Klasifikasi Berita Bahasa Indonesia Menggunakan *Mutual Information* dan *Support Vector Machine*

Lalu Gias Irham[1], Adiwijaya[2], Untari Novia Wisesty[3]

[1,2,3]Fakultas Informatika, Universitas Telkom, Bandung
[1]giasirham@students.telkomuniversity.ac.id, [2]adiwijaya@telkomuniversity.ac.id,
[3]untarinw@telkomuniversity.ac.id

Abstract
News is a source of information disseminated in various types of media. In order to make it easier for news readers to obtain the desired news, the news needs to be classified. The large number of scattered news creates difficulties in classifying the news based on the topic. Therefore the author conducted a study to classify news into 12 classes (culture, economy, entertainment, law, health, life, automotive, education, politics, sports, technology, and tourism) automatically against 360 Indonesian news data. In this study several test scenarios were conducted to see the effect of stopword removal and stemming methods on data preprocessing, the effect of mutual information in selecting features, and performance of Support Vector Machine in classifying news data. The test results showed that the data using only stemming without stopword removal, using the MI selection feature and SVM classification method produced the best results of 94.24%, compared to the other methods.

Keywords: news, text classification, support vector machine, feature selection, mutual information