan produk biologi yang mengandung antigen yang mana ketika diberikan kepada manusia, secara aktif akan mengembangkan sistem kekebalan tubuh khusus terhadap penyakit tertentu (COVID-19 Komite Penanganan, 2020). Negara-negara yang terdampak pandemi COVID-19, termasuk Indonesia sedang mengembangkan vaksin yang cocok untuk pencegahan infeksi *SARS-CoV-2*, yaitu vaksin virus yang dilemahkan, vaksin hidup dilemahkan, vaksin vektor virus, vaksin asam nukleat, seperti virus. Tujuan dibuatnya vaksin adalah untuk mengurangi penyebaran COVID-19, menurunkan

angka kesakitan dan kematian akibat COVID-19, mencapai imunitas kelompok dan melindungi masyarakat dari COVID-19, sehingga dapat menjaga produktivitas sosial dan ekonomi (Kemenkes RI Dirjen P2P, 2020c).

Menurut Menteri Kesehatan, vaksin COVID-19 memiliki tiga manfaat, termasuk di dalamnya untuk menambah kekebalan setiap orang yang divaksinasi secara langsung. Jika penduduk yang divaksinasi jumlahnya banyak, maka sistem kekebalan penduduk akan memberikan perlindungan bagi mereka yang belum divaksinasi atau belum menjadi populasi sasaran vaksin (Yudho Winanto, 2020). Pemerintah telah membagi populasi sasaran vaksin menjadi lima kelompok masyarakat, yaitu kelompok lansia, kelompok masyarakat umum dan rentan, kelompok petugas publik, kelompok SDM kesehatan, dan kelompok usia 12-17 tahun. Masing-masing kelompok masyarakat sudah ditetapkan target dan tahapan vaksinasinya beserta jenis vaksin apa yang akan diberikan.

#### **a. Tahapan Vaksinasi**

Melalui laman situs resmi [covid19.go.id](https://covid19.go.id), pemerintah telah mengeluarkan petunjuk teknis terkait vaksinasi COVID-19 yang ada dalam Keputusan Direktur Jenderal Pencegahan dan Pengendalian Penyakit Nomor HK.02.02/4/1/2021 tentang Petunjuk Teknis Pelaksanaan Vaksinasi dalam Rangka Penanggulangan Pandemi COVID-19. Adapun tahapan yang dilakukan pemerintah sebagai berikut:

##### 1) Tahap 1 (Januari-April 2021)

Sasaran vaksinasi COVID-19 tahap pertama antara lain SDM Kesehatan, seperti tenaga kesehatan, asisten tenaga kesehatan, tenaga penunjang serta

mahasiswa yang sedang menjalani pendidikan profesi kedokteran yang bekerja pada Fasilitas Pelayanan Kesehatan (Fasyankes).

2) Tahap 2 (Januari-April 2021)

Sasaran vaksinasi COVID-19 tahap kedua yaitu petugas pelayanan publik, seperti Tentara Nasional Indonesia (TNI)/Kepolisian Negara Republik Indonesia (POLRI), aparat hukum, dan petugas pelayanan publik lainnya yang meliputi petugas di bandara/ pelabuhan/ stasiun/ terminal, perbankan, perusahaan listrik negara, dan perusahaan daerah air minum, serta petugas lain yang terlibat secara langsung memberikan pelayanan kepada masyarakat. Tahap kedua ini juga ditargetkan kepada kelompok usia lanjut yang berusia  $\geq 60$  tahun.

3) Tahap 3 (April 2021-Maret 2022)

Sasaran vaksinasi COVID-19 tahap ketiga ditujukan kepada masyarakat umum dan rentan dari aspek geospasial, sosial, dan ekonomi.

4) Tahap 4 (April 2021-Maret 2022)

Sasaran vaksinasi COVID-19 tahap keempat diberikan kepada kelompok usia 12-17 tahun dan masyarakat pelaku perekonomian lainnya dengan pendekatan klaster sesuai dengan ketersediaan vaksin.

**b. Jenis-Jenis Vaksin COVID-19**

Menteri Kesehatan, Terawan Agus Putranto menyatakan bahwa pemerintah telah menetapkan enam jenis vaksin COVID-19 yang akan digunakan di Indonesia (Kemenkes RI, 2020a). Di antaranya ialah:

1) Vaksin Merah Putih

Vaksin Merah Putih adalah hasil kerjasama BUMN PT Bio farma (Persero) dengan Lembaga Eijkman dan dengan perusahaan vaksin *China Sinovac Biotech*. Pemerintah mengharapkan vaksin ini dapat selesai pada akhir 2021.

