

BAB I

PENDAHULUAN

1.1. Latar Belakang

Dalam laporan yang ditulis oleh We Are Social ada 204,7 juta pengguna internet di Indonesia pada awal tahun 2022. Angka ini naik tipis 1,03% dibandingkan awal tahun sebelumnya yang tercatat berjumlah 202.6 juta [1]. Naiknya jumlah pengguna internet sangat bisa dipahami terlebih akibat adanya wabah Covid 19 yang membuat masyarakat harus melakukan aktivitas seperti belajar, maupun bekerja secara daring. Selain aktivitas belajar dan bekerja, 95 persen internet digunakan untuk mengakses konten pada media sosial.

Indonesia merupakan negara di urutan ke 5 dalam penggunaan sosial media Twitter[2]. Adapun twitter banyak digunakan sebagai media promosi baik promosi dalam bidang politik, ekonomi, agama, dan sosial budaya. Proses promosi yang dilakukan biasanya dari *tweet-ke-tweet*, semakin banyak *tweet* yang dibuat maka akan membuat *trending* suatu topik tersebut. Hal ini karena tingginya tingkat interaksi di twitter dan berbeda dari sosial media seperti Facebook, di Twitter kita bisa melakukan *jump into conversation* antar sesama pengguna tanpa khawatir dianggap aneh.

Banyak diskusi yang dilakukan di Twitter salah satunya tentang bergabungnya dua perusahaan teknologi asal Indonesia Gojek dan Tokopedia menjadi satu entitas bisnis raksasa bernama GoTo. Keduanya resmi menggabungkan layanan *e-commerce*, logistik, transportasi serta layanan keuangan. Munculnya GoTo sempat masuk *trending* topik pada tahun 2021. Para pengguna twitter memberikan opini baik pro maupun kontra terhadap mergernya dua perusahaan ini. Opini-opini tersebut dapat diolah, dicari dan dianalisis orientasi sentimennya. Analisis sentimen

adalah proses untuk menentukan apakah suatu tulisan seseorang termasuk dalam kategori positif, atau negatif. Hal ini bertujuan untuk mendapatkan pendapat dan sikap masyarakat.

Dalam penelitian ini penulis akan melakukan analisis sentiment terhadap pengguna Twitter mengenai mergernya perusahaan teknologi Gojek dan Tokopedia. Hal ini dilakukan dengan memasukan data *tweet* dalam Bahasa Indonesia, klasifikasi dilakukan dengan algoritma KNN (*K-Nearest Neighbors*). Algoritma ini menjadi salah satu metode yang digunakan untuk mengelompokkan objek berdasarkan contoh pelatihan terdekat di ruang fitur dan untuk menentukan apakah *tweet* tersebut berorientasi pada sentiment positif, negatif atau netral.

1.2. Rumusan Masalah

Berdasarkan latar belakang diatas, dirumuskan permasalahan yang diselesaikan dalam penelitian ini adalah bagaimana mengamati dan menganalisis opini pengguna twitter mengenai bergabungnya dua perusahaan teknologi Gojek dan Tokopedia, kemudian mengklasifikasikan data *tweets* tersebut, bagaimana tingkat akurasi KNN (*K-Nearest Neighbors*) dalam sentimen analisis Bahasa Indonesia.

1.3. Batasan Masalah

Batasan masalah dalam penelitian ini adalah :

1. Data *tweets* yang digunakan sebanyak 2000 yang telah di filter dan diambil pada tanggal 19 Mei 2022 sampai 27 Mei 2022.
2. Data *tweets* yang akan dinilai adalah *tweets* yang mengandung keyword berikut: GoTo, Gojek, IPO Tokopedia, IPO Gojek, IPO GoTo.
3. Metode pengklasifikasian yang digunakan dalam penelitian ini adalah metode *K-Nearest Neighbors (KNN)*.

4. Hasil keluaran penelitian ini dari pengklasifikasian sentiment menggunakan *K-Nearest Neighbors (KNN)* berupa sentimen positif dan negatif serta akurasi dari metode ini.
5. Data yang digunakan pada penelitian ini adalah data *tweets* dalam Bahasa Indonesia.

1.4. Maksud dan Tujuan Penelitian

1.4.1. Maksud Penelitian

Berdasarkan permasalahan yang dibahas dalam penelitian ini, adapun maksud dari penelitian ini adalah untuk menentukan hasil analisis sentiment data *tweet* apakah *tweet* tersebut berorientasi pada sentiment positif, negatif atau netral.

Serta sebagai salah satu syarat dalam menyelesaikan studi pada program sarjana Informatika Universitas AMIKOM Yogyakarta.

1.4.2. Tujuan Penelitian

Tujuan dari pembuatan Skripsi ini adalah :

1. Menerapkan Algoritma *K-Nearest Neighbors (KNN)* untuk klasifikasi sentimen masyarakat dari data twitter dengan topik bergabungnya dua perusahaan teknologi asal Indonesia Gojek dan Tokopedia.
2. Mengetahui sentimen pengguna twitter bergabungnya dua perusahaan teknologi asal Indonesia Gojek dan Tokopedia
3. Mengetahui seberapa cocok Algoritma *K-Nearest Neighbors (KNN)* untuk analisis sentimen Bahasa Indonesia.

