

BAB II

TINJAUAN PUSTAKA

2.1. Teori

Teori merupakan kumpulan pernyataan yang menjelaskan suatu fenomena melalui hubungan antara konsep-konsep yang saling terkait dalam rangkaian sebab-akibat. Dalam konteks penelitian sosial, teori menjadi dasar untuk memahami realitas empiris. Fungsinya meliputi penjelasan, prediksi, serta memberikan kerangka bagi interpretasi data. Teori juga dapat digunakan untuk meramalkan kemungkinan gejala yang akan muncul, menyediakan struktur untuk menghubungkan studi dengan yang lain, dan memberikan interpretasi yang lebih luas pada hasil penelitian. Selain itu, teori membantu dalam pembuatan hipotesis, penyusunan instrumen, dan mengklarifikasi persoalan yang dihadapi dalam penelitian sosial (Marliyah, 2021). Penelitian ini merujuk pada sejumlah teori dasar yang menjadi landasan bagi penulis dalam pelaksanaannya. Berikut adalah penjelasan mengenai teori-teori yang akan diterapkan dalam penelitian.

2.1.1. Teori Umum

Beberapa teori umum yang memiliki relevansi dengan isu penelitian diuraikan sebagai berikut.

2.1.1.1 Data, Informasi, dan Big Data

Data adalah representasi dari peristiwa yang terjadi dan merupakan bentuk yang belum terstruktur yang memerlukan pengolahan lebih lanjut melalui model

untuk menghasilkan informasi. (Nawassyarif et al., 2020). Data adalah serangkaian kejadian yang diambil dari fakta. Data dapat berupa angka, huruf, atau simbol tertentu, atau kombinasi dari ketiganya. Namun, data tidak dapat langsung diterima sehingga perlu diproses untuk menghasilkan informasi yang berguna untuk pembuatan keputusan. Hal ini menunjukkan bahwa proses data menjadi informasi adalah langkah penting dalam memahami dan menggunakan data dengan lebih bermakna (Candra & Wulandari, 2021). Informasi adalah hasil dari pemrosesan data yang memberikan nilai penting bagi penerima dan berfungsi sebagai dasar untuk pengambilan keputusan yang memiliki dampak langsung pada saat itu atau tidak langsung di masa mendatang (Yusman, 2023).

Big data didefinisikan sebagai informasi yang digunakan untuk menghasilkan pemahaman dan membuat keputusan, yang ditandai dengan volume, kecepatan, dan variasi yang tinggi (N. S. Putra et al., 2023). Big Data merujuk pada kumpulan informasi digital yang disimpan atau dikirim melalui sistem komputer. Istilah ini juga mencakup jumlah data yang sangat besar dan beragam sehingga aplikasi yang ada tidak mampu mengolahnya dengan efisien. Ini juga mencakup teknologi dan alat yang digunakan dalam konteks Big Data (Nainggolan & Nasution, 2023).

2.1.1.2 Data Mining

Data mining merupakan ekstraksi informasi berharga dari data yang besar, kompleks, dan bervariasi. Tujuannya adalah untuk menemukan pola tersembunyi, korelasi, dan wawasan yang sulit ditemukan melalui analisis biasa (Pratama & Jaman, 2023). Data Mining menggunakan berbagai teknik seperti statistik,

matematika, kecerdasan buatan, dan pembelajaran mesin untuk mengekstrak informasi berharga dari basis data besar. Praktik ini melibatkan serangkaian proses untuk mengungkap pengetahuan baru dari data (Sebastian & Juliane, 2023). Ini adalah bagian dari proses *Knowledge Discovery in Databases* (KDD) yang mencakup pemilihan data, *preprocessing*, transformasi, *data mining*, dan evaluasi hasil (Zai, 2022).

Tahapan dalam Data Mining meliputi Pembersihan data (*data cleaning*), Integrasi data (*data integration*), Seleksi data (*data selection*), Transformasi data (*data transformation*), Proses *mining*, Evaluasi pola yang ditemukan untuk menemukan informasi dan pengetahuan yang penting, dan terakhir Proses Pola (*pattern evaluation*) (Triandini et al., 2021). Dalam Data Mining, terdapat tiga aspek: Asosiasi, Klasifikasi, dan *clustering*. Asosiasi adalah proses yang digunakan untuk menemukan hubungan antara nilai atribut dalam sekumpulan data. Klasifikasi adalah teknik untuk memprediksi kelas atau properti dari setiap instansi data. *Clustering*, di sisi lain, melibatkan pengelompokan data ke dalam kelompok yang serupa tanpa memperhatikan kelas data tertentu, sesuai dengan topik yang dibahas (Wijaya & Dwiasnati, 2020).

