

BAB I

PENDAHULUAN

1.1 Latar Belakang

Media sosial adalah saluran berbasis Internet yang memungkinkan pengguna untuk berinteraksi secara oportunistik dan selektif, dengan secara *real-time* maupun secara *asynchronously* (atau komunikasi yang dalam penyampaian dan penerimaan pesan dengan menggunakan computer dan layanan daring sebagai media untuk komunikasi) dengan audiens yang luas dan sempit untuk mendapatkan nilai dari konten yang dibuat pengguna dan persepsi interaksi dengan orang lain[1].

Dalam penelitian ini, peneliti akan memanfaatkan salah satu media sosial untuk pengambilan data yaitu Twitter. Twitter dipilih dikarenakan memiliki 100 juta pengguna aktif, yang menyimpan pengetahuan berharga yang seringkali tersembunyi dibalik nilai konten yang dibuat oleh para pengguna yang tidak dapat diproses dengan mudah[2]. Twitter diartikan sebagai salah satu website *Micro-blogging* terbesar didunia yang telah berevolusi menjadi sumber berbagai jenis data, dimana pada media Twitter semua orang bisa memposting kejadian yang seseorang alami dengan rentang waktu sama dengan kejadian dan juga didalam Twitter seseorang dapat berbagi pandangan mereka tentang berbagai cara pandangan terhadap isu politik ataupun sebuah produk dapat menimbulkan sebuah sentimen[3].

Setelah mengekstraksi sentimen dari sebuah web *micro-blogging* seperti Twitter, seseorang dapat menggunakan sentimen untuk rekomendasi, ringkasan

system baru, pemasaran produk baru, karenanya menganalisis sentimen itu penting.[4]

Data sentimen yang berhasil dikumpulkan melalui Twitter inilah yang akan digunakan oleh peneliti untuk menjadi bahan penelitian perbandingan antara KNN dan SVM, dan menemukan sentimen dari kedua buah objek.

Ada beberapa pendekatan analisis sentimen di Twitter, salah satunya adalah pembelajaran mesin. Terdapat banyak algoritma yang telah menerapkan implementasinya pada teks atau dokumen, seperti NBC, SVM, KNN, Decision Tree. Dipenelitian ini peneliti hanya akan menggunakan dua algoritma, yaitu SVM dan KNN. SVM merupakan metode supervised learning yang mengenali pola-pola yang digunakan untuk klasifikasi[5]. SVM adalah sebuah klasifikasi algoritma linear yang memiliki peranan utama menentukan pemisah linear yang paling utama di ruang pencarian terbaik yang dapat memisahkan dua kelas berbeda. Kelebihan SVM adalah mampu mengidentifikasi hyperplane yang terpisah yang memaksimalkan margin antara dua kelas[6].

K-Nearest Neighbor adalah metode sederhana namun efektif untuk teks kategorisasi[5]. Sebuah metode yang sederhana namun memiliki tingkat akurasi yang tinggi[7]. Metode K-Nearest Neighbor dinilai tangguh dalam menghadapi noise[8].

Dengan alasan diatas algoritma klasifikasi SVM dan KNN akan mencari tahu sentimen kedua buah objek, waktu proses dan tingkat akurasi pengklasifikasian teks dokumen. Dan menilai peforma mana yang lebih baik antara SVM dan KNN dalam pengklasifikasian teks dokumen.

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah yang telah dipaparkan di atas, maka rumusan masalah penelitian ini adalah

1. Diantara dua algoritma, yaitu K-Nearest Neighbor dan Support Vector Machine, manakah algoritma yang memiliki akurasi dan waktu proses terbaik.
2. Lebih condong kemanakah sentimen yang didapatkan Gojek dan Grab.

1.3 Batasan Masalah

Batasan masalah pada penelitian ini antara lain:

1. Penelitian ini hanya membandingkan dua model algoritma klasifikasi, yaitu Support Vector Machine dan K-Nearest Neighbor.
2. Penelitian ini hanya menggunakan sentimen terhadap Gojek dan Grab di Twitter sebagai objek pembuatan model algoritma.
3. Penelitian ini hanya mengukur nilai akurasi dan waktu proses model SVM dan KNN.
4. Jumlah *tweet* yang dipakai pada penelitian ini berjumlah 2000 *tweets*, dimana 1600 digunakan untuk data latih dan 400 digunakan untuk data uji, yang diambil dari tanggal 4 November 2018 – 21 Januari 2019.
5. *Tweet* yang diambil hanya yang berbahasa Indonesia.

1.4 Maksud dan Tujuan Penelitian

1. Penelitian ini bertujuan untuk membandingkan antara SVM dan KNN, dengan menggunakan tolak ukur akurasi dan waktu proses pada klasifikasi teks.
2. Untuk mendapatkan sentimen terhadap kedua perusahaan tersebut dari hasil penelitian yang akan dilakukan.

