

BAB II

TINJAUAN PUSTAKA

1.1 Data Mining

Data mining adalah serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual. Perlu diingat bahwa kata *mining* sendiri berarti usaha untuk mendapatkan sedikit data berharga dari sejumlah data besar. Karena itu *data mining* sebenarnya memiliki akar yang panjang dari bidang ilmu seperti kecerdasan buatan (*artificial intelligent*), *meachine learning*, statistik dan basis data. Secara sederhana, data mining dapat diartikan sebagai proses mengekstrak atau mencari pengetahuan yang ada pada sekumpulan data. Banyak orang yang menyebut data mining adalah sinonim dari *Knowledge Discovery in Database* atau yang biasa disebut KDD. Beberapa teknik yang sering disebut-sebut dalam literatur data mining antara lain yaitu *association rule mining, clustering, klasifikasi, neural network*, dan lain-lain. (Kurniawan, 2017)

1.2 Analisis Cluster

Menurut Johnson & Wichern, (2007) Analisis cluster merupakan salah satu teknik dalam analisis multivariat metode independent yang mengelompokkan data. Analisis ini memiliki tujuan utama untuk mengelompokkan objek-objek yang paling dekat kesamaannya dengan objek

lain berada dalam kelompok yang sama, serta mempunyai kemiripan satu dengan yang lain. Menurut Mongi (2015) Pada analisis cluster terdapat 2 (dua) macam metode yang ada dalam proses analisis cluster yaitu metode berhirarki dan metode tak berhirarki. Metode berhirarki biasanya digunakan untuk data berukuran kecil dan hanya terdiri satu jenis variabel saja sedangkan metode non berhirarki biasanya digunakan untuk data yang berukuran besar dan terdiri dari banyak variabel. Dalam menerapkan analisis cluster permasalahan utamanya yaitu apabila variabel dasar pengelompokan bersifat kategorik maupun campuran (numerik dan kategorik) untuk mengatasi masalah pengelompokan tersebut, metode yang dapat digunakan adalah *Two Step Cluster*.

1.3 Ukuran Jarak

Menurut Hapsari *et al.* (2020) Jarak merupakan ukuran kedekatan antara objek yang satu dengan yang lain. Ukuran jarak dibutuhkan untuk setiap pasang objek yang akan dikelompokkan. Ukuran jarak yang diterapkan dalam metode *Two Step Cluster* sebagai berikut :

1. Jarak *Euclidean*

Jarak ini merupakan jarak yang umum digunakan dan dapat digunakan apabila semua variabelnya berskala numerik. Jarak ini harus memenuhi asumsi bahwa peubah-peubah yang diamati tidak berkorelasi dan antar peubah memiliki satuan yang sama. Jarak *Euclidean* dapat dirumuskan sebagai berikut :

$$d_{a,b} = \sqrt{\sum_{k=1}^p (x_{ab} - x_{bc})^2}$$

dengan :

$d_{a,b}$ = jarak antara objek a dengan objek b

x_{ab} = nilai objek a pada peubah ke- b

x_{bc} = nilai objek b pada peubah ke- c

P = banyaknya peubah yang diamati

2. Jarak *Log-Likelihood*

Jarak ini digunakan untuk peubah berskala kontinu dan kategorik. Jarak antar gerombol j dengan gerombol s dapat dilihat dari persamaan berikut :

$$d(j, s) = \xi_j + \xi_s - \xi_{(j,s)}$$

dengan,

$$\xi_j = -N \left(\sum_{k=1}^{k^A} \frac{1}{2} \log \right) (\hat{\sigma}_k^2 + \hat{\sigma}_{jk}^2) - \sum_{K=1}^{K^B} \sum_{l=1}^{L_K} \frac{N_{jkl}}{N_j} \log \frac{N_{jkl}}{N_j}$$

$$\xi_s = -N \left(\sum_{k=1}^{k^A} \frac{1}{2} \log \right) (\hat{\sigma}_k^2 + \hat{\sigma}_{sk}^2) - \sum_{K=1}^{K^B} \sum_{l=1}^{L_K} \frac{N_{skl}}{N_j} \log \frac{N_{skl}}{N_j}$$

$$\xi_{(js)} = -N \left(\sum_{k=1}^{k^A} \frac{1}{2} \log \right) (\hat{\sigma}_k^2 + \hat{\sigma}_{(js)k}^2) - \sum_{K=1}^{K^B} \sum_{l=1}^{L_K} \frac{N_{jkl}}{N_j} \log \frac{N_{(js)kl}}{N_j}$$

