

BAB II

TINJAUAN PUSTAKA

2.1 Tinjauan Pustaka

Penelitian menurut Yulita, W (2021) melakukan analisis sentimen terhadap vaksin covid-19 menggunakan *Algoritma Naïve Bayes*. Penelitian ini bertujuan untuk menganalisis pendapat tentang covid-19 di Indonesia. Penelitian ini melibatkan 3780 *tweet* yang terkait dengan vaksinasi dan menggunakan algoritma Naïve Bayes Classifier. Hasil analisis menunjukkan bahwa mayoritas *tweet* menunjukkan sikap positif sebanyak 60,3%, sementara *tweet* yang bersikap netral sebanyak 34,4%, dan *tweet* yang menentang sebanyak 5,4%. Tingkat akurasi yang berhasil dicapai mencapai sebesar 93% (0,93).

Menurut Salsabila et al. (2022) melakukan penelitian mengenai analisis sentimen pelanggan tokopedia dengan menggunakan Metode Naïve Bayes Classifier terhadap komentar pelanggan yang mencakup beberapa sentimen positif dan negatif. Penelitian ini fokus pada komentar pelanggan terkait aplikasi Tokopedia di *Playstore*. Hasil eksperimen yang dilakukan oleh *rapid evaluator* pada 143 data uji menunjukkan tingkat akurasi 95,10% dengan *class recall*: 91,46% (*positive class: Positive*), serta *class precision*: 100.00% (*positive class: Positive*). Nilai *AUC(Area Under the Curve)* untuk kelas positif adalah 0.99 (*positive class: Positive*).

Penelitian menurut Utami et al. (2021) mengenai analisis sentimen terhadap xiaomi Indonesia menggunakan Metode *Naïve Bayes*. Penelitian ini bertujuan untuk mengetahui sentimen masyarakat terhadap brand Xiaomi menggunakan

metode Naïve Bayes Classifier (NBC) dengan menggunakan data sebanyak 2.078 yang diambil dari media sosial *twitter* dengan menggunakan teknik *scrapping*. Data yang dipergunakan dalam penelitian ini terdiri dari *tweet* berbahasa Indonesia yang dikumpulkan dari platform media sosial *twitter*. Pengumpulan data dilakukan menggunakan *twitter* API dengan kata kunci "xiaomi". Selanjutnya, data yang berhasil dikumpulkan dibagi menjadi dua jenis, yaitu data latih dan data uji, dengan rasio 80:20. Pembagian ini bertujuan untuk menguji kinerja algoritma Naive Bayes. Dalam tahap analisis ini, digunakan *library TextBlob* untuk memperoleh label secara otomatis dari data *tweet*. Hasil penelitian menunjukkan bahwa analisis sentimen terhadap merek Xiaomi mencapai tingkat akurasi sebesar 71.88%, dengan persentase data polaritas positif sebesar 39%, polaritas netral sebesar 51%, dan polaritas negatif sebesar 10%.

Penelitian dilakukan juga oleh Muslimin & Lusiana, (2023). Penelitian yang dilakukan mengenai analisis sentimen masyarakat terhadap harga bahan pokok menggunakan Metode Naive Bayes Classifier. Penelitian ini memanfaatkan sebanyak 2070 data *tweet*, diambil dari situs *netlytic.org* dengan menggunakan kata kunci "bahan pokok". Hasil analisis menunjukkan bahwa terdapat prevalensi lebih tinggi dari sentimen negatif dibandingkan sentimen positif, dengan 2,8% dari data menunjukkan sentimen positif dan 97,2% menunjukkan sentimen negatif. Kemudian, dilakukan pembagian data pada dataset dengan perbandingan 6:4. Selanjutnya dilakukan pembagian data dengan alokasi 60% untuk data latih dan 40% untuk data uji. Penggunaan ukuran 40% untuk data uji

menghasilkan tingkat akurasi optimal. Hasil pengujian model menggunakan metode Naïve Bayes Classifier menunjukkan evaluasi tertinggi, dengan akurasi mencapai 94,38%, presisi sebesar 59,67%, recall sebesar 67,93%, dan F-measure sebesar 62,32%.

