

ABSTRAK

Penelitian ini melakukan analisis sentimen terhadap opini publik terkait isu kenaikan tunjangan DPR. Tujuan utama penelitian ini adalah mengklasifikasikan sentimen masyarakat dan mengevaluasi performa algoritma *Support Vector Machine* (SVM) dalam menangani masalah ketidakseimbangan kelas *class imbalance*. Berdasarkan hasil praproses dan pelabelan data, diperoleh total 26.074 data bersih yang sangat didominasi oleh sentimen *negatif* sebanyak 19.453 komentar 74,61%, berbanding sentimen *positif* sebanyak 6.621 komentar 25,39%. Analisis menunjukkan bahwa dominasi narasi negatif ini berakar dari keresahan sosial ekonomi, yang secara spesifik direpresentasikan oleh tingginya kemunculan kata kunci seperti *rakyat*, *gaji*, *korupsi*, dan *susah*. Klasifikasi dilakukan menggunakan ekstraksi fitur TF-IDF yang diproses dengan model SVM. Hasil pengujian membuktikan bahwa model ini sangat andal dengan tingkat *akurasi* mencapai 97,20%. Model juga sukses mengatasi *class imbalance*, divalidasi oleh capaian metrik evaluasi yang tinggi dan seimbang pada setiap kelasnya, yaitu *presisi* 98,45% untuk kelas negatif dan 93,63% untuk kelas positif, *recall* 97,79% untuk kelas negatif dan 95,47% untuk kelas positif, serta *F1-score* yakni 98,10% untuk kelas negatif dan 94,50% untuk kelas positif. Secara komparatif, penggunaan *Kernel Linear* 97,20% terbukti mengungguli *Kernel RBF* 91,52% karena karakteristik data ekstraksi TF-IDF yang berdimensi tinggi dan berongga *sparse* jauh lebih optimal dipisahkan secara linear *linearly separable* tanpa memicu risiko *overfitting*. Meskipun akurasi sangat tinggi, evaluasi model menemukan adanya 146 data salah klasifikasi 2,80%, yang terdiri dari 86 *False Positive* dan 60 *False Negative*. Misklasifikasi tersebut umumnya disebabkan oleh keterbatasan kamus *leksikon* dalam mendeteksi kata slang seperti *laknat*, serta bias ambiguitas pada kata dominan seperti *dpr* dan *rakyat*. Untuk pengembangan selanjutnya, penelitian ini merekomendasikan perluasan pustaka *leksikon*, eksplorasi fitur semantik, serta pengujian guna mengukur ketahanan atau *robustness* model ber-kernel Linear ini pada dataset kebijakan publik lainnya yang memiliki dimensi tinggi.

Kata kunci: Analisis sentimen, *Support Vector Machine*, tunjangan DPR, TF-IDF, *class imbalance*.