

BAB V PENUTUP

5.1 Kesimpulan

Berdasarkan hasil analisis data dan pembahasan yang dilakukan dalam penelitian ini, dapat disimpulkan bahwa pemodelan topik melalui media X menggunakan *Apache Spark* studi kasus COVID-19 di Indonesia tahun 2020-2022 menunjukkan beberapa temuan sebagai berikut :

1. Media X terbukti efektif dalam menyebarkan informasi terkait COVID-19, dengan analisis topik yang menunjukkan isu terkait kesehatan, vaksinasi, kebijakan pemerintah, serta persepsi masyarakat terhadap pandemi.
2. Pemodelan topik yang dilakukan dengan *Apache Spark* berhasil mengidentifikasi sepuluh tema dari topik, yaitu :
 - Topik 0 : Pemulihan Pasien COVID-19 di Indonesia
 - Topik 1 : Tindakan Pemerintah dalam Menghadapi Pandemi COVID-19
 - Topik 2 : Protokol Kesehatan sebagai Upaya Pencegahan COVID-19
 - Topik 3 : Kebijakan Pemerintah Terkait Angka Positif dan Kematian COVID-19
 - Topik 4 : Analisis Angka Kematian dan Upaya Pemerintah untuk Keselamatan Masyarakat
 - Topik 5 : Normalisasi Kehidupan Pasca-Pandemi di Indonesia
 - Topik 6 : Peran Masyarakat dalam Memerangi Penyebaran COVID-19
 - Topik 7 : Vaksinasi dan Penanganan Penyebaran COVID-19
 - Topik 8 : Imbauan Pemerintah untuk Memutus Rantai Penyebaran COVID-19
 - Topik 9 : Sinergi dan Gotong Royong dalam Penanganan Pandemi COVID-19 di Indonesia

3. Evaluasi model topik dilakukan menggunakan metrik *Perplexity* dan *Coherence Score*. Hasil evaluasi menunjukkan nilai *perplexity* sebesar 5.5987700238050815 dan *coherence score* sebesar 0.6061784364959958. Nilai *perplexity* yang rendah mengindikasikan bahwa model memiliki kemampuan prediksi yang baik terhadap data. Sementara itu, nilai *coherence score* menunjukkan bahwa topik-topik yang dihasilkan memiliki tingkat koherensi yang cukup baik dan dapat diinterpretasikan dengan relatif mudah. Dengan hasil evaluasi ini, dapat disimpulkan bahwa model topik yang dihasilkan cukup representatif untuk mengidentifikasi dan memahami topik-topik utama yang terkandung dalam kumpulan data *tweet* terkait Covid-19 di Indonesia.

5.2 Saran

Saran penulis untuk penelitian selanjutnya yang serupa adalah peningkatan kualitas data pelatihan. Disarankan untuk mengumpulkan dan menggunakan data pelatihan yang lebih beragam dan representatif. Hal ini dapat mencakup teks dari berbagai sumber, kategori, dan konteks untuk meningkatkan kemampuan model dalam memahami berbagai jenis bahasa dan situasi. Selain itu, pra-pemrosesan data yang mendalam sangat disarankan, termasuk pembersihan data tingkat lanjut, normalisasi teks komprehensif, dan ekstraksi fitur yang relevan. Penanganan variasi bahasa dan pengayaan data semantik juga dapat dipertimbangkan untuk meningkatkan kualitas data. Tidak lupa untuk menyesuaikan parameter dengan karakteristik dataset yang digunakan agar selaras dengan proses yang dilakukan.