

BAB II

TINJAUAN PUSTAKA

2.1 Analisis Sentimen

Analisis sentimen adalah bidang studi yang menganalisis pendapat, perasaan, penilaian, penilaian, sikap, dan perasaan seseorang tentang suatu produk, layanan, organisasi, orang, masalah, peristiwa, atau topik. *Opinion mining* digunakan untuk menampilkan opini tentang produk dan layanan, peristiwa sosial dan politik, dan aktivitas spesifik lainnya. Analisis sentiment berfokus pada pengidentifikasian opini seseorang tentang suatu masalah atau objek, apakah cenderung positif maupun negatif (Ananda & Pristyanto, 2021). Sentimen akan diproses untuk menghasilkan informasi atau pola yang dapat diubah sebagai pengetahuan baru. (Putra & Nugroho, 2021).

2.1.1 Text Mining

Text mining adalah teknik yang digunakan untuk menangani klasifikasi, pengelompokan, ekstraksi informasi, dan pencarian informasi. Perbedaan antara *text mining* dan *data mining* adalah bahwa pola yang digunakan dalam *text mining* berasal dari kumpulan bahasa alami yang tidak terstruktur, sedangkan pola *data mining* berasal dari *database* terstruktur (Agustina et al., 2021).

2.1.2 Text Preprocessisng

Untuk melakukan penambangan informasi atau *text mining*, perlu dilakukan beberapa tahapan untuk mengolah sumber data terstruktur dan tidak terstruktur dari berbagai sumber. Proses awal pengolahan data atau disebut juga *text preprocessing* bertujuan untuk menangani berbagai jenis data yang masih belum teratur supaya

dapat diterapkan beberapa metode *text mining* yang ada (Ika et al., 2021). Proses ini digunakan untuk membersihkan data dari *noise* dan siap digunakan pada proses selanjutnya. Proses *text preprocessing* memiliki beberapa tahapan yaitu (Ananda & Pristyanto, 2021):

1. *Cleansing*: Proses *cleansing* membersihkan kumpulan tanda baca, *URL*, nama pengguna, tagar (#), dan *emoticon*.
2. *Case Folding*: Ini adalah proses mengubah huruf dari campuran huruf (huruf kecil dan huruf kapital) menjadi huruf kecil.
3. *Tokenization*: Proses *Tokenization* bertujuan untuk mengubah kalimat menjadi *token-token* atau sepotong kata.
4. *Stopword Removal*: *Stopword Removal* adalah penghapusan kata yang tidak berbobot, meskipun kata itu muncul berkali-kali dalam sebuah kalimat.
5. *Stemming*: *Stemming* adalah proses mengubah kata-kata yang terdapat dalam suatu dokumen menjadi kata-kata sederhana.

2.1.3 Pembobotan Kata

Pembobotan kata atau *term weighting* proses penghitungan bobot setiap kata yang akan dicari di seluruh dokumen untuk menentukan ketersediaan dan kesamaan kata dalam dokumen (Putra & Nugroho, 2021). *Term Frequency Inverse Document Frequency* (TF-IDF) adalah metode untuk menghitung bobot setiap kata yang diurai. Model pembobotan TF-IDF merupakan metode yang mengintegrasikan model *Term Frequency* (TF) dan *Inverse Document Frequency* (IDF) (Aliyah et al., 2020). Pembobotan kata menggunakan metode TF-IDF bertujuan untuk mengubah

data *tweet* menjadi data numerik. Untuk menghitung bobot pada metode TF-IDF digunakan persamaan berikut (Pratiwi et al., 2021):

$$IDF = \log \frac{D}{df} \quad (2.1)$$

$$w_{ij} = tf_{ij} \times \log \frac{D}{df} \quad (2.2)$$

Keterangan:

IDF : nilai IDF dari setiap kata

D : total dokumen yang tersedia

df : jumlah kemunculan kata pada semua dokumen

w_{ij} : nilai bobot kata

tf_{ij} : jumlah kemunculan kata pada suatu dokumen

2.1.4 Klasifikasi

Klasifikasi merupakan proses pengelompokan dan memprediksi data baru dari suatu variabel berdasarkan standar data yang ada. Klasifikasi memiliki dua jenis data, yaitu data latih (*training data*) dan data uji (*testing data*). Data latih adalah data yang ada berdasarkan fakta yang telah terjadi dan berfungsi untuk membentuk tabel probabilitas. Data uji itu sendiri adalah data yang sudah memiliki kelas atau sebutan dan digunakan untuk menguji atau menghitung nilai akurasi dari data atau tabel probabilitas yang telah dibentuk (Nitami & Februariyanti, 2022).

2.1.5 *Decision Tree*

Decision tree merupakan suatu metode klasifikasi dimana variabel dependen yang bertipe kategori maka menghasilkan pohon klasifikasi (*classification trees*). Dimana node paling atas disebut *root* dan cabangnya berisi nilai dari atribut. (Ernamia & Herliana, 2022).

Decision tree merupakan *flowchart* yang berstruktur pohon (*tree*), dimana setiap *node* (simpul *internal*) mempresentasikan atribut dan cabangnya menandakan hasil tes atau nilai dari atribut, sedangkan daunnya digunakan untuk mempresentasikan kelas atau distribusi dari kelas. (Harun & Ananda, 2021).