2) *AstraZeneca*

Pengujian vaksin *AstraZeneca* yang dilakukan oleh perusahaan farmasi *AstraZeneca* dan *Oxford University* menunjukkan efisiensi rata-rata produksi vaksin *coronavirus* sebesar 70%. Saat ini, tahap uji coba masih berlanjut pada 20.000 relawan. Vaksin ini dianggap mudah untuk dikeluarkan karena tidak perlu disimpan pada suhu yang sangat dingin.

3) *China National Pharmaceutical Group Corporation (Sinopharm)*

Vaksin *Sinopharm* masih dalam tahap pengujian, namun telah digunakan oleh lebih dari satu juta penduduk China berdasarkan izin penggunaan darurat. Sebelum *Sinopharm* terbukti sukses, maka vaksin ini hanya digunakan untuk pejabat China, pekerja keliling, dan juga pelajar. Negara pertama di luar China yang menyetujui penggunaan vaksin ini adalah Uni Emirat Arab pada September 2020.

4) *Moderna*

*Moderna* mengklaim tingkat efektivitas produksi vaksinnnya sebesar 94,5%. Pada akhir November, *Moderna* mengaku telah mengajukan permohonan penggunaan darurat vaksin COVID-19 ke badan regulasi di AS dan Eropa.



*Moderna* yakin bahwa vaksinnnya memenuhi syarat penggunaan darurat yang ditetapkan oleh *Food and Drug Administration (FDA)* AS.

5) *Pfizer Inc and BioNTech*

Vaksin *Pfizer Inc and BioNTech* telah menyaraknkn BPOM di AS dan Eropa untuk segera menggunakan vaksin *coronavirus* mereka. Pada uji coba terakhir tanggal 18 November 2020, mereka mengklaim bahwa 95% vaksin tersebut efektif melawan *coronavirus* dan tidak ada bahaya keamanan.

6) *Sinovac Biotech Ltd*

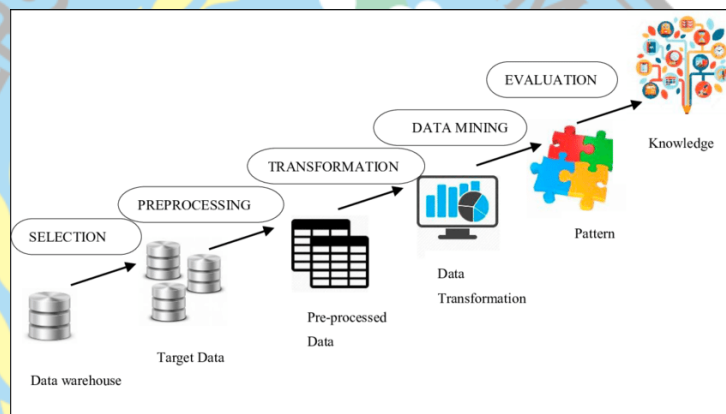
Saat ini, *CoronaVac* sedang memasuki uji coba fase tiga. *Sinovac* sedang menguji vaksinnnya di Brazil, Indonesia, dan Bangladesh. Seperti yang ditunjukkan pada hasil awal pada monyet yang dipublikasikan di jurnal *Science*, antibodi yang dihasilkan oleh vaksin tersebut dapat menetralkan 10 *strain Sars-CoV-2*.

### 2.3 **Data Mining**

*Data Mining* adalah rangkaian kegiatan untuk menemukan pola yang menarik dari data dalam jumlah besar, kemudian data-data tersebut dapat disimpan dalam *database*, *data warehouse*, atau penyimpanan informasi. Konsep *Data Mining* sudah banyak diterapkan dalam berbagai macam bidang keilmuan, seperti pendidikan, pemerintahan, kesehatan, dan sebagainya. Pada umumnya, *Data Mining* dapat diaplikasikan dari banyaknya jumlah data yang tersimpah dalam suatu *database* untuk digali informasinya guna memperoleh suatu pengetahuan yang bermanfaat. *Data Mining* adalah studi yang mengumpulkan,

membersihkan, mengolah, menganalisis, dan memperoleh manfaat dari data (Charu & Angarwal, 2015).

