

BAB I

PENDAHULUAN

1.1 Latar Belakang

Twitter merupakan salah platform media sosial yang berbasis *microblogging*. *Microblogging* adalah jenis media sosial yang memungkinkan pengguna untuk menulis dan membagikan aktivitas serta pendapat mereka (Bagus et al., 2022). dimana pengguna dapat mengirimkan pesan yang disebut tweet untuk memberikan dan menerima informasi di lini masa. Pada tahun 2024, media sosial Twitter memiliki sekitar 335,7 juta pengguna di seluruh dunia (Jack Shepherd, 2025). Angka pengguna yang sangat besar ini menyebabkan cuitan atau pendapat yang dibagikan sering kali mencakup berbagai topik, seperti hiburan, pendidikan, ekonomi, politik, dan banyak lagi. Salah satu topik yang telah menjadi tren dalam beberapa tahun terakhir adalah mengenai *BRICS*, khususnya terkait dengan bergabungnya Indonesia dalam kerjasama ekonomi *BRICS*.

BRICS adalah organisasi yang dibentuk untuk mewadahi negara-negara berkembang, yang pertama kali diperkenalkan oleh ekonom Amerika Serikat, Jim O'Neill, pada tahun 2001. Akronim *BRICS* terdiri dari Brasil, Rusia, India, China, dan Afrika Selatan. Para pemimpin *BRICS* berkomitmen untuk menjadi kekuatan perubahan dan suara bagi negara-negara berkembang. Saat ini, *BRICS* secara kolektif mewakili hampir 20% perekonomian global, dengan karakteristik populasi besar, wilayah luas, dan pertumbuhan ekonomi di atas rata-rata. Empat negara dalam BRIC mencakup lebih dari 25% luas daratan dunia, 40% populasi global, dan sekitar 18% dari total perekonomian dunia (Alfiyah et al., 2024). *BRICS* bertujuan untuk mendukung negara-negara berkembang dalam berbagai aspek, termasuk ekonomi, militer, teknologi, dan hubungan diplomatik antar negara. Selama masa kepemimpinan Presiden Jokowi, terdapat wacana mengenai potensi kerja sama Indonesia dengan *BRICS*, khususnya dalam konteks penguatan ekonomi global dan peluang investasi. Berita-berita tersebut tidak hanya menarik perhatian

media, tetapi juga memicu beragam pandangan dari masyarakat global yang aktif menyampaikan pendapatnya di media sosial, khususnya di Twitter atau X. Platform ini berfungsi sebagai ruang diskusi terbuka bagi pengguna untuk membahas manfaat dan tantangan yang mungkin dihadapi Indonesia jika bergabung dengan *BRICS*. Sebagian pengguna menyambut baik langkah ini dengan harapan dapat memperkuat posisi Indonesia di tingkat global, sementara yang lain mengemukakan kritik terkait potensi risiko yang mungkin timbul. Pendapat dari masyarakat yang menghasilkan berbagai pandangan ini dapat dimanfaatkan sebagai data untuk menguji algoritma analisis sentimen, yaitu *Multinomial Naïve Bayes*. Algoritma *Multinomial Naïve Bayes* adalah salah satu metode pembelajaran probabilistik yang berlandaskan pada teorema Bayes dan sering diterapkan dalam bidang *Natural Language Processing* (NLP). Metode ini beroperasi berdasarkan konsep *Term Frequency*, yang mengacu pada frekuensi kemunculan suatu kata dalam sebuah dokumen (Yuyun et al., 2021). Algoritma ini dapat mengklasifikasikan sentimen ke dalam tiga kategori, yaitu sentimen positif, netral, dan negatif. Melalui analisis sentimen tersebut, diharapkan dapat diperoleh gambaran yang lebih terstruktur mengenai opini publik terhadap potensi Indonesia bergabung dengan *BRICS*. Selain itu, hasil analisis ini dapat menjadi dasar untuk memahami persepsi masyarakat secara lebih mendalam, sehingga dapat memberikan kontribusi pada pengambilan keputusan strategis terkait isu ini. Pada penelitian yang dilakukan oleh (Bessou & Aberkane, 2019) menunjukkan bahwa algoritma *Multinomial Naïve Bayes* memiliki akurasi tertinggi dalam memprediksi polaritas topik dibandingkan dengan metode machine learning lainnya, seperti *Support Vector Machines*, *Random Forest*, dan *Logistic Regression*. Hal ini dibuktikan dalam studi yang menggunakan fitur *uni-gram* dan *bi-gram* dengan representasi *count vector*, yang berhasil mencapai akurasi 85,57% pada klasifikasi dua kelas dan 65,64% pada klasifikasi tiga kelas. Oleh karena itu, penelitian ini memilih menggunakan model algoritma Multinomial Naïve Bayes sebagai metode untuk analisis sentimen guna memperoleh hasil yang akurat dan efisien.

