

## Intisari

Pertumbuhan jumlah kendaraan bermotor di DIY tidak sebanding dengan pertumbuhan luas jalan, sehingga semakin lama jalan-jalan di DIY menjadi semakin macet. Penelitian ini bertujuan untuk mengembangkan sistem pemantauan arus lalu lintas dengan mengklasifikasi *tweet* dan memetakan lokasi yang dirujuk oleh *tweet* berbahasa Indonesia secara waktu nyata.

Pada penelitian ini, konten *tweet* diolah dengan tahapan pengolahan awal, ekstraksi ciri, klasifikasi *tweet*, pemberian label lokasi, pemotongan frasa lokasi, dan kemudian hasilnya ditampilkan dalam bentuk halaman situs web. Dalam proses klasifikasi, penelitian ini mencoba membandingkan tiga algoritme *machine learning*, di antaranya adalah *Naive Bayes* (NB), *Support Vector Machine* (SVM), dan *Decision Tree* (DT).

Sistem pemantauan arus lalu lintas yang dikembangkan bekerja dengan baik. Sistem ini dapat memberi tahu kondisi arus lalu lintas di berbagai tempat di DIY secara waktu nyata dengan menggunakan data dari jejaring sosial Twitter. Secara umum, algoritme SVM dikombinasikan dengan penggunaan semua kata sebagai ciri mempunyai performa terbaik dalam mengklasifikasikan data *tweet*. Dengan data berimbang 35.184 *tweet* serta penggunaan semua kata sebagai ciri, model klasifikasi NB, SVM, dan DT secara berturut-turut memiliki akurasi 99,37%, 99,77%, dan 99,48%. Dengan data tidak berimbang 110.449 *tweet* serta penggunaan semua kata sebagai ciri, model klasifikasi NB, SVM, dan DT secara berturut-turut memiliki akurasi 99,76%, 99,87%, dan 99,70%.

**Kata kunci :** Lalu Lintas, Twitter, Bahasa Indonesia, Klasifikasi, *Machine Learning*

### ***Abstract***

*Growth in the number of vehicles in DIY were not proportional to the growth of the area of roads, sooner or later roads in DIY would be increasingly jammed. This research aimed to develop a system of monitoring traffic flow by classifying text and mapped the locations referenced by the Bahasa Indonesia text in real-time using Twitter data.*

*In this study, tweet contents were processed to the stages of preprocessing, feature extraction, tweet classification, locations labeling, chunking location phrases, and then the results were displayed in the form of web site pages. In the process of classification, this research tried to compare three machine learning algorithms, among which are Naive Bayes (NB), Support Vector Machine (SVM), and Decision Tree (DT).*

*Traffic flow monitoring system developed in this study worked well. This system could give information about real-time traffic conditions in the various places in DIY by using data from social networking site Twitter. In general, SVM algorithm combined with the usage of all words as features had the best performance in classifying the data tweet. With balanced data 35,184 tweets and with the usage of all words as features, classification model NB, SVM, and DT had accuracy of 98.02%, 98.31% and 98.41% respectively. With imbalanced data 110,449 tweets and with the usage of all words as features, classification model NB, SVM, and DT had accuracy of 99.23%, 99.23% and 99.42% respectively.*

**Keywords :** *Traffic, Twitter, Bahasa Indonesia, Classification, Machine Learning*