

BAB II

TINJAUAN PUSATAKA

2.1 Tokopedia

Tokopedia merupakan e-commerce yang saat ini terkenal. Tokopedia.com memberikan wadah kepada pedagang dan konsumen melakukan penukaran jual beli barang secara gratis, tanpa dipungut biaya apapun. Selain gratis, Tokopedia.com juga menyediakan sistem layanan bersama yang dijamin aman. William Tanuwijaya (Chief Tokopedia.com) menginformasikan kepada Beat bahwa lebih dari 1.000.000 masyarakat Indonesia telah memulai bisnisnya di Tokopedia. Secara konsisten lebih dari 16,5 juta barang dikirimkan ke pembeli yang tersebar dari Sabang hingga Merauke, dengan nilai tukar mencapai triliunan rupiah setiap bulannya. William juga mengatakan perilaku konsumen telah berubah akhir-akhir ini. Pada tahun 2014, jumlah kunjungan dari serbaguna mencapai 56%, namun komitmen terhadap pertukaran sekitar 29%. Menjelang akhir semester utama 2016, 79,55 persen kunjungan datang dari portabel, dan yang menarik, komitmen pertukaran sudah mencapai 73,58 persen. (Dharmawan, 2016, Para. 2)

2.2 Google Play

Play store adalah layanan pemasok konten terkomputerisasi yang diklaim oleh Google yang menyediakan berbagai toko produk berbasis web seperti aplikasi, game, film atau musik, dan buku dengan kelas berbeda. Google Play Store dapat diakses melalui situs, aplikasi android dan Google televisi. Di Google Play Store terdapat beberapa elemen, salah satunya adalah rating dan survei highlight dari klien terhadap aplikasi atau layanan yang dapat diakses. Survei atau audit merupakan suatu pesan atau kalimat yang berisi penilaian atau komentar terhadap hasil karya seseorang. Pentingnya survei ini sering kali digunakan sebagai tolok ukur suatu aplikasi terlepas dari apakah survei tersebut disarankan untuk klien baru. (Herlinawati et al., 2020)

2.3 Analisis Sentimen

Sentimen analisis atau yang sering disebut dengan pemeriksaan perasaan atau penambangan penilaian adalah suatu bidang ilmu yang membedah anggapan, pendapat, penilaian, evaluasi, cara pandang, perasaan masyarakat terhadap suatu barang, berita, data atau administrasi perkumpulan dan masyarakat. Pemeriksaan perasaan sebagian besar berpusat pada pernyataan atau kesimpulan yang mengungkapkan suatu asumsi, dimana asumsi tersebut dapat berupa pernyataan yang pesimistis atau baik. Pemeriksaan perasaan ini mengolah informasi berbasis teks secara alami untuk melihat perasaan yang terkandung dalam suatu penilaian atau penilaian (Diyah Hepatika dan Nurlaela, 2021).

Sentimen analisis dilakukan untuk melihat sentimen terhadap suatu isu. Tujuan dari pemeriksaan perasaan adalah untuk mengkarakterisasi perangkat terkomputerisasi yang cocok untuk mengekstraksi data emosional dari pesan dalam bahasa biasa, seperti anggapan dan pendapat, dengan cara ini membuat informasi terorganisir dan penting untuk digunakan oleh jaringan pendukung emosional pilihan atau pemimpin. (Studi et al., 2019)

2.4 Text mining

Penambangan teks, atau disebut penambangan informasi teks atau pencarian

informasi dalam kumpulan data tercetak, adalah proses semi-komputerisasi untuk menghilangkan desain dari informasi. Jenis pekerjaan penambangan pesan meliputi klasifikasi, pengumpulan pesan, ekstraksi ide/zat, penyelidikan perasaan, penjumlahan laporan, dan demonstrasi hubungan unsur (misalnya berkonsentrasi pada hubungan antar zat). Sumber informasi yang digunakan dalam text mining adalah kumpulan teks yang memiliki desain tidak terstruktur atau semi terorganisir. Tujuan dari penambangan teks adalah untuk mendapatkan data berharga dari berbagai macam laporan. Dimana sumber informasi yang digunakan dalam text mining adalah teks yang bermacam-macam. (DARUSSALAM, 2021)

2.5 Algoritma K-Nearest Neighbor (K-NN)

Algoritma K-Nearest Neighbors adalah perhitungan untuk memutuskan pengelompokan informasi berdasarkan model fundamental yang tidak berkembang, gambaran kelas yang jelas dan jelas. Perhitungan K-Closest Neighbors merupakan pengelompokan dalam rangka menyiapkan informasi. K-Closest Neighbors memiliki aturan langsung yang berfungsi mengingat pemisahan paling terbatas dari tes hingga informasi persiapan tes. (Juanita & Luhur, 2021).

