

## ABSTRAK

Penelitian ini bertujuan untuk membangun sebuah sistem pengelolaan surat masuk yang dikelola oleh sekretariat Fakultas Sains dan Teknologi Informatika Universitas Sanata Dharma. Sistem yang dibangun memiliki fungsi klasifikasi dan pencarian surat. Data yang digunakan dalam penelitian adalah arsip (surat) fisik. Surat – surat tersebut akan diproses menjadi arsip digital untuk kemudian dapat diolah oleh sistem yang akan dibangun. Arsip digital yang dimaksudkan adalah berupa file PDF dan teks.

Fungsi klasifikasi yang dimiliki sistem mengelompokkan surat ke dalam delapan kategori. Kedelapan kategori tersebut, yaitu Administrasi Umum, Akademik, Kemahasiswaan, Kepegawaian, Keuangan, Perlengkapan, Yayasan, dan Lain – lain. Proses klasifikasi surat pada sistem menggunakan algoritma Multinomial Naive Bayes. Melalui pengujian menggunakan *3-fold cross validation* terhadap 278 data didapat persentase rata – rata benar sebesar 79% dan salah sebesar 20%. Kemudian pada pengujian menggunakan *5-fold* terhadap data yang sama didapat persentase rata – rata benar sebesar 83% dan salah sebesar 16%. Kesalahan klasifikasi yang terjadi dapat disebabkan oleh penentuan kelas terhadap *data training* yang mungkin mengalami kesalahan. Hal ini dikarenakan penentuan kelas tersebut tidak dilakukan oleh pakar.

Fungsi pencarian pada sistem memanfaatkan *library* Apache Lucene. Selain mencari surat, fungsi pencarian juga dapat menampilkan surat yang berhasil ditemukan. Fungsi pencarian memiliki filter kategori surat yang dapat dipilih oleh *user*.

## ABSTRACT

This research aims to build a management system of archive letters, which are managed by the secretariat of Science and Technology Faculty of Sanata Dharma University. The system has functions to classify and search the letter. The data that will be used in this research are physical archives (letters). These letters will be made into digital archives for which can be processed by the system. The digital archives will be in the form of PDF and text file.

The classification function of the system will classify letters into eight categories. Those eight categories are *Administrasi Umum, Akademik, Kemahasiswaan, Kepegawaian, Keuangan, Perlengkapan, Yayasan, dan Lain – lain*. The process of these letters classification is using Multinomial Naive Bayes algorithm. The average percentage results based on the validation using 3-fold cross validation to 278 data are 79% of correct classification and 20% of wrong classification. Then, based on the validation using 5-fold cross validation to same data, the average percentage results are 83% of correct classification and 16% of wrong classification. Those false classifications could be caused by uncorrectly classified training data. The classification of training data was not performed by expert.

The search function in the system is using the Apache Lucene library. Beside to search the letters, the search function also could open the found letters. This search function has letter categories filter that can be choosed by the user.