

Perbandingan Support Vector Machine dan Naïve Bayes dalam Pengklasifikasian ujaran SARA di Twitter

Amirul Mahdi, Yuliant Sibaroni²

^{1,2,3}Fakultas Informatika, Universitas Telkom, Bandung

¹primumtertium@students.telkomuniversity.ac.id, ²yuliant@telkomuniversity.ac.id

Abstrak

SARA (Suku, Agama, Ras, dan Antargolongan) SARA merupakan pikiran sentimen mengenai identitas diri yang menyangkut, keturunan, agama, kebangsaan atau kesukuan dan golongan. Pada zaman seperti sekarang ini banyak orang yang menyalahgunakan media sosial sebagai sarana untuk menyebarkan kebencian melalui ujaran SARA, dan media yang cukup umum digunakan adalah *Twitter*. SARA tidak hanya akan menyebarkan kebencian kepada target tapi juga dapat menyebabkan ketersinggungan kepada orang yang berhubungan dengan ujaran SARA tersebut. Adapun dilakukan penelitian ini adalah menerapkan klasifikasi *Sentiment analysis* guna mencari tau kebiasaan orang Indonesia yang masih menggunakan *Twitter* sebagai media untuk menyebarkan kebencian melalui ujaran SARA. Apa yang dilakukan dalam jurnal ini adalah perbandingan antara metode klasifikasi *Support Vector Machine (SVM)* dengan *Naïve Bayes* untuk mencari tingkat akurasi yang lebih tinggi dalam kasus pengklasifikasian ujaran SARA dan dilakukan juga pembobotan kata dengan metode *TF-IDF*. Adapun guna lainnya dari penelitian ini adalah untuk menentukan performansi dari *Support Vector Machine* dalam melakukan proses klasifikasi *Tweet SARA*. Dari hasil pengujian sebanyak 2x menunjukkan bahwa implementasi *Multinomial Naïve Bayes* pada kasus “pengklasifikasian ujaran sara di twitter” memiliki tingkat *accuracy*, *recall*, *F1-score* dan *precision* lebih tinggi pada percobaan kedua dengan 82,97%, 82,97%, 81,82%, 83,68% dan pada percobaan pertama *SVM Linear* dengan $C = 1$ dan $\gamma = 1$ lebih unggul dengan 56,95%, 56,95%, 56,84%, 56,83% dan juga pengukuran dilakukan dengan 2 kernel *SVM* lain yaitu *RBF* dan *Polynomial* dan memperoleh tingkat akurasi 59.57%, *recall* 59.57%, *F1-score* 55.14%, dan *precision* 72.90% untuk *RBF* dan untuk akurasi 29.78%, *recall* 29.78%, *F1-score* 16.22%, dan *precision* 46.11% pada percobaan ketiga *SVM linear* mendapat hasil terbaik saat $C = 2$ dan $\gamma = 2$ dan $C = 1$ dan $\gamma = 3$ yaitu akurasi 80.85%, *recall* 80.85%, *F1-score* 80.84% dan *precision* 81.81% .

Kata kunci : SARA, *Support Vector Machine*, *Naïve Bayes*, *Sentiment analysis*