Decision tree merupakan salah satu metode untuk mengklasifikasikan data. Model *decision tree* merupakan sebuah *tree* yang terdiri dari *root node*, *internal node* dan *terminal node*. Sementara *root node* dan *internal node* merupakan variabel/fitur, *terminal node* adalah label kelas. Dalam melakukan klasifikasi, sebuah data *query* akan menelusuri *root node* dan *internal node* sampai mencapai *terminal node*. Pelabelan kelas data *query* berdasarkan label di *internal node*. Pada *Decision Tree* tradisional, data yang digunakan adalah data dengan nilai fitur yang sudah pasti. (Tri Romadloni et al., 2019).

Konsep *entropi* digunakan untuk penentuan pada atribut mana sebuah pohon akan terbagi (*split*). Semakin tinggi *entropy* sebuah sampel, semakin tidak murni sampel tersebut. Rumus yang digunakan untuk menghitung *entropy* sampel S adalah tersebut. Rumus yang di gunakan untuk menghitung *entropy* sample s adalah.

$$Entropy(S) = \sum_i^c -P_i \log_2 P_i \quad (2.3)$$

Dimana c adalah jumlah nilai yang terdapat pada atribut target (jumlah kelas). Sedangkan p_i menyatakan porsi atau rasio antara jumlah sample di kelas i dengan jumlah semua sample pada himpunan data. (Marutho, 2019).

2.1.6 Naïve Bayes

Naïve Bayes Classifier adalah algoritma *Bayesian Learning* yang paling cepat dan sederhana. Algoritma klasifikasi *Naïve Bayes* cocok untuk diterapkan pada data dengan jumlah yang banyak dan mampu menangani data yang kosong (*missing value*) serta menggunakan perhitungan probabilitas dan statistik yang memprediksi peluang dimasa depan berdasarkan pengalaman sebelumnya. (Ernamia & Herliana, 2022).

Algoritma *Naïve Bayes* merupakan salah satu algoritma yang terdapat dalam teknik klasifikasi. *Naïve Bayes* merupakan pengklasifikasian dengan bentuk model probailistik dan statistik yang disederhanakan dengan berdasar pada *teorema Bayes* dengan asumsi bahwa setiap atribut bersifat bebas (*independence*). Dengan kata lain, algoritma ini mengasumsikan bahwa ada atau tidak ciri tertentu dari sebuah kelas tidak ada hubungannya dengan ciri dari kelas lainnya. (Fikri et al., 2020).

$$P(C_i|X) = \frac{P(X|C_i) \cdot P(C_i)}{P(X)} \quad (2.4)$$

$$P(X) = (P(X_1|C_{i2})P(C_i)) + (P(X_2|C_{i2})P(C_i)) + \dots + (P(X_n|C_{in})P(C_i)) \quad (2.5)$$

dimana:

$P(C_i|X)$: Probabilitas hipotesis C berdasarkan kondisi X (*posterior probability*)

$P(C_i)$: Probabilitas hipotesis C (*prior probability*)

$P(X|C_i)$: Probabilitas X berdasarkan kondisi pada hipotesis C (*likelihood*)

$P(X)$: Probabilitas X (*predictor prior probability*)

2.1.7 Confusion Matrix

Confusion matrix adalah tabel yang memberikan informasi tentang perbandingan hasil klasifikasi berdasarkan hasil prediksi yang dibuat oleh 6ulfil dengan hasil klasifikasi yang sebenarnya. Tabel dalam *confusion matrix* menunjukkan jumlah data uji yang diklasifikasikan dengan benar dan jumlah data uji yang diklasifikasikan salah (Fikri et al., 2020). Tabel *confusion matrix* berisikan empat kemungkinan keluaran sebagai bahan acuan dalam membandingkan antara kejadian yang sebenarnya (aktual) dengan kejadian yang terprediksi. Metode ini menggunakan tabel matriks seperti pada Tabel 2.1 (Ruhyana & Rosiyadi, 2020).

Tabel 2.1 Model *Confusion Matrix*

		Aktual	
		<i>Complaint</i>	<i>Not Complaint</i>
Prediksi	<i>Complaint</i>	TP	TN
	<i>Not Complaint</i>	FP	FN

Akurasi dapat dihitung menggunakan rumus :

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2.6)$$

Keterangan:

TP: jumlah data *complaint* yang diprediksi *complaint*.

TN: jumlah data *complaint* yang diprediksi *not complaint*.

FP: jumlah data *not complaint* yang diprediksi *complaint*.

FN: sama dengan jumlah data *not complaint* yang diprediksi *not complaint*.

2.2 Media Sosial

Media sosial adalah media *online* dan penggunaannya dapat dengan mudah berpartisipasi, berbagi, dan membuat konten termasuk *blog*, media sosial, *wiki*, forum, dan dunia maya (Rafiq, 2020). Menurut hasil riset yang dilakukan oleh ‘*We Are Social*’, dari sekian banyak jenis media sosial, yang saat ini diminati masyarakat Indonesia adalah *Youtube*, *Facebook*, *Instagram* dan *Twitter* (Puspitarini & Nuraeni, 2019). Media sosial juga merupakan sarana untuk menyampaikan atau mendapatkan informasi yang aktual bagi para pengguna media sosial itu sendiri.

