

BAB 1

PENDAHULUAN

1.1. Latar Belakang

Analisis *cluster* merupakan salah satu alat yang penting dalam pengolahan data statistik untuk melakukan analisis data. Analisis *cluster* merupakan seperangkat metodologi yang secara otomatis mengelompokkan pola/objek ke dalam sebuah *cluster* berdasarkan kemiripannya. Secara intuitif, pola/objek yang berada pada satu *cluster* lebih memiliki kemiripan antar satu sama lain dibandingkan pola/objek *cluster* lain. Analisis *cluster* memiliki aplikasi yang luas seperti data *mining* (penambangan data), pencarian informasi, biologi, kesehatan, *marketing* dan segmentasi gambar (Pande dkk, 2012).

Ahuja dan Bal (2014) mengungkapkan bahwa analisis *cluster* dapat digunakan sebagai sebuah alat untuk mendapatkan pengetahuan mengenai penyebaran/pola data atau dapat digunakan sebagai langkah *pra process* untuk algoritma lain. Analisis *cluster* juga dapat digunakan untuk menyimpulkan data sehingga menemukan kelompok yang "natural" atau "real". Secara umum terdapat dua jenis data dalam analisis *cluster* yang digunakan yakni data *metric* dan *non metric* atau kombinasi dari kedua jenis data tersebut sehingga akan menambah kerumitan data dalam analisis *cluster*. Analisis *cluster* dapat diaplikasikan pada berbagai masalah, salah satu diantaranya adalah riset pasar. Dalam riset pasar, analisis *cluster* dapat membantu bagian *marketing* untuk menemukan perbedaan antar konsumen berdasarkan kebiasaan membeli. Informasi perbedaan ini dapat digunakan sebagai bahan pertimbangan dalam membuat sistem rekomendasi untuk melakukan kegiatan promosi. Selain itu, analisis *cluster* juga dapat digunakan untuk mendeteksi adanya penyimpangan data yang ekstrim dari sekelompok data.

Seperti telah diungkapkan sebelumnya, dalam analisis *cluster* dapat menggunakan jenis data *metric*, *non metric* atau bahkan keduanya. Data *metric* merupakan data yang berupa angka. Sedangkan data *non metric* merupakan data yang tidak berupa angka. Dalam skala pengukuran, data *metric* terdiri dari interval dan rasio. Sedangkan data *non metric* terdiri dari nominal dan ordinal. Untuk mengolah data *non metric*, data tersebut harus diubah terlebih dahulu menjadi angka. Sebagai contoh, dalam sebuah kuisisioner terdapat pertanyaan mengenai jenis kelamin. Jenis kelamin terdiri dari laki-laki atau perempuan.

Ketika akan diolah datanya secara statistik, maka data ini harus diubah terlebih dahulu. Perubahan dilakukan dengan menggunakan angka sebagai simbol pembeda. Misal untuk menunjukkan jenis kelamin laki-laki disimbolkan dengan angka 1 dan untuk menunjukkan jenis kelamin perempuan disimbolkan angka 0. Ketika diaplikasikan dalam analisis *cluster*, data *non metric* yang digunakan sudah diubah menjadi bentuk angka. Akan tetapi sebenarnya, angka-angka ini hanyalah simbol yang tidak terukur. Perubahan ke dalam bentuk angka ini dilakukan agar dapat dilakukan perhitungan sesuai dengan metode analisis *cluster* yang digunakan.

Dalam melakukan analisis *cluster* juga terdapat teknik iterasi saat mengaplikasikan algoritma metode *clustering*. Teknik iterasi merupakan sebuah prosedur yang dilakukan secara berulang-ulang. Teknik iterasi dalam analisis *cluster* digunakan untuk mengumpulkan objek ke dalam sebuah kelompok yang sama. Jika analisis *cluster* diaplikasikan secara manual terutama dalam jumlah data yang banyak, tentunya akan membutuhkan waktu yang lama.

Saat ini terdapat sebuah *software* yang digunakan sebagai alat bantu dalam melakukan analisis *cluster*. *Software* SPSS merupakan salah satu *tools* analisis *cluster* yang sangat terkenal. Kelemahan dari *software* ini adalah hanya dapat melakukan analisis *cluster* dengan data yang bersifat *metric* atau merupakan data hasil kombinasi antara *metric* dan *non metric*. Hal ini dikarenakan hanya terdapat tiga metode yang dapat digunakan pada *software* ini. Secara garis besar, metode dalam analisis *cluster* terbagi menjadi dua yakni metode hirarki dan metode *non* hirarki. Diantara kedua metode tersebut, metode yang paling terkenal dan sering digunakan adalah metode *non* hirarki yakni *K-Means*. Pada *software* SPSS, hanya terdapat tiga metode analisis *cluster* yang dapat digunakan yakni *K-Means*, *Two-Step Cluster* dan metode hirarki. Menurut Agusta (2007), *K-Means* merupakan salah satu metode data *clustering non* hirarki yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih *cluster*/kelompok. Sedangkan Mongi (2015) mengungkapkan bahwa metode *Two-Step Cluster* merupakan suatu metode penggerombolan yang dapat mengatasi masalah skala pengukuran, data berukuran besar dengan peubah yang memiliki tipe data kategorik dan kontinyu serta mengetahui gerombol optimal yang terbentuk. Dapat disimpulkan metode *Two-Step Cluster* merupakan metode analisis *cluster* yang digunakan untuk menangani data yang bersifat campuran.

