

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan informasi di Indonesia yang semakin pesat membuat kemajuan teknologi semakin canggih. Di bidang teknologi, masyarakat dapat terhubung dan berinteraksi satu sama lain menggunakan *Internet* tanpa batasan waktu. Masyarakat yang aktif menggunakan teknologi, *website*, dan media sosial serta dapat mengumpulkan informasi dengan lebih cepat dan mudah. Komunikasi khususnya di Indonesia nampaknya semakin mudah dengan berkembangnya teknologi, jejaring sosial merupakan alternatif untuk mengumpulkan dan menyebarkan informasi dengan cepat, salah satu media sosial yang saat ini digunakan adalah *Twitter*.

Kementerian Informasi dan Komunikasi (Kemkominfo) menyebutkan saat ini terdapat 82 juta pengguna *internet* di Indonesia. Dengan prestasi tersebut, Indonesia menempati peringkat ke-8 dunia. Dari total jumlah pengguna *internet*, 80% adalah remaja berusia antara 15 dan 19 tahun. “Untuk pengguna Facebook, Indonesia menduduki peringkat ke-4 dunia”, Asosiasi Penyelenggara Jasa Internet Indonesia (APJII) mengumumkan jumlah pengguna *internet* Indonesia mencapai 221 pada tahun 2024. Berdasarkan hasil survei penetrasi *internet* Indonesia tahun 2024 yang dirilis APJII, tingkat penetrasi *internet* di Indonesia mencapai 79,5%. Dibandingkan periode sebelumnya, meningkat 1,4%. Pada tahun 2018, tingkat penetrasi *internet* di Indonesia mencapai 64,8%. Kemudian masing-masing 73,7% pada tahun 2020, 77,01% pada tahun 2022 dan 78,19% pada tahun 2023. Data diambil dari http://www.learningandteaching.info/learning/behaviour_mod.html Artikel APJII, diakses pada tanggal 20 Januari 2024.

Tweet adalah postingan online yang dibuat oleh pengguna *Twitter* atau “*mikroblog*”. *Twitter* merupakan media yang memungkinkan penggunanya untuk saling berinteraksi. Dengan menggunakan *Twitter*, pengguna dapat mengirim pesan kepada pengguna lain secara pribadi atau publik. *Twitter* menghubungkan penggunanya di berbagai belahan dunia, kapan pun, di mana pun dengan akses

internet. Twitter merupakan salah satu sosial media yang banyak digemari dan digunakan masyarakat untuk beropini atau menyampaikan pendapat tentang bahasan tertentu. Pengguna *Twitter* juga sangat luas mulai dari kalangan umum, pejabat negara, *publik figur*, hingga *artis* Putra dkk (2023).

Text Mining merupakan suatu penemuan baru yang sebelumnya informasinya tidak diketahui. Informasi yang diambil dari berbagai sumber daya tertulis dilakukan secara otomatis. Elemen kuncinya adalah menghubungkan beberapa informasi yang diekstraksi menjadi satu sehingga dapat membentuk fakta baru atau *hipotesis* baru untuk dieksplorasi lebih lanjut. Di dalam *Twitter* sendiri belum dapat dihubungkan dengan *analisis sentiment*, tidak terdapatnya fitur untuk mendeteksi cuitan yang berkaitan dengan sentimen positif dan negatif. Salah satu cara untuk mengolah data berupa teks dengan *text mining* Lillah dkk (2023).

Analisis sentimen merupakan proses di mana pengguna *texts mining* mendapatkan berbagai sumber data dari *internet* maupun platform media sosial. Tujuan dari analisis sentimen, biasanya disebut sebagai penambangan opini, adalah untuk mengidentifikasi sentimen atau nada emosional yang direpresentasikan dalam sebuah teks. Dalam penelitian Analisis Sentimen Relokasi Ibukota Nusantara metode yang digunakan adalah *Naïve Bayes* (NB) dan *K-Nearest Neighbor* (KNN). Hasil penelitian memberikan hasil komparatif terhadap kinerja metode-metode tersebut, yaitu metode *Naïve Bayes* memberikan tingkat akurasi analisis sentimen sebesar 82,27%, nilai presisi sebesar 86,36% dan nilai analisis sentimen sebesar 86,36%. Kinerja metode KNN juga memberikan hasil analisis dengan tingkat akurasi sebesar 88,12%, presisi sebesar 93,98% dan nilai recall sebesar 81,53%. Berdasarkan hasil analisis tersebut, prosedur analisis menggunakan metode *K-Nearest Neighbor* mengungguli metode NB dalam mengukur sentimen terhadap pemindahan ibu kota Indonesia Supriyanto dkk (2023).

