

BAB V

SIMPULAN DAN SARAN

5.1 Simpulan

Penelitian ini menunjukkan bahwa algoritma SVM pada skenario ketiga merupakan model paling unggul dengan akurasi rata-rata 75%, melampaui performa Naive Bayes. Temuan ini diperkuat dengan adanya korelasi positif sedang ($r = 0,538$) antara ulasan positif dan rating aplikasi, yang membuktikan keselarasan antara narasi pengguna dengan penilaian bintang. Meskipun selisih akurasi tertinggi antara kedua metode hanya sebesar 2%, algoritma SVM terbukti jauh lebih unggul dan stabil dibandingkan Naive Bayes dalam menangani variasi ulasan. Keunggulan SVM terlihat pada selisih rata-rata *F1-Score* yang mencapai 0,8, sebuah parameter yang lebih reliabel karena menunjukkan keseimbangan antara presisi dan *recall*. Dengan demikian, SVM secara signifikan lebih efektif dalam mengatasi masalah ketidakseimbangan data (*data imbalance*) pada ulasan aplikasi OTT. Algoritma Naive Bayes bekerja berdasarkan probabilitas kemunculan kata yang menawarkan efisiensi dan kecepatan tinggi, sementara SVM menggunakan pendekatan geometris dengan mencari *hyperplane* melalui optimasi parameter seperti nilai *C*, *kernel*, dan *gamma*. Algoritma SVM mampu menciptakan batas keputusan yang tegas pada ruang dimensi tinggi. Hal ini membuat SVM lebih unggul dalam menangani data yang kompleks dan memberikan akurasi yang lebih presisi dibandingkan Naive Bayes yang hanya mengandalkan perhitungan frekuensi fitur. Penelitian ini memberikan kontribusi nyata bagi industri layanan OTT di Indonesia sebagai instrumen *Business Intelligence*. Melalui model klasifikasi yang dihasilkan, pengembang dapat memantau kepuasan pelanggan secara otomatis dan *real time*.

Pemilihan *hyperparameter* pada *Grid Search* SVM yang meliputi variasi nilai *C* [0.1, 1, 10], penggunaan kernel '*linear*' dan '*rbf*', serta

parameter *gamma* 'scale' dan 'auto', merupakan konfigurasi yang optimal karena mencakup spektrum pencarian yang luas untuk menyeimbangkan antara kompleksitas model dan kemampuan generalisasi. Pengujian nilai *C* yang variatif memungkinkan model menentukan tingkat toleransi terhadap *error* atau *noise* pada ulasan, sementara perbandingan antara kernel linear dan RBF berfungsi untuk memastikan apakah pola sentimen dalam korpus lebih efektif dipisahkan secara linear atau memerlukan transformasi ruang fitur yang lebih kompleks. Selain itu, penggunaan parameter *gamma* yang adaptif terhadap varians fitur (TF-IDF) sangat krusial dalam menjaga stabilitas model saat menangani data berdimensi tinggi, sehingga kombinasi ini secara sistematis menjamin ditemukannya titik kerja model yang paling tangguh dalam menghadapi masalah data imbalance dan ambiguitas leksikal ulasan pengguna.

5.2 Saran

Meskipun SVM menunjukkan keunggulan performa daripada Naive Bayes, pada penelitian ini mencatat adanya kendala pada klasifikasi kelas sentimen Netral yang disebabkan oleh ketidakseimbangan data (*Imbalance data*). Ketimpangan jumlah sampel antara sentimen Positif, Negatif dan Netral menyebabkan model cenderung mengalami bias. Oleh karena itu, pada penelitian ini digunakan teknik resampling seperti SMOTE dan melakukan *tuning* konfigurasi parameter SVM untuk nilai *svm_C*, *svm_kernel*, & *svm_gamma* guna meningkatkan sensitivitas model terutama terhadap kelas minoritas (Kelas Netral). Untuk penelitian kedepan (*Future Work*) disarankan melakukan beberapa skenario *tuning* atau konfigurasi parameter SVM untuk nilai *svm_C*, *svm_kernel*, & *svm_gamma* guna mendapatkan hasil klasifikasi yang lebih variatif.