*Data Mining* merupakan bagian internal dari *Knowledge Discovery in Database (KDD)*, yaitu sebuah langkah dalam proses mencari pola-pola yang terdapat dalam setiap informasi. Langkah-langkah untuk menggambarkan proses KDD dalam menghasilkan *knowledge* dan terdiri dari beberapa tahapan, antara lain:



**Gambar 2.1 Proses KDD**

Sumber: sis.binus.ac.id

### 1. *Data Selection*

Seleksi atau pemilihan data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam KDD dimulai. Data yang didapatkan dari proses seleksi digunakan untuk proses *Data Mining*, kemudian disimpan dalam berkas terpisah dari baris data operasional. *Data selection* merupakan proses meminimalkan jumlah data untuk proses *Data Mining* dengan tetap merepresentasikan data aslinya. *Data Selection* dapat berupa *sampling*, *denoising*, dan *feature extraction* (Sulastri & Gufroni, 2017).

## 2. *Pre-processing/ Cleaning*

Proses *cleaning* pada menjadi fokus KDD sebelum proses *Data Mining* dilakukan. *Cleaning* mencakup proses membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data, serta dilakukan proses *enrichment*, yaitu proses "memperkaya" data yang sudah ada dengan data atau informasi lain yang relevan dan diperlukan untuk KDD.

## 3. *Transformation*

*Transformation* merupakan deskripsi tentang *coding*. *Coding* yaitu proses *transformasi* pada data yang telah dipilih sehingga data tersebut dapat sesuai untuk proses *Data Mining*. Proses *coding* dalam KDD merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam basis data dan merupakan proses perubahan atau penggabungan data ke dalam format yang sesuai untuk diproses dalam *Data Mining*. Beberapa metode *Data Mining* membutuhkan format data yang khusus sebelum bisa diaplikasikan. Contohnya, beberapa metode standar seperti analisis asosiasi dan *clustering* yang hanya dapat menerima input data kategorikal. Maka dari itu, data yang berupa angka numerik yang berlanjut perlu dibagi-bagi menjadi beberapa interval (Eska, 2016a).

## 4. *Data Mining*

*Data Mining* adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode, dan algoritma dalam *Data Mining* sangat bervariasi, karena itu pemilihannya



yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan. Data yang telah dikumpulkan harus dianalisis, diproses, dan diubah ke dalam hasil yang dapat menginformasikan, menginstruksi, menjawab, ataupun memberikan pemahaman, dan pembuatan keputusan.

#### 5. *Interpretation/ Evaluation*

Pola informasi yang dihasilkan dari proses *Data Mining* perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. *Interpretation* mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya. Evaluasi pola digunakan untuk mengidentifikasi pola-pola menarik kedalam *knowledge based* yang ditemukan. Pada tahap *interpretation/ evaluation*, hasil dari *Data Mining* berupa pola-pola yang khas maupun model prediksi dievaluasi untuk menilai apakah tercapainya hipotesa yang ada (Eska, 2016b).

*Data Mining* memiliki lima peran utama, yaitu estimasi, prediksi, klasifikasi, klusterisasi, dan asosiasi. Peran *Data Mining* yang sering digunakan adalah klasifikasi dan klusterisasi karena dapat digunakan untuk atribut yang banyak (Tan, et. al., 2014b). Penjelasan lebih lengkap mengenai lima peran pada *Data Mining*, sebagai berikut:

##### a) Estimasi

Estimasi mirip seperti klasifikasi, yang membedakan adalah estimasi memuat variabel target berupa numerik sedangkan klasifikasi berupa kategorik.

Model estimasi dibangun menggunakan *record* lengkap yang menyediakan nilai variabel target sebagai nilai prediksi.

b) Prediksi

Prediksi berguna untuk memperkirakan nilai di masa mendatang. Prediksi dapat dilakukan jika ada data historisnya atau data yang telah terjadi sebelumnya. Kemudian dengan metode yang diinginkan, diterapkan pada data tersebut untuk mengetahui kebiasaan yang terjadi sehingga dapat memperkirakan nilai yang belum terjadi di masa mendatang.

c) Klasifikasi

Klasifikasi memuat variabel target berupa kategorik, misalnya penggolongan pendapatan yang dipisahkan dalam tiga kategori, yaitu tinggi, sedang, dan rendah. Klasifikasi berguna untuk memprediksi suatu data akan tergolong ke dalam kategori tertentu.

d) Klasterisasi

Klasterisasi merupakan pengelompokan *record*, pengamatan, atau memperhatikan dan membentuk kelas objek-objek yang memiliki kemiripan. Umumnya, klasterisasi melibatkan penghitungan jarak yang dilakukan untuk mengetahui tingkat kemiripan yang berada dalam data.

e) Asosiasi

Asosiasi atau lebih dikenal sebagai analisis keranjang belanja, berguna untuk mendeteksi kumpulan atribut yang muncul secara bersamaan (*co-occur*)

dalam frekuensi yang sering, dan membentuk sejumlah kaidah dari kumpulan-kumpulan tersebut.