2.1.1.3 Database

Database adalah kumpulan data yang disimpan secara sistematis melalui sistem data, menggunakan berbagai format seperti teks, gambar, video, dan file. Ini memungkinkan penyimpanan dan manajemen data yang cepat dan efisien, serta memberikan akses mudah untuk informasi yang akurat tanpa manipulasi sistem.

Tujuannya adalah menyederhanakan proses bisnis, menyimpan data secara permanen, dan memfasilitasi pengorganisasian sistem dalam perusahaan atau organisasi (Pulungan et al., 2022).

Dalam jurnalnya (Syahputri & Nasution, 2023) menjelaskan bahwa *Database* atau Basis data adalah kumpulan data terstruktur yang saling terhubung, disimpan dalam media penyimpanan komputer, dan dikelola melalui perangkat lunak khusus. Fungsinya adalah untuk menyimpan, mengelola, dan mengorganisir data dengan tujuan memberikan akses yang terstruktur, aman, dan efisien terhadap informasi. Sedangkan Sistem basis data merujuk pada struktur yang melibatkan pengguna yang menggunakan basis data bersama-sama, personel yang merancang dan mengelola basis data, teknik desain dan pengelolaan basis data, serta infrastruktur komputer yang mendukungnya (Ritonga et al., 2023).

2.1.1.4 Text Mining

Apa yang dimaksud dengan *Text Mining*? *Text mining* adalah proses memeriksa sejumlah besar dokumen untuk menemukan informasi baru atau menjawab pertanyaan penelitian. Ini melibatkan identifikasi fakta, hubungan, dan pernyataan penting dalam dokumen. Setelah diekstraksi, informasi tersebut diubah menjadi bentuk terstruktur yang dapat dianalisis atau disajikan dengan berbagai cara, seperti tabel HTML, peta pikiran, atau bagan. Metode yang digunakan termasuk Pemrosesan Bahasa Alami (NLP). Data terstruktur yang dihasilkan dapat diintegrasikan ke dalam *database* atau gudang data untuk analisis lebih lanjut, termasuk analisis preskriptif atau prediktif (Tojimamatov et al., 2023).

Text mining merupakan teknik yang digunakan untuk menangani masalah klasifikasi, pengelompokan, ekstraksi informasi, dan pencarian informasi. Ini merupakan langkah dalam menghasilkan informasi agar pengguna dapat berinteraksi dengan koleksi dokumen dari waktu ke waktu menggunakan alat analisis. Proses utamanya adalah mengidentifikasi kata-kata kunci yang mewakili isi dokumen dan kemudian menganalisis hubungan antar dokumen tersebut (Afdal & Elita, 2022).

Definisi lain dari *Text Mining*, adalah proses pencarian dan penggalian pengetahuan dari *database* dalam bentuk teks, juga dikenal sebagai *knowledge discovery in textual database* (KDT). Ini melibatkan analisis kumpulan teks yang tidak terstruktur untuk menemukan pola atau pengetahuan baru. Meskipun memiliki nilai komersial yang tinggi karena sebagian besar informasi perusahaan disimpan dalam bentuk teks, proses ini kompleks karena pola teks yang tidak terstruktur (Firdaus & Firdaus, 2021).

2.1.1.5 Analisis Sentimen

Analisis sentimen adalah cabang dari *data mining* yang menitikberatkan pada penelitian tentang pendapat, perasaan, evaluasi, sikap, dan sentimen terhadap berbagai hal seperti produk, layanan, peristiwa, atau masalah. Selain itu, proses analisis sentimen dapat membantu dalam memahami, mengekstrak, dan mengolah teks secara otomatis untuk mendapatkan informasi yang berharga (Alifia et al., 2023). Dalam konteks penggunaan internet, analisis sentimen juga memainkan peran penting karena melibatkan identifikasi pendapat individu yang diekspresikan

dalam tulisan. Pengguna internet sering kali menggunakan platform seperti media sosial untuk mengekspresikan pendapat mereka tentang berbagai topik, yang bisa mencakup perasaan positif, netral, atau negatif (Sarimole & Kudrat, 2024).

