BAB V
PENUTUP

5.1. Kesimpulan

Berdasarkan hasil penelitian, pengujian, dan evaluasi yang telah dilakukan mengenai perbandingan metode *Ward's Method*, DBSCAN, dan K-Means dalam mengelompokkan aduan pelayanan masyarakat di Dispendukcapil Kota Surabaya, dapat ditarik beberapa kesimpulan sebagai berikut:

a. Penerapan *Ward's Method*, DBSCAN, dan K-Means dalam melakukan klasterisasi data aduan masyarakat dilakukan melalui tahapan *Text Mining* yang sistematis, dimulai dari *preprocessing* teks yang meliputi *cleaning, tokenizing, normalization, stopword removal,* dan *stemming*, kemudian pembobotan kata menggunakan TF-IDF untuk mengubah data teks menjadi representasi numerik. Setelah itu dilakukan proses pemodelan sesuai karakteristik masing-masing metode. *Ward's Method* diterapkan menggunakan pendekatan *hierarchical agglomerative clustering* berbasis minimisasi variansi dengan jarak *Euclidean* dan divisualisasikan melalui dendrogram untuk menentukan jumlah klaster optimal. DBSCAN diterapkan menggunakan pendekatan berbasis kepadatan yang mampu mengidentifikasi *noise* sehingga memerlukan proses penyaringan data valid sebelum evaluasi dilakukan, sedangkan K-Means diterapkan dengan pendekatan partisi dan penentuan jumlah klaster optimal menggunakan *Variance Ratio Criterion* (VRC) untuk mengurangi subjektivitas pemilihan nilai k. Dengan demikian, ketiga metode berhasil diimplementasikan secara terstruktur dalam sistem klasterisasi berbasis teks berdasarkan kepadatan teks.

b. Hasil klasterisasi menunjukkan adanya perbedaan jumlah dan struktur klaster pada masing-masing metode. *Ward's Method* menghasilkan dua klaster utama, dengan satu klaster besar berisi 1.255 dokumen yang merepresentasikan aduan administratif prosedural dan satu klaster kecil berisi 2 dokumen yang bersifat lebih spesifik dan sensitif. DBSCAN

menghasilkan 10 klaster utama dan 1 klaster yang dianggap sebagai *noise*, sementara K-Means menghasilkan 3 klaster dengan distribusi yang secara visual tampak lebih terbagi, namun kurang kuat secara struktur. Analisis lanjutan terhadap distribusi emosi menunjukkan bahwa mayoritas aduan didominasi oleh emosi sedih, yang mengindikasikan bahwa dataset memiliki karakteristik relatif homogen dan sebagian besar berisi keluhan administratif. Temuan ini menunjukkan bahwa meskipun segmentasi klaster dapat dilakukan dengan berbagai metode, struktur alami data cenderung terkonsentrasi pada satu kelompok besar dengan karakteristik topik yang serupa.

c. Validitas metode klasterisasi dievaluasi menggunakan *Silhouette Coefficient* dan *Index Davies-Bouldin* sebagai ukuran kekompakan intra-klaster dan pemisahan antar-klaster. Hasil evaluasi menunjukkan bahwa *Ward's Method* memperoleh nilai *Silhouette Coefficient* tertinggi sebesar 0,76 dan nilai *Index Davies-Bouldin* terendah sebesar 1,02, yang menunjukkan struktur klaster yang kuat, kompak, dan memiliki pemisahan yang jelas. DBSCAN memperoleh nilai *Silhouette Coefficient* sebesar 0,52 dan IDB sebesar 1,12, yang menunjukkan kualitas klaster lemah, sedangkan K-Means memperoleh nilai *Silhouette Coefficient* terendah sebesar 0,49 dan IDB tertinggi sebesar 3,47, yang mengindikasikan struktur klaster yang lemah dan tingkat overlap yang tinggi. Sistem komparasi otomatis pada program juga menetapkan Ward sebagai metode terbaik berdasarkan prioritas *Silhouette Coefficient* tertinggi dan IDB terendah. Dengan demikian, dapat disimpulkan bahwa *Ward's Method* merupakan metode yang paling optimal dan paling sesuai untuk karakteristik data aduan pelayanan masyarakat dalam penelitian ini.

5.2. Saran Pengembangan

Berdasarkan hasil penelitian dan keterbatasan yang ditemukan selama proses pengerjaan, terdapat beberapa saran yang dapat diajukan untuk pengembangan penelitian selanjutnya, antara lain:

a. Metode ekstraksi fitur saat ini masih menggunakan TF-IDF, yang hanya memperhatikan frekuensi kata tanpa memahami konteks semantik atau makna kata dalam kalimat. Oleh karena itu, disarankan untuk menggunakan metode representasi teks yang lebih canggih seperti *Word Embedding* (*Word2Vec* atau *FastText*) agar hasil pengelompokan bisa lebih akurat dalam menangkap sinonim atau konteks keluhan.
b. Penelitian Selanjutnya dapat mempertimbangkan penggunaan algoritma *Fuzzy C-Means* yaitu metode yang menerapkan *soft clustering*, yang mana setiap dokumen memiliki nilai derajat keanggotaan pada beberapa klaster sekaligus sehingga informasi topik tidak hilang akibat pemaksaan pengelompokan tunggal.
c. Metode analisis emosi yang diterapkan saat ini masih menggunakan pendekatan berbasis aturan (*rule-based*) sederhana dengan daftar kata kunci yang didefinisikan secara manual. Pendekatan ini memiliki keterbatasan dalam menangkap konteks kalimat yang kompleks. Penelitian selanjutnya disarankan untuk beralih ke metode Analisis Sentimen Berbasis *Machine Learning*, seperti *Naive Bayes* atau SVM yang mampu memahami konteks semantik dengan lebih baik. Selain itu, akurasi deteksi dapat ditingkatkan dengan mengintegrasikan kamus sentimen standar bahasa Indonesia, seperti *InSe*t atau *SentiStrengthID* serta memperkaya leksikon dengan kosa kata lokal (slang) khas Surabaya agar cakupan emosi yang terdeteksi menjadi lebih luas dan presisi.

Halaman ini sengaja dikosongkan