## 2.4 *Clustering*

*Clustering* adalah salah satu metode eskplorasi data yang digunakan untuk mencari pola yang ada pada suatu dataset. Umumnya, pola tersebut dapat dilihat dari kesamaan sifat, karakteristik, atau ciri dari *record-record* pada dataset (Yusuf dkk., 2012a). Proses *clustering* akan mengelompokkan item data ke dalam sejumlah kecil kelompok sedemikian sehingga setiap kelompok mempunyai suatu kesamaan yang esensial dimana nantinya akan mempermudah dalam pencarian data berdasarkan kesamaan yang ada. Dengan kata lain, sebuah *cluster* memiliki sekumpulan objek yang digabung bersama karena persamaan atau kedekatannya. Pengelompokan atau *clustering* adalah salah satu metode yang sifatnya tanpa arahan (*unsupervised*) (Putra dkk., 2018).

*Clustering* tidak memuat variabel target, hal inilah yang membedakannya dengan klafisikasi. *Clustering* memiliki banyak metode yang telah dikembangkan oleh para ahli. Masing-masing metode mempunyai karakter, kelebihan, dan kekurangan. Metode *clustering* menurut strukturnya dibagi menjadi dua, yaitu pengelompokan hirarki dan partitioning. Pengelompokan hirarki memiliki aturan dimana satu data tunggal bisa dianggap sebagai sebuah kelompok, dua tau lebih kelompok kecil dapat digabung menjadi satu kelompok besar, dan begitu seterusnya hingga semua data dapat bergabung menjadi satu kelompok. Sedangkan metode *clustering* partitioning mengelompokkan data kedalam sejumlah kelompok yang tidak tumpang tindih (*overlap*) antara satu kelompok

dengan kelompok lain, artinya setiap data hanya menjadi anggota satu kelompok (Salsabila, 2019a).

*Clustering* memiliki suatu analisa statistik untuk memisahkan objek kedalam beberapa kelompok yang mempunyai sifat berbeda antar kelompok yang satu dengan yang lain. Pada analisa *cluster* ini, tiap-tiap kelompok bersifat homogen antar anggota kelompok atau variasi objek dalam kelompok yang terbentuk sekecil mungkin. Tujuan utama analisa *cluster* adalah menggabungkan objek-objek yang mempunyai kesamaan kedalam sebuah kelompok. Pengambilan keputusan dengan analisa cluster memiliki enam tahapan, yaitu menentukan tujuan analisa *cluster*, menentukan desain penelitian analisa *cluster*, menentukan asumsi analisa *cluster*, menurunkan *cluster-cluster* dan memperkirakan, menginterpretasi hasil analisis *cluster*, dan mengukur tingkat validasi hasil analisis *cluster* (Salsabila, 2019b).

## 2.5 *Spectral Clustering*

*Spectral Clustering* merupakan algoritma untuk menemukan  $k$  *cluster* dengan menggunakan *eigenvector* dari suatu matriks. Matriks ini diturunkan dari matriks similaritas antara setiap data yang akan dikelompokkan secara berpasangan (Meila, 2016). *Spectral Clustering* adalah teknik yang populer karena kesederhanaannya, intuisi, dan kemampuannya untuk mengelompokkan titik data yang tidak dapat diakses secara linier. Disamping itu juga dapat memberikan hasil perhitungan yang sebanding atau lebih baik dari metode lainnya (Luxburg, 2007).

Pada pengertian lain, *Spectral Clustering* adalah salah satu metode *clustering* yang melakukan pengelompokan berdasarkan kesamaan setiap datanya

yang dilihat dari keterkaitan antara data satu dengan yang lainnya. Ada dua tahap penting harus dijalankan, yaitu (a) membuat graf dari *dataset* dan (b) melakukan *cluster* data dengan menemukan sebuah partisi optimal dari graf (Hong dkk., 2014). Graf tersebut memiliki verteks yang merupakan setiap *record* pada data dan *edge* berupa hubungan antar data yang biasanya berupa jarak dari dua *record* yang berhubungan (Yusuf dkk., 2012b). Perlu diketahui konsep dari suatu graf  $G$  didefinisikan sebagai pasangan himpunan  $(V, E)$ , ditulis dengan notasi  $G = (V, E)$ .  $V$  adalah himpunan tidak kosong dari simpul-simpul (*vertices*) dan  $E$  adalah himpunan sisi (*edges*) yang menghubungkan sepasang simpul (Munir, 2010).