Langkah-langkah analisis sentimen meliputi pengumpulan data teks dari berbagai sumber, pembersihan data dari elemen tidak relevan, dan pemecahan menjadi unit terkecil. Setiap teks diberi label sentimen sesuai konteksnya. Kemudian, menggunakan metode analisis sentimen seperti Naive Bayes, teks diidentifikasi berdasarkan sentimennya. Hasil analisis dievaluasi untuk memahami tren sentimen dan menghasilkan wawasan yang berguna dalam konteks yang bersangkutan. Indikator atau langkah-langkah dalam analisis sentimen dijelaskan sebagai berikut:

1. Data Collection

Pengumpulan data dilakukan untuk mengambil data agar bisa digunakan dalam penelitian. Salah satu metode pengumpulan data adalah *crawling*. *Crawling* merupakan teknik yang secara otomatis mengambil data dari suatu situs web berdasarkan kata kunci yang ditentukan pengguna. Prosesnya melibatkan kunjungan ke setiap dokumen di situs web tersebut, mencatat semua URL, dan mengumpulkan data yang relevan sesuai dengan permintaan (Dikiyanti et al., 2021).

2. Data Preprocessing

Data yang diperoleh melalui proses *crawling* tidak bisa langsung diklasifikasikan, oleh karena itu, perlu dilakukan *preprocessing* data

agar menjadi lebih terstruktur dan bersih sehingga dapat diklasifikasikan. *Preprocessing* melibatkan langkah-langkah seperti *case folding*, *tokenisasi*, *filtering* dan *stemming* (Silalahi, 2023).

3. Data Modelling

Tahap ketiga dalam *text mining* yang bertujuan untuk menyiapkan data teks agar siap untuk diproses menggunakan teknik *data mining*. Langkah ini melibatkan penyesuaian data dengan algoritma yang akan digunakan dalam proses Data Mining (Budiarti et al., 2023).

4. Data Visualisation

Visualisasi data adalah cara untuk membantu orang memahami signifikansi data dengan menampilkan informasi secara visual. Penggunaan data memungkinkan organisasi membuat keputusan yang lebih cerdas dan mendekati waktu nyata serta melihat tren dari penyimpanan data besar (Ariandi & Puteri, 2022).

2.1.2. Teori Khusus

Beberapa teori khusus yang memiliki relevansi dengan isu penelitian diuraikan sebagai berikut.

2.1.2.1 Platform Twitter

Twitter mulai beroperasi pada bulan Maret 2006, didirikan oleh Jack Dorsey, dan situs jejaring sosialnya diluncurkan pada bulan Juli. Sejak itu, Twitter telah menjadi salah satu dari sepuluh situs yang paling sering dikunjungi di Internet dan

dikenal sebagai platform pesan singkat terkemuka di dunia maya (Djamiludin et al., 2022). Twitter merupakan platform *microblogging* yang membatasi penggunaannya hingga 280 karakter, dengan fitur unggulannya yaitu menampilkan *Trending Topic* yang mempermudah pengguna dalam memantau topik-topik populer baik dalam maupun luar negeri, menjadikan mereka merasa selalu terkini. Selain itu, Twitter dianggap sebagai media sosial yang ramah karena memfasilitasi interaksi dan sosialisasi tanpa membatasi pertemanan (Girmanfa & Susilo, 2022). Twitter adalah sebuah platform media sosial yang memiliki potensi besar di Indonesia, di mana pengguna dapat dengan cepat berpendapat atau mengkritik suatu hal secara langsung (*real-time*) (Raisa & Riza, 2023).

2.1.2.2 *Electronic Traffic Law Enforcement*

Ketidakpatuhan terhadap aturan lalu lintas tidak boleh diabaikan karena sebagian besar kecelakaan disebabkan oleh pengemudi yang tidak patuh terhadap peraturan lalu lintas. Untuk mengurangi masalah tersebut, pemerintah mulai menerapkan ETLE (*electronic traffic law enforcement*), sebuah aplikasi tilang elektronik yang merupakan inovasi dalam penegakan hukum lalu lintas di era digital. Penerapan ETLE merupakan langkah yang baik, dan urgensi dari ETLE saat ini adalah menciptakan transparansi antara kepolisian dan masyarakat untuk menghindari praktik pungutan liar (pungli) (Nurmansyah et al., 2023).