Penelitian selanjutnya juga dilakukan oleh Pintoko & Lhaksana (2018). analisis sentimen jasa transportasi pada *twitter* menggunakan Metode Naive Bayes Classifier. Penelitian ini dilakukan dengan tujuan menganalisis sentimen terhadap layanan transportasi *online*. Pengumpulan data dilakukan menggunakan *platform* media sosial, terutama pada *Twitter*, dengan keharusan menggunakan data yang valid. Data *tweet* yang terkumpul kemudian dianalisis dan diklasifikasikan menjadi kategori positif dan negatif menggunakan metode Naive Bayes Classifier. Hasil dari sistem yang dikembangkan menunjukkan bahwa sentimen positif mencapai 88.60%, sementara sentimen negatif sebesar 11.40%, dengan tingkat akurasi sebesar 86.80%. Hasil menunjukkan bahwa tingkat sentimen positif dalam *tweet* masyarakat lebih tinggi dibandingkan dengan tingkat sentimen negatif.

Pada penelitian ini, penulis menganalisis sentimen masyarakat terhadap penutupan TikTok Shop menggunakan metode *Naïve Bayes Classifier*, dan mengklasifikasikan sentimen tersebut ke dalam tiga kategori yaitu Positif, Negatif, dan Netral. Proses analisis sentimen ini mencakup pengambilan data dengan menggunakan *tweet harvest*, pelabelan dengan *TextBlob*, serta *Preprocessing* yang meliputi *case folding*, *tokenizing*, *normalization*, *stopword*

removal, dan *stemming*. Data *tweet* diklasifikasikan menggunakan metode *Naïve Bayes Classifier* dengan evaluasi model dilakukan melalui *Confusion Matrix*.

Perbandingan tinjauan pustaka dari peneliti sebelumnya dapat dilihat pada tabel 2.1.

Tabel 2. 1 Tinjauan Pustaka

Peneliti	Objek	Metode	Keterangan
Winda, Y. (2021)	Vaksin Covid-19	Algoritma Naïve Bayes Classifier	Tingkat akurasi yang dicapai sebesar 93% (0,93), dari analisis penelitian ini menunjukkan bahwa mayoritas <i>tweet</i> menunjukkan sikap positif sebanyak 60,3%, sementara <i>tweet</i> yang bersikap netral sebanyak 34,4%, dan <i>tweet</i> yang menentang sebanyak 5,4%.
Salsabila et al. (2022)	Pelanggan Tokopedia	Naïve Bayes Classifier	Tingkat akurasi 95,10% dengan class recall: 91,46% (positive class: Positive), serta class precision: 100.00% (positive class: Positive). Nilai AUC (Area Under the Curve) untuk kelas positif adalah 0.99 (positive class: Positive).
Utami et al. (2021)	Xiaomi Indonesia	Naïve Bayes	Tingkat akurasi sebesar 71.88%, dengan persentase data polaritas positif sebesar 39%, polaritas netral sebesar 51%, dan polaritas negatif sebesar 10%.
Muslimin & Lusiana. (2023)	Kenaikan Harga Bahan Pokok	Naïve Bayes Classifier	Hasil analisis menunjukkan bahwa sentimen negatif lebih banyak dibandingkan sentimen positif, dengan persentase 2,8% sentimen positif dan 97,2% untuk sentimen negatif. Pembagian data pada dataset dengan perbandingan 6:4. Data dibagi menjadi 60% data latih dan 40% data uji. Ukuran data uji sebanyak 40% dari keseluruhan data menghasilkan tingkat akurasi yang paling baik. Pengujian model dengan metode Naïve Bayes Classifier diperoleh hasil nilai evaluasi yaitu score accuracy tertinggi yaitu sebesar 94,38%, precision sebesar 59,67%, recall sebesar 67,93%, dan F-measure sebesar 62,32%.

