

BAB 2

TINJAUAN PUSTAKA DAN DASAR TEORI

2.1. Tinjauan Pustaka

Tinjauan Pustaka yang digunakan dalam penelitian ini didapat dari beberapa penelitian terdahulu, diantaranya:

Penelitian yg dilakukan oleh Nur Hayatin, 2021 “Implementasi Multinomial *Naïve Bayes* buat pembagian terstruktur mengenai Data Tweets Mengandung Term Depresi”. Penelitian ini telah berhasil mengklasifikasi data tweet yang mengandung terms depresi menggunakan memakai prosedur pemecahan Multinomial *Naïve Bayes*. asal hasil pengujian dihasilkan nilai akurasi sebesar 70%, dengan nilai precision recall masing-masing 72% beserta 65%, serta nilai f-measure sebesar 68%. buat kedepan, perlu dilakukan uji coba menggunakan menggunakan prosedur pemecahan yang lain supaya bisa mempertinggi keakuratan berasal hasil pembagian terstruktur mengenai.

Sari, M. (2020) pada jurnal "Analisis Sentimen Pengguna media sosial Twitter terhadap Pemerintah Kota Surabaya", penggunaan metode *Naive Bayes* pada analisis sentimen pengguna media umum sudah terbukti efektif pada mengidentifikasi sentimen positif serta negatif berasal teks yg terdapat. Hal ini membagikan bahwa metode *Naive Bayes* dapat sebagai pilihan yang sempurna pada menganalisis sentimen pengguna TikTok terhadap produk fashion.

Nurul Anizah (2023) dalam jurnalnya yang berjudul “Analisis Sentimen Terhadap Event Big Sale 11.11 Shopee pada media umum Instagram memakai Metode Naïve Bayes” . Sistem Analisis Sentimen di event Big Sale 11.11 Shopee menggunakan menggunakan pembagian terstruktur mengenai *Naive Bayes* bisa menganalisa sentimen komentar pada Instagram asal Akun Shopee sesuai kategorinya yaitu Positif, Netral, serta Negatif.

Aditia Rahman, Fadhil Rahmat, Muhammad Yoga Fariqi serta Sumarni Adi, 2020, “Metode *Naive Bayes* buat Menganalisis Akurasi Sentiment Komentar di Youtube”. pada penelitian ini diketahui bahwa akibat analisis sentimen positif lebih mungil asal sentimen negatif dengan nilai sentimen sebanyak 39,3% serta sentimen positif 60,6% berasal data penguji sejumlah 1500 data, yang akan terjadi tersebut didapatkan berasal pengujian memakai prosedur pemecahan *Naive Bayes* sebesar 78,17%.

Maulana Sura Dikuha Syaifullah (2022), dalam penelitiannya membuat sistem pendukung keputusan buat seleksi atlet beladiri JIU JITSU UNESA sesuai data fisik memakai Naïve Bayes. Penelitian ini dilakukan buat membantu seleksi atlet jujitsu yang dilakukan sang instruktur. Peneliti melakukan pembagian terstruktur mengenai sesuai data yg diambil oleh pelatih. selesainya melakukan percentage split memakai software, serta sudah mendapatkan yang akan terjadi menggunakan nilai yang cukup stabil mulai 70% pada percentage split 60%, nilai 85.71 di percentage split 70%, dan terakhir menerima nilai 80% pada percentage split 80%. menggunakan demikian mampu dikatakan bahwa metode penjabaran Naïve Bayes memiliki akurasi yang tinggi dan stabil buat penelitian ini.

Perbedaan Penelitian ini difokuskan pada analisis sentimen pengguna TikTok terhadap produk fashion menggunakan Naïve Bayes. Tujuannya adalah untuk mengidentifikasi trend dan preferensi pengguna TikTok dalam konteks fashion, sehingga hasilnya dapat membantu perusahaan dalam merumuskan strategi pemasaran dan pengembangan produk yang sesuai dengan kebutuhan konsumen.