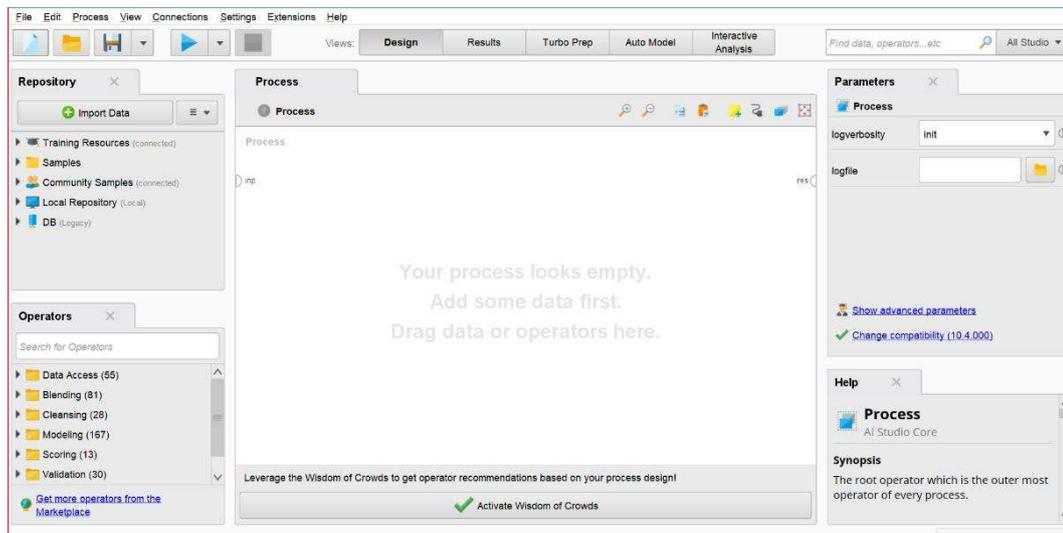
Penerapan sistem tilang elektronik atau yang dikenal dengan *electronic traffic law enforcement* (ETLE) di beberapa wilayah merupakan upaya pemerintah dan instansi terkait untuk mengurangi kemacetan lalu lintas yang disebabkan oleh

pelanggaran peraturan. Program ini diharapkan dapat mengurangi jumlah pelanggaran lalu lintas di jalan. Meskipun begitu, sistem tilang elektronik belum diterapkan secara nasional di Indonesia, hanya berlaku di sejumlah rute kota. Agar penegakan hukum berjalan dengan adil, Kepolisian Negara Republik Indonesia harus memerhatikan prinsip praduga tak bersalah, keseimbangan, kesesuaian, dan legalitas (Dahlan et al., 2023).

2.1.2.3 Rapid Miner

RapidMiner merupakan perangkat lunak *open-source* yang dapat diakses oleh siapa pun. Fungsinya sebagai alat analisis data yang menggunakan berbagai metode, termasuk metode deskriptif dan prediktif. Dalam penelitian ini, RapidMiner digunakan untuk memproses data dan melakukan klasifikasi menggunakan algoritma Naïve Bayes (Muflih et al., 2023).

Aplikasi RapidMiner memiliki fungsi untuk menganalisis dan mengimplementasikan metode *data mining* pada data set yang menjadi dasar penelitian ini (Andry et al., 2023). RapidMiner dirancang untuk mempelajari konsep *data mining*, menganalisis data, melakukan pemrosesan data, dan mengembangkan model prediktif. Perangkat lunak ini ditampilkan dengan antarmuka yang ramah pengguna, sehingga mempermudah pengguna dalam mengoperasikannya (Hartono & Widiatoro, 2024).



Gambar 2. 1 Tampilan Antarmuka Software RapidMiner
(Sumber: AI Studio RapidMiner 2024)

2.1.2.4 Python

Bahasa pemrograman Python, yang kini sangat populer, awalnya dikembangkan oleh Guido Van Rossum di Stichting Mathematisch Centrum (CWI), Amsterdam pada tahun 1991. Inspirasi pengembangan Python berasal dari bahasa pemrograman ABC yang sedang berkembang pada masa itu. Salah satu perbedaan utama Python dengan bahasa pemrograman lainnya adalah proses pengembangannya yang melibatkan jutaan *programmer*, peneliti, dan pengguna dari berbagai latar belakang, bukan hanya dari bidang IT, karena Python bersifat *open source* (Rahman et al., 2023).

Python adalah bahasa pemrograman berbasis objek yang interaktif dan memiliki struktur data tingkat tinggi. Dikenal sebagai bahasa interpretatif dengan banyak fungsi, Python dirancang untuk kejelasan dan kemudahan pemahaman kode. Fleksibilitasnya memungkinkan penggunaan untuk aplikasi mandiri atau sebagai pemrograman skrip, dengan fokus pada efisiensi waktu, kemudahan

pengembangan, dan kompatibilitas sistem (Triono et al., 2023). Python dikenal dengan berbagai kerangka kerjanya yang beragam, menyediakan banyak potongan kode yang sudah ada yang memungkinkan pengembang untuk meningkatkan kualitas proyek mereka. Hal yang serupa berlaku untuk penambahan data (Runimeirati et al., 2023).

