ABSTRAK

Penelitian ini digunakan dalam klasifikasi bahasa Jawa. Hasil yang dikeluarkan berupa informasi mengenai kategori dokumen, yaitu ekonomi, kesehatan, pendidikan atau politik. Proses awal, yaitu menginputkan dokumen yang akan digunakan sebagai data *training* ke dalam sistem, berdasarkan kategori yang telah diketahui. Kemudian dilakukan proses *pre-processing* berupa *tokenisasi* (pemenggalan kata dan penghapusan tanda baca dan karakter), *case folding* (mengubah kata kedalam huruf kecil), *stopword* (penghapusan kata yang dianggap tidak penting), *stemming* (pengembalian kata kebentuk dasar), dan menghitung *term frequency*. Setelah menghasilkan kata unik, diolah untuk dihitung W (bobot kata) dan *Laplace Smoothing* dan digunakan dalam proses klasifikasi. Pada data *testing*, dokumen juga melewati proses *pre-processing*. Dari kedua data, dilakukan proses *matching*, yaitu mendapatkan kata – kata yang sama dari data *training* dan *testing*. Jika data *matching* telah diperoleh, maka akan digunakan untuk menjalankan proses klasifikasi menggunakan metode Naïve Bayesian. Pada penelitian ini dilakukan pengujian *cross validation* kemudian dilakukan uji presisi. Data yang digunakan sebanyak 40 dokumen. Tingkat akurasi untuk 3 *fold* mencapai 69,78 %, untuk 5 *fold* mencapai 77,5%.

.

Kata kunci : klasifikasi dokumen bahasa Jawa, *Naïve Bayesian*, pemerolehan informasi

ABSTRACT

This research is used for javanese classification. The output are information about document category, there are economic, health, education, or politic. The first process is inputing document that will be used for training data into the system, based on known category. Then the process continue with preprocessing for make model of documents collection that inputted like tokenizing (slice of words and erasing punctuation and character), case folding (change word into lower case), stopwords (erasing unimportant words), stemming (returning the word into first form), and counting term frequency. After producing unique word and will processed to count W (word weight) and Laplace smoothing and used for classification process. At testing data, documents also need preprocessing. From both process, will be doing matching process, that is accuiring the same words from training data and testing. If matching data is done, then it will be used for classification process using Naïve Bayesian method. At this research will be using cross validation. Data that is used are 40 documents. Accuration for 3 fold reach 69,78%, and for 5 fold reach 77,5%.

Keywords : Javanese languange classification, Naïve Bayesian, Information Retrieval