Tabel 2. 1 Tinjauan Pustaka

No	Peneliti	Metode	Objek Data	Hasil
1	(Hayati 2021)	Naïve bayes	Implementasi Multinomial Naïve Bayes Untuk Klasifikasi Data Tweets Mengandung Term Depresi	Penelitian ini telah berhasil mengklasifikasi data tweet yang mengandung terms depresi dengan menggunakan algoritma Multinomial Naïve Bayes. Dari hasil pengujian didapatkan nilai akurasi sebesar 70%, dengan nilai precision recall masing-masing 72% dan 65%, dan nilai f- measure sebesar 68.
2	(Sari, M. 2020)	<i>Naive Bayes</i>	Analisis Sentimen Pengguna Media Sosial Twitter terhadap Pemerintah Kota Surabaya	penggunaan metode <i>Naive Bayes</i> dalam analisis sentimen pengguna media sosial telah terbukti efektif dalam mengidentifikasi sentimen positif dan negatif dari teks yang ada.

3	(Nurul Anizah 2023)	Naïve bayes	Analisis Sentimen Terhadap Event Big Sale 11.11 Shopee di Media Sosial Instagram Menggunakan Metode Naïve Bayes	dengan menggunakan klasifikasi <i>Naive Bayes</i> mampu menganalisa sentimen komentar pada Instagram dari Akun Shopee berdasarkan kategorinya yaitu Positif, Netral, dan Negatif.
4	(Rahman et al. 2020)	Naïve bayes	Metode <i>Naive Bayes</i> Untuk Menganalisis Akurasi Sentiment Komentar di Youtube	Pada penelitian ini diketahui bahwa hasil analisis sentimen positif lebih kecil dari sentimen negatif dengan nilai sentimen sebesar 39,3% dan sentimen positif 60,6% dari data pengujian sejumlah 1500 data, hasil tersebut didapatkan dari pengujian menggunakan algoritma <i>Naive Bayes</i> sebesar 78,17%
5	Maulana Sura Dikuha Syaifullah, Natalia Damastuti (2022)	Naïve Bayes	Beladiri Jiu Jitsu Unesa	70% pada percentage split 60%, nilai 85.71 pada percentage split 70%, dan 80% pada percentage 80%.

2.2. Dasar Teori

2.2.1. TikTok

TikTok adalah platform media sosial yang memungkinkan penggunanya untuk membuat dan melihat video pendek dengan durasi 10 menit. Video-video tersebut dapat disertai dengan music, filter, dan fitur kreatif lainnya, TikTok menyediakan berbagai fitur interaksi sosial, seperti komentar, duet, serta fitur *stitch*, yang memungkinkan pengguna untuk berkolaborasi atau menanggapi video kreator lainnya.

TikTok dikembangkan oleh ByteDance, sebuah perusahaan teknologi asal Tiongkok, dan pertama kali diluncurkan pada september 2016. Sejak diluncurkan, aplikasi ini dengan cepat meraih popularitas di dunia, terutama di kalangan generasi muda. Pada oktober 2020. TikTok telah diunduh lebih dari 2 miliar kali di seluruh dunia, menjadikannya salah satu aplikasi media sosial yang paling banyak digunakan. Platform ini kini berfungsi sebagai alat penting tidak hanya untuk hiburan, tetapi juga untuk pemasaran, tran budaya, serta berbagai kampanye sosial dan bisnis, termasuk di dunia fashion.

2.2.2. Analisis sentimen

Analisis sentimen adalah proses yang digunakan untuk mengidentifikasi opini atau sentimen dalam teks berdasarkan pola tertentu, dengan tujuan pengelompokan sentimen tersebut ke dalam kategori yang targetkan, seperti positif, negatif, atau netral (Arsi, kusuma, & Nurhakim, 2021). Metode ini memungkinkan kita untuk memahami bagaimana pemasaran atau tanggapan seseorang terhadap suatu topik tertentu.

Menurut Muhthia (2024), analisis sentimen bertujuan untuk menentukan opini atau perilaku seorang penulis terhadap suatu topik tertentu. Perilaku tersebut dapat mencerminkan penilaian, opini, serta alasan di balik sentimen yang digunakan. Dalam konteks analisis teks, perilaku ini ditelusuri melalui kata-kata atau frasa yang menunjukkan pandangan atau sikap penulis terhadap subjek yang dibahas.