1.5. Manfaat Penelitian

Adapun manfaat penelitian ini adalah :

1. Mendapatkan hasil akurasi dari Algoritma *K-Nearest Neighbors (KNN)* pada klasifikasi sentimen dengan topik bergabungnya dua perusahaan teknologi asal Indonesia Gojek dan Tokopedia.

2. Memberikan informasi tentang sentimen twitter terhadap rencana bergabungnya dua perusahaan teknologi asal Indonesia Gojek dan Tokopedia.
3. Menjadi referensi untuk penelitian selanjutnya yang berhubungan dengan Sentimen Analisis Bahasa Indonesia.

1.6. Metodologi Penelitian

Untuk menghasilkan laporan dengan data yang akurat serta relevan, metode penelitian yang digunakan dalam penelitian ini meliputi beberapa tahap yaitu sebagai berikut :

1.6.1. Metode Pengumpulan Data

1. Metode Deskriptis

Penulis melakukan crawling data *tweets* pada trending twitter tentang bergabungnya dua perusahaan teknologi asal Indonesia Gojek dan Tokopedia menggunakan tools software Rapidminer dengan mengarsipkan data dalam bentuk file csv/xlsx. Data yang telah diambil selanjutnya difilter untuk mendapatkan data yang siap untuk dianalisis.

2. Studi Pustaka

Penulis menghimpun informasi dengan mengumpulkan referensi diantaranya: buku, jurnal-jurnal, dokumen skripsi maupun website mengenai metode K-Nearest Neighbors (KNN) guna menunjang penelitian.

1.6.2. Metode Analisis

Tahap analisis digunakan untuk menganalisis dan mempelajari data maupun informasi yang sudah didapatkan pada proses pengumpulan data. Data maupun informasi yang telah di analisis digunakan untuk mengembangkan metode yang dapat mengklasifikasi data twitter. Data yang telah didapatkan akan melewati serangkaian *preprocessing text* (*Cleaning, Case Folding, Tokenization, Filtering, Stemming, Normalization*). Data yang telah melalui tahap *preprocessing* akan dibagi menjadi 2, yaitu data latih dan

data uji. Dalam penelitian ini, proses *K Nearest Neighbors* meliputi 3 proses, yaitu : Menghitung bobot kata (TF-IDF), Menghitung tingkat kemiripan (*Cosine Similarity*) dan Menentukan nilai sentiment. Selanjutnya impelementasi metode K-Nearest Neighbors dalam bentuk program computer untuk mengetahui kemampuan dari metode K-NN tersebut.

1.6.3. Metode Perancangan

Tahapan dalam merancang aplikasi ini terdiri dari beberapa tahap yaitu:

1. Perancangan Sistem berisikan tahapan yang dilewati (input data, *preprocessing*, klasifikasi K-NN) sehingga menghasilkan akurasi.
2. Perancangan Antar Muka (*Interface*) Agar terjalin hubungan baik antara pengguna (*user*) dan aplikasi, maka perlu dirancang tampilan antarmuka.

1.6.4. Metode Pengujian

Metode pengujian yang dilakukan untuk menguji system dengan dilakukan perhitungan akurasi dari model pengklasifikasian yang dihasilkan. Maka tahap ini menggunakan metode *confusion matrix*. *Confusion matrix* merupakan salah satu metode yang dapat digunakan untuk mengukur kinerja suatu metode klasifikasi dalam perhitungan akurasi pada konsep data mining.

1.7. Sistematika Penelitian

Sistematika penulisan laporan ini terdiri dari pokok-pokok permasalahan yang dibahas pada masing-masing yang diuraikan menjadi beberapa bagian:

BAB I PENDAHULUAN

Bab ini membahas tentang gambaran umum isi tugas akhir yang meliputi latar belakang yang menguraikan konsep-konsep dasar tentang penelitian ini. Pada pendahuluan ini juga di uraikan rumusan masalah, batasan masalah, tujuan penelitian, metode penelitian dan sistematika penelitian.

BAB II LANDASAN TEORI

Pada bab ini akan dibahas mengenai segala sesuatu hal yang berkaitan dengan proses penulisan tugas akhir ini dan dijadikan sebagai sebuah landasan dalam penulisan dan penelitian.

BAB III METODE PENELITIAN

Bab ini berisi tentang penjelasan metode dalam menganalisa perhitungan yang digunakan dalam penelitian, permodelan data yang digunakan untuk penelitian, dan tahapan penelitian.

BAB IV HASIL DAN PEMBAHASAN

Bab ini berisi implementasi algoritma yang diusulkan terhadap dataset, penguraian tahapan peneliti dalam melakukan penelitian, dan hasil dari penelitian.

BAB V PENUTUP

Bab ini berisi kesimpulan dari penelitian yang sudah dilakukan, penarikan hasil dari penelitian, dan berisi saran yang dapat digunakan peneliti lain untuk mengembangkan hasil dari penelitian yang sudah dilakukan.

DAFTAR PUSTAKA

Berisikan tentang sumber-sumber, jurnal ataupun buku-buku yang menjadi refrensi sebagai acuan dalam penyusunan skripsi ini.