Algoritma K-Nearest neighbors memiliki tujuan untuk mengurutkan objek berdasarkan ciri-ciri informasi persiapan tes. Pengklasifikasi tidak menggunakan apa pun untuk mengoordinasikan dan hanya berdasarkan kumpulan data saat ini. Dengan adanya titik pertanyaan, ia akan melacak k artikel atau (fokus persiapan) yang paling dekat dengan titik pertanyaan. Nilai K pada K-NN berarti nilai K terdekat dari data pengujian. Salah satu permasalahan yang diperhatikan KNN adalah pemilihan nilai K yang tepat. Metode pemberian suara sebagian besar dari tetangga K untuk kualitas K yang sangat besar dapat menyebabkan perubahan informasi yang sangat besar. Untuk mengatasi masalah pemungutan suara sebagian besar, penggunaan muatan biasanya ditambahkan untuk menentukan pesaing kelas yang harus diambil berdasarkan informasi tes dari tetangga terdekat K. Beratnya masing-masing tetangga terdekat ditentukan oleh resepnya.

2.6 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 motivasi di balik penelitian ini adalah untuk membedah pendapat tentang survei pemanfaatan aplikasi. dengan jumlah survei yang diperlukan sebanyak 998 yang terdiri dari 484 opini positif dan 514 perasaan negatif mempersiapkan dan menguji informasi dengan menyampaikan nilai matematika berdasarkan variabel akurasi dengan review yang dihasilkan dari setiap aplikasi yaitu 85,14%, 91,91% dan 74,44% sedangkan untuk dokumen yaitu 81,15%, 87,15% . (Juanita & Luhur, 2021).

Para ahli mengarahkan penilaian terhadap sudut pandang tersebut baik sebagai pendapat tertentu maupun pendapat negatif. Siklus pemeriksaan diselesaikan dengan metode preprocessing dan diberi cap negatif dan positif. Mengingat data

teks terbuka, siklus pengumpulan diselesaikan dengan estimasi knn. Rapidminer digunakan untuk menyelidiki berbagai jalan sehubungan dengan data pesan dengan perhitungan KNN secara lengkap bertujuan untuk menemukan nilai presisi, ketepatan dan audit. Berdasarkan hasil penyelidikan yang diperoleh, presisinya adalah 87,00% dan AUC yang dipertimbangkan adalah 0,916. Skor tersebut sangat tinggi untuk menggambarkan data evaluasi siswa mengenai pandemi ini dengan tujuan untuk mengumpulkan investigasi ini (Mara et al., 2021)

Dalam peninjauan ini dilakukan perhitungan KNN (K-Nearest Neighbor pada musim pemilu pemanfaatan Twitter bertepatan pemilu provinsi DKI tahun 2017. Informasi tweet yang digunakan sebanyak 2.000 informasi tweet berbahasa Indonesia. Memanfaatkan perhitungan KNN dengan pembobotan kata tf-idf dan kemampuan cosine-similarity, dalam melakukan pengelompokan nilai opini menjadi dua kelas yaitu yakin dan negatif. Dari hasil percobaan diketahui nilai ketepatan yang paling tinggi adalah 67.2% dengan $k = 5$, akurasi paling tinggi adalah 56,94%, saat $k=5$, dan 78,24% review dengan $k=15$. (Deviyanto, 2018)

Dalam penilaian ini; kita memerlukan prosedur yang dapat mengatur tujuan dengan tegas. Derajat penilaian ini adalah meninjau data pengurus penyelenggara pergerakan dengan menggunakan perhitungan K-Nearest Neighbor (K-NN) yang menggunakan 100 kajian positif dan 100 tinjauan negatif dengan enam kata yang dikaitkan dengan penilaian yaitu: Cepat, Luar Biasa, Luar Biasa, Mengerikan, Cencil, dan Tangguh. Titik akurasi gambaran pengatur pergerakan dengan menggunakan estimasi K-Nearest Neighbor (K-NN) sudah muncul sebesar 87.00% dan titik AUC sebesar 0.916, titik AUC tersebut telah disubmit dengan paket Wonderful Request sehingga dikomunikasikan bahwa K-Nearest Neighbor (K-NN) memiliki hasil yang cermat dalam menganalisis penilaian ulasan penyelenggara perjalanan (Ernawati & Wati, 2018).

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 audit yang diharapkan adalah 998 yang terdiri dari 484 opini positif dan 514 perasaan persiapan negatif dan pengujian informasi dan menghasilkan ketepatan bernilai 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)	Informasi tweet yang digunakan adalah 2.000 informasi tweet berbahasa Indonesia. Memanfaatkan pembobotan tf-idf dan kemampuan cosine likeness, nilai opini dibagi menjadi dua kelas menjadi spesifik positif dan negatif. Dari hasil eksperimen itu bisa jadi signifikan 67.2% dengan nilai k = 5, akurasi tertinggi sebesar 56.94%, pada saat k = 5, dan 78.24% review dengan nilai k-15
4.	(Ernawati & Wati, 2018)	Algoritma K-Nearest Neighbor (KNN)	Dayang yang di gunakan 100 riview 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 KNearest Neighbor (KNN) dan Levenshtein Distance di tunjukan pada nilai k=3 yaitu sebesar 65.625%.