Permasalahan dari penggunaan *software* SPSS ini adalah adanya keterbatasan metode analisis *cluster* yang dapat digunakan. Ketiga metode analisis *cluster* yang ada pada *software* hanya dapat diaplikasikan pada data yang bersifat *metric*. Sebagai contoh, metode *K-Means* hanya dapat digunakan untuk mengelompokkan data yang bersifat *metric*. Hal ini dikarenakan pengukuran jarak yang digunakan adalah *Euclidean Distance* dimana merupakan pengukuran jarak untuk data *metric*. Bagaimana jika metode *K-Means* digunakan untuk mengelompokkan data yang bersifat *non metric*. Metode ini dapat digunakan akan tetapi akan memunculkan keraguan pada interpretasi *cluster*. Hal ini dikarenakan, data *non metric* yang telah diubah menjadi angka merupakan data yang menjadi simbol atau urutan dari sebuah atribut. Ketika dilakukan pengukuran jarak dengan menggunakan *Euclidean Distance*, angka-angka yang digunakan belum tentu menggambarkan jarak antar objek secara kuantitatif karena data *non metric* hanya menggambarkan sebuah objek.

Dyan (2010) melakukan pengelompokkan *handphone* dengan menggunakan *Basic Sequential Algorithmic Scheme* (BSAS). Pengelompokkan *handphone* ini dilakukan sebagai usulan perbaikan untuk menentukan tata letak *handphone* pada toko tersebut. Selain itu, dengan penyimpanan *handphone* yang sudah berada dalam satu kelompok membuat penjual mudah untuk merekomendasikan jenis *handphone* yang berbeda-beda pada konsumen. Data atribut dalam penelitian ini merupakan data hasil pengkategorisasian (*non metric*). Karena data yang digunakan bersifat *non metric* maka dikembangkan metode untuk melakukan pengelompokkan data tersebut dengan menggunakan *proximity matrix* sebagai pengukuran jarak antar vektor dan metode BSAS sebagai metode *clustering* yang digunakan. Dikarenakan belum terdapat *software* analisis *cluster* yang menggunakan metode BSAS, maka perhitungan dalam penelitian ini dilakukan secara manual. Kelemahan dari perhitungan secara manual ini adalah dari segi waktu dan ketelitian. Semakin banyak data yang dikelompokkan tentunya semakin lama pula waktu yang dibutuhkan untuk melakukan perhitungan. Ketelitian juga penting dalam perhitungan karena akan mempengaruhi hasil dari *cluster* yang terbentuk.

Basic Sequential Algorithmic Scheme (BSAS) merupakan salah satu metode analisis *cluster* yang paling dasar dan mudah digunakan. Theodoridis dkk (2003) mengungkapkan bahwa metode ini merupakan metode yang cepat dan mudah. *Sequential clustering algorithms* sendiri terbagi menjadi tiga jenis yakni *Basic*

Sequential Algorithmic Scheme (BSAS), *Modified Basic Sequential Algorithmic Scheme (MBSAS)* dan *A Two – Threshold Sequential Scheme*. Dari ketiga jenis algoritma ini, yang paling dasar adalah BSAS. Dalam BSAS juga dapat mengatasi kelemahan dari metode hirarki dan non hirarki. Metode BSAS tidak memerlukan parameter jumlah *cluster* yang diinginkan. Parameter yang dibutuhkan dalam metode ini hanyalah batas nilai pengukuran jarak yang diizinkan.

Berdasarkan latar belakang diatas maka dapat dilihat bahwa saat ini *software* SPSS hanya dapat digunakan untuk menganalisis data yang bersifat *metric*. Selain itu, untuk data *non metric* belum terdapat *software* yang dapat digunakan untuk membantu mempercepat proses perhitungan. Karena itu, penelitian yang akan dilakukan sekarang ini adalah mengembangkan sebuah perangkat lunak untuk melakukan analisis *cluster* dengan data yang bersifat *non metric*. Pengembangan perangkat lunak ini dilakukan untuk membantu dilakukannya analisis *cluster* dengan data yang murni bersifat *non metric*. Adanya pengembangan perangkat lunak inipun tentunya dapat membantu pengguna sehingga tidak perlu mengkhawatirkan jika data yang digunakan merupakan data *non metric*. Untuk memudahkan analisis *cluster* ini maka dipilih metode BSAS. Metode ini dipilih karena merupakan metode analisis *cluster* yang berurutan dan mudah untuk digunakan serta merupakan metode analisis *cluster* yang paling dasar.

1.2. Perumusan Masalah

Berdasarkan latar belakang di atas, rumusan masalah dalam penelitian ini adalah bagaimana mempermudah perhitungan analisis *cluster* untuk data *non metric* sehingga tidak perlu dilakukan perhitungan secara manual.

1.3. Tujuan Penelitian

Adapun tujuan dari dilakukannya penelitian ini adalah sebagai berikut.

- a. Membuat perangkat lunak untuk mempermudah perhitungan analisis *cluster* data *non metric* dengan menggunakan BSAS.
- b. Mengetahui kelemahan dan kelebihan dari perangkat lunak yang dibuat.

1.4. Batasan Masalah

Dalam setiap penelitian, tentunya dibutuhkan batasan masalah untuk memfokuskan penelitian dan menghindari hal-hal yang tidak berkaitan dengan penelitian yang akan dilakukan. Adapun batasan masalah dalam penelitian yang akan dilakukan adalah sebagai berikut :

- a. Perangkat lunak dapat melakukan perhitungan dengan jumlah variabel pembeda maksimal 655 kolom dan jumlah case maksimal 24.844 baris.
- b. Data *non metric* yang akan dikelompokkan telah dikategorikan dalam bentuk angka.
- c. Khusus untuk *file excel* yang dibuka, pada kolom paling pertama merupakan nama dari variabel pembeda.