Dalam penelitian ini, akan diterapkan beberapa teknik untuk menganalisis sentimen terkait topik *BRICS*. Pertama, analisis akan dilakukan dengan menggunakan algoritma *Multinomial Naïve Bayes* secara tunggal. Kedua, penelitian ini akan menerapkan teknik *Laplace Estimator/Smoothing*. *Laplace smoothing* merupakan salah satu metode *Smoothing* yang dapat diterapkan pada *Naïve Bayesian Classifier*. Konsep dasar dari metode ini cukup sederhana, yaitu menghindari munculnya nilai nol dalam model probabilitas. Ketiga, akan digunakan teknik *SMOTE* untuk menangani masalah ketidakseimbangan data. Terakhir, analisis akan mengintegrasikan kedua teknik, yaitu *Laplace estimator/smoothing* dan *SMOTE*, ke dalam algoritma *Multinomial Naïve Bayes*.

Penelitian serupa dilakukan oleh (Gaa Ngilo, 2023) yang menganalisis sentimen opini pengguna Twitter terhadap aplikasi Bibit. Tujuan dari penelitian ini adalah untuk melakukan analisis sentimen terhadap opini pengguna Twitter mengenai aplikasi Bibit dengan menggunakan metode *Multinomial Naïve Bayes*. Dataset yang digunakan terdiri dari 2.764 data yang telah dilabeli secara otomatis menggunakan pendekatan *Lexicon Based*. Dalam penelitian ini, data dibagi menjadi dua bagian, yaitu data latih sebanyak 2.211 tweet dan data validasi sebanyak 553 tweet. Hasil pengujian menunjukkan tingkat akurasi pelatihan sebesar 91,50%, akurasi validasi sebesar 85,35%, dan akurasi pengujian sebesar 88%. Analisis sentimen yang dilakukan divisualisasikan dalam bentuk *pie chart*, grafik, dan *word cloud*, yang mengungkap bahwa sentimen pengguna terhadap aplikasi Bibit cenderung positif dengan persentase 52% untuk sentimen positif dan 48% untuk sentimen negatif.

Kemudian terdapat pula penelitian serupa yang dilakukan oleh (Khotimah et al., 2025) yang menggabungkan teknik *SMOTE* dan *Laplace smoothing* untuk menganalisis sentimen ulasan aplikasi Pintu di Google Play Store. Data ulasan dikumpulkan melalui metode *web scraping* dan diproses melalui berbagai tahap pra-pemrosesan teks, termasuk pembersihan, normalisasi, penghapusan *stopwords*, tokenisasi, dan translasi. Klasifikasi sentimen ulasan dilakukan menggunakan *TextBlob*, dengan mengeliminasi ulasan netral untuk menyederhanakan klasifikasi

menjadi dua kategori, yaitu sentimen positif dan negatif. Untuk mengatasi masalah ketidakseimbangan data, penelitian ini menerapkan teknik oversampling *SMOTE*, sedangkan *Laplace smoothing* diterapkan pada model *Naïve Bayes* untuk meningkatkan akurasi prediksi. Hasil evaluasi menunjukkan bahwa model yang dikembangkan mencapai akurasi sebesar 95,07%, dengan presisi dan *recall* yang tinggi untuk kelas positif (masing-masing 97% dan 98%). Namun, model masih menghadapi tantangan dalam mengenali sentimen negatif, dengan presisi 62% dan *recall* 58%.

Berdasarkan permasalahan yang telah diuraikan sebelumnya, penelitian ini mengusulkan untuk melakukan analisis sentimen terhadap opini pengguna Twitter terkait *BRICS* menggunakan metode *Multinomial Naïve Bayes*. Dalam penelitian ini, beberapa teknik akan diterapkan untuk membandingkan tingkat akurasi dari setiap pengujian yang dilakukan. Pendekatan pertama adalah dengan menerapkan metode *Multinomial Naïve Bayes* tanpa modifikasi tambahan. Pendekatan kedua adalah dengan mengombinasikan metode *Multinomial Naïve Bayes* dengan teknik *Laplace smoothing* untuk meningkatkan distribusi probabilitas pada data yang jarang muncul. Pendekatan ketiga adalah mengintegrasikan metode *Multinomial Naïve Bayes* dengan teknik *SMOTE* (*Synthetic Minority Oversampling Technique*) untuk menangani masalah ketidakseimbangan data. Terakhir, pendekatan keempat mengombinasikan metode *Multinomial Naïve Bayes* dengan kedua teknik tersebut, yaitu *Laplace smoothing* dan *SMOTE*, untuk mengoptimalkan akurasi model. Melalui penerapan dan pengujian dengan berbagai kombinasi teknik ini, diharapkan dapat diperoleh wawasan yang lebih mendalam mengenai pengaruh masing-masing pendekatan terhadap tingkat akurasi model analisis sentimen. Selain itu, penelitian ini juga bertujuan untuk menentukan kombinasi metode dan teknik yang menghasilkan tingkat akurasi terbaik dalam mengklasifikasikan sentimen pengguna Twitter terhadap topik *BRICS*. Pendekatan ini diharapkan dapat memberikan kontribusi signifikan dalam pengembangan model analisis sentimen yang lebih akurat dan efektif.