Dimana :

N : banyaknya objek

N_j : jumlah objek di dalam kelompok j

N_{jkl} : jumlah objek di kelompok j untuk peubah kategorik ke- k dengan kategori ke- l

$\hat{\sigma}_k^2$: ragam dugaan untuk peubah kontinu ke- k untuk keseluruhan objek

$\hat{\sigma}_{jk}^2$: ragam dugaan untuk peubah kontinu ke- k untuk keseluruhan objek dalam gerombol j

k^A : banyaknya peubah kontinu

k^B : banyaknya peubah kategorik

L_K : banyaknya kategori untuk peubah kategorik ke- k

1.4 Analisis Two Step Cluster

Two Step Cluster merupakan analisis pengelompokkan yang dirancang untuk menangani data dengan ukuran yang lebih besar, terutama pada permasalahan objek dengan skala pengukuran yang berbeda yaitu peubah numerik dan kategorik. Jarak antara dua kelompok didefinisikan sebagai jarak antar pusat dari masing-masing kelompok tersebut. Ukuran jarak yang digunakan dalam metode *Two Step Cluster* adalah jarak *Euclidean* dan jarak *log-likelihood*. Prosedur pengelompokkan dengan metode *Two Step Cluster* mempunyai dua tahapan yaitu tahap *preclustering* (pengelompokkan awal dan pengelompokkan akhir). (Mongi, 2015)

1.4.1 Pengelompokkan Awal

Langkah pertama di metode *Two Step Cluster* ialah pembentukan *cluster* awal dilakukan dengan pendekatan secara sekuensial, yaitu

setiap objek yang diamati satu persatu yang bersumber dari ukuran jarak. Prosedur ini dapat dilakukan dengan cara memilih satu amatan awal secara random untuk diukur jaraknya masing-masing amatan dengan amatan lain berdasarkan ukuran jarak yang sudah ditetapkan. Matriks yang terbentuk dari perhitungan jarak tersebut kemudian dibentuk *cluster* dengan pendekatan *Complete Linkage* yang mengklaster data berdasarkan hasil jarak terbesar yang diperoleh. Rumus *Complete Linkage* dapat didefinisikan sebagai berikut :

$$d_{w(u,v)} = \max(d_{wu}, d_{wv})$$

dengan :

wu : anggota *cluster* ke- u

wv : anggota *cluster* ke- v .

1.4.2 Pengelompokan Akhir (Pembentukan Cluster Optimal)

Pada tahap ini dilakukan penentuan jumlah cluster optimal dari data yang diolah dengan menghitung *schwarz's bayesian criterion* (*BIC*) untuk tiap cluster. hasil perhitungan tersebut digunakan untuk menduga jumlah cluster. (Lathifaturrahmah, 2014). Rumus *BIC* dan *AIC* untuk kelompok j adalah sebagai berikut :

$$BIC(J) = -2 \sum_{j=1}^J \xi_j + m_j \log(N)$$

$$AIC(J) = -2 \sum_{j=1}^J \xi_j + 2m_j$$

Dengan :

$$m_j = J \left\{ 2K^A + \sum_{k=1}^{K^B} (L_k - 1) \right\}$$

$$\xi_j = -N \left(\sum_{K=1}^{K^A} \frac{1}{2} \log(\hat{\sigma}_k^2 + \hat{\sigma}_{jk}^2) \sum_{K=1}^{K^B} \sum_{l=1}^{L_k} \frac{N_{jkl}}{N_j} \log \left(\frac{N_{jkl}}{N_j} \right) \right)$$