2.2.1 Twitter

Twitter merupakan salah satu jejaring sosial terpopuler yang berperan sebagai forum komunikasi di masyarakat. *Twitter* memungkinkan orang di seluruh dunia untuk terhubung dengan keluarga, teman, dan orang-orang terkasih dari komputer atau ponsel mereka. Salah satu layanan yang ditawarkan *Twitter* kepada penggunaannya adalah pembuatan pesan status (*tweets*) yang dapat dibaca oleh pengguna *Twitter* lain dan umumnya berisi pendapat pengguna tentang berbagai topik dengan batas hingga 140 karakter, sehingga *Twitter* adalah salah satu situs

web yang mengumpulkan data opini dari orang-orang di seluruh dunia (Fikri et al., 2020).

2.2.2 Twitter API

Twitter API (Application Programming Interface) adalah akses programatik kepada perusahaan, pengembang, dan pengguna ke data *Twitter* (Agustina et al., 2021). Untuk menambang *tweet*, diperlukan empat kode akses yaitu *API key*, *API secret*, *access token*, dan *access token secret* (Aliyah et al., 2020).

2.3 Marketplace Tokopedia

Perkembangan toko daring atau *marketplace* kini semakin cepat dan beragam. Persaingan bisnis antar *marketplace* semakin ketat. *Marketplace* adalah wadah komunitas bisnis interaktif secara elektronik yang menyediakan pasar dimana perusahaan dapat ambil andil dalam B2B *e-commerce* dan atau kegiatan *e-business* lain. (Alfiah & Damayanti, 2020). Terlihat dari banyaknya *marketplace* di Indonesia seperti Tokopedia, Shopee, Lazada, Bukalapak, Blibli, JD.ID, Zalora, dan lain-lain.

Tokopedia adalah sebuah perusahaan *e-commerce* yang memiliki beberapa layanan lain seperti *marketplace* yang mempertemukan penjual dengan pembeli, layanan teknologi finansial dan pembayaran yang dapat memenuhi kebutuhan keuangan pengguna mulai dari investasi, pinjaman sampai asuransi, layanan mitra tokopedia yang menjembatani kesenjangan antara *online* dan *offline* yang akan menambah nilai ke toko fisik tradisional melalui perluasan jangkauan pelanggan *online* dan mengekspos penjual ke lebih banyak pemasok dengan harga yang lebih

baik, dan layanan logistik dan *fulfilment* yang dapat memastikan barang sampai ke tangan pelanggan dalam keadaan baik dan cepat.

2.4 Penelitian Terkait

Penyusunan tugas akhir ini mengambil beberapa referensi penelitian sebelumnya, termasuk jurnal-jurnal yang berhubungan dengan penelitian ini, dengan menentukan sumber yang tertera pada tabel 2.2

Tabel 2.2 Penelitian Terkait

No.	Nama Penulis	Judul	Isi Ringkasan	Hasil
1.	Zuda Pradana Putra, Aryo Nugroho (2022)	Pebandingan Performa <i>Naïve Bayes</i> dan KNN pada Klasifikasi Teks Sentimen Jasa Ekspedisi	Penelitian ini bertujuan untuk mengetahui performa dari model <i>Naïve Bayes</i> dan KNN (<i>K-Nearest Neighbor</i>) dalam mengklasifikasikan <i>dataset</i> teks sentimen jasa ekspedisi. Ulasan pengguna twitter pada akun @jntexpressid, @JNE_ID, dan @posindonesia akan di tambang dan diklasifikan berdasarkan sentimen positif, netral, dan negatif.	Hasil dari penelitian ini menyatakan penggunaan teknik <i>resampling</i> untuk dataset yang tidak seimbang, sukses menaikan akurasi baik <i>naïve bayes</i> maupun K-NN dibanding sebelum menggunakan metode <i>over-sampling</i> . Diketahui akurasi pada algoritma <i>Naïve Bayes</i> sedikit lebih baik dalam pengklasifikasian terhadap dataset berupa teks sentimen jasa ekspedisi.

Tabel 2.2 Penelitian Terkait (Lanjutan 1)

No.	Nama Penulis	Judul	Isi Ringkasan	Hasil
2.	Herliana, A. (2022).	Analisis Sentimen Kuliah Daring Dengan Algoritma <i>Naïve Bayes</i> , K-NN Dan <i>Decision Tree</i>	Penelitian ini dilakukan dengan menggunakan algoritma <i>Naïve Bayes</i> , <i>K-Nearest Neighbor</i> dan <i>Decision Tree</i> untuk membandingkan tingkat akurasi dari ketiga algoritma tersebut. Dalam mengklasifikasi data sentimen terhadap kuliah daring dan di klasifikasikan dengan berdasarkan sentiment positif dan negatif.	Pada penelitian ini menunjukkan hasil akurasi algoritma <i>naïve bayes</i> menghasilkan nilai akurasi sebesar 81,57% pada algoritma <i>K-Nearest Neighbor</i> didapatkan akurasi sebesar 62,10%. Dan untuk Algoritma <i>Decision Tree</i> hasil akurasi sebesar 51,89%. Dapat dibuktikan algoritma <i>Naïve Bayes</i> merupakan algoritma yang memiliki nilai akurasi paling tinggi dibandingkan kedua metode lainnya dengan tingkat akurasi sebesar 81,57%.

