

BAB I

PENDAHULUAN

1.1 Latar belakang

Perkembangan internet Indonesia kian hari semakin meningkat. Berdasarkan data survei Asosiasi Penyelenggara Jasa Internet Indonesia (APJII) tahun 2016, pengguna Internet mencapai 132,7 juta orang dan mencapai 1.4326 juta pada tahun 2017 (news.detik.com). Hal ini juga terlihat dari survei APJII tentang penggunaan internet di berbagai bidang dimana pengguna internet cenderung membaca berita. Salah satunya seperti penggunaan Internet dalam bidang gaya hidup meraih hasil tertinggi dalam aktivitas media sosial mencapai 87,13%, kemudian unduhan musik 71,10%, unduhan film 70%, dan di ikuti bidang sosial politik. Hasil tertinggi yaitu kegiatan sosial membaca dan membaca. Lingkungan berkurang 50,26%. Terlihat dari survei ini bahwa penggunaan media sosial di Indonesia memiliki tingkat penggunaan internet yang tinggi, diikuti oleh media baca massa online yang menggerakkan banyak Media massa mulai aktif di media sosial. (<https://apjii.or.id/survei2017>).

Shopee merupakan website e-commerce yang menjual berbagai kebutuhan antara lain makanan, pakaian, alat masak, perlengkapan sekolah, dll. Shopee biasanya digunakan oleh berbagai kalangan, sehingga sangat membantu dalam berbelanja.



Gambar 1. 1 Survei Pengguna Media Sosial (Sumber, Alim, 2015).

Media sosial sendiri mengalami peningkatan pengguna aktif seperti facebook, twitter, dan youtube, dari data 2015 memaparkan jumlah pengguna twitter lebih dari 285 juta jiwa. Mayoritas penggunaanya dari umur 16 tahun hingga 65 tahun keatas. Twitter pada waktu belakangan ini kerap kali digunakan dalam ranah sosial – politik misalnya seperti pergerakan sosial bersama, melaporkan informasi tentang kemacetan, politik, kondisi cuaca, bencana alam dan informasi tertentu seperti peringatan terhadap suatu kejadian. Beberapa jurnal pun meneliti mengenai keutamaan twitter dalam menganalisis dan melakukan klasifikasi terhadap data contohnya seperti jurnal dengan judul analysis of tweet related to cyberbullying yang isinya tentang analisa sentimen terkait cyberbullying (Alim, 2015).

Media sosial twitter juga belakangan ini mengalami perubahan fungsi yang cukup drastis. Pada awal mula twitter berdiri media ini digunakan sebagai alat berbalas pesan dan tempat berbagi, namun akibat popularitas yang cukup tinggi

penggunaan fungsinya bertambah banyak, seperti pencarian trend, promosi produk, dan pemberian rating. Akibatnya fungsi pada twitter yang awal mula sebagai tempat tanya jawab dan berbagi informasi menjadi sarana yang digunakan diberbagai aspek salah satunya sebagai media penelitian berupa text mining, sentimen analisis, dan artificial intelligence (Hartanto,2017).

Analisis sentimen atau opinion mining merupakan proses memahami, mengekstrak dan mengolah data tekstual secara otomatis untuk mendapatkan informasi sentimen yang terkandung dalam suatu kalimat. Minimumnya di dalam analisis sentimen terdapat 2 kategori dalam menentukan sentimen yaitu kategori sentimen positif dan negatif. Sentimen analisis sendiri merupakan salah satu model klasifikasi data dengan pendekatan supervised learning di dalam machine learning. Machine learning mempunyai 2 tipe teknik yaitu supervised learning dan unsupervised learning. (Brownlee, 2016).

Supervised learning adalah salah satu tipe algoritma machine learning yang menggunakan data latih yang dikenal (training dataset) untuk membuat prediksi. Sedangkan unsupervised learning adalah salah satu tipe algoritma machine learning yang digunakan untuk menarik kesimpulan dari dataset yang terdiri dari input data labeled response. Metode unsupervised learning yang paling umum adalah analisa cluster, yang digunakan pada analisa data untuk mencari pola-pola tersembunyi atau pengelompokan dalam data, sedangkan pada supervised learning contoh yang paling umum adalah klasifikasi data. Mayoritas praktis dari machine learning sendiri

menggunakan supervised learning. Karena besarnya pengaruh dan manfaat dari klasifikasi data menyebabkan penelitian dan aplikasi berbasis klasifikasi data seperti text mining, analisis sentiment justru berkembang pesat (Ghulam, 2016).