2.1.2.5 Naïve Bayes

Naïve Bayes merupakan algoritma yang populer dalam *data mining* karena kemudahannya dalam penggunaan, pemrosesan yang cepat, dan implementasi yang sederhana. Algoritma ini efektif dalam mengklasifikasi variabel dengan menggunakan probabilitas dan statistik. Dengan model probabilitasnya, Naïve Bayes Classifier dapat dilatih secara efektif dalam *supervised learning*. Keunggulan lainnya adalah bahwa algoritma ini tidak membutuhkan jumlah data *training* yang besar (Khotimah & Utami, 2022).

Naïve Bayes Classifier adalah suatu teknik klasifikasi yang menggunakan probabilitas sederhana dengan menerapkan teorema Bayes dan mengasumsikan tingkat tidak ketergantungan yang tinggi. Penelitian ini berfokus pada ukuran *dataset* yang memungkinkan metode ini untuk memberikan performa yang cepat dan akurat dalam proses klasifikasi. Naïve Bayes Classifier hanya membutuhkan *dataset* pelatihan yang relatif kecil untuk menentukan estimasi parameter yang dibutuhkan dalam klasifikasi. Selama tahap klasifikasi, kategori nilai ditentukan berdasarkan kemunculan term dalam data, menggunakan persamaan yang telah ditentukan (Singgalen, 2022).

Dalam menerapkan Naïve Bayes, terdapat beberapa langkah yang perlu dilakukan oleh peneliti, antara lain:

1. Pengumpulan data, di mana data dikumpulkan dari sumber yang relevan atau yang akan diteliti.
2. *Preprocessing* data, tahap ini melibatkan membersihkan data mentah agar menjadi lebih mudah dipahami.
3. Pelabelan data untuk mengetahui sentimen dalam *dataset* kalimat. Label sentimen dapat dibagi menjadi dua jenis: positif dan negatif.
4. K-Fold Cross Validation merupakan metode pengujian di mana data *training* dan data testing dibagi secara merata menjadi K subset. Data *training* digunakan untuk membangun model, sementara data testing untuk memvalidasi model. Metode ini meningkatkan keandalan pengujian karena seluruh data dimanfaatkan.

Algoritma Naïve Bayes, adalah sebuah pendekatan klasifikasi berbasis statistik dan probabilitas. Kelebihan Naïve Bayes adalah kemampuannya untuk memberikan estimasi dengan menggunakan jumlah data *training* yang relatif kecil. Teorema Bayes digunakan untuk menghitung probabilitas dari dua kejadian terpisah, dirumuskan (Erfina & Lestari, 2023):

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \quad \text{Rumus 2. 1 Rumus Naive Bayes}$$

Keterangan dari rumus di atas:

P : Probabilitas

X : Data dengan *class* yang belum ditentukan

H : Hipotesis data X merupakan suatu *class* spesifik

$P(H|X)$: Probabilitas hipotesis H berdasarkan kondisi X

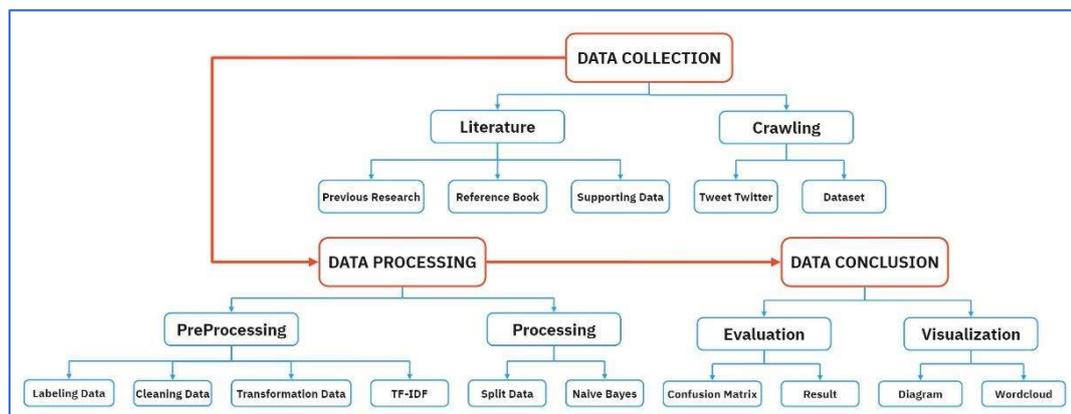
$P(X|H)$: Probabilitas X berdasarkan kondisi hipotesis H

$P(H)$: Probabilitas hipotesis H

$P(X)$: Probabilitas X

2.2. Kerangka Pemikiran

Kerangka pemikiran ini mencakup serangkaian proses berpikir yang diorganisir sebagai panduan untuk menangani masalah penelitian dengan cara yang logis dan teratur. Dalam konteks penelitian ini, kerangka pemikiran digambarkan sebagai berikut.