Tabel 2.2 Penelitian Terkait (Lanjutan 2)

No.	Nama Penulis	Judul	Isi Ringkasan	Hasil
3.	Hasri, C. F., & Alita, D. (2022)	Penerapan Metode <i>Naïve Bayes Classifier</i> Dan <i>Support Vector Machine</i> Pada Analisis Sentimen Terhadap Dampak Virus <i>Corona</i> Di <i>Twitter</i>	Penelitian ini bertujuan menganalisis opini publik untuk menemukan klasifikasi <i>positif</i> , <i>negatif</i> dan netral terhadap kehidupan masyarakat baik itu sebelum dan sesudah wabah pandemi <i>Corona</i> . Adapun metode klasifikasi data yang digunakan oleh penulis yaitu <i>Naïve Bayes Classifier</i> dikarenakan metode ini memiliki tingkat akurasi, penarikan dan presisi yang lebih baik daripada metode SVM.	Pada pengujian algoritma <i>Naive Bayes Classifier</i> menghasilkan nilai akurasi data sebesar 81.07%. sedangkan algoritma <i>support vector machine</i> menghasilkan nilai akurasi data sebesar 79,96%.

Tabel 2.2 Penelitian Terkait (Lanjutan 3)

No.	Nama Penulis	Judul	Isi Ringkasan	Hasil
4.	Anam, M. K., Pikir, B. N., & Firdaus, M. B. (2021).	Penerapan <i>Naïve Bayes Classifier</i> , <i>K-Nearest Neighbor (KNN)</i> dan <i>Decision Tree</i> untuk Menganalisis Sentimen pada Interaksi <i>Netizen</i> dan Pemerintah	Penelitian ini bertujuan membandingkan keakuratan ketiga metode <i>Naive Bayes Classifier</i> , <i>K-Nearest Neighbor (KNN)</i> dan <i>Decision Tree</i> . Dalam mengklasifikasikan sentimen analisis tentang pendapat atau kecenderungan opini <i>netizen</i> terhadap pemerintah Pekanbaru yang mengandung sentimen positif, negatif, dan netral dalam komentar di media sosial <i>Twitter</i>	Hasil dari penelitian ini menunjukkan, data yang diperoleh dari <i>tweet</i> yang telah diolah menggunakan sentiment analisis dengan metode <i>Naive Bayes</i> menghasilkan akurasi sebesar 100%, metode KKN menghasilkan akurasi 98,25%, dan metode <i>Decision Tree</i> menghasilkan akurasi sebesar 62,28%

Tabel 2.2 Penelitian Terkait (Lanjutan 4)

No.	Nama Penulis	Judul	Isi Ringkasan	Hasil
5.	Puspita, R., & Widodo, A. (2021).	Perbandingan Metode KNN, <i>Decision Tree</i> , dan <i>Naïve</i> <i>Bayes</i> Terhadap Analisis Sentimen Pengguna Layanan BPJS.	Penelitian ini bertujuan untuk mengetahui opini dari masyarakat Indonesia mengenai layanan dari BPJS dengan cara mengumpulkan data dari <i>Twitter</i> menggunakan metode algoritma <i>KNN</i> , <i>Decision Tree</i> dan <i>Naïve Bayes</i> . Serta mengetahui tingkat, akurasi, dari, tiga metode tersebut.	Hasil dari penelitian ini dapat diketahui bahwa metode <i>Decision Tree</i> adalah metode yang tingkat akurasinya lebih tinggi dibandingkan kedua metode lainnya dengan tingkat akurasi sebesar 96.13%.

Tabel 2.2 Penelitian Terkait (Lanjutan 5)

No.	Nama Penulis	Judul	Isi Ringkasan	Hasil
6.	Watmah, S., Suryanto, S., & Martias, M. (2021).	Komparasi Metode K- NN, <i>Support Vector Machine</i> , Dan <i>Random Forest</i> Pada <i>E-Commerce Shopee</i>	Pada penelitian ini bertujuan untuk menguji seberapa kepuasan pengguna akun shopee berdasarkan <i>review</i> pada <i>google playstore</i> menggunakan 3 metode klasifikasi yang berbeda diantaranya <i>K-NN</i> , <i>Support Vector Machine</i> , dan <i>Random Forest</i> . untuk menggali kecenderungan dari komentar pengguna lalu membandingkan hasilnya untuk mengetahui rasio dan akurasi menggunakan <i>data mining</i> .	Dari hasil studi yang dilakukan dengan menggunakan 265 data komentar pada <i>google playstore</i> menunjukkan nilai akurasi, presisi, dan <i>recall</i> pada metode K-NN 89,0%, 89,7% 87,5%. <i>Random Forest</i> 83,0%, 85,7%, 81,4%, dan SVM 89,4%, 89,5%, 89,7%. Dengan demikian klasifikasi terbaik pada studi ini adalah metode SVM dengan nilai 89,4% presisi 89,5% dan <i>recall</i> 89,7%.

