

INTISARI

PERBANDINGAN NAIVE BAYES TERMODIFIKASI UNTUK DETEKSI ARTIKEL CLICKBAIT BERBAHASA INDONESIA

Oleh

Lalu Tarangga Arief Gunawan

18/424190/PA/18295

Judul *clickbait* telah memberikan dampak negatif ke masyarakat. Masyarakat merasa ditipu dan termanipulasi karena berita tersebut. Di lain pihak sudah banyak penelitian yang menggunakan pembelajaran mesin untuk mendeteksi judul *clickbait*. Penelitian-penelitian tersebut sayangnya hanya berfokus kepada kata-kata yang digunakan tanpa mempertimbangkan penerbit dari artikel tersebut. Pada penelitian ini dataset yang digunakan adalah CLICK-ID dataset yang terdiri dari 15000 data dengan 8710 data non-clickbait dan 6290 artikel clickbait yang memiliki 12 penerbit yang berbeda. Dataset pada penelitian ini akan dipreprocessing dengan menggunakan lowercase conversion, stemming, dan stopword removal. Dataset yang telah dibersihkan dimasukkan kedalam 2 algoritma Naive Bayes yang telah dilakukan oleh penelitian sebelumnya yaitu Multinomial Naive Bayes (MNB) dan Tree Augmented Naive Bayes (TANB) serta 2 algoritma Naive Bayes yang dimodifikasi yaitu Publisher Multinomial Naive Bayes (PMNB) dan Title Publisher Multinomial Naive Bayes (TPMNB). Hasil dari penelitian ini didapati bahwa TPMNB memiliki akurasi, presisi, dan recall tertinggi dengan nilai berturut-turut 0.909826, 0.936554, dan 0.886057 untuk pelatihan dan 0.774, 0.766667, dan 0.720528 untuk pengujian sedangkan TANB memiliki nilai recall tertinggi dengan nilai 0.876067 untuk pelatihan dan 0.7173 untuk pengujian. Selain itu juga diperoleh bahwa MNB, PMNB, dan TPMNB memiliki kompleksitas pelatihan dan prediksi yang sama yaitu $O(nt)$ dan $O(t)$ sedangkan TANNB memiliki kompleksitas yang lebih besar dibandingkan algoritma lain dalam penelitian ini.

Kata kunci: naive bayes, *clickbait*, judul, penerbit, klasifikasi teks