

## Abstraksi

Klasifikasi teks adalah proses pengelompokan dokumen ke dalam kelas yang berbeda, dalam tahapannya tiap dokumen menunjuk pada satu kelas tertentu dan dibutuhkan proses untuk menggali informasi dari dokumen tersebut. Teks *preprocessing* yang dilakukan pada Skripsi ini mencakup teknik dalam *retrieval* data di antaranya *word tokenization*, *feature selection*, dan *term weighting* hingga merepresentasikan dokumen dalam *Vector Space Model*. Dalam *feature selection*, akan dihitung bobot dari semua *keyword* kemudian diambil *keyword* yang lebih besar dari nilai *threshold*. Untuk menghitung bobot *keyword*, terdapat beberapa metode yang dapat digunakan seperti *Term Frequency* (TF), *TF\*IDF* (*Inverse Document Frequency*) dan *Information Gain* (IG). Dalam melakukan pembobotan term, konsep perhitungan bobot TF dan TFIDF digunakan dengan adanya tambahan teknik normalisasi seperti L1 dan L2.

Peningkatan kinerja dan efisiensi klasifikasi teks dokumen khususnya klasifikasi *multi-class* dengan sejumlah besar kategori dilakukan dengan menggunakan *error-correcting output codes* (ECOC). ECOC mengurangi masalah *multi-class* pada sekumpulan *binary classification* dan menggabungkan hasil *binary classification* tersebut untuk memprediksi kelas / label pada *multiclass*.

Hasil pengujian menunjukkan bahwa, Pertama, penggabungan metode Support Vector Machine (SVM) dengan Error Correcting Output Coding (ECOC) dapat meningkatkan akurasi jika dibandingkan dengan metode Support Vector Machine (SVM) pada saat jumlah data yang digunakan adalah 10%, 25%, 75% dan 100% terhadap jumlah data pada tiap kategori. Kedua, penggabungan metode Support Vector Machine (SVM) dengan Error Correcting Output Coding (ECOC) dapat meningkatkan akurasi jika dibandingkan dengan metode Support Vector Machine (SVM) pada saat nilai *threshold-nya* lebih besar dari 0.03. Ketiga, penggabungan metode Support Vector Machine (SVM) dengan Error Correcting Output Coding (ECOC) dapat meningkatkan akurasi jika dibandingkan dengan metode Support Vector Machine (SVM) pada saat *Term Weighting* tanpa adanya normalisasi dan keempat, penggabungan metode Support Vector Machine (SVM) dengan Error Correcting Output Coding (ECOC) dapat meningkatkan akurasi jika dibandingkan dengan metode Support Vector Machine (SVM) pada saat penggunaan parameter SVM (C) yang besar, kernel polynomial dengan degree yang kecil dan RBF kernel dengan nilai gamma yang besar.

**Kata kunci:** Klasifikasi, Support Vector Machine, ECOC, *preprocessing*, akurasi