Tabel 2.2 Penelitian Terkait (Lanjutan 6)

No.	Nama Penulis	Judul	Isi Ringkasan	Hasil
7.	Hendrik Setiawan, Ema Utami, Sudarmawan (2021)	Analisis Sentimen <i>Twitter</i> Kuliah <i>Online</i> Pasca <i>Covid-19</i> Menggunakan Algoritma <i>Support</i> <i>Vector</i> <i>Machine</i> dan <i>Naïve Bayes</i>	Penelitian ini melakukan analisis sentimen terhadap tanggapan para mahasiswa terkait peraturan kuliah <i>online</i> dikala pandemic <i>Covid-19</i> . Pada penelitian ini menggunakan algoritma <i>Support Vector Machine</i> dan <i>Naïve Bayes</i> . Penelitian ini bertujuan untuk membandingkan performa algoritma <i>Support Vector Machine</i> dan <i>Naïve Bayes</i> .	Penelitian ini mendapatkan hasil bahwa performa terbaik algoritma <i>Naïve Bayes</i> didapatkan pada iterasi ke 1 mendapatkan hasil akurasi 81,20%, waktu 9,00 detik, <i>recall</i> 79,60% dan presisi 79,40%. Untuk hasil terbaik pada algoritma SVM didapatkan pada iterasi ke 423 dikarenakan proses <i>training</i> data untuk mendapatkan hasil optimal dengan hasil akurasi 85%, waktu 31,60 detik, <i>recall</i> 84% dan presisi 83,60%.

Tabel 2.2 Penelitian Terkait (Lanjutan 7)

No.	Nama Penulis	Judul	Isi Ringkasan	Hasil
8.	Nuraini Ika Pratiwi Kalingara, Oktariani Nurul Pratiwi, Hilman Dwi Anggana (2021)	Analisis Sentimen <i>Review Customer Terhadap Layanan Ekspedisi JNE Dan J&T Express</i> Menggunakan Metode <i>Naïve Bayes</i>	Penelitian ini melakukan analisis sentimen terhadap komentar pengguna terhadap layanan Ekspedisi Jne Dan J&T <i>Express</i> pada sisal media <i>Twitter</i> . Komentar diklasifikasikan kedalam label negatif, netral, dan positif. Meted yang digunakan pada penelitian ini adalah <i>Naïve Bayes</i> . Penelitian ini bertujuan untuk mengukur performa <i>Naïve Bayes</i> dalam klasifikasi data menggunakan <i>confusion matrix</i> .	Hasil dari penelitian ini menunjukkan <i>Naïve Bayes</i> memiliki akurasi yang paling baik dengan mengambil rasio testing 75:25. Pada penelitian ini, penerapan <i>k-fold cross validation</i> menghasilkan score 76% untuk ekspedisi JNE dan score 75% untuk ekspedisi J&T <i>Express</i> . Dari keseluruhan persentase yang diperoleh maka model masuk ke dalam klasifikasi cukup baik.

Tabel 2.2 Penelitian Terkait (Lanjutan 8)

No.	Nama Penulis	Judul	Isi Ringkasan	Hasil
9.	Harun, A., & Ananda, D. P. (2021).	Analisa Sentimen Opini Publik Tentang Vaksinasi Covid-19 di Indonesia Menggunakan <i>Naïve Bayes</i> dan <i>Decission</i> <i>Tree</i>	Penelitian ini bertujuan untuk mengetahui analisis opini masyarakat terhadap vaksinasi <i>COVID-19</i> di Indonesia berdasarkan sentimen positif dan negatifnya, serta membandingkan tingkat akurasi menggunakan algoritma <i>Naïve Bayes Classifier (NBC)</i> dan <i>Decision Tree</i> .	Hasil penelitian analisa sentimen opini masyarakat tentang vaksinasi <i>COVID-19</i> yang telah dilakukan, cenderung ke tanggapan negatif dengan nilai akurasi 100.00% menggunakan algoritma NBC dan 50.39% menggunakan algoritma <i>Decision Tree</i> .

Tabel 2.2 Penelitian Terkait (Lanjutan 9)

No.	Nama Penulis	Judul	Isi Ringkasan	Hasil
10.	Budiyantara, A., Irwansyah, I., Prengki, E., Pratama, P. A., & Wiliani, N. (2020).	Komparasi Algoritma <i>Decision Tree</i> , <i>Naive Bayes</i> Dan <i>K-Nearest Neighbor</i> Untuk Memprediksi Mahasiswa Lulus Tepat Waktu	Pada penelitian ini dilakukan komparasi antara algoritma <i>Decision Tree</i> , <i>Naive Bayes</i> dan <i>K-Nearest Neighbor (K-NN)</i> untuk menemukan hasil tingkat akurasi yang tertinggi dan kinerja yang baik menggunakan klasifikasi dari data Akademik Mahasiswa STMIK Dipanegara Makassar untuk memprediksi mahasiswa lulus tepat waktu.	Dari hasil penelitian tersebut menunjukkan nilai akurasi komparasi antara algoritma <i>Decision Tree (C4.5)</i> memperoleh hasil akurasi tertinggi sebesar 98.04% pada pengujian ke 3. Metode <i>Naive Bayes</i> memperoleh hasil akurasi tertinggi sebesar 96.00% pada pengujian ke 4, dan Metode <i>K-Nearest Neighbor (K-NN)</i> memperoleh hasil akurasi tertinggi sebesar 90.00% pada pengujian ke 2.

