

BAB II

TINJAUAN PUSATAKA

Shopee

Shopee merupakan aplikasi E-commerce yang didirikan sejak tahun 2019 yang membantu para penjual dan pembeli untuk melakukan transaksi penjualan dan pembeli barang maupun jasa. Aplikasi ini sudah terdownload sebanyak 100jt lebih pengguna dengan rating 4+ dengan kapasitas 160MB, dan memiliki 11jt ulasan, baik ulasan positif maupun ulasan negatif (Masripah & Utami, 2020) . Shopee merupakan aplikasi marketplace yang tentunya dapat di download hanya bermodalkan dengan kuota internet. Tujuan didirikannya shopee adalah ingin memperkuat transformasi dari teknologi dan ingin mengubah dunia dengan menyediakan platform untuk menghubungkan pembeli dan penjual dalam satu komunitas yang lebih baik lagi. Shopee menawarkan belanja online komprehensif, dari berbagai pilihan produk sampai ke sebuah komunitas sosial untuk bereksplorasi, dan layanan untuk selalu memenuhi kebutuhan konsumen tanpa hambatan.

Analisis Sentimen

Sentimen analisis atau yang sering disebut dengan istilah analisis sentimen atau opinion mining merupakan salah satu bidang ilmu yang menganalisiskan pendapat, sentimen, evaluasi, penilaian, sikap, emosi public terhadap suatu produk, berita, informasi ataupun jasa organisasi maupun individu . Analisis Sentimen biasanya fokus pada komentar atau pendapat yang mengungkapkan suatu ekspetasi, dimana ekspetasi tersebut bisa berupa komentar negatif maupun positif. Sentimen analisis ini mengolah data tekstual secara otomatis untuk melihat sentimen yang terkandung di dalam suatu opini atau pendapat (Diyah Hepatika & Nurlaela, 2021).

Analisis sentimen dilakukan untuk melihat pendapat terhadap sebuah masalah. Tujuan dari analisis sentimen adalah untuk mendefinisikan alat otomatis yang mampu mengekstrak informasi subjektif dari teks dalam bahasa alami, seperti pendapat dan sentimen, sehingga menciptakan pengetahuan yang terstruktur dan dapat di tindaklanjuti untuk di gunakan yang baik oleh decision support system atau decision maker. (Studi et al., 2019)

Text mining

Text mining juga dikenal dengan text data mining ataupun pencarian pengetahuan di basis data tekstual adalah proses yang semi otomatis melakukan ekstraksi dari pola data. Tipe pekerjaan text mining meliputi kategorisasi, text clustering, ekstraksi konsep/entitas, analisis sentimen, document summarization, dan entity-relation modeling (yaitu, hubungan pembelajaran antara entitas). Sumber data yang digunakan pada text mining adalah kumpulan text yang memiliki format yang tidak terstruktur atau minim semi terstruktur. Tujuan text mining adalah untuk mendapatkan informasi yang berguna dari sekumpulan dokumen. Dimana sumber data yang di gunakan dalam text mining adalah sekumpulan teks (DARUSSALAM, 2021).

Algoritma K-Nearest Neighbor (K-NN)

Algoritma K-Nearest Neighbors merupakan algoritma penentuan klasifikasi data berdasarkan contoh dasar yang tidak membangun, representasi deklaratif eksplisit kategori. Algoritma K-Nearest Neighbors adalah sebuah klasifikasi berdasarkan data training K-Nearest Neighbors memiliki prinsip sederhana bekerja berdasarkan jarak terpendek dari sampel uji ke sampel data latih (Juanita & Luhur, 2021).

Algoritma K-Nearest neighbors memiliki tujuan untuk mengklasifikasikan objek berdasarkan atribut data training sample. Classifier tidak menggunakan apapun untuk dicocokkan dan hanya berdasarkan pada dataset yang ada. Diberikan titik query, akan di temukan sejumlah k objek atau (titik training) yang paling dekat dengan titik query. Nilai K pada K-NN berarti K-data terdekat dari data uji. Salah satu masalah yang dihadapi K-NN adalah pemilihan nilai K yang tepat. Cara voting mayoritas dari K-tetangga untuk nilai K yang besar bisa mengakibatkan distorsi data yang besar. Untuk menangani masalah voting mayoritas tersebut biasanya di tambahkannya penggunaan bobot untuk menghitung kandidat kelas yang sebaiknya diambil oleh data uji dari K-tetangga terdekat. Bobot dari setiap tetangga terdekat dihitung dengan formula

Studi Literatur

Dalam penyusunan penelitian ini menggunakan penelitian terdahulu sebagai bahasan referensi dan juga sebagai bahan acuan untuk mempermudah dalam pembuatan laporan penelitian. Berikut ini merupakan hasil dari penelitian terdahulu dengan menggunakan algoritma K – Nearest Neighbor :

Terdapat banyak aplikasi yang sudah di unduh sehingga dengan tujuan penelitian ini untuk menganalisa sentimen pada ulasan penggunaan aplikasi . dengan jumlah ulasan yang di perlukan sebanyak 998 yang terdiri dari 484 sentimen positif dan 514 sentimen negatif training dan data testing dengan menghasilkan nilai akurasi presision dan recall yang di hasilkan dari tiap aplikasi yaitu 85,14 %, 91,91% dan 74,44% sedangkan untuk berkas yaitu 81,15%,87,15% (Juanita & Luhur, 2021).

