

BAB V

KESIMPULAN DAN SARAN

5.1. Kesimpulan

Kesimpulan dari penelitian ini didapatkan dari hasil penelitian yang telah dilakukan oleh penulis adalah sebagai berikut.

1. *Dataset* data profil mahasiswa baru masih belum cukup untuk dijadikan sumber data tunggal dalam melakukan proses *data mining*. Oleh karena itu, diperlukannya data tambahan agar data yang tersedia dapat bekerja lebih baik lagi. Data kuesioner mahasiswa mampu menutup kekurangan dari data profil mahasiswa. Dalam proses *data selection* akhirnya dikumpulkan 12 atribut yang akan digunakan dalam penelitian ini. Atribut ini dipilih karena sesuai dengan kebutuhan penelitian yang dilakukan. Atribut tersebut mencakup *Timestamp*, NIM, Nama, KOTA, PROVINSI, PRODUCT, PRICE, PLACE, PROMOTION, PEOPLE, PROCESS, PHYSICAL EVIDENCE.
2. *Data preprocessing* yang dilakukan meliputi *handling missing value* untuk mengatasi data yang hilang, *handling outliers* untuk menangani data *outliers* yang terdapat pada tujuh atribut numerik (PRODUCT, PRICE, PLACE, PROMOTION, PEOPLE, PROCESS, PHYSICAL EVIDENCE). Terakhir adalah *handling duplicate data* yang akan menangani data duplikat yang terdapat pada *dataset*.

3. *Data transformation* melibatkan proses *feature scaling* pada kolom numerik (PRICE, PLACE, PROMOTION, PEOPLE, PROCESS, PHYSICAL EVIDENCE). *Feature scaling* ini menggunakan fitur *standard scaler* yang terdapat pada *sklearn* pada *python*. Selain itu, transformasi yang dilakukan adalah pengubahan nama kota menjadi angka numerik untuk dilakukan filter terhadap kota yang berada dalam top 10 kota asal mahasiswa terbanyak.
4. *Data mining* menggunakan penerapan teknik *KMeans* untuk melakukan *clustering* data. Hasil dari penerapan teknik ini disajikan dalam bentuk grafik *scatter* yang memperlihatkan hasil sebaran data dalam dua dimensi. Nilai *k* yang digunakan adalah tiga, yang berarti klaster yang dibentuk berjumlah tiga buah. Tiga klaster yang terbentuk mewakili perspektif mahasiswa baru dalam memilih Universitas Komputer Indonesia sebagai tempat untuk melanjutkan pendidikan. Klaster 0 merupakan klaster dengan ekspektasi yang tinggi pada semua aspek, mungkin mewakili pelanggan atau entitas yang sangat menuntut. Klaster 1 lebih rata-rata dan mungkin lebih fleksibel atau kurang peduli terhadap setiap atribut dibandingkan klaster lain. Klaster 2 berada di tengah-tengah, menunjukkan perhatian lebih pada atribut seperti proses dan bukti fisik, tetapi tidak setinggi klaster 1 dalam hal ekspektasi.
5. Interpretasi atau evaluasi menggunakan nilai dari *Davies-Bouldin Indeks*. Di mana nilai ini akan memberikan skor untuk setiap klaster yang terbentuk sehingga bisa terlihat klaster yang paling optimal bentuknya.

Nilai *Davies-Bouldin Indeks* untuk nilai $k=3$ adalah 1.6645786314264834.

6. Implementasi web menggunakan *framework streamlit*. Dengan mengimpor kode dari *Google Colaboratory* sebuah *web app* segmentasi mahasiswa baru dapat dibentuk.

5.2. Saran

Saran yang dapat disampaikan sebagai bahan pertimbangan pengembangan sistem informasi segmentasi mahasiswa baru Universitas Komputer Indonesia adalah sebagai berikut.

1. Pengembangan segmentasi mahasiswa baru ini baru hanya memakai satu teknik algoritma *clustering* yaitu *KMeans*. Sehingga masih ada algoritma lain yang bisa dicobakan seperti *DBSCAN (Density-Based Spatial Clustering of Applications with Noise)* atau *GMM (Gaussian Mixture Models)*
2. Untuk penelitian selanjutnya, penentuan nilai k yang optimal dapat dilakukan dengan menggunakan berbagai teknik lain seperti Indeks Calinski-Harabasz dan Dendrogram, untuk mendapatkan hasil yang lebih akurat dan komprehensif.
3. Pengembangan web yang akan dilakukan untuk proses *data mining* dapat menggunakan *framework* atau bahasa pemrograman lain seperti PHP dan *JavaScript*.