ABSTRAK

klasifikasi merupakan salah satu teknik *data mining* dan juga *text mining* yang digunakan dalam proses mencari set model atau fungsi-fungsi yang menjelaskan atau membedakan kelas-kelas data dan konsep yang bertujuan memungkinkan untuk menggunakan model tersebut dalam melakukan testing terhadap data testing.

Dalam klasifikasi, dilakukan pembelajaran terhadap data training dimana data training tersebut memiliki atribut-atribut yang salah satu diantaranya ditetapkan sebagai kelas untuk baris-baris data pada data training itu. Hasil pembelajaran tersebut berupa suatu model atau yang sering disebut dengan classifier. Model ini nantinya akan digunakan untuk memprediksi dan menguji atribut yang menjadi kelas dari data testing.

Proses pengklasifikasian yang sering dilakukan saat ini adalah pengklasifikasian terhadap dokumen teks bahasa inggris. Oleh karena itu, pada penelitian tugas akhir ini, penulis akan mengklasifikasikan dokumen teks berbahasa indonesia. Klasifikasi teks merupakan suatu proses pengklasifikasian dokumen-dokumen ke dalam satu atau lebih kategori yang telah didefenisikan sebelumnya atau ke dalam kelas-kelas dari dokumen yang sama. Dokumen-dokumen yang telah terklasifikasi ke dalam kelas yang telah didefenisikan sebelumnya merupakan tujuan utama dari klasifikasi teks. Masing-masing dokumen dapat dikategorikan ke dalam banyak kategori, tepat satu kategori, atau bahkan tidak termasuk ke dalam kategori manapun.

Pada pengklasifikasian dokumen teks berbahasa indonesia ini, penulis menggunakan dua metode atau pendekatan yang terdapat dalam *information retrieval* dan *machine learning*. Kedua metode tersebut adalah *Latent Semantic Indexing* dan *Support Vector Machine*. Dengan menggunakan kedua metode ini diharapkan dapat mempermudah dalam proses pengklasifikasian khususnya untuk dokumen teks bahasa indonesia.

Kata kunci : klasifikasi, text mining, data mining, machine learning, information retrieval, latent semantic indexing, support vector machine.