

## BAB 5 KESIMPULAN

### 5.1. Kesimpulan

Berdasarkan hasil penelitian dan analisis data yang telah dilakukan, berikut kesimpulan yang menjawab rumusan masalah penelitian.

Pertama, hasil ekstraksi aspek ulasan pengguna aplikasi MySiloam menggunakan metode BERTopic dengan perbandingan tiga algoritma klusterisasi menunjukkan bahwa ketiga algoritma menghasilkan karakteristik yang berbeda. HDBSCAN menghasilkan jumlah topik yang lebih banyak dan tidak stabil — 5 topik pada skenario tanpa *stemming* dan 9 topik pada skenario dengan *stemming* — dengan tingkat *noise* yang lebih tinggi (2,5% dan 5,2%), sehingga menghasilkan beberapa topik yang kurang representatif secara semantik. BIRCH menghasilkan 3 topik tanpa *noise* pada kedua skenario namun dengan kualitas koherensi yang lebih rendah. K-Means menghasilkan 3 topik tanpa *noise* pada kedua skenario dengan distribusi dokumen yang paling seimbang. Dari perbandingan keenam kombinasi, K-Means dengan *stemming* menghasilkan ekstraksi aspek terbaik dengan tiga aspek utama yang berhasil diidentifikasi secara otomatis, yaitu Fitur Medis Aplikasi (713 dokumen) yang mencakup ulasan terkait janji temu dokter, antrian, jadwal, dan obat; Pelayanan & Kepuasan (574 dokumen) yang mencakup ulasan terkait kualitas pelayanan, keramahan, dan kepuasan pasien; serta Teknis Aplikasi (412 dokumen) yang mencakup ulasan terkait *error*, *login*, *booking*, dan pembaruan aplikasi.

Kedua, evaluasi kualitas topik menggunakan metrik  $C_v$  sebagai metrik utama, didukung oleh UMass, NPMI, dan *Topic Diversity* sebagai metrik pendukung, menunjukkan bahwa K-Means dengan *stemming* menghasilkan kualitas topik terbaik di antara keenam kombinasi skenario yang diuji, dengan nilai  $C_v=0,4113$ , UMass=-3,8012 (melampaui *threshold* -5,0), dan NPMI=-0,0200 yang paling mendekati *threshold* 0,0, serta *Topic Diversity*=0,9000. Algoritma HDBSCAN menghasilkan kualitas terendah dengan  $C_v=0,2348$  dan UMass=-16,1669 pada

skenario tanpa *stemming*, bahkan menghasilkan nilai NPMI=NaN pada skenario dengan *stemming* akibat fragmentasi topik yang berlebihan. BIRCH berada di posisi menengah namun nilai UMass yang sangat rendah (-13,4242 dan -14,5805) menunjukkan bahwa kata-kata dalam topik yang dihasilkan kurang konsisten muncul bersama dalam korpus. Secara keseluruhan, *stemming* terbukti meningkatkan kualitas topik pada algoritma K-Means namun berdampak negatif pada HDBSCAN, menunjukkan bahwa pengaruh *preprocessing* terhadap kualitas *topic modeling* sangat bergantung pada algoritma *clustering* yang digunakan.

Ketiga, performa model SVM dalam mengklasifikasikan sentimen pengguna pada setiap aspek layanan aplikasi MySiloam menunjukkan hasil yang baik, dengan *F1-score* sebagai acuan utama penilaian karena lebih representatif dalam kondisi distribusi kelas yang tidak merata. Skenario 1 (pendekatan dua tahap) secara konsisten mengungguli Skenario 2 (klasifikasi gabungan empat kelas) pada seluruh aspek. *F1-score* sentimen tertinggi diperoleh pada aspek Fitur Medis Aplikasi sebesar 93,49% (akurasi 93,48%), diikuti Pelayanan & Kepuasan sebesar 90,69% (akurasi 91,57%), dan Teknis Aplikasi sebesar 84,42% (akurasi 84,51%), dengan rata-rata *F1-score* sentimen keseluruhan mencapai 89,53%. Skenario 1 secara konsisten mengungguli Skenario 2 yang hanya mencapai rata-rata *F1-score* 82,74%, karena pemisahan tugas klasifikasi memungkinkan setiap model lebih terfokus dalam mempelajari pola yang spesifik. Keterbatasan yang ditemukan adalah rendahnya performa klasifikasi kelas Netral pada kedua skenario akibat jumlah sampel yang sangat sedikit, terutama pada aspek Fitur Medis Aplikasi yang hanya memiliki 2 sampel Netral dalam dataset, sehingga metrik evaluasi untuk kelas ini tidak dapat diinterpretasikan secara andal secara statistik.