Gambar 2. 2 Kerangka Pemikiran
(Sumber: Peneliti 2024)

Keterangan dari kerangka pemikiran:

1. *Data Collection*

Data Collection atau pengumpulan data, tahapan ini dilakukan literatur studi dengan mengumpulkan jurnal penelitian terdahulu sebagai bahan rujukan penelitian. Selain jurnal, buku dan dokumen lain juga dikumpulkan sebagai referensi peneliti. Untuk data yang dijadikan bahan penelitian bersumber dari

database Twitter yang diperoleh dengan cara *crawling* data dari *database* Twitter.

2. **Data Processing**

Pengolahan data dilakukan dengan 2 tahap yaitu *preprocessing* data dan *processing* data. Data yang diperoleh dari proses *crawling* data akan dilakukan labeling data sentiment secara manual. Kemudian *preprocessing* data dimulai dari proses cleaning data dan diakhiri dengan transformasi data (case folding, tokenize, stopword removal, filter by length). Setelah itu data dibagi dua menjadi data latih dan data uji yang digunakan untuk *processing* data menggunakan algoritma Naive Bayes Classifier.

3. **Data Conclusion**

Setelah data diolah menggunakan algoritma Naive Bayes, evaluasi data (confusion matrix) perlu dilakukan untuk menentukan hasil akurasi data sebelum penarikan kesimpulan. Data yang diperoleh dari penarikan hasil uji divisualisasikan menggunakan diagram atau grafik dan *wordcloud*.

2.3. **Hipotesis Penelitian**

Penelitian ini akan memanfaatkan Algoritma Naive Bayes untuk mengategorikan *tweet* ke dalam kategori sentimen positif atau negatif. Analisis ini bertujuan untuk memperoleh pemahaman mengenai pandangan masyarakat terhadap penerapan ETLE dan menyediakan wawasan tentang penerimaan masyarakat terhadap sistem tersebut. Hasil analisis dapat digunakan untuk

meningkatkan strategi komunikasi dan kebijakan terkait ETLE di masa mendatang.

Hipotesis yang dapat diajukan adalah sebagai berikut:

1. Penerapan *electronic traffic law enforcement* (ETLE) berdampak pada sentimen pengguna Twitter, mayoritas *tweet* memiliki sifat menunjukkan kecenderungan sentimen positif
2. Penerapan *electronic traffic law enforcement* (ETLE) memiliki dampak pada sentimen pengguna Twitter, dengan mayoritas *tweet* menunjukkan sentimen negatif.

2.4. Penelitian Terdahulu

Penelitian terdahulu merupakan cara bagi peneliti untuk membandingkan dan mengambil inspirasi baru untuk penelitian berikutnya. Selain itu, studi sebelumnya membantu dalam menempatkan penelitian secara kontekstual dan menunjukkan keorisinalan penelitian tersebut. Berikut ini cakupan berbagai hasil penelitian terdahulu yang relevan dengan topik yang akan diteliti, disertai dengan rangkumannya yang telah dipublikasikan.

1. Shella Ayudya Aryani Armand, Moch Hafid T, dan M. Rafi Muttaqin

Judul Penelitian “Analisis Sentimen Sistem E-Tilang Pada Platform Twitter Menggunakan Metode Naive Bayes” (Armand et al., 2023). Penelitian ini menganalisis pendapat masyarakat terhadap sistem e-Tilang dengan tujuan mengelompokkan komentar-komentar tersebut menjadi sentimen positif, netral, dan negatif. Algoritma Naive Bayes dipilih karena memiliki probabilitas tertinggi

untuk mengklasifikasikan data. Proses pelabelan data dilakukan menggunakan Bahasa Pemrograman Python dan Lexicon Based dengan menggunakan Google Colab. Data melalui empat tahapan *preprocessing*: transformasi, *tokenizing*, *filtering*, dan *stemming*. Setelah *preprocessing*, dilakukan pembobotan TF-IDF untuk mengetahui frekuensi kata dalam setiap komentar. Hasil klasifikasi menggunakan Naive Bayes dievaluasi dengan *Confusion Matrix* di Google Colab. Evaluasi menunjukkan presisi positif 43%, presisi negatif 38%, dan presisi netral 85%, serta recall positif 43%, recall negatif 14%, dan recall netral 95%, dengan akurasi keseluruhan mencapai 80%. Dari analisis ini, disimpulkan bahwa tanggapan masyarakat terhadap sistem e-Tilang cenderung netral.

