

Klasifikasi Tweets Argumen dan Non Argumen pada Twitter dengan Metode k-Nearest Neighbors

Fikri Maulana Hanafi¹, Angelina Prima Kurniati², Widi Astuti³

^{1,2,3}Fakultas Informatika, Universitas Telkom, Bandung

¹fmaulanah@students.telkomuniversity.ac.id, ²angelina@telkomuniversity.ac.id,

³widiwdu@telkomuniversity.ac.id

Abstrak

Informasi adalah salah satu hal yang penting untuk didapatkan oleh manusia. Perkembangan teknologi yang pesat ini membuat proses pencarian informasi menjadi semakin mudah untuk didapatkan. Media sosial adalah salah satu media yang memudahkan untuk mendapatkan informasi untuk saat ini. Oleh karena itu media sosial saat ini sudah tidak dapat dipisahkan dengan kehidupan sehari-hari manusia. Hal tersebut membuat pengguna media sosial akan sangat cepat mendapatkan informasi yang tersebar sangat luas di internet. *Twitter* adalah platform media sosial yang berbasis *computer-mediated* untuk komunikasi secara *online*. Media sosial *Twitter* memiliki sekitar 1,3 miliar akun dan ada sekitar 336 juta pengguna aktif yang memposting sekitar 500 juta *tweets* per hari. Dengan menggunakan metode klasifikasi *k-Nearest Neighbor* penulis membuat analisis mengenai klasifikasi *tweets* pada *Twitter* ke dalam kelompok argumen atau non-argumen. Alasan dipilihnya metode *k-Nearest Neighbors* karena metode *k-NN* tidak memerlukan parameter yang terlalu spesifik. Parameter yang digunakan adalah bukan parameter pasti, melainkan dari banyaknya jumlah data latih. Penelitian ini menggunakan 12 skema pengujian dengan dikombinasikan dengan penggunaan *feature extraction* TF-IDF dan *Bag of Words* serta *distance measure* pada *k-NN* yaitu *Euclidean*, *Manhattan*, dan *Minkowski*. Model terbaik yang didapatkan pada penelitian ini adalah skema yang menggunakan *feature extraction* TF-IDF serta *distance measure* menggunakan *Euclidean* dengan nilai $n = 3$ (tiga), lalu skema menggunakan *feature extraction* TF-IDF serta *distance measure* menggunakan *Manhattan* dengan nilai $n = 3$ (tiga), dan terakhir yaitu skema menggunakan *feature extraction* TF-IDF serta *distance measure* menggunakan *Manhattan* dengan nilai $n = 4$ (empat). Hasil dari ketiga skema tersebut didapatkan nilai akurasi dengan nilai 91,5 %, *precision* dengan nilai 85,7 %, *recall* dengan nilai 27,3 %, dan terakhir *F1-Score* dengan nilai 41,4 %.

Kata kunci : *k-nearest neighbors, twitter, argumen, klasifikasi, media sosial*