Pintoko & Lhaksana. (2018)	Jasa Transportasi Online	Naïve Bayes Classifier	Sistem yang dikembangkan menunjukkan bahwa sentimen positif mencapai 88.60%, sementara sentimen negatif sebesar 11.40%, dengan tingkat akurasi sebesar 86.80%. Hasil menunjukkan bahwa tingkat sentimen positif dalam <i>tweet</i> masyarakat lebih tinggi dibandingkan dengan tingkat sentimen negatif.
Mauboi, F.T. (2024)	Penutupan TikTok Shop	Naïve Bayes Classifier	Penelitian ini menganalisis sentimen masyarakat terhadap penutupan TikTok Shop menggunakan metode <i>Naïve Bayes Classifier</i> , dan mengklasifikasikan sentimen tersebut ke dalam kategori tiga yaitu Positif, Negatif, dan Netral. Proses analisis sentimen ini mencakup pengambilan data dengan menggunakan <i>tweet harvest</i> , pelabelan dengan <i>TextBlob</i> , serta <i>Preprocessing</i> yang meliputi <i>case folding</i> , <i>tokenizing</i> , <i>normalization</i> , <i>stopword removal</i> , dan <i>stemming</i>

2.2 Landasan Teori

2.2.1 Analisis Sentimen

Analisis Menurut Liu (2012), analisis sentimen adalah teknik pemrosesan bahasa alami yang digunakan untuk mengekstrak, mengidentifikasi, atau mengklasifikasikan sentimen atau opini yang terkandung dalam teks. Tujuannya adalah untuk memahami sikap, perasaan, atau penilaian subjek terhadap suatu topik atau entitas.

Analisis sentimen, disebut juga penambangan opini, adalah bidang studi yang menganalisis opini, sentimen, evaluasi, penilaian, sikap, dan emosi terhadap objek atau topik, seperti produk, layanan, organisasi, individu, isu, peristiwa, dan atributnya. Tujuan dari sentimen analisis dan penambangan opini adalah menganalisis dan mengidentifikasi pandangan, sikap dan emosi orang mengenai suatu hal tertentu (Zhang et al. 2020).

Sentimen atau yang biasa dikenal dengan istilah opinion mining merupakan salah satu cabang penelitian dari text mining yang bertujuan untuk menentukan persepsi atau subjektivitas publik (khalayak) terhadap suatu topik pembahasan, kejadian, ataupun permasalahan (Rachman & Pramana, 2020).

2.2.2 TikTok Shop

TikTok Shop merupakan salah satu fitur yang berfungsi sebagai marketplace atau tempat belanja online. Menurut peneliti Sa'adah didalam (Supriyanto, dkk 2023) Tiktok Shop merupakan “Suatu *E-commerce*, jadi dapat dikatakan bahwa Tiktok Shop ini adalah suatu sistem informasi bisnis karena penjualan yang dilakukan melalui media elektronik yang didalamnya bisa memberi, menyediakan bahkan menyimpan suatu informasi khususnya mengenai jual beli atau bisnis”. Tiktok Shop juga pada umumnya menawarkan layanan yang sama dengan marketplace dan *e-commerce*, namun Tiktok Shop menawarkan harga yang sangat ekonomis dibandingkan dengan marketplace dan *e-commerce* yang sudah ada, serta menawarkan potongan harga menarik (Sa'adah, 2022).

2.2.3 X (Twitter)

Salah satu *platform* yang termasuk media sosial adalah X, yang sebelumnya dikenal dengan nama Twitter. X adalah sebuah layanan jejaring sosial yang memberikan layanan memperbarui status atau tulisan dengan maksimal 140 karakter (Pramudawardani, 2021).

Twitter merupakan media sosial yang hadir dengan format yang berbeda, dimana twitter memiliki konsep yaitu menyebarkan informasi pesan secara singkat, padat dan real time dengan kalimat yang kurang dari 140 karakter kepada

pembacanya di seluruh dunia yang bisa digunakan sebagai sarana penyebar informasi kepada semua orang baik yang dikenal maupun tidak, untuk memberitahukan keberadaan penggunanya. *Twitter* adalah sebuah situs web yang dimiliki dan dioperasikan oleh *Twitter Inc*, yang menawarkan jaringan sosial berupa *microblog* sehingga memungkinkan penggunanya untuk mengirim dan membaca pesan yang disebut kicauan atau tweets (Puspitadewi et al. 2016).