## 5.2. Saran

Berdasarkan hasil penelitian yang telah dilakukan, terdapat beberapa saran yang ditujukan kepada dua pihak, yaitu saran untuk pengembangan penelitian selanjutnya dan rekomendasi praktis bagi pihak MySiloam sebagai pemangku kepentingan utama penelitian ini.

**a. Saran untuk Penelitian Selanjutnya**

Pertama, penelitian selanjutnya disarankan untuk mengatasi keterbatasan klasifikasi kelas Netral yang masih rendah pada penelitian ini dengan menambah jumlah sampel data berlabel Netral melalui teknik *data augmentation* seperti *synonym replacement* atau *back-translation*, sehingga distribusi kelas menjadi lebih seimbang dan model dapat mengenali pola sentimen netral dengan lebih akurat.

Kedua, eksplorasi penggunaan model *transformer* berbasis bahasa Indonesia seperti IndoBERT atau IndoRoBERTa sebagai pengganti SVM untuk tahap klasifikasi sentimen sangat direkomendasikan, mengingat model *transformer* mampu memahami konteks kalimat secara bidireksional dan berpotensi menghasilkan akurasi yang lebih tinggi terutama pada ulasan yang bersifat kontradiktif atau sarkastik.

Ketiga, perluasan cakupan data dengan mempertimbangkan ulasan multibahasa dari App Store secara lebih optimal, serta memperluas rentang waktu pengumpulan data secara berkala agar model dapat mengikuti perkembangan fitur dan perubahan persepsi pengguna terhadap aplikasi MySiloam dari waktu ke waktu.

Keempat, penelitian selanjutnya dapat mempertimbangkan penambahan aspek analisis yang lebih granular, misalnya memecah aspek Teknis Aplikasi menjadi sub-aspek yang lebih spesifik seperti performa, antarmuka pengguna, dan keamanan data, sehingga wawasan yang dihasilkan lebih *actionable* bagi tim pengembang aplikasi kesehatan digital.

**b. Rekomendasi untuk Pihak MySiloam**

Pertama, berdasarkan temuan bahwa aspek Fitur Medis Aplikasi mendominasi ulasan dengan 713 dokumen dan mengandung sentimen negatif yang signifikan terkait kata kunci seperti *janji temu*, *antrian*, dan *jadwal*, MySiloam disarankan untuk memprioritaskan peningkatan stabilitas dan kemudahan penggunaan fitur inti tersebut. Keluhan yang berulang pada fitur janji temu dokter dan antrian digital mengindikasikan adanya hambatan pengalaman pengguna yang perlu segera ditangani melalui pembaruan antarmuka maupun peningkatan kapasitas sistem.

Kedua, pada aspek Teknis Aplikasi, tingginya frekuensi kata kunci seperti *error*, *login*, *booking*, dan *update* dalam ulasan negatif mengindikasikan perlunya

peningkatan stabilitas dan keandalan sistem secara berkelanjutan. MySiloam disarankan untuk memperkuat proses pengujian kualitas (*quality assurance*) sebelum setiap pembaruan versi aplikasi dirilis, sehingga potensi gangguan teknis yang berdampak pada pengalaman pengguna dapat diminimalkan.

Ketiga, MySiloam disarankan untuk mengimplementasikan sistem *dashboard* analisis sentimen berbasis *machine learning* yang dibangun dalam penelitian ini sebagai alat pemantauan ulasan pengguna secara otomatis dan berkelanjutan. Sistem ini dapat diintegrasikan ke dalam alur kerja tim produk MySiloam sehingga setiap ulasan baru yang masuk dari Google Play Store maupun App Store dapat secara otomatis diklasifikasikan berdasarkan aspek dan sentimennya tanpa perlu membaca ulasan satu per satu secara manual. Dengan demikian, tim manajemen MySiloam dapat memperoleh wawasan yang terstruktur dan berbasis data mengenai aspek layanan mana yang perlu segera ditingkatkan.

Keempat, pengembangan sistem *deployment* ke depan dapat diperkaya dengan fitur analisis tren sentimen berbasis waktu (*time-series sentiment analysis*) sehingga tim pengembang MySiloam dapat memantau perubahan persepsi pengguna secara periodik, terutama setelah peluncuran pembaruan versi aplikasi baru. Dengan fitur ini, MySiloam dapat mengukur secara objektif apakah setiap pembaruan yang dirilis berdampak positif atau justru memunculkan keluhan baru dari pengguna, sehingga pengambilan keputusan pengembangan produk menjadi lebih berbasis data dan terukur.