1.5 Manfaat Penelitian

Manfaat Penelitian ini adalah sebagai berikut:

1. Penelitian ini atau yang sejenis berguna untuk mengetahui sentimen masyarakat khususnya pengguna Twitter terhadap objek. Hasil yang didapatkan dapat dijadikan referensi untuk meningkatkan kredibilitas dan memperoleh kepercayaan dan berkembang menjadi lebih baik kedepannya.
2. Bagi pengguna internet dapat menjadikan hasil penelitian sebagai pertimbangan *cita* dari objek yang diteliti.
3. Hasil penelitian dapat dijadikan referensi bagi para peneliti berikutnya yang akan membahas mengenai analisis sentimen (*opinion mining*) terutama pada jejaring sosial Twitter.

1.6 Metode Penelitian

1. Pengumpulan Data
 - a. Studi Literatur

Studi literatur dilakukan untuk mempelajari penelitian-penelitian dengan tema sejenis yang pernah dilakukan sebelumnya.

Penelitian-penelitian ini dijadikan referensi pada penelitian ini. Sumber yang dipakai dapat berupa buku, jurnal, *paper*, *website*, dan lain-lain. Pada penelitian ini, referensi yang dicari adalah mengenai analisis sentimen di Twitter dan metode-metode klasifikasi yang dipakai.

b. Pengumpulan Data lewat Twitter

Data *input* yang akan dipakai pada penelitian ini berupa *tweet* dengan kata kunci “Gojek dan Grab”. Data yang sudah dikumpulkan kemudian akan dibagi menjadi data latih dan data uji. Data latih akan diberikan porsi sebesar 80% dan data uji sebesar 20%. Data latih akan diberi label secara manual, sedangkan data uji tidak diberi label.

2. Eksperimen

a. Alat dan Bahan

Penelitian ini menggunakan bahasa Python 3.6 dengan IDE JetBrains Pycharm Community Edition 2018 ver. 2.3, dengan *library* Scikit-learn untuk *machine learning* dan *library* NLTK untuk pra-proses data, serta menggunakan Twitter API. Sedangkan bahan untuk penelitian ini adalah data *tweet* yang diambil dari Twitter.

b. Tahap penelitian

Pada tahap ini akan dilakukan perancangan model yang akan dibuat. Tahapan-tahapan dalam proses perancangan model ini antara lain *pre-processing*, pembobotan dengan metode TF dan TF-IDF,

perancangan model dari data latih dengan menggunakan metode Support Vector Machine dan K-Nearest Neighbor.

c. Pengujian

Model yang telah dibangun kemudian diuji dan dievaluasi. Pengujian dilakukan untuk mengetahui nilai akurasi dan *process time* dari model yang telah dibangun. Selain menguji model, kedua model yang telah dibangun akan dibandingkan, untuk mengetahui model yang memiliki akurasi yang lebih baik. Pengujian menggunakan metode F1 Score berbasis *confusion matrix* dengan mencari *accuracy*, *precision*, *recall* dan *f1 score*. Kemudian model akan divalidasi menggunakan metode K-Fold Cross Validation dengan K=10 untuk memeriksa apakah model melakukan klasifikasi secara *overfitting*.

3. Hasil dan Pembahasan

Pada tahap ini memaparkan dan membahas hasil dari eksperimen yang telah dilakukan. Pembahasan yang dilakukan mengenai akurasi model dan kecepatan waktu proses dari kedua algoritma.

4. Penulisan Laporan

Hasil penelitian kemudian dituliskan ke dalam sebuah laporan. Dalam penulisan laporan penelitian akan ditarik pula suatu kesimpulan berdasarkan hasil pembahasan dan pengujian, serta saran untuk penelitian berikutnya.

1.7 Sistematika Penulisan

Rincian sistematika penulisan yang digunakan dalam penelitian ini adalah sebagai berikut:

1. BAB I PENDAHULUAN

Bab ini memuat paparan latar belakang penelitian, rumusan masalah, batasan masalah dalam penelitian, tujuan penelitian, manfaat penelitian, metodologi penelitian, dan sistematika penulisan.

2. BAB II TINJAUAN PUSTAKA

Bab ini berisi pembahasan tentang penelitian sebelumnya yang dijadikan referensi pada penelitian ini., serta teori-teori yang digunakan dalam penelitian ini.

3. BAB III METODOLOGI PENELITIAN

Bab ini berisi analisis dari penelitian serta perancangan model yang akan dibuat.

4. BAB IV HASIL DAN PEMBAHASAN

Bab ini berisi hasil dan pembahasan dari eksperimen yang telah dilakukan.

5. BAB V KESIMPULAN DAN SARAN

Bab ini membahas kesimpulan dari penelitian serta saran untuk penelitian selanjutnya.