

ABSTRAK

Kalimat sitasi berperan penting dalam penulisan jurnal ilmiah. Kalimat sitasi dapat diidentifikasi dengan mengekstraksi fitur-fiturnya. Pada penelitian ini digunakan 5 fitur utama dan juga akan dikombinasikan. Fitur-fitur yang kita gunakan adalah unigram, bigram, proper noun, cue phrase, dan pronoun. Untuk mengklasifikasi kita menggunakan *naive bayes* (NB) dan *support vector machine* (SVM). Penelitian ini menggunakan 500 makalah ilmiah yang diambil dari acl-arc. Hasil dari penelitian ini adalah fitur yang terbaik untuk mengidentifikasi kalimat sitasi adalah “*proper noun*, dan *cue phrase*” dengan 59,069% f-measure, dan 92,157% akurasi, jika menggunakan naive bayes, dan 51,234% f-measure, dan 92,503% akurasi jika menggunakan SVM.

Kata kunci: *supervised learning*, ekstraksi fitur, identifikasi, kalimat sitasi