Pada algoritma *Spectral Clustering*, terdapat beberapa metode berdasarkan perbedaan rumus graf *Laplacian* yang digunakan. Di antaranya adalah metode *Un-Normalized Laplacian* dan *Normalized Laplacian*. Perbedaan keduanya terletak pada proses penghitungan nilai matriks *laplacian*, dimana metode *Un-Normalized Laplacian* menggunakan rumus yang tidak ternormalisasi dan *Normalized Laplacian* menggunakan rumus yang telah ternormalisasi. Langkah-langkah dalam melakukan *Spectral Clustering* adalah sebagai berikut:

A. Membentuk *Similarity Matrix*

Matriks similaritas berguna untuk menunjukkan ada atau tidaknya hubungan antara data satu dengan yang lainnya. Apabila terdapat hubungan, maka terdapat nilai antara data satu dengan yang lainnya, sedangkan apabila tidak terdapat hubungan akan bernilai nol. Nilai diagonal pada matriks similaritas akan bernilai nol karena tidak ada hubungan terhadap data itu sendiri. Pengukuran similaritas atau kemiripan dapat dilakukan dengan menggunakan *k-Nearest Neighbor* (k-NN)



dengan perhitungan *Euclidean Distance* yang merupakan metode untuk mencari jarak antara dua titik data, dimana semakin kecil jarak tersebut maka suatu data akan semakin dekat dan mirip (Yudhana dkk, 2020). Perhitungan nilai similaritas dihitung dengan *Euclidean Distance* menggunakan rumus pada halaman berikutnya:

$$d(x, y) = W = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad 2.1$$

Dimana:

x dan y = data yang dibandingkan

Gambaran *similarity matrix*:

$$W = \begin{bmatrix} x_1y_1 & \cdots & x_1y_j \\ \vdots & \ddots & \vdots \\ x_iy_1 & \cdots & x_iy_j \end{bmatrix}$$

B. Membentuk *Degree Matrix*

*Degree Matrix* adalah sebuah matriks diagonal yang menggambarkan tingkat derajat simpul (jumlah sisi/edge yang terhubung dengan setiap data yang berupa simpul/vertex) dalam suatu graf. *Similarity matrix* dan *degree matrix* ini digunakan untuk membentuk *laplacian matrix* (Ramdan & Hermawan, 2012).

Rumus menghitung nilai diagonal adalah:

$$D_i = \sum_{j=1}^n W \quad 2.2$$

Dimana:

- D = matriks diagonal
- n = jumlah data
- i dan j = nomor baris dan kolom data

- $W$  = matriks similaritas

Gambaran *degree matrix*:

$$D_i = \begin{bmatrix} D_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & D_i \end{bmatrix}$$

C. Membentuk *Un-Normalized Laplacian Matrix* dan *Normalized Laplacian Matrix*

*Laplacian Matrix* merupakan matriks yang merepresentasikan suatu graf dan berguna untuk menemukan banyak *insights* yang bermanfaat dari graf. *Laplacian Matrix* terbagi menjadi beberapa metode, diantaranya adalah *Un-Normalized Laplacian* dan *Normalized Laplacian*. Perbedaan dari kedua metode tersebut terletak pada rumus yang diaplikasikan, jika *Un-Normalized Laplacian* menggunakan rumus yang tidak ternormalisasi maka *Normalized Laplacian* menggunakan rumus yang telah dinormalisasi. *Normalized Laplacian* disebut juga dengan *Symmetric Laplacian*. Berikut rumus *Laplacian Matrix* yang akan diaplikasikan:

$$L = D_i - W \tag{2.3}$$

$$L_{sym} = D_i^{-1/2} \cdot W \cdot D_i^{-1/2} \tag{2.4}$$

Dimana

- $L$  adalah nilai *Un-Normalized Laplacian*
- $L_{sym}$  adalah nilai *Normalized Laplacian*

Gambaran *Laplacian Matrix*:

$$L_{i,j} = \begin{bmatrix} L_{1,1} & \cdots & L_{1,j} \\ \vdots & \ddots & \vdots \\ L_{i,1} & \cdots & L_{i,j} \end{bmatrix}$$