Text Mining merupakan suatu penemuan baru yang sebelumnya informasinya tidak diketahui. Informasi yang diambil dari berbagai sumber daya tertulis dilakukan secara otomatis. Elemen kuncinya adalah menghubungkan beberapa informasi yang diekstraksi menjadi satu sehingga dapat membentuk fakta

baru atau *hipotesis* baru untuk dieksplorasi lebih lanjut. Di dalam *Twitter* sendiri belum dapat dihubungkan dengan *analisis sentiment*, tidak terdapatnya fitur untuk mendeteksi cuitan yang berkaitan dengan sentimen positif dan negatif. Salah satu cara untuk mengolah data berupa teks dengan *text mining* Firdaus dkk (2021).

Pada penelitian selanjutnya mengenai penelitian terhadap kinerja metode-metode tersebut, yaitu metode *Naïve Bayes* memberikan tingkat akurasi sebesar 82,27%, nilai presisi sebesar 86,36%. Kinerja metode KNN juga memberikan hasil analisis dengan tingkat akurasi sebesar 88,12%, presisi sebesar 93,98%. Berdasarkan hasil analisis tersebut, prosedur analisis menggunakan metode *K-Nearest Neighbor* mengungguli metode NB dalam mengukur sentimen terhadap pemindahan ibu kota Indonesia Syahril dkk (2023).

Menurut penelitian yang telah disebutkan, hal ini dapat disimpulkan perhitungan menggunakan *K-Nearest Neighbor* memiliki presisi akurasi yang lumayan tinggi sesuai untuk melakukan klasifikasi. Oleh karena itu, penelitian ini berfokus untuk menganalisis *Tweet* terhadap pendapat publik dengan pengelompokan analisis sentimen pengguna *Twitter* peneliti memilih menggunakan metode perhitungan *K-Nearest Neighbor*. Peneliti dapat mengetahui akurasi dan presisi yang tertinggi serta hasil sentimen pada *tweet* yang mengacu pada opini publik atas kasus Rafael Alun Trisambodo yang menjabat sebagai Ketua Kepala Bagian Umum Kantor Wilayah Direktorat Jenderal Pajak Jakarta Selatan. Kasus sebagai terdakwa menerima suap dan melakukan tindak pidana pencucian uang (TPPU). Dengan input berupa data *tweet* dalam Bahasa Indonesia, akan dilakukan klasifikasi dengan algoritma KNN (*K-Nearest Neighbor*) untuk menentukan apakah *tweet* tersebut termasuk sentimen positif atau negatif.

1.2 Rumusan Masalah

Berdasarkan latar belakang, maka dapat dituliskan rumusan masalah sebagai berikut :

1. Berapakah K optimal pada metode *K-Nearest Neighbor* atas kasus korupsi Rafael Alun Trisambodo?

2. Berapa tingkat akurasi dan presisi dari metode *K-Nearest Neighbor* pada *tweet* yang mengacu pada opini publik atas kasus korupsi Rafael Alun Trisambodo?

1.3 Tujuan Penelitian

Tujuan dari penelitian ini adalah :

1. Untuk mengetahui berapa nilai K Optimal pada metode *K-Nearest Neighbor* untuk kasus korupsi Rafael Alun Trisambodo.
2. Untuk menentukan hasil klasifikasi akurasi dan presisi pada *tweet* pengguna *Twitter* dengan menggunakan metode *K-Nearest Neighbor*.
3. Untuk mengklasifikasikan komentar pengguna *twitter* pada kelas kelas *sentiment positive* dan *sentiment negative*.

1.4 Manfaat Penelitian

Dengan adanya penelitian ini manfaat yang diharapkan yaitu :

1. Mendapatkan informasi hasil kinerja dari metode *K-Nearest Neighbor*.
2. Memperoleh hasil penelitian yang selanjutnya untuk rujukan penelitian lain dan sebagai informasi secara umum dengan metode *K-Nearest Neighbor*.

1.5 Batasan Penelitian

Batasan masalah dari penelitian ini adalah :

1. Penelitian ini menggunakan klasifikasi hanya pada komentar pengguna *Twitter* menurut Bahasa Indonesia yang baku menurut KKBI.
2. Data yang telah terkumpul dengan *crawling* selama kurang lebih satu bulan, mulai dari tanggal 3 Januari 2024 – 31 Januari 2024.
3. Data dikumpulkan untuk dijadikan sampel sebanyak 500 komentar *tweet* mendapatkan hasil dengan data positif= 0 dengan 251 data dan negatif=1 dengan 249 data.

4. Data teks komentar *Twitter* diklasifikasikan menjadi dua kelas yaitu *sentiment positive* juga *sentiment negative*.
5. Penelitian ini menggunakan metode KNN dengan nilai $K=5,7$ dan 9 .
6. Tidak menyertakan komentar yang menggunakan *emoji* atau emotikon.
7. Kata kunci yang penelitian ini gunakan adalah korupsi rafael alun.

