

ABSTRAK

IDENTIFIKASI TWEET KEMACETAN LALU LINTAS MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE DAN K-NEAREST NEIGHBORS

Oleh

Ikhsan Budiyanto
12/331431/PA/14685

Salah satu permasalahan yang sering terjadi dan menjadi topik pembicaraan pengguna Twitter di Indonesia adalah kemacetan lalu lintas. Tak heran karena berdasarkan salah satu hasil survei Indonesia menempati urutan ke-11 sebagai negara dengan indeks kemacetan tertinggi di dunia. Dampak negatif yang ditimbulkan dari kemacetan lalu lintas ini dapat meliputi berbagai aspek, mulai dari kerugian waktu, ekonomi, bahkan hingga masalah psikologi. Oleh karena itu, perlu dikembangkan sebuah metode bagaimana mengidentifikasi *tweet* kemacetan lalu lintas yang diposting melalui Twitter sehingga mampu membedakan *tweet* tersebut menginformasikan kejadian kemacetan lalu lintas atau bukan.

Pada penelitian ini, proses identifikasi *tweet* kemacetan lalu lintas menggunakan algoritma klasifikasi Support Vector Machine dan K-Nearest Neighbors. Data yang digunakan merupakan *tweet* yang mengandung kata kunci “macet” dan dikategorikan kedalam 3 kelas yaitu macet lalu lintas sebagai kejadian, macet lalu lintas bukan sebagai kejadian dan macet bukan dalam konteks lalu lintas. Jumlah *tweet* yang digunakan sebagai data *training* sebanyak 1709 dan 419 sebagai data *testing* yang diambil pada 1 Juni - 31 Juli 2018.

Penelitian dilakukan dengan membandingkan algoritma Support Vector Machine dan K-Nearest Neighbors dalam mengidentifikasi *tweet* kemacetan lalu lintas. Evaluasi model dilakukan dengan menggunakan metode *10-fold cross validation*. Dari hasil pengujian didapatkan nilai akurasi 97.14% untuk algoritma Support Vector Machine dengan parameter $C = 1$ dan 92.36% untuk algoritma K-Nearest Neighbors dengan parameter $K = 50$.

Kata kunci: macet, identifikasi, support vector machine, k-nearest neighbors

ABSTRACT

TRAFFIC JAM TWEETS IDENTIFICATION USING SUPPORT VECTOR MACHINE AND K-NEAREST NEIGHBORS ALGORITHMS

Oleh

Ikhsan Budiyanto
12/331431/PA/14685

One of the most recurring and most discussed topic by Indonesian Twitter users is traffic jam. It is not surprising that Indonesia sits on the 11th place in countries with worst traffic congestion in the world, based on a survey. The negative impacts from traffic jam comprises certain aspects, from time and economical loss, even psychological issues. Therefore, a method to identify traffic jam tweets generated on Twitter is necessary to be developed, to distinguish between tweets that are informing traffic jam and tweets that are not.

In this research, Support Vector Machine and K-Nearest Neighbors classification algorithm are used to identify tweets corresponding to traffic jam. Data used are tweets that possess “macet” keyword, and categorized in 3 classes: traffic jam as an occurrence, traffic jam in any context except as an occurrence, and congestion that is not traffic-related. Tweets used as data training is 1709 tweets, and 419 tweets used as data testing, extracted between 1 June and 31 July 2018.

The research compares Support Vector Machine algorithm with K-Nearest Neighbors in identifying traffic jam tweets. Model evaluation is done with 10-fold cross validation method. The testing result comes with 97.14% accuracy for Support Vector Machine algorithm with $C = 1$ parameter, and 92.36% accuracy for K-Nearest Neighbors algorithm with $K = 50$ parameter.

Keywords: traffic jam, identification, support vector machines, k-nearest neighbors