Tabel 2.2 Penelitian Terkait (Lanjutan 10)

No.	Nama Penulis	Judul	Isi Ringkasan	Hasil
11.	Adinegoro, R. W., Wahyuni, E. D., & Arifiyanti, A. A. (2020).	Aplikasi Website Sentiment Analysis Ulasan Tokopedia	<p>Pada penelitian ini dilakukan analisis sentimen yang terkandung dalam sebuah ulasan yang diberikan pengguna Sentimen yang dianalisa merupakan sentimen positif, netral, dan negatif. Untuk mendapatkan ulasan produk pada Tokopedia dilakukan pengambilan data dengan cara mengakses API Tokopedia.</p> <p>Penelitian ini dilakukan dengan algoritma <i>Naive Bayes</i> menggunakan 4 algoritma <i>naive bayes</i> yang berbeda untuk mengetahui algoritma dengan <i>performa</i> yang terbaik.</p>	<p>Dari penelitian yang dilakukan, algoritma <i>Naive Bayes</i> dengan performa terbaik adalah <i>Complement</i> dimana menghasilkan tingkat akurasi prediksi sebesar 83,66%, dengan akurasi sebesar 84,37% pada pengujian sistem pada data ulasan sebanyak 32 buah.</p>

Tabel 2.2 Penelitian Terkait (Lanjutan 11)

No.	Nama Penulis	Judul	Isi Ringkasan	Hasil
12.	Syarifuddin, M. (2020).	Analisis Sentimen Opini Publik Terhadap Efek PSBB Pada Twitter Dengan Algoritma <i>Decision Tree-Knn- Naïve Bayes</i>	Pada penelitian ini dilakukan analisis sentimen terhadap opini publik terhadap PSBB pada media sosial <i>Twitter</i> . Selain itu dilakukan perbandingan antara algoritma <i>Decision Tree</i> , <i>K-Nearest Neighbors (K-NN)</i> , dan <i>Naïve Bayes Classifier</i> dengan tujuan menemukan akurasi terbaik.	Hasil penelitian menunjukkan bahwa <i>accuracy</i> dari <i>Decision Tree</i> , KNN, dan <i>Naïve Bayes</i> sebesar 83,3%, 80,80%, dan 80,03%. Hasil untuk <i>precision</i> nya sebesar 81,06%, 82,72%, dan 87,54%. Sementara hasil untuk <i>Recall</i> -nya adalah 87,17%, 74,41%, dan 62,71%. Sehingga dapat disimpulkan bahwa <i>classifier Decision Tree</i> adalah <i>classifier</i> terbaik. karena memberikan prediksi yang lebih akurat dan tepat.

Tabel 2.2 Penelitian Terkait (Lanjutan 12)

No.	Nama Penulis	Judul	Isi Ringkasan	Hasil
13.	Apriani, R., & Gustian, D. (2019).	Analisis Sentimen Dengan <i>Naïve</i> <i>Bayes</i> Terhadap Komentar Aplikasi Tokopedia	Penelitian ini menggunakan metode <i>Naive Bayes</i> untuk menghasilkan sentimen positif dan negatif terhadap komentar pengguna aplikasi Tokopedia di <i>Playstore</i> .	Hasil dari penelitian ini, pengujian berdasarkan nilai <i>class negative</i> , <i>class positive</i> , <i>recall</i> , dan <i>accuracy</i> pada analisis sentimen. dengan nilai <i>accuracy performance</i> yang baik sebesar 97,13%, dengan nilai <i>precision</i> 1 Sementara pada <i>Class Recall</i> dihasilkan nilai 95,49% (<i>positive class: negative</i>). Dan nilai AUC 0,980.

Tabel 2.2 Penelitian Terkait (Lanjutan 13)

No.	Nama Penulis	Judul	Isi Ringkasan	Hasil
14.	Romadloni, N. T., Santoso, I., & Budilaksono, S. (2019).	Perbandingan Metode <i>Naive</i> <i>Bayes</i> , <i>Knn</i> Dan <i>Decision</i> <i>Tree</i> Terhadap Analisis Sentimen Transportasi KRL <i>Commuter</i> <i>Line</i>	Penelitian ini melakukan analisis sentimen terhadap komentar pengguna terhadap layanan transportasi KRL <i>Commuter Line</i> media sosial <i>Twitter</i> . Komentar diklasifikasikan kedalam label negatif, netral, dan positif. Dengan membandingkan 3 metode <i>Naive Bayes</i> , <i>KNN</i> dan <i>Decision Tree</i> . Penelitian ini bertujuan untuk mengukur tingkat akurasi <i>Naive Bayes</i> , <i>KNN</i> dan <i>Decision Tree</i> .	Hasil dari penelitian ini menunjukkan dapat dilakukan dengan metode <i>Naive Bayes</i> , dengan akurasi 80,00% , <i>precision</i> 66,67 %, <i>sensitifity</i> 100% dan <i>Specifity</i> 66,67 %. Pada metode <i>KNN</i> akurasi 80%, <i>precision</i> 100%, <i>sensitivity</i> 50%, <i>specificity</i> 100% dan pada metode <i>Decision Tree</i> akurasi 100%, <i>precision</i> 100%, <i>sensitivity</i> 100%, <i>specificity</i> 100%.