Lebih lanjut, Mahendrajaya, Buntoro, & Setyawan (2019) menjelaskan bahwa analisis sentimen digunakan untuk mengetahui apakah opini yang terdapat dalam suatu teks bersifat positif atau negatif dengan cara mengelompokkan opini berdasarkan kalimat atau keseluruhan teks. Proses ini memungkinkan kita untuk mendapatkan gambaran lebih jelas tentang bagaimana sekelompok orang memproses suatu topik, atau layanan, yang sangat berguna dalam berbagai bidang seperti pemasaran, analisis konsumen, atau riset sosial.

2.2.3. Apify

Apify adalah sebuah platform yang menyediakan layanan untuk web scraping, otomatisasi, dan pemrosesan data dari internet. Dengan Apify, pengguna dapat membuat, menjelajahi, dan mengelola apa yang disebut sebagai "actor", yaitu skrip atau program otomatis yang dirancang untuk mengakses situs web, mengekstrak data, dan melakukan berbagai tugas otomatis di internet. Platform ini memungkinkan pengguna untuk mengambil data dari berbagai sumber web, melakukan otomatisasi proses, seperti pengisian formulir, serta menjalankan tugas lain yang biasanya memerlukan interaksi manual dengan situs web.

Apify sangat berguna dalam konteks pengumpulan data besar (big data) untuk keperluan analisis, riset pasar, pelacakan harga, hingga pengambilan data yang diperlukan untuk analisis sentimen atau aplikasi kecerdasan buatan (AI). Dengan fitur yang fleksibel dan kemampuan integrasi yang baik, Apify memungkinkan pengguna untuk melakukan otomatisasi skala besar tanpa harus berurusan dengan infrastruktur teknis yang rumit.

2.2.4. Livehaf

Livehaf adalah merek fashion e-commerce Indonesia yang fokus pada pakaian sehari-hari. Semua produk Livehaf diproduksi secara lokal di Indonesia dengan bekerja sama dengan pabrik besar dan usaha kecil menengah (UKM) di seluruh negeri. Arif Hidayat, pendiri Livehaf, mendapatkan inspirasi untuk menciptakan merek ini setelah melihat banyak merek luar negeri yang diproduksi di pabrik-pabrik di Indonesia. Ia bercita-cita menciptakan merek lokal yang bisa bersaing dengan merek internasional, dengan fokus utama pada kenyamanan dan gaya pakaian sehari-hari yang cocok untuk pasar Indonesia.

2.2.5. Text mining

Text mining merupakan sebuah proses di mana data dalam bentuk teks ditambang dan dianalisis untuk menemukan keterhubungan antar dokumen dengan cara mencari kata-kata yang dapat mewakili isi dari dokumen tersebut. Tujuan utama dari text mining adalah untuk mengekstrak informasi yang berguna dari sumber data yang tidak terstruktur, seperti dokumen teks, artikel, atau laporan (Soepomo, 2014).

Dalam proses ini, data yang digunakan berasal dari kumpulan dokumen dengan format yang tidak terstruktur. Melalui eksplorasi pola-pola unik dan identifikasi hubungan dalam teks, text mining berfungsi untuk mengkategorikan dan mengelompokkan teks. Proses ini sangat berguna dalam analisis besar-besaran dari teks seperti analisis sentimen, penemuan tren, atau klasifikasi dokumen berdasarkan topik yang relevan.

2.2.6. Python

Python adalah bahasa pemrograman taraf tinggi yang ditafsirkan, berorientasi objek, dengan semantik dinamis. Struktur data bawaan tingkat tinggi, dikombinasikan dengan pengetikan dinamis serta pengikat bergerak maju, membentuk sangat menarik buat Rapid *Application* Development, dan untuk dipergunakan menjadi bahasa skrip atau lem buat menghubungkan komponen yg terdapat bersama-sama.

Python yang sederhana dan mudah dipelajari menekankan keterbacaan dan karenanya mengurangi biaya pemeliharaan program. *Python* mendukung modul dan paket, yang mendorong modularitas program dan pengguna kembali kode, penerjemah python dan perpustakaan standar yang luas tersedia dalam bentuk sumber atau biner tanpa biaya untuk semua *platform* utama, dan dapat didistribusikan secara bebas (*Python (programming language) – Wikipedia*).

