

# **APLIKASI DETEKSI KEMIRIPAN DOKUMEN MENGGUNAKAN TEXT MINING DENGAN METODE COSINE SIMILARITY**

Oleh  
**Bagus Kurniawan**  
**12.622.032**

Diajukan kepada Program Studi Teknik Informatika Fakultas Teknik  
Universitas Muhammadiyah Gresik pada tanggal 20 Juli 2017  
untuk memenuhi sebagian persyaratan untuk menjadi sarjana S-1 Program Studi  
Teknik Informatika

## **INTISARI**

Dokumen teks merupakan suatu bentuk penyampaian laporan dalam bentuk tulisan yang terkadang memiliki kemiripan dari sumber teks yang lain. *Text mining* merupakan salah satu tahapan pengolahan dokumen teks untuk mengetahui *similarity* dokumen satu dengan yang lainnya dengan beberapa tahapan seperti: *stopword*, *case folding*, *stemming*, *tokenizing*, dan pembobotan kata serta *analyzing* dokumen, *case folding* merupakan tahapan awal *text mining* mengubah seluruh kata menjadi huruf kecil, *stopword* merupakan tahapan lanjutan dengan menghapus kata yang termasuk dalam penghubung kata, *stemming* merupakan tahapan lanjutan untuk mengubah kata menjadi kata dasar tanpa imbuhan algoritma yang digunakan nazief & adriani, *tokenizing* merupakan tahapan selanjutnya untuk memotong kata perkata, *analyzing* merupakan tahapan untuk melakukan analisis kata atau melakukan pembobotan kata dalam hal ini pembobotan yang digunakan adalah TF-IDF sedangkan untuk menghitung nilai *similarity* menggunakan metode *cosine similarity*. Penelitian ini membuat sistem untuk mengetahui *similarity* suatu dokumen dan juga memanfaatkan pembobotan TF-IDF untuk mengetahui pengaruh nilai ketelitian nilai *similarity* dokumen yang nantinya sebagai acuan nilai dengan dibuktikan dengan letak kemiripan kata yang sama.

**Kata Kunci:** *text mining*, *cosine similarity*, *tf-idf*.

Pembimbing I : Harunur Rosyid, ST., M.Kom.