

Pemantauan Lalu Lintas dan Kondisi Jalan dengan Pendekatan Ekstraksi Informasi dari Media Sosial Twitter = Monitoring Traffic and Road Conditions with Information Extraction Approach from Twitter Social Media

Prabu Kresna Putra, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20510361&lokasi=lokal>

Abstrak

Padatnya aktivitas lalu lintas yang terjadi setiap harinya menimbulkan gangguan yang dapat mengganggu proses pelaksanaan kegiatan berlalu lintas. Pemerintah Provinsi DKI Jakarta telah mengembangkan sistem pemantauan guna memantau situasi lalu lintas dalam mendukung kegiatan berlalu lintas dan mengurangi gangguan situasi lalu lintas. Walaupun terdapat beberapa sistem pemantauan lalu lintas, namun Informasi yang diterima oleh pengguna atau penyelenggara lalu lintas masih tidak lengkap / kurang lengkap. Penelitian ini berfokus memanfaatkan data media sosial Twitter untuk digunakan terkait pemantauan lalu lintas dan kondisi jalan. Pendekatan text mining digunakan untuk mengekstraksi informasi dari data media sosial. Jenis analisa yang digunakan pada penelitian ini adalah klasifikasi untuk penyaringan data relevan dan klasifikasi untuk mengkategorikan data relevan, ekstraksi informasi lokasi untuk ekstraksi informasi lokasi dan geocoding untuk mengkonversi informasi lokasi menjadi informasi geografis. Algoritma klasifikasi yang digunakan pada penelitian ini adalah Naïve Bayes, Random Forest, Logistic Regression dan Support Vector Machine. Metode ekstraksi fitur yang digunakan pada penelitian ini adalah Bag Of Words (BOW) dan Term Frequency - Inverse Document Frequency (TF-IDF). Metode ekstraksi informasi lokasi yang digunakan pada penelitian ini adalah Named Entity Recognition (NER) dan Part Of Speech Tagger. Metode geocoding yang digunakan memanfaatkan library ArcPy. Hasil evaluasi menunjukkan model klasifikasi terbaik untuk menyaring data relevan dihasilkan dengan menggunakan algoritma Logistic Regression dan gabungan tiga fitur ekstraksi BOW Unigram dan TF-IDF Word Trigram dan N-Gram Char dengan nilai F1-score 93%. Model klasifikasi untuk mengkategorikan data relevan dihasilkan dengan menggunakan algoritma Logistic Regression dan gabungan dua fitur ekstraksi BOW Unigram dan TF-IDF N-Gram Char dengan nilai F1-score 96%. Metode ekstraksi informasi lokasi terbaik dihasilkan dengan menggunakan metode Stanford NER dengan nilai F1-score sebesar 48% dan presisi sebesar 84%. Tingkat keberhasilan tahapan geocoding untuk hasil ekstraksi informasi lokasi menggunakan metode NER adalah sebesar 59%. Disamping itu, dilakukan juga visualisasi sederhana berbasis web untuk menampilkan informasi data yang telah diolah dalam bentuk spasial.

.....Dense traffic activities that occur every day cause disruption that can interfere with the process of carrying out traffic activities. The DKI Jakarta Government has developed a monitoring system to monitor the traffic situation in support of traffic activities and reduce traffic disruption. Although there are some traffic monitoring system, but the information received by the user or traffic organizers are still not complete / incomplete. This study focuses on utilizing Twitter social media data for use in monitoring traffic and road conditions. The process of extracting information related to traffic situations uses a method with a text mining approach. The type of analysis used in this study is the classification for the filtering of relevant data and classification for categorizing relevant data, location extraction for the extraction of location information and geocoding to convert location information into geographic information. The classification

algorithm used in this study is Naïve Bayes, Random Forest, Logistic Regression and Support Vector Machine. Feature extraction methods used in this study are Bag of Words (BOW) and Term Frequency - Inverse Document Frequency (TF-IDF). The location extraction method used in this study is Named Entity Recognition (NER) and Part Of Speech Tagger. The geocoding method used utilizes the ArcPy library. The evaluation results show that the best classification model to filter out relevant data is generated using a Logistic Regression algorithm and a combination of three extraction features, namely BOW Unigram and TF-IDF Word Trigram and N-Gram Char with an F1-score of 93%. Classification models to categorize relevant data are generated using Logistic Regression algorithm and a combination of two features, namely Unigram BOW extraction and TF-IDF N-Gram Char with F1-score value of 96%. Best location extraction method produced using Stanford NER with F1-score value of 48% and a precision of 84%. The success rate of the geocoding stage for location extraction results using the NER method was 59%. Besides that, a simple web-based visualization is also produced to display information that has been processed in spatial interface.