2.2.7. Text Preprocessing

Text preprocessing adalah salah satu implementasi penting dalam text mining yang bertujuan untuk mengubah data teks tidak terstruktur menjadi lebih terstruktur dan siap

untuk dianalisis. Proses ini dilakukan melalui beberapa tahapan untuk menyiapkan teks mentah agar dapat diproses lebih lanjut dalam analisis teks atau klasifikasi sentimen (Tineges, 2021). Berikut adalah tahapan-tahapan dalam text preprocessing :

1. *Cleansing*

Proses ini bertujuan untuk menghapus karakter yang tidak diperlukan dari teks, seperti simbol, hashtag, emoji, URL, atau elemen-elemen lain yang tidak relevan dalam analisis.

2. *Case Folding*

Pada tahap ini, semua huruf dalam teks diubah menjadi huruf kecil (non-kapital) untuk memastikan keseragaman dalam proses pencarian dan analisis. Ini dilakukan untuk menghindari perbedaan akibat penggunaan huruf kapital atau non-kapital.

3. *Tokenizing*

Proses tokenizing bertujuan untuk memecah teks atau kalimat menjadi unit-unit kecil yang disebut "token". Misalnya, kalimat "kualitas produk bagus" akan dipecah menjadi tiga token: [kualitas, produk, bagus].

4. *Normalisasi*

Normalisasi dilakukan untuk mengubah kata-kata singkatan atau istilah slang menjadi bentuk baku. Dalam Bahasa Indonesia, normalisasi juga bisa mengubah kata-kata yang tidak baku menjadi bentuk yang benar, sehingga lebih konsisten dalam analisis.

5. *Stopword Removal*

Pada tahap ini, kata-kata yang dianggap tidak relevan dalam proses pengklasifikasian, seperti kata sambung (dan, atau, yang), dihapus. Stopword removal penting untuk memastikan bahwa hanya kata-kata yang signifikan dalam konteks analisis yang digunakan.

Proses text *preprocessing* ini adalah langkah awal dan sangat penting dalam text mining, yang memungkinkan teks mentah dari berbagai dokumen dianalisis lebih lanjut untuk menemukan pola, kata kunci, atau hubungan antar dokumen. Dengan melalui serangkaian tahapan ini, teks dapat diolah menjadi data yang lebih mudah dipahami dan diinterpretasikan untuk tujuan analisis lebih lanjut.

2.2.8. *Naïve Bayes*

Klasifikasi *Naïve Bayes* adalah suatu metode klasifikasi yang menggunakan perhitungan probabilitas. Dengan menerapkan *teorema bayes* terhadap data statistic dapat mengasumsikan bahwa keberadaan atau ketiadaan dari sesuatu kelas dengan fitur lainnya.

Statistic yang diklasifikasikan juga dapat digunakan untuk memprediksi probabilitas keanggotaan sesuatu kelas (Christianto, Andjarwirawan and Tjondrowiguno, 2020) berikut merupakan persamaan teorema bayes :

$$p(X|H) = \frac{P(X|H)P(H)}{P(X)} \text{ (Rumus 2.1)}$$

Keterangan:

X : Data dengan kelas yang belum diketahui

H : Hipotesis data X adalah tidak dikenal

$P(H | X)$: Probabilitas hipotesis H didasarkan pada kondisi X (*posterior probability*)

$P(H)$: Probabilitas hipotesis H (*prior probability*)

$P(X | H)$: Probabilitas hipotesis X didasarkan pada kondisi H

$P(X)$: probabilitas X

Metode multinomial *Naive Bayes Terkenal* dengan kesederhanaan serta efektifitasnya pada memproses klasifikasi kategori dokumen teks menggunakan cara membuat probabilitas dari adonan kata-kata dan kategori dibanding menggunakan metode *Naive Bayes*.

2.2.9. Accuracy

Accuracy adalah salah satu metrik evaluasi yang umum digunakan dalam analisis performa model machine learning. Akurasi mengukur sejauh mana prediksi model sesuai dengan hasil sebenarnya, dalam bentuk rasio antara jumlah prediksi yang benar dengan total jumlah data. Formula akurasi dapat dirumuskan sebagai:

$$Accuracy = \frac{Jumlah\ Prediksi\ Benar}{Jumlah\ Total\ data} \times 100\% \text{ (Rumus 2.2)}$$

Hidayat dan Taufik (2021), *Accuracy* adalah metrik yang sederhana tetapi sangat efektif untuk digunakan ketika data seimbang. Namun, jika data tidak seimbang .misalnya, jumlah data positif jauh lebih banyak daripada data negatif.