Penelitian K-Nearest Neighbor (K-NN), pada tahun 2014 dibuat oleh Veverly Widyastuti Palinoan dengan judul “Sistem klasifikasi dokumen bahasa Jawa dengan Metode K-Nearest Neighbor (K-NN) “. Pada penelitian ini dilakukan pengujian dengan Cross Validation kemudian dilakukan uji presisi data yang digunakan sebanyak 40 dokumen dengan tingkat akurasi K=4 mencapai 95% , K=8 mencapai 92%. Penelitian selanjutnya oleh Dedy Santoso, Dian Eka Ratnawati, dan Indriati (2015) berjudul ‘Perbandingan Kinerja Metode Naïve Bayes, K-Nearest Neighbor, dan Metode Gabungan KMeans dan LVQ dalam Pengkategorian Buku Komputer Bahasa Indonesia Berdasarkan Judul dan Sinopsis’. Pada hasil penelitian tersebut menunjukkan bahwa rata-rata akurasi KNN mencapai 96% dan Naïve Bayes mencapai 98%.

Berdasarkan uraian diatas, peneliti tertarik untuk melakukan analisis sentimen menggunakan metode *K- Nearest Neighbor* dengan menggunakan pembobot *Term Frequency–Inverse Document Frequency* (TF-IDF) untuk mengetahui seberapa besar peningkatan akurasi dianggap berpengaruh. Hal ini dilakukan karena pola perubahan data *Time Series* pada periode lalu diyakini akan kembali terulang pada masa kini atau masa mendatang.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan di atas, maka rumusan masalah dalam penelitian ini adalah :

1. Bagaimana klasifikasi sentimen pada twitter Shopee menggunakan metode K-Nearest Neighbor dengan menggunakan pembobot Term frequency-inverse document frequency (TF-IDF)?
2. Bagaimana hasil akurasi menggunakan metode K- Nearest Neighbor dengan menggunakan pembobot Term frequency-inverse document frequency (TF-IDF)?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah di atas, maka tujuan penelitian ini adalah:

1. Mengklasifikasikan sentimen pada twitter Shopee menggunakan metode *Nearest Neighbor* dengan menggunakan pembobot *Term frequency-inverse document frequency* (TF-IDF)
2. Mendapatkan hasil akurasi menggunakan metode *K- Nearest Neighbor* dengan menggunakan pembobot (TF-IDF)

1.4 Manfaat Penelitian

Adapun manfaat yang bisa dipetik dari penelitian ini diantaranya adalah:

1. Manfaat teoritis

Membantu perkembangan ilmu pengetahuan mengenai metode peramalan sehingga dapat digunakan sebagai bahan bacaan dan referensi bagi pembaca dalam melakukan peramalan terutama menggunakan metode K-Nearest Neighbor.

2. Manfaat praktis

- a. Bagi peneliti, dapat mengembangkan ilmu statistik yang diperoleh selama perkuliahan pada kasus yang terjadi di masyarakat. Dalam hal ini yaitu penerapan model K-Nearest Neighbor pada Sentimen
- b. Sedangkan bagi pihak terkait seperti pemerintah dan masyarakat, penelitian ini dapat menghasilkan sentiment positif dan negatif terkait Shopee44 menggunakan metode K-Nearest Neighbor.

1.5 Batasan Penelitian

Berdasarkan rumusan masalah dan tujuan penelitian, batasan penelitian ini adalah :

1. Data yang digunakan adalah data Shopee44 yang diambil melalui Twitter dengan #Shopee44
2. Nilai K yang digunakan adalah K=1, K=3, K=5, K=7, K=9
3. Tingkat keakurasian didapatkan menggunakan metode K-Nearest Neighbor