2.2.4 Text Mining

Menurut Feldman & Sanger didalam (Hakim, 2018) Text mining (penambangan teks) merupakan suatu penambangan yang dilakukan oleh komputer untuk mendapatkan sesuatu yang baru, sesuatu yang tidak diketahui sebelumnya atau menemukan kembali informasi yang tersebar, yang berasal dari informasi yang diekstrak secara otomatis dari sumber-sumber data teks yang berbeda-beda. Text mining, juga dikenal sebagai text data mining atau text *analytics* yaitu sebuah proses ekstraksi informasi berharga atau pengetahuan yang terkandung dalam dokumen teks. Tujuan utama dari text mining adalah untuk menganalisis dan memahami struktur, pola, dan makna dari teks, serta mengungkap informasi tersembunyi yang mungkin sulit diakses secara manual.

Text Mining adalah penggalian data untuk menyelesaikan masalah kebutuhan informasi dengan menerapkan teknik data *mining*, *machine learning*, *natural language processing*, pencarian informasi, dan manajemen pengetahuan. *Text mining* melibatkan pra-proses dokumen seperti kategorisasi teks, ekstraksi informasi, dan ekstraksi kata (Novianti & Wibowo, 2022).

2.2.5 *Tweet Harvest (Twitter Crawler)*

Tweet Harvest adalah alat baris perintah yang menggunakan playwright untuk mengambil *tweet* dari hasil pencarian twitter berdasarkan kata kunci dan rentang tanggal tertentu. *Tweet* yang berhasil diambil akan disimpan ke file CSV. Untuk menggunakan *tweet harvest* diperlukan token autentikasi yang didapatkan dengan masuk ke akun twitter anda di browser dan mengekstrak token otentikasi (Vincent, 2023).

2.2.6 *TextBlob*

TextBlob adalah perpustakaan Python untuk memproses data tekstual. Ini menyediakan API sederhana untuk mempelajari tugas-tugas pemrosesan bahasa alami (NLP) yang umum seperti penandaan bagian ucapan, ekstraksi frasa kata benda, analisis sentimen, klasifikasi (Loria, 2018).

Python *TextBlob* merupakan library Python yang digunakan untuk memproses data tekstual. *TextBlob* menyediakan API yang dapat digunakan untuk pemrosesan bahasa alami (NLP) seperti ekstraksi frase kata benda, analisis sentimen, klasifikasi, penerjemah. Hasi objek *TextBlob* digunakan untuk memproses pembelajaran bahasa alami dan library *TextBlob* hanya dapat mengenali Bahasa Inggris (Parlika et al. 2020).

2.2.7 *Naïve Naïve Bayes Classifier*

Naïve Bayes Classifier adalah salah satu algoritma yang digunakan dalam klasifikasi teks serta menggunakan metode *Machine Learning* dengan perhitungan probabilitas dan statistik yang dikemukakan oleh Thomas Bayes. Metode ini dapat mengklasifikasikan opini positif dan negatif (Sundara et al. 2020).

Algoritma NBC merupakan klasifikasi probabilitas sederhana dengan menggunakan model statistik untuk menghitung peluang kelas yang memiliki setiap kelompok dengan atribut yang tampil, dimana label kelas diambil dari beberapa himpunan berhingga (Maulana et al. 2023).