Solusi banyaknya gerombol yang optimal adalah yang memiliki nilai BIC terbesar, tetapi ada beberapa kasus dalam pengelompokan dimana BIC akan terus meningkat nilainya bila jumlah kelompok semakin meningkat. Dalam situasi tersebut *ratio BIC changes* (rasio perubahan BIC) dan *ratio of distance measure changes* (rasio perubahan jarak) digunakan untuk mengidentifikasi solusi banyaknya kelompok optimal. Jumlah kelompok yang terbentuk dapat diketahui dengan menggunakan perbandingan antar jarak untuk k gerombol, dengan rumus perbandingannya sebagai berikut :

$$R(k) = d_{k-1}/d_k$$

$$d_k = l_{k-1} - l_k$$

Dengan :

$R(k)$ = rasio perubahan jarak

$l_v = (r_v \log - BIC_v)/2$, atau

$l_v = (2r_v - AIC_v)/2$

$v = k, k - 1$

d_{k-1} = jarak jika k gerombol digabungkan dengan k-1 kelompok

1.5 Universitas Muhammadiyah Semarang

Universitas Muhammadiyah Semarang (UNIMUS) merupakan salah satu perguruan tinggi swasta yang terletak di Semarang, Jawa Tengah, Indonesia. Pendirian Universitas Muhammadiyah Semarang (UNIMUS) mulai dirintis oleh Pimpinan Wilayah Muhammadiyah Jawa Tengah sejak tahun 1996. Proses tersebut ditindaklanjuti dengan pembentukan panitia pendiri Universitas Muhammadiyah Semarang berdasarkan Surat Keputusan Pimpinan Wilayah Muhammadiyah Jawa Tengah Nomor: SK.PW/III.B/9.6/001/1997. Tanggal 4 Agustus 1999 Menteri Pendidikan dan Kebudayaan Republik Indonesia menerbitkan Ijin Operasional Universitas Muhammadiyah Semarang dengan nomor : 139/D/O/1999. Saat ini Unimus telah memiliki 8 fakultas dan 30 program studi yang terdiri dari 4 (empat) Program Profesi, 19 (Tujuh Belas) Program Strata Satu (S-1), 4 (empat) program Studi Diploma 3 (D3), 1(satu) Program Studi Diploma (D-4), dan 2 (dua) Program Megister (S-2). Untuk usia yang relative sangat muda ini, Unimus memperoleh mahasiswa baru menjadi perjuangan yang sangat tinggi. Sebab, Unimus harus berkompetisi dengan perguruan tinggi lainnya. Kondisi tersebut menuntut Unimus untuk melakukan usaha-usaha pemasaran, khususnya usaha promosi dengan berbagai macam cara. Oleh karena itu Social Media Marketing, Promosi, dan Brand Image sangat berperan dan perlu diperhatikan untuk menunjang minat calon mahasiswa baru (Unimus.ac.id).

1.6 Promosi

Promosi berasal dari kata *promote* didalam bahasa inggris yang artinya adalah mengembangkan atau meningkatkan, pengertian tersebut kalau disinggung dengan bidang Universitas atau Perguruan tinggi berarti sebagai alat untuk menambah Jumlah Mahasiswa baru. Promosi merupakan proses pemasaran dalam menginformasikan dan mendorong peminat terhadap Universitas, jasa, dan ide dari organisasi dengan memengaruhi minat calon mahasiswa baru agar berminat untuk bergabung dalam Universitas tersebut. Tujuan utama promosi ialah untuk menyebarluaskan informasi suatu produk/jasa serta untuk mendapatkan konsumen baru dengan memberikan informasi kepada konsumen mengenai produk yang ditawarkan, dimana konsumen bisa tertarik dengan kualitas dan harga produk yang ditawarkan. Untuk menarik minat mahasiswa baru dalam memilih perguruan tinggi perlu diadakan promosi baik melalui media elektronik, cetak, dan lain-lain yang kegunaannya untuk menyampaikan informasi dan menarik perhatian agar jumlah mahasiswa meningkat. (Garaika dan Winda, 2018)