



Intisari

Komparasi Metode *Multiclass Classifier* untuk Pengklasifikasian Artikel Berita Keterbukaan Pemerintah

Oleh:

Adira Kurniawan Firdaus

15/378052/PA/16527

Open Government Indonesia (OGI) adalah suatu inisiatif untuk mengkonsolidasikan gerakan keterbukaan pemerintah dalam mendorong pembuatan kebijakan yang transparan dan inklusif, serta mendorong keterlibatan aktif dari warga negara sesuai dengan prinsip keterbukaan pemerintah. Dalam penelitian ini, akan diteliti potensi penggunaan algoritma untuk membangun model pengkategorian berita keterbukaan pemerintah sesuai dengan prinsip keterbukaan pemerintah.

Klasifikasi dokumen berita berdasarkan kategori di penelitian ini akan membandingkan algoritma klasifikasi *Multinomial Naïve Bayes* (MNB), *Random Forest Classifier* (RFC), dan *Support Vector Machine* (SVM). Pengujian dilakukan pada hasil ekstraksi fitur TF-IDF dari teks berita keterbukaan pemerintah berjumlah 1500 yang dibagi menjadi 5 kategori.

Hasil penelitian ini menunjukkan penggunaan *Grid Search K-Fold Cross Validation* untuk mengoptimalkan pengklasifikasi menghasilkan hasil yang kompetitif di antara ketiganya. Implementasi SVM tanpa *stemming* memiliki hasil akurasi dengan performa terbaik dengan akurasi sebesar 88,00%, presisi 87,98, recall sebesar 87,66%, dan F1-score sebesar 87,75%. Penggunaan RFC dengan *stemming* memiliki performa lebih baik dibandingkan implementasi tanpa *stemming* dengan nilai terbaik yaitu 85,67%, presisi sebesar 84,52%, recall sebesar 84,52%, dan F1-score sebesar 84,45%.

Kata Kunci: *Pengkategorian Berita, SVM, Naïve Bayes, Multiclass Categorization*



Abstract

Comparison Of Multiclass Classifier Methods to Classify Open Government News Article

By:

Adira Kurniawan Firdaus

15/378052/PA/16527

Open Government Indonesia (OGI) is an initiative to consolidate open government movement in guiding transparent and inclusive policy making, and driving active public participation based on open government principles. In this research, the possibilities to develop open government news categorization model based on open government's principles will be explored.

By classifying based on categories, this research will compare Multinomial Naïve Bayes (MNB), Random Forest Classifier (RFC), and Support Vector Machine (SVM) as multiclass classifier. Research will be applied on TF-IDF features extraction from 1500 open government news divided to 5 categories.

The results of the training show that by utilizing Grid Search K-Fold Cross Validation to optimize classifiers, competitive results can be achieved between the three. Implementation of SVM without stemming shows best performance with accuracy score of 88,00%, precision of 87,98%, recall of 87,66%, and F1-score of 87,75%. Usage of RFC with stemming shows better result than without stemming with accuracy score of 85,67%, precision of 84,52%, recall of 84,52%, and F1-score of 84,45%.

Key Words: *News Categorization, MNB, RFC, SVM*