Secara umum, teorema Bayes dapat dinyatakan secara matematis dalam persamaan [1].

$$P(H|X) = \frac{P(H)P(X|H)}{P(X)} \quad (1)$$

X = data dengan kelas tidak dikenal

H = hipotesis data X adalah kelas khusus

P(H|X) = probabilitas hipotesis H didasarkan pada kondisi X

P(H) = probabilitas hipotesis H

P(X|H) = probabilitas hipotesis X didasarkan pada kondisi H

P(X) = probabilitas X

2.2.8 Text Preprocessing

Text processing adalah proses mengubah data mentah menjadi data yang sesuai dengan prosedur mining yang akan dilakukan dan merupakan tahap yang paling penting dalam data *mining*. Dalam tahap ini dilakukan proses pembersihan pada data yang masih kotor (Fikri et al. 2023). Pada tahap *preprocessing* ini dilakukan dengan lima tahap yaitu :

1. *Case folding*

Case folding merupakan tahap penyeragaman bentuk kata dari huruf besar menjadi huruf kecil, atau pengubahan seluruh karakter pada teks menjadi

huruf kecil serta menghilangkan tanda baca serta angka. Prosedur *case folding* yaitu memproses huruf alphabet dari “a” sampai dengan “z” saja agar karakter selain huruf tersebut akan dihapus (Ridwansyah, 2022).

2. *Tokenizing*

Tokenizing merupakan proses pemotongan string input berdasarkan tiap kata yang menyusunnya. Atau dengan kata lain adalah proses untuk membagi teks yang berasal dari kalimat atau paragraf menjadi bagian-bagian tertentu (Hildayah & Sahibu, 2021).

3. *Normalization*

Spelling Normalization merupakan proses memperbaiki dari kata-kata yang mengalami salah pengejaan maupun disingkat dalam bentuk tertentu (Septianningrum et al. 2021).

4. *Stopword Removal*

Stopword Removal merupakan proses pemilihan kata-kata penting yang mempunyai arti dan tidak, sehingga kata yang tidak mempunyai arti atau tidak relevan akan dibuang, contohnya “ini”, “itu”, “yang”, “di” dan lain-lain (Putu & Amarullah, 2021).

5. *Stemming*

Stemming merupakan proses mengubah suatu kata menjadi kata dasar atau menghapus suatu imbuhan kata. Dalam tahapan *stemming* menggunakan *library* Sastrawi atau perpustakaan (*library*) yang dikembangkan untuk bahasa pemrograman *python* dengan tujuan melakukan stemming dalam bahasa Indonesia (Soleqah, 2023).

2.2.9 Google Collaboratory

Google Collab, singkatan dari *Google Collaboratory*, adalah platform cloud yang disediakan oleh Google untuk menjalankan dan menulis kode Python melalui browser tanpa memerlukan konfigurasi tambahan. Ini memanfaatkan infrastruktur cloud Google dan memberikan lingkungan pengembangan yang kuat dengan akses ke GPU dan TPU (unit pemrosesan tensor) secara gratis (Nazar, 2024).

2.2.10 Python

Python merupakan bahasa pemrograman tingkat tinggi yang mendukung pemrograman berorientasi objek. *Python* sering digunakan dalam aplikasi *web*, pengembangan perangkat lunak, ilmu data, dan machine learning (ML). *Python* merupakan sebuah bahasa pemrograman tingkat tinggi yang bersifat general-purpose untuk mendukung pemrograman berorientasi objek. *Python* memiliki perbedaan dengan bahasa pemrograman lain yaitu dalam penulisan sintaks yang mudah dipahami dan bersifat ekspresif (Putri et al. 2022).

2.2.11 Machine Learning

Machine learning merupakan kecerdasan buatan (Artificial Intelligence) yang bisa membuat sistem mempunyai keahlian belajar sendiri secara otomatis serta meningkatkan kemampuannya berdasarkan pengalaman tanpa perlu diprogram oleh manusia (Mayang, 2021). *Machine Learning* juga merupakan salah satu cabang dari ilmu Kecerdasan Buatan, khususnya yang mempelajari tentang bagaimana komputer mampu belajar dari data untuk meningkatkan kecerdasannya, *Machine learning* memiliki fokus pada pengembangan sebuah sistem yang mampu belajar

sendiri untuk memutuskan sesuatu, tanpa harus berulang kali diprogram oleh manusia. Dengan metode tersebut, mesin tidak hanya bisa menemukan aturan untuk perilaku optimal dalam pengambilan keputusan, namun juga bisa beradaptasi dengan perubahan yang terjadi (Wahyono, 2018).