2. Rakhmi Khalida dan Siti Setiawati

Penelitian “Analisis Sentimen Sistem E-Tilang Menggunakan Algoritma Naive Bayes Dengan Optimalisasi Information Gain” (Khalida & Setiawati, 2020). Kehadiran e-Tilang juga dianggap sebagai solusi untuk mengurangi praktik pungutan liar oleh penegak hukum, menyelaraskan terminologi yang digunakan, dan meningkatkan transparansi dalam penggunaan denda. Penelitian ini mengadopsi analisis sentimen terhadap sistem e-Tilang, atau yang dikenal sebagai *opinion mining*, untuk mengelompokkan beragam komentar masyarakat menjadi positif, negatif, atau netral. Media sosial Twitter dipilih sebagai platform untuk mengungkapkan pendapat karena kemudahan penggunaan, topik yang selalu diperbarui, dan keterbukaan dalam mengakses *tweet*. Pendapat dari Twitter dikumpulkan, kemudian melalui tahapan *preprocessing*, dilanjutkan dengan seleksi

fitur menggunakan information gain untuk mengurangi gangguan yang disebabkan oleh label yang tidak relevan. Langkah berikutnya adalah klasifikasi sentimen menggunakan algoritma Naïve Bayes, diikuti oleh penentuan polaritas sentimen. Penelitian ini mencapai akurasi sebesar 41,82%, presisi sebesar 50,51%, dan recall sebesar 45,45%.

3. Hendriansyah dan Wahyu Pramusinto

Judul Penelitian “Analisis Sentimen Masyarakat Terhadap Tilang Elektronik Pada Twitter Dengan Algoritma K-Nearest Neighbor” (Hendriansyah & Pramusinto, 2023). Penelitian ini bertujuan untuk mengeksplorasi pandangan masyarakat Indonesia terhadap tilang elektronik selama periode 20 Maret hingga 29 Mei 2023. Fokus penelitian ini adalah menilai akurasi algoritma K-Nearest Neighbor dalam menganalisis sentimen dan mencari cara untuk menganalisis sentimen berdasarkan pendapat masyarakat Indonesia di media sosial Twitter. K-Nearest Neighbor (K-NN) merupakan salah satu algoritma klasifikasi yang digunakan, yang mengoperasikan prinsip mencari jarak terdekat antara data yang dievaluasi dan tetangga terdekatnya. Metode yang digunakan melibatkan ekstraksi fitur dengan CountVectorizer dan penggunaan algoritma klasifikasi K-Nearest Neighbor. Penelitian ini menggunakan library Tweet-harvest di Google Colabs dengan bahasa pemrograman Python. Data dikumpulkan dari Twitter dengan mencari tweet dalam bahasa Indonesia yang berhubungan dengan "tilang elektronik". Analisis dari 509 tweet menunjukkan bahwa 82,71% komentar bersifat positif dan 17,29% bersifat negatif antara tanggal 20 Maret dan 29 Mei 2023. Hasil

uji menunjukkan kinerja terbaik diperoleh dengan nilai $K = 5$, dengan akurasi 68,85% dan ingatan (recall) sebesar 77%.

4. Syarif Hidayatulloh, Wazir Putra, dan Dimas Febriawan

Penelitian dengan judul “Analisis Sentimen Ulasan Aplikasi Digital Korlantas POLRI Menggunakan Naïve Bayes pada Google Play Store” (Hidayatulloh et al., 2024). Penelitian ini menggunakan algoritma Naïve Bayes sebagai metode klasifikasi dengan dataset sebanyak 1500 ulasan pengguna aplikasi Digital Korlantas POLRI, yang terdiri dari 536 data sentimen positif dan 964 data sentimen negatif. Data tersebut dibagi dengan rasio 80:20, menghasilkan 1200 data latih dan 300 data uji. Hasil akhir dari penelitian ini menunjukkan nilai akurasi, presisi, dan recall untuk mengevaluasi kinerja algoritma Naïve Bayes dalam klasifikasi. Dari evaluasi terhadap 300 data uji, diperoleh nilai akurasi sebesar 70,33%, presisi 56,34%, dan recall 74,77%. Hasil evaluasi ini menunjukkan bahwa algoritma Naïve Bayes berhasil dalam melakukan proses klasifikasi.