Peneliti melakukan analisis terhadap pandangan baik berupa pendapat positif ataupun negatif. Proses analisis dilakukan dengan metode preprocessing dan di beri label negatif dan positif. Berdasarkan data teks yang tersedia dilakukan proses klasifikasi dengan algoritma knn. Rapidminer digunakan untuk eksperimen data teks dengan algoritma knn dengan tujuan mencari nilai akurasi, presisi dan recall. Dari hasil penelitian yang di peroleh mendapatkan nilai 87.00% untuk accuracy dan 0.916 untuk nilai AUC. Nilai nilai yang cukup tinggi untuk klasifikasi data opini mahasiswa terhadap pandemi ini sehingga penelitian ini digolongkan (Mara et al., 2021)

Pada penelitian ini dibuat untuk mengimplementasikan algoritma KNN (K-Nearest Neighbor) dalam analisis sentimen penggunaan twitter tentang topik pilkada DKI 2017. Data tweet yang digunakan sebanyak 2000 data tweet berbahasa Indonesia. Menggunakan

algoritma knn dengan pembobotan kata tf-idf dan fungsi cosine similarity, di lakukan klasifikasi nilai sentimen ke dalam dua kelas yaitu positif dan negatif. Dari hasil pengujian diketahui bahwa nilai akurasi terbesar adalah 67,2% dengan nilai $k=5$, presisi tertinggi 56,94%, ketika $k=5$, dan recall 78,24% dengan $k=15$ (Deviyanto, 2018)

Dalam penelitian ini; dibutuhkan suatu metode yang mampu mengklasifikasikan pendapat secara akurat. Ruang lingkup penelitian ini adalah peninjauan agen perjalanan pengolahan data menggunakan algoritma K-Nearest Neighbor (K-NN) yang menggunakan 100 review positif dan 100 review negatif dengan enam kata yang berhubungan dengan sentimen yaitu: Fast, Good, Great, Buruk, Cencil, dan Tunggu. Titik akurasi peninjauan agen perjalanan menggunakan K-Nearest Neighbor (K-NN) algoritma telah mencapai 87,00% dan titik AUC adalah 0,916, titik AUC milik kelompok Klasifikasi Excellent sehingga dinyatakan bahwa K-Nearest Neighbor (K-NN) memiliki hasil yang akurat dalam menganalisis sentimen ulasan agen perjalanan (Ernawati & Wati, 2018).

Berdasarkan rumusan masalah dan pembahasan yang telah peneliti deskripsikan pada bab sebelumnya, kombinasi K-nearest Neighbor dan Levenshtein Distance dapat di terapkan pada analisis sentimen. Tingkat akurasi dari kombinasi K-nearest Neighbor dan Levenshtein Distance ditunjukkan pada saat nilai $k = 3$ yaitu sebesar 65.625%. Kesimpulan yang dapat diambil adalah peningkatan hasil akurasi dengan kombinasi K-nearest Neighbor dan Levenshtein Distance (Studi et al., 2019).



Tabel 2.1 Tabel Peneliti Terdahulu

| Jurnal Terdahulu | | | |
|------------------|-------------------------|---|---|
| No | Nama Peneliti | Metode | Hasil |
| 1. | (Juanita & Luhur, 2021) | Algoritma K-Nearest Neighbor (KNN) | Jumlah ulasan yang diperlukan sebanyak 998 yang terdiri dari 484 sentimen positif dan 514 sentimen negatif training dan data testing dan menghasilkan nilai akurasi 85,14%, 91,91% dan 74,44%. |
| 2. | (Mara et al., 2021) | Algoritma K-Nearest Neighbor (KNN) | Proses dilakukan dengan menggunakan metode preprocessing dan di beri label negatif dan positif. Berdasarkan data teks yang tersedia di peroleh nilai 87.00% untuk accuracy dan 0.916 untuk nilai AUC. |
| 3. | (Deviyanto, 2018) | Algoritma K- Nearest Neighbor (KNN) | Data twit yang digunakan sebanyak 2000 data twit berbahasa indonesia. Menggunakan pembobotan tf-idf dan fungsi cosine similarity, di lakukan nilai sentimen ke dalam dua kelas yaitu positif dan negatif. Dari hasil pengujian yang di dapat bernilai 67,2% dengan nilai k=5, persisi tertinggi 56,94%, ketika k=5, dan recall 78,24% dengan nilai k-15 |
| 4. | (Ernawati & Wati, 2018) | Algoritma K-Nearest Neighbor (KNN) | Dayang yang di gunakan 100 review positif negatif dengan enam kata yang berhubungan dengan sentimen. Titik akurasi peninjauan agen perjalanan mencapai 87.00% dan titik AUC 0.916%. |
| 5. | (Studi et al., 2019) | Algoritma K-Nearest Neighbor (KNN) dan Levenshtein Distance | Tingkat akurasi dari kombinasi K-Nearest Neighbor (KNN) dan Levenshtein Distance di tunjukan pada nilai k=3 yaitu sebesar 65.625%. |

