

Abstrak

Internet telah membawa banyak dampak positif dalam kehidupan manusia. Akan tetapi tidak dapat dipungkiri beberapa dampak negatif juga muncul akibat kebebasan informasi yang ditawarkan melalui media internet. Diantaranya adalah maraknya konten porno yang tersebar dengan bebas di dunia maya.

Untuk itulah diperlukan sebuah mekanisme pemfilteran konten porno yang secara cerdas dapat mengenali apakah suatu *web page* terindikasi mengandung konten porno atau tidak. Mekanisme pemfilteran yang dibahas pada tugas akhir ini menggunakan *naive bayes classifier* sebagai metode pengklasifikasian *web page*. Diharapkan dengan menggunakan *naive bayes classifier* didapatkan akurasi hasil klasifikasi yang tinggi meskipun masing-masing *web page* mempunyai bahasa yang berbeda yaitu Bahasa Indonesia dan *English*.

Dalam penerapan *naive bayes classifier*, *web page* harus di-*preprocess* terlebih dahulu. *Web page* akan diekstraksi menjadi token/atribut. Permasalahan muncul ketika ada token/atribut yang mempunyai varian nol yang dapat menyebabkan ketidakakuratan pada perhitungan *posterior probability naive bayes classifier*. Untuk mengatasi hal tersebut penulis menawarkan tiga alternatif penanganan atribut bervariasi nol, yaitu pengabaian atribut bervariasi nol, penambahan instance dummy, dan inisialisasi standar deviasi minimum. Selain itu juga dianalisis pengaruh penggunaan *stopword removal* pada preprocessing dalam meningkatkan hasil akurasi klasifikasi.

Ditunjukkan dalam tugas akhir ini bahwa alternatif terbaik dalam penanganan atribut bervariasi nol adalah dengan menggunakan inisialisasi standar deviasi minimum dan penanganan *stopword removal*. Alternatif ini dapat menghasilkan nilai akurasi klasifikasi yang relatif tinggi yaitu 96%.

Kata kunci: *naive bayes classifier*, pemfilteran, porno, preprocess, *stopword removal*, atribut bervariasi nol