5. Ni Luh Putu Merawati, Ahmad Zuli Amrullah, dan Ismarmiaty

Penelitian dengan judul “Analisis Sentimen dan Pemodelan Topik Pariwisata Lombok Menggunakan Algoritma Naive Bayes dan Latent Dirichlet Allocation” (Merawati et al., 2021). Tujuan penelitian ini adalah untuk mengelompokkan opini-opini wisatawan menjadi dua kategori, yaitu positif dan negatif, serta melakukan analisis topik pada kedua kategori tersebut. Analisis topik bertujuan untuk mengidentifikasi topik yang paling umum dibahas dalam setiap kategori. Langkah-

langkah penelitian mencakup pengumpulan data, pembersihan data, transformasi data, klasifikasi data menggunakan metode Naive Bayes, dan penggunaan metode Latent Dirichlet Allocation (LDA) untuk analisis topik. Hasil pengujian model menggunakan algoritma Naive Bayes menunjukkan tingkat akurasi, *accuracy*, *recall*, dan *spesificity* masing-masing sebesar 92%, 100%, 83,84%, dan 100%. Hasil analisis topik menggunakan metode LDA untuk kategori positif dan negatif dapat dilihat dari nilai koherensi, di mana semakin tinggi nilai koherensi suatu topik, semakin mudah topik tersebut diinterpretasikan oleh manusia. Nilai koherensi tertinggi untuk kategori positif terdapat pada topik ke-8 dengan nilai 0,613, sedangkan untuk kategori negatif terdapat pada topik ke-12 dengan nilai 0,528. Penggunaan algoritma Naive Bayes dan LDA dinilai efektif untuk analisis sentimen dan analisis topik pada konteks pariwisata di Lombok.

6. Wiga Maulana Baihaqi dan Arif Munandar

Judul penelitian “Sentiment Analysis of Student Comment on the College Performance Evaluation Questionnaire Using Naïve Bayes and IndoBERT” (Baihaqi & Munandar, 2023). Penelitian ini bertujuan untuk menganalisis penilaian kinerja perguruan tinggi oleh mahasiswa berdasarkan sentimen positif dan negatif, dengan maksud untuk mengevaluasi tingkat kepuasan mahasiswa terhadap seluruh aspek dan operasi universitas. Metode penelitian ini menggunakan algoritma Naïve Bayes dan model IndoBERT untuk membangun model klasifikasi berdasarkan data kuesioner, meliputi proses pengumpulan data, *preprocessing* data, ekstraksi fitur, pemodelan, dan evaluasi. Hasil dari model IndoBERT menunjukkan kinerja terbaik

dengan tingkat akurasi mencapai 85%. Model IndoBERT berhasil mengidentifikasi sentimen dalam teks secara efektif, dengan mampu membedakan antara komentar positif dan negatif mengenai kinerja universitas.

7. Rini, Retno, Syifa, Diki, dan Dedi

Penelitian ini berjudul “Analisis Sentimen Persepsi Masyarakat Terhadap Penggunaan Aplikasi My Pertamina Pada Media Sosial Twitter Menggunakan Metode Naïve Bayes Classifier” (Maria et al., 2023). Manfaat dari penelitian ini adalah untuk memahami tren komentar atau tweet pengguna Twitter terkait dengan kebijakan pembelian *Pertalite* melalui Aplikasi MyPertamina dengan menganalisis sentimen. Oleh karena itu, diperlukan suatu metode untuk mengklasifikasikan komentar publik tersebut menggunakan analisis sentimen di platform media sosial Twitter. Algoritma yang digunakan dalam analisis sentimen ini adalah metode Naïve Bayes Classifier (NBC). Analisis sentimen dilakukan dengan mengumpulkan 1000 *tweet* untuk membentuk *dataset*, di mana sentimen akan diberi label berdasarkan dua kategori, yaitu keluhan (*complaint*) dan bukan keluhan (*non-complaint*). Tahapan dalam analisis sentimen ini mencakup *preprocessing* data, pengolahan data, klasifikasi, dan evaluasi. Dari hasil analisis, ditemukan tingkat akurasi sebesar 82.96%, presisi sebesar 81.17%, *recall* sebesar 86.07%, dan nilai AUC sebesar 0.906. Berdasarkan hasil presentasi tersebut, dapat disimpulkan bahwa metode ini telah berhasil mencapai hasil optimal dalam mengklasifikasikan komentar publik.