

ABSTRAK

Sistem yang dikembangkan oleh penulis merupakan sistem yang dapat menentukan jalan terbaik apabila salah satu jalan mengalami kepadatan kendaraan atau macet. Jalan yang ditentukan pada sistem ini sangat terbatas, hanya jalan seputaran bundaran HI Jakarta.

Dalam penelitian Tugas Akhir ini, data kemacetan diambil dari media sosial Twitter karena banyaknya *tweet* pada twitter yang menyatakan keadaan lalulintas di Jakarta. Sebelum diklasifikasi, data akan melalui proses *pre-processing* yang terdiri dari *Case Folding*, *Cleaning*, *Tokenization*, dan *Data Transformation*. Sistem ini menggunakan algoritma *K-Nearest Neighbor* (KNN) untuk mengklasifikasikan data Twitter. Jumlah data yang digunakan pada penelitian ini adalah 600 data dengan data diuji sebanyak tiga kali. Pada pengujian pertama data dibagi sebanyak 50% *data training* dan 50% *data testing*, sedangkan pada pengujian kedua, data dibagi sebanyak 67% *data training* dan 33% *data testing*, Terakhir, data dibagi menjadi 80% *data training* dan 20% *data testing*. Dari pengujian tersebut didapat nilai *precision* tertinggi pada data *training* sebanyak 80% dan data *testing* sebanyak 20% dengan nilai *Accuracy* = 84,16%, *Precision*= 96,00%, dan *Recall* 84,00% dengan jumlah tetangga (K) adalah 27.

Kata Kunci : Kemacetan, *Data Mining*, *Pre-processing*, Klasifikasi