Tabel 2.2 Penelitian Terkait (Lanjutan 14)

No.	Nama Penulis	Judul	Isi Ringkasan	Hasil
15.	Nancy Ria Silvani Huaturuk, H., Rima Dias Ramadhani, R., & Dwi Januarita AK, J. (2018)	Komparasi Akurasi <i>Naïve</i> <i>Bayes</i> dan <i>Support</i> <i>Vector</i> <i>Machine</i> (SVM) untuk Rekomendasi Produk in <i>Fashion</i> <i>Dress</i>	Penelitian ini bertujuan untuk mengetahui tingkat akurasi yang lebih tinggi dan lebih baik dari algoritma <i>Naïve Bayes</i> dan <i>Support Vector Machine</i> (SVM). Serta mampu memberikan rekomendasi algoritma yang sesuai dengan kebutuhan data untuk dapat diterapkan pada sistem klasifikasi untuk rekomendasi <i>fashion dress</i> masa kini.	Pada penelitian ini dapat disimpulkan bahwa <i>Naive Bayes</i> mendapatkan hasil akurasi lebih tinggi dari SVM karena dalam tahapan klasifikasiannya <i>Naive Bayes</i> memproses satu persatu data atribut. Berbeda dengan SVM yang melakukan klasifikasi secara general sehingga cakupan SVM lebih luas. Hasil yang diperoleh <i>Naive Bayes</i> sebesar 74% sedangkan SVM memiliki nilai akurasi yaitu 66%.

2.5 Matriks Penelitian

Tabel 2.3 Matriks Penelitian

No.	Penulis	Algoritma	Tujuan Penelitian	Subjek Penelitian	Sumber Data
1.	Zuda Pradana Putra, Aryo Nugroho (2022).	<ul style="list-style-type: none"> • <i>Naïve Bayes</i> • <i>K-NN</i> 	<ul style="list-style-type: none"> • Klasifikasi opini • Membandingkan kinerja algoritma 	<ul style="list-style-type: none"> • JNE • <i>J&T Express</i> • Pos Indonesia 	<i>Twitter</i>
2.	Herliana, A. (2022).	<ul style="list-style-type: none"> • <i>Naive Bayes</i> • <i>K-Nearest Neighbor (KNN)</i> • <i>Decision Tree</i> 	<ul style="list-style-type: none"> • Klasifikasi opini • Membandingkan kinerja algoritma 	<ul style="list-style-type: none"> • Kuliah Daring 	<i>Twitter</i>
3.	Hasri, C. F., & Alita, D. (2022)	<ul style="list-style-type: none"> • <i>Naïve Bayes</i> • <i>Support Vector Machine (SVM)</i> 	<ul style="list-style-type: none"> • Klasifikasi opini • Membandingkan kinerja algoritma 	<ul style="list-style-type: none"> • COVID-19 	<i>Twitter</i>

Tabel 2.3 Matriks Penelitian (Lanjutan 1)

No.	Penulis	Algoritma	Tujuan Penelitian	Subjek Penelitian	Sumber Data
4.	Anam, M. K., Pikir, B. N., & Firdaus, M. B. (2021).	<ul style="list-style-type: none"> • <i>Decision Tree</i> • <i>KNN</i> • <i>Naive Bayes</i> 	<ul style="list-style-type: none"> • Membandingkan kinerja algoritma • Klasifikasi opini 	<ul style="list-style-type: none"> • Shopee 	<i>Google Play Store</i>
5.	Puspita, R., & Widodo, A. (2021).	<ul style="list-style-type: none"> • <i>KNN</i> • <i>Decision Tree</i> • <i>Naive Bayes</i> 	<ul style="list-style-type: none"> • Membandingkan kinerja algoritma • Klasifikasi opini 	<ul style="list-style-type: none"> • Layanan BPJS 	<i>Twitter</i>
6.	Watmah, S., Suryanto, S., & Martias, M. (2021).	<ul style="list-style-type: none"> • <i>KNN</i> • <i>SVM</i> • <i>Random Forest</i> 	<ul style="list-style-type: none"> • Pengukuran kinerja algoritma • Klasifikasi opini 	<ul style="list-style-type: none"> • KRL Commuter Line 	<i>Twitter</i>

Tabel 2.3 Matriks Penelitian (Lanjutan 2)

No.	Penulis	Algoritma	Tujuan Penelitian	Subjek Penelitian	Sumber Data
7.	Hendrik Setiawan, Ema Utami, Sudarmawan (2021)	<ul style="list-style-type: none"> • <i>SVM</i> • <i>Naive Bayes</i> 	<ul style="list-style-type: none"> • Membandingkan kinerja algoritma • Klasifikasi opini 	<ul style="list-style-type: none"> • Pelajar • Mahasiswa 	<i>Twitter</i>
8.	Nuraini Ika Pratiwi Kalingara, Oktariani Nurul Pratiwi, Hilman Dwi Anggana (2021)	<ul style="list-style-type: none"> • <i>Naive Bayes</i> 	<ul style="list-style-type: none"> • Membandingkan kinerja algoritma • Klasifikasi opini 	<ul style="list-style-type: none"> • JNE • <i>J&T Express</i> 	<i>Twitter</i>
9.	Harun, A., & Ananda, D. P. (2021).	<ul style="list-style-type: none"> • <i>Naive Bayes</i> • <i>Decision Tree</i> 	<ul style="list-style-type: none"> • Membandingkan kinerja algoritma • Klasifikasi opini 	<ul style="list-style-type: none"> • Vaksinasi COVID-19 	<i>Facebook</i>