#### D. Menghitung *Eigenvalue* dan *Eigenvector*

*Eigenvalue* dan *eigenvector* berguna untuk membantu memudahkan proses pengelompokan dari data yang mengandung banyak *outlier*/pencilan sehingga dapat menemukan pola data yang memiliki kemiripan (Rohmah & Sugiyarto, 2021). Penentuan seberapa banyak *Eigenvalue* dan *eigenvector* yang akan dihitung tergantung pada parameter jumlah cluster, disimbolkan dengan  $k$ . *Eigenvalue* dapat diperoleh dengan rumus:

$$\det(A - \lambda I) = 0 \quad 2.5$$

sedangkan *Eigenvector* diperoleh dengan:

$$(A - \lambda I)x = 0 \quad 2.6$$

Dimana:

- $A = \text{Laplacian Matrix}$
- $\lambda = \text{Eigenvalue}$
- $x = \text{Eigenvector}$
- $k = \text{parameter jumlah cluster}$

Gambaran *Eigenvector*:

$$\begin{bmatrix} x_{\lambda_{1,1}} & \cdots & x_{\lambda_{1,k}} \\ \vdots & \ddots & \vdots \\ x_{\lambda_{i,1}} & \cdots & L_{i,k} \end{bmatrix}$$

E. Data *eigenvector* kemudian dikelompokkan menggunakan *K-Means Clustering*.

## 2.6 *K-Means Clustering*

Algoritma *K-means Clustering* merupakan salah satu algoritma yang digunakan untuk partisi dimana pengolompokannya didasarkan pada penentuan jumlah awal kelompok dengan mendefinisikan nilai titik tengah (*center*) cluster direpresentasikan dengan nilai rata-rata (*mean*) dari objek dalam cluster. *K-Means* menggunakan proses secara berulang kali untuk mendapatkan basis data *cluster* (Putra dkk., 2018).

Langkah-langkah dalam melakukan *K-means Clustering* adalah sebagai berikut:

1. Menentukan banyak *cluster* yang ingin dibentuk.
2. Menentukan nilai awal sebagai pusat cluster awal (*centroid*) sebanyak yang diinginkan.
3. Menghitung jarak setiap data terhadap semua *centroid* menggunakan rumus *Euclidean Distance* pada rumus 2.1 sehingga mendapatkan jarak yang paling dekat antara data dengan *centroid*.
4. Memasukkan setiap data berdasarkan jarak dengan *centroid* dari kecil ke besar.
5. Membuat *centroid* baru yang diperoleh dengan menghitung rata-rata *cluster* menggunakan rumus:

$$\mu_j(t + 1) = \frac{1}{N_{sj}} \sum_{j \in S_j} x_j \quad 2.7$$

Dimana:

- $\mu_j(t + 1)$  adalah *centroid* baru pada iterasi  $(t + 1)$
  - $N_{sj}$  adalah banyak data pada *cluster*  $S_j$
6. Melakukan kembali langkah 2 sampai 5, sehingga data di dalam *cluster* tetap.

## 2.7 *Davies-Bouldin Index (DBI)*

*Davies-Bouldin Index* adalah metode untuk mengevaluasi cluster yang diperkenalkan oleh David L. Davies dan Donald W. Bouldin (Narwin, et al., 2017). Metode ini merupakan salah satu metode untuk mengukur validitas *cluster* pada suatu metode pengelompokan yang didasarkan atas skema evaluasi internal *cluster*, dimana skema ini memuat dua faktor yaitu kohesi dan separasi. Kohesi didefinisikan sebagai kedekatan suatu data terhadap titik pusat data dalam *cluster* yang sama, sedangkan separasi didasarkan pada jarak antar titik pusat *cluster* yang satu terhadap titik pusat *cluster* lainnya.

Pengukuran dengan *Davies-Bouldin Index* ini memaksimalkan jarak inter-*cluster* antara *cluster*  $C_i$  dan  $C_j$  dan pada waktu yang sama akan meminimalkan jarak antar titik dalam sebuah *cluster*. Apabila jarak kedekatan inter-*cluster* semakin maksimal, maka kesamaan karakteristik antar *cluster* akan semakin sedikit sehingga perbedaannya terlihat lebih jelas. Sebaliknya, apabila jarak intra-*cluster* semakin minimal, maka artinya tingkat kesamaan masing-masing objek dalam *cluster* tersebut semakin tinggi (Wani & Riyaz, 2017).