BAB 1

PENDAHULUAN

1.1 Latar Belakang

Teknologi mengalami perkembangan dari tahun ke tahun terutama di bidang teknologi informasi dan komunikasi. Informasi sudah menjadi bagian kebutuhan keseharian masyarakat. Oleh karena itu, tidak heran jika Kementerian Komunikasi dan Informatika menyebutkan dalam situs web resmi bahwa di era revolusi industri 4.0 ini menjadikan teknologi informatika dan telekomunikasi sebagai hal yang sangat dibutuhkan tidak hanya dalam kehidupan masyarakat tetapi juga dalam hal perkembangan ekonomi negara.

Dengan semakin berkembangnya teknologi, informasi yang dapat kita akses dan kita terima menjadi lebih mudah. Tidak heran jika di *website* banyak sekali informasi yang dapat diakses dengan menggunakan internet. Mulai dari informasi penting tentang pemerintahan hingga informasi keseharian seperti resep masakan. Dalam mengelola informasi tersebut, perlu dilakukan pengumpulan dan pengelompokan informasi dalam jumlah yang besar yang sering disebut dengan data.

Pada era teknologi sekarang, data sudah menjadi hal penting yang tidak bisa lepas dalam sebuah organisasi maupun perusahaan. Karena dengan data tersebut, kita bisa mengambil kesimpulan yang akan berguna untuk pengembangan dan kemajuan baik organisasi maupun perusahaan. Sudah

banyak perusahaan yang mencari seorang ahli di bidang data. Tentunya seorang ahli data dapat mengelola data sebaik mungkin dengan berbagai metode atau teknik dalam pengolahan data. Bahkan sudah banyak *software* yang dapat membantu dalam pengolahan data. Salah satu metode analisis yang digunakan oleh seorang ahli data adalah analisis klaster (pengelompokan).

Menurut Mattjik & Sumertajaya (2011), analisis klaster merupakan teknik peubah ganda yang memiliki tujuan untuk mengelompokkan objek-objek berdasarkan kesamaan karakteristik objek yang dimiliki. Salah satu metode klaster adalah metode *k-means*. Menurut Somasundaram & Mary Shanthi Rani (2012), metode *k-means* merupakan salah satu teknik partisi paling umum yang memiliki banyak aplikasi di bidang kompresi gambar dan video, segmentasi gambar, pengenalan pola dan *data mining*. Metode *k-means* merupakan metode populer yang memiliki algoritma membentuk klaster berdasarkan objek-objek yang saling berdekatan.

Dalam menentukan klaster terbaik untuk metode *k-means* diperlukan teknik lain. Sudah banyak teknik penentuan klaster terbaik, salah satu tekniknya adalah metode *Elbow*. Menurut Dewi & Pramita (2019), metode *Elbow* merupakan metode untuk menentukan jumlah klaster yang tepat melalui presentase hasil perbandingan antara jumlah klaster, dimana jumlah klaster yang terbaik berdasarkan titik yang membentuk siku pada grafik. Tetapi metode *Elbow* dapat digunakan setelah proses metode *k-means* selesai dan perlu perhitungan untuk setiap jumlah klaster yang ditentukan.

Dalam hal ini penulis ingin meneliti sebuah teknik untuk mengetahui jumlah klaster terbaik melalui data yang ada dan dapat dilakukan sebelum dimulainya proses metode *k-means*, sehingga tidak perlu perhitungan untuk setiap jumlah klaster yang ditentukan dengan metode *k-means*. Salah satu teknik tersebut adalah dengan mencari tingkat similaritas antar objek pada data dengan menggunakan fungsi kernel. Similaritas antar objek pada analisis pengelompokan adalah jarak terdekat antar suatu objek dengan objek lainnya berdasarkan indikator yang sama. Sehinngga semakin kecil

jarak antar suatu objek dengan objek lain, maka semakin tinggi tingkat similaritas antar dua objek tersebut. Menurut Cheng et al. (2017), fungsi kernel merupakan fungsi yang memproyeksikan data asli yang linier atau non linier ke dalam ruang fitur yang berdimensi tinggi, dimana semua data dapat disajikan secara linier. Kemudian jumlah klaster dapat ditentukan dengan mencari nilai eigen dan vektor eigen pada hasil fungsi kernel tersebut.