Tabel 2.3 Matriks Penelitian (Lanjutan 3)

No.	Penulis	Algoritma	Tujuan Penelitian	Subjek Penelitian	Sumber Data
10.	Budiyantara, A., Irwansyah, I., Prengki, E., Pratama, P. A., & Wiliani, N. (2020).	<ul style="list-style-type: none"> • <i>Decission Tree</i> • <i>Naive Bayes</i> • <i>KNN</i> 	<ul style="list-style-type: none"> • Pengukuran kinerja algoritma 	<ul style="list-style-type: none"> • Pelajar • Mahasiswa 	Data Akademik Mahasiswa STMIK Dipanegara Makassar
11.	Adinegoro, R. W., Wahyuni, E. D., & Arifiyanti, A. A. (2020).	<ul style="list-style-type: none"> • <i>Naive Bayes</i> 	<ul style="list-style-type: none"> • Membandingkan kinerja algoritma • Klasifikasi opini 	<ul style="list-style-type: none"> • Tokopedia 	Aplikasi Tokopedia

Tabel 2.3 Matriks Penelitian (Lanjutan 4)

No.	Penulis	Algoritma	Tujuan Penelitian	Subjek Penelitian	Sumber Data
12.	Syarifuddin, M. (2020).	<ul style="list-style-type: none"> • <i>Naive Bayes Classifier</i> (NBC) • <i>KNN</i> • <i>Decision Tree</i> 	<ul style="list-style-type: none"> • Klasifikasi opini • Membandingkan kinerja algoritma 	<ul style="list-style-type: none"> • PSBB 	<i>Twitter</i>
13.	Apriani, R., & Gustian, D. (2019).	<ul style="list-style-type: none"> • <i>Naive Bayes</i> 	<ul style="list-style-type: none"> • Klasifikasi opini • Pengukuran kinerja algoritma 	<ul style="list-style-type: none"> • Tokopedia 	<i>Google Play Store</i>
14.	Romadloni, N. T., Santoso, I., & Budilaksono, S. (2019).	<ul style="list-style-type: none"> • <i>Naive Bayes</i> • <i>KNN</i> • <i>Decision Tree</i> 	<ul style="list-style-type: none"> • Klasifikasi opini • Membandingkan kinerja algoritma 	<ul style="list-style-type: none"> • Komentar Netizen 	<i>Twitter</i>

Tabel 2.3 Matriks Penelitian (Lanjutan 5)

No.	Penulis	Algoritma	Tujuan Penelitian	Subjek Penelitian	Sumber Data
15.	Nancy Ria Silvani Huaturuk, H., Rima Dias Ramadhani, R., & Dwi Januarita AK, J. (2018)	<ul style="list-style-type: none"> • <i>Naïve Bayes</i> • <i>SVM</i> 	<ul style="list-style-type: none"> • Membandingkan kinerja algoritma • Klasifikasi data 	<ul style="list-style-type: none"> • Fashion Dress 	<i>UCI Repository</i>
16.	Penelitian yang dilakukan (2022)	<ul style="list-style-type: none"> • <i>Decision Tree</i> • <i>Naïve Bayes</i> 	<ul style="list-style-type: none"> • Membandingkan kinerja algoritma 	<ul style="list-style-type: none"> • Tokopedia 	<i>Twitter</i>

2.6 *State Of the Art*

Penelitian yang akan dilakukan dengan judul “Komparasi Tingkat Akurasi Algoritma *Decision Tree* dan *Naïve Bayes* Dalam Analisis Sentimen Persepsi Terhadap Tokopedia”, bertujuan untuk menganalisis dan mengklasifikasikan opini-opini pengguna *marketplace* tersebut berdasarkan sentimen yang terkandung dalam persepsi tersebut. Penelitian ini dilakukan untuk mengetahui tingkat akurasi yang tertinggi dari algoritma *Decision Tree* dan *Naïve Bayes* yang banyak digunakan dalam analisis sentimen yang didalam penelitian ini dilakukan untuk membandingkan tingkat akurasi dari kedua algoritma tersebut berdasarkan permasalahan peningkatan jumlah opini serta persepsi berisi sentimen yang dapat mempengaruhi kualitas perusahaan tersebut.

Penelitian ini memiliki persamaan dengan penelitian yang telah dilakukan sebelumnya yaitu, menggunakan klasifikasi algoritma dari *Decision Tree* serta *Naïve Bayes*. Namun penelitian ini juga memiliki perbedaan pada subjek penelitian dan algoritma klasifikasi yang digunakan untuk mencari tingkat akurasinya. Penelitian ini menggunakan komparasi klasifikasi algoritma *Decision Tree* dan *Naïve Bayes* dengan tujuan membandingkan tingkat akurasi dari kedua algoritma tersebut dengan menggunakan subjek analisis sentimen *marketplace* Tokopedia. Dengan dilakukannya penelitian ini diharapkan dapat menjadi bahan evaluasi untuk menentukan algoritma sentimen yang tepat dan akurat, serta sebagai peningkatan kualitas perusahaan dan sebagai informasi mengenai *marketplace* tersebut.