

BAB V PENUTUP

5.1 Kesimpulan

Berdasarkan hasil penelitian dan pembahasan pada Bab IV, dapat diambil beberapa hasil penelitian klasifikasi topik berita pada akun Instagram Merapi Uncover menggunakan algoritma Support Vector Machine (SVM), dapat diambil beberapa kesimpulan berdasarkan rumusan masalah yang diajukan, yaitu "Bagaimana performa algoritma *Support Vector Machine* dalam melakukan klasifikasi topik berita pada Merapi Uncover", penelitian ini berhasil menjawabnya melalui analisis komprehensif terhadap 1.501 postingan akun Instagram *merapi_uncover* periode 10–15 November 2025 yang dipetakan dari 77 kategori awal menjadi lima kelas besar.

Performa algoritma *Support Vector Machine* (SVM) dengan representasi fitur TF-IDF dalam klasifikasi topik berita Merapi Uncover dapat disimpulkan sebagai berikut:

1. Performa Keseluruhan Model

Model SVM linear menghasilkan akurasi sebesar 81% pada data uji dengan *weighted F1-score* 0,79 dan *macro F1-score* 0,69, menunjukkan bahwa mayoritas caption Instagram berhasil diklasifikasikan ke kategori topik yang tepat. Nilai *weighted F1-score* yang mencapai 0,79 membuktikan bahwa SVM efektif menangani data teks pendek berbahasa Indonesia yang berasal dari media sosial, meskipun terdapat variasi gaya bahasa informal, singkatan, dan emoji.

2. Performa Per Kategori Topik

Kinerja model bervariasi signifikan antar kelas, dengan kategori Hiburan/Budaya/Wisata mencatat *F1-score* tertinggi 0,91 (*precision* 0,87; *recall* 0,96), diikuti Lalu Lintas & Transportasi dengan *F1-score* 0,82 (*precision* 0,73; *recall* 0,94). Kategori Kriminal & Keamanan Publik menunjukkan performa moderat dengan *F1-score* 0,72 (*precision* 0,74; *recall* 0,70). Sebaliknya, kategori

Bencana & Kebencanaan dan Sosial & Pemerintahan memiliki *F1-score* relatif rendah, masing-masing 0,51 dan 0,48, yang mengindikasikan model mengalami kesulitan dalam mendeteksi dan membedakan kedua topik tersebut dari kategori lain.

3. Faktor Penentu Performa

Ketidakseimbangan distribusi data menjadi faktor utama variasi performa antar kelas, dengan kelas mayoritas Hiburan/Budaya/Wisata (800 sampel) memiliki representasi jauh lebih besar dibandingkan kelas minor seperti Kriminal & Keamanan Publik (101 sampel). Tahapan *preprocessing* yang meliputi *cleaning* (penghapusan URL, *mention*, *hashtag*, angka, tanda baca, emoji) dan *stopword removal* menggunakan pustaka Sastrawi, serta ekstraksi fitur TF-IDF dengan konfigurasi *max_features* 50.000 dan *n-gram range* (1,2), terbukti menghasilkan representasi teks yang cukup baik untuk keperluan klasifikasi.

4. Kontribusi Penelitian

Penelitian ini membuktikan bahwa kombinasi SVM dan TF-IDF merupakan solusi yang efisien dan efektif untuk klasifikasi topik berita pada platform media sosial berbahasa Indonesia, khususnya untuk teks pendek dengan karakteristik informal seperti caption Instagram. Model klasik SVM tetap kompetitif dan tidak memerlukan sumber daya komputasi tinggi seperti *deep learning*, menjadikannya alternatif praktis untuk implementasi sistem klasifikasi berita otomatis pada akun komunitas lokal seperti Merapi Uncover.

5.2 Saran

Berdasarkan temuan dan keterbatasan penelitian ini, beberapa saran untuk pengembangan lebih lanjut adalah:

1. Penyeimbangan Kelas dan Perluasan Data

Penelitian selanjutnya disarankan untuk menerapkan teknik penyeimbangan kelas seperti *oversampling* (SMOTE), *class_weight*, atau *undersampling* pada kelas mayoritas guna meningkatkan *recall* dan *F1-score* kategori minor seperti Bencana & Kebencanaan dan Sosial & Pemerintahan. Perluasan periode pengumpulan data menjadi rentang waktu yang lebih panjang (minimal 6 bulan hingga 1 tahun) akan memberikan representasi topik yang lebih kaya dan mengurangi bias temporal pada model.

2. Optimasi Hyperparameter dan Eksplorasi Kernel

Perlu dilakukan *tuning* sistematis terhadap parameter SVM seperti nilai C (regularisasi), gamma, dan pemilihan kernel (linear vs RBF vs polynomial) melalui teknik *grid search* atau *random search* dengan validasi silang stratifikasi (*stratified cross-validation*) untuk menemukan konfigurasi optimal yang meningkatkan performa pada seluruh kelas.

3. Pengayaan Representasi Fitur

Eksplorasi representasi fitur alternatif seperti *n-gram* karakter, kombinasi TF-IDF dengan *word embeddings* (Word2Vec, FastText), atau reduksi dimensi menggunakan Latent Semantic Analysis (LSA) dapat membantu menangkap variasi bahasa slang, singkatan, dan kesalahan ejaan yang sering muncul pada caption Instagram. Penambahan fitur metadata seperti jumlah komentar, tipe media (foto vs video), atau *engagement rate* juga berpotensi meningkatkan akurasi klasifikasi.

4. Perbandingan dengan Model Alternatif

Studi komparatif antara SVM dengan algoritma klasik lainnya (Naïve Bayes, Random Forest, Logistic Regression) serta model berbasis *deep learning* (LSTM, IndoBERT, transformer ringan) pada dataset yang sama akan memberikan pemahaman menyeluruh tentang *trade-off* antara akurasi, efisiensi komputasi, dan kemudahan implementasi. Penelitian terkini menunjukkan bahwa model transformer seperti IndoBERT mampu mencapai akurasi lebih tinggi namun dengan biaya komputasi yang jauh lebih besar.

5. Implementasi Sistem Otomatis dan Pemantauan Berkelanjutan

Hasil penelitian dapat dikembangkan menjadi sistem penandaan topik otomatis (*automated tagging system*) yang terintegrasi dengan alur publikasi akun Merapi Uncover, memungkinkan pengguna menyaring berita berdasarkan kategori yang diminati. Perlu dilakukan pemantauan rutin terhadap *model drift* akibat perubahan pola bahasa atau munculnya topik baru, serta mekanisme *retraining* berkala untuk menjaga performa model tetap optimal.

Dengan mempertimbangkan saran-saran tersebut, diharapkan penelitian lanjutan dapat menghasilkan sistem klasifikasi topik berita yang lebih akurat dan robust, sehingga semakin mendukung pemanfaatan data media sosial seperti Merapi Uncover sebagai sumber informasi yang terstruktur dan mudah diakses.