



ABSTRACT

Research abstract retrieving in library of is research reports encountered many constraints. The constraints caused research reports not organized and categorized by their knowledge. As a result, abstracts of research reports are difficult to find when their needed. Though abstract is one of text data so abstract difficult to processed with algorithms. Therefore automation research abstract categorization is expected to help solve these problems.

This study concentrated to use a text mining statistic, classification, and clustering methods. In this study, the algorithm statistic used is Naïve Bayes. In this study, the algorithm classification used is SVM and MLP whereas the algorithm's clustering used is EM.

This research result is Naïve Bayes algorithm have higher accuracy than Algorithms MLP, SVM and EM. That indicated from correct rate's Naïve Bayes Algorithm by 90 % and 86,6667 %. MLP, SVM, and EM Algorithm showed lower performance than Naïve Bayes Algorithm. MLP acquire correct rate by 80 % and 78,3333 % and SVM by 81,6667 % and 76,6667 %. While EM Algorithm received incorrect rate of 56,6667 % dan 63,3333 %

Keywords : Text mining, abstract research, categorization, clasification, *clustering*



INTISARI

Pencarian abstrak laporan penelitian di perpustakaan sering kali menemui banyak kendala. Kendala tersebut terjadi akibat penyimpanan laporan penelitian yang tidak terorganisasi dan dikategorikan berdasarkan kategori bidang ilmu. Akibatnya, abstrak laporan penelitian tersebut sulit ditemukan saat dibutuhkan. Padahal abstrak penelitian merupakan salah satu data teks yang sulit diolah secara algoritme. Oleh karena itu otomatisasi kategorisasi abstrak penelitian diharapkan dapat membantu memecahkan persoalan tersebut.

Penelitian ini menekankan pada kategorisasi abstrak penelitian menggunakan metode statistika, klasifikasi, dan kluster. Pada metode statistika, algoritme yang digunakan pada penelitian ini adalah Naïve Bayes. Pada metode klasifikasi menggunakan SVM (Support Vector Machine) dan MLP (Multilayer Perceptron) sedangkan pada kluster, algoritme yang digunakan adalah EM (Expectation Maximization).

Hasil penelitian ini ialah Algoritme Naïve Bayes memiliki akurasi yang lebih tinggi daripada algoritme MLP, SVM, dan EM. Hal ini ditunjukkan *correct rate* Algoritme Naïve Bayes sebesar 90 % dan 86,6667 %. Algoritme MLP, SVM, dan EM menunjukkan akurasi yang lebih rendah dibanding Naïve Bayes. MLP memperoleh *correct rate* sebesar 80 % dan 78,3333 % dan SVM memperoleh 81,6667 % dan 76,6667 %. Sedangkan EM memperoleh *incorrect rate* sebesar 56.666 % dan 63,3333 % .

Kata kunci teks mining, abstrak penelitian, kategorisasi, klasifikasi, *clustering*.