

## INTISARI

### KLASIFIKASI KONDISI LALU LINTAS MENGGUNAKAN ALGORITME NAÏVE BAYES BERBASIS DATA TWITTER

Oleh

Mohammad Afrizal

14/364104/PA/15891

Dewasa ini Twitter banyak digunakan untuk mendeteksi keadaan lalu lintas. Pendekatan menggunakan metode Naïve Bayes dilakukan untuk mengklasifikasi *tweet* apakah mengandung sentimen lalu lintas atau tidak dengan hasil akurasi cukup baik. Namun, metode Naïve Bayes belum dikembangkan untuk mengklasifikasi keadaan lalu lintas dari sekumpulan data Twitter. Tujuan dari penelitian ini adalah untuk melihat akurasi dari pengklasifikasian dua keadaan lalu lintas (macet dan tidak macet) berbasis data Twitter menggunakan algoritme Naïve Bayes.

Implementasi algoritme Naïve Bayes dilakukan dalam beberapa tahapan, yaitu pengambilan data, pra pemrosesan, pemilihan ciri, pelatihan data, dan evaluasi. Pengambilan data menghasilkan 4246 *tweets* yang berasal dari berbagai akun. Setelah dilakukan pra pemrosesan didapatkan 3231 data untuk diproses. Ciri waktu dipilih karena merupakan informasi identik dari setiap *tweet*. Kemudian dilakukan pelatihan model klasifikasi menggunakan algoritme Naïve Bayes. Pelatihan model menghasilkan bobot probabilitas prior dan posterior pada setiap ciri.

Hasil pengujian model klasifikasi menggunakan teknik *cross validation* menghasilkan rata-rata akurasi sebesar 77,5%, presisi 90,0%, *recall* 79,1%, dan *f-score* sebesar 82,8%.

Kata kunci— Twitter, Naïve Bayes, Lalu lintas

## ***ABSTRACT***

### ***TRAFFIC CONDITION CLASSIFICATION USING NAÏVE BAYES ALGORITHM BASED ON TWITTER DATA***

*By*

Mohammad Afrizal

14/364104/PA/15891

*Today Twitter is widely used to detect traffic conditions. The approach using the Naïve Bayes method is done to classify tweets whether they contain traffic sentiments or not with accurate results. However, the Naïve Bayes method has not been developed to classify traffic conditions from a set of Twitter data. The purpose of this study is to see the accuracy of the classification of two traffic conditions (jammed and not jammed) based on Twitter data using the Naïve Bayes algorithm.*

*Naïve Bayes algorithm implementation is done in several stages, namely data retrieval, pre-processing, feature selection, data training, and evaluation. Data collection generated 4246 tweets from various accounts. After pre-processing obtained 3231 data to be processed. Time characteristic is chosen because it is identical information from each tweet. Then carried out the classification model training using the Naïve Bayes algorithm. Model training produces both prior and posterior probability weights on each feature.*

*The result of classification test using cross validation technique resulted in average accuracy of 77,5%, precision 90,0%, recall 79,1%, and f-score equal to 82,8%.*

*Keywords— Twitter, Naïve Bayes, Traffic*