Untuk mengetahui penggunaan teknik tersebut pada semua kasus pengelompokan (*clustering*) pada metode *k-means*, perlu dilakukan analisis serta validasi untuk metode *k-means*. Validasi dilakukan dengan menggunakan *davies bouldin index*. Menurut Jaelanie (2016), DBI ini digunakan untuk mengevaluasi klaster dengan menghitung kuantitas dan fitur turunan dari kumpulan data. Dalam proses awal metode *k-means* terlebih dahulu menentukan jangkauan banyaknya klaster, jangkauan yang biasanya diambil rentang 2 sampai banyaknya data. Kemudian dilakukan proses analisis dengan metode *k-means*. Dari hasil metode *k-means* pada semua klaster, selanjutnya dilakukan evaluasi klaster terbaik dengan DBI. Hasil DBI terendah adalah hasil klaster terbaik. Proses ini efisien apabila jumlah klaster yang digunakan kecil, tetapi tidak efisien untuk jumlah klaster yang besar. Hal ini dikarenakan harus melakukan percobaan terhadap semua klaster yang mungkin dan membutuhkan proses pengulangan yang panjang dan lama. Maka dari itu diperlukan teknik yang dapat menentukan jumlah klaster sebelum menggunakan metode *k-means*.

Dalam menentukan jumlah klaster sebelum menggunakan *k-means*, dapat ditentukan berdasarkan nilai dan vektor eigen pada similaritas antar objek dengan menggunakan fungsi kernel. Pada analisis klaster, fungsi kernel dapat digunakan untuk mengetahui nilai similaritas antara masing-masing objek. Semakin tinggi nilai similaritas antar objek maka objek tersebut memiliki banyak kesamaan yang berarti memiliki kemungkinan untuk menjadi satu kelompok. Maka dari itu dalam penulisan ini mengambil judul "Penentuan Jumlah Klaster Pada Metode *K-Means* Berdasarkan Nilai Dan Vektor Eigen Dari Matriks Kernel".

1.2 Perumusan Masalah

Berdasarkan latar belakang, maka terdapat pertanyaan penelitian sebagai berikut:

1. Bagaimana jumlah klaster yang dihasilkan dengan nilai dan vektor eigen dari matriks kernel berdasarkan jumlah indikator pada data yang digunakan?

2. Apakah jumlah klaster yang dihasilkan berdasarkan nilai dan vektor eigen dari matriks kernel lebih baik dari jumlah klaster yang dihasilkan metode *Elbow*?

3. Bagaimana hasil perbandingan jumlah klaster yang dihasilkan dengan nilai dan vektor eigen dari matriks kernel dan metode *Elbow* berdasarkan nilai dari *Davies Bouldin Index*?

1.3 Tujuan Penelitian

Sesuai dengan rumusan masalah di atas, maka tujuan yang ingin dicapai dalam penelitian ini sebagai berikut:

1. Mengetahui jumlah klaster yang dihasilkan dari nilai dan vektor eigen dari matriks kernel berdasarkan jumlah indikator yang berbeda.

2. Mengetahui seberapa baik jumlah klaster yang dihasilkan berdasarkan nilai dan vektor eigen dari matriks kernel terhadap metode *Elbow*.

3. Mengetahui hasil perbandingan jumlah klaster yang dihasilkan dengan nilai dan vektor eigen dari matriks kernel dan metode *Elbow* berdasarkan nilai dari *Davies Bouldin Index*.

1.4 Manfaat Penelitian

Penelitian ini diharapkan dapat bermanfaat bagi pihak yang terkait, yaitu:

1. Bagi perusahaan

   Bagi perusahaan terutama pada divisi IT dapat dijadikan referensi dalam pengolahan data yang besar dalam upaya efisiensi proses analisis data terutama dalam analisis klaster.

2. Bagi peneliti selanjutnya

   Bagi peneliti selanjutnya dapat dijadikan referensi penelitian mengenai penentuan jumlah klaster.

UNIVERSITAS NEGERI JAKARTA