1.2 Identifikasi Masalah

Berdasarkan latar belakang yang telah diuraikan diatas, terdapat beberapa permasalahan utama yang mendasari penelitian ini:

1. Teknik yang digunakan dalam penelitian ini, yaitu *Laplace smoothing* dan *SMOTE*, perlu dievaluasi efektivitasnya dalam meningkatkan akurasi model analisis sentimen terhadap opini publik mengenai topik *BRICS*..
2. Banyaknya komentar dari masyarakat Indonesia dan luar negeri tentang *BRICS* di platform X menunjukkan adanya beragam pendapat.
3. Minimnya penelitian yang menggabungkan teknik *SMOTE* dan *Laplace smoothing* secara bersamaan dalam analisis sentimen membuka peluang bagi penelitian ini untuk berkontribusi dalam memperkaya wawasan akademik dalam bidang *machine learning*.
4. Tingkat ketidakseimbangan data sentimen pada platform X dapat mempengaruhi kinerja dari model algoritma .

Melalui identifikasi permasalahan ini, penelitian ini bertujuan untuk mengembangkan model analisis sentimen yang lebih akurat dan efisien, serta memberikan kontribusi dalam pengembangan teknik pengolahan data teks dalam ranah akademik dan praktis.

1.3 Pembatasan Masalah

Batasan masalah dalam penelitian ini ditetapkan sebagai berikut:

1. Data yang digunakan dalam penelitian ini hanya mencakup tweet yang membahas topik *BRICS* dalam periode waktu dari Desember 2023 hingga Desember 2024.
2. Data yang dianalisis merupakan data publik yang diambil dari platform media sosial Twitter atau X melalui proses *crawling*, yang dapat diakses secara legal melalui Twitter API.
3. Penelitian ini berfokus pada analisis opini masyarakat global, tanpa membatasi pada masyarakat Indonesia saja.

4. Analisis sentimen dalam penelitian ini dilakukan dengan menggunakan metode *Multinomial Naïve Bayes*, yang dilengkapi dengan penerapan teknik *Laplace smoothing* dan *SMOTE (Synthetic Minority Oversampling Technique)*.
5. Sentimen yang dianalisis dikategorikan ke dalam tiga klasifikasi utama, yaitu positif, negatif, dan netral.
6. Evaluasi model dalam penelitian ini difokuskan pada pengukuran tingkat akurasi, presisi, *recall*, dan *F1-score* dari model *Multinomial Naïve Bayes*, baik dengan maupun tanpa kombinasi teknik *Laplace smoothing* dan *SMOTE*.

1.4 Perumusan Masalah

Berdasarkan latar belakang, identifikasi masalah, dan pembatasan masalah yang telah dijelaskan, perumusan masalah dalam penelitian ini adalah: "Bagaimana hasil akurasi pengujian analisis sentimen mengenai topik *BRICS* Indonesia menggunakan algoritma *Multinomial Naïve Bayes* dengan Teknik *Laplace smoothing* untuk menghindari nilai 0 dan *SMOTE* untuk imbalancing data?".

1.5 Tujuan Penelitian

Tujuan penelitian ini adalah untuk mengevaluasi efektivitas algoritma *Multinomial Naïve Bayes* dalam analisis sentimen terhadap opini pengguna Twitter mengenai topik *BRICS*, dengan membandingkan tingkat akurasi dari empat pendekatan model: penggunaan *Multinomial Naïve Bayes* murni, dengan *Laplace smoothing*, dengan *SMOTE*, dan kombinasi keduanya.

1.6 Manfaat Penelitian

Berdasarkan rumusan masalah dan tujuan penelitian yang telah dijelaskan sebelumnya, penelitian ini diharapkan dapat memberikan manfaat sebagai berikut:

1. Menyumbangkan kontribusi terhadap pengembangan ilmu pengetahuan di bidang *Natural Language Processing* (NLP), khususnya dalam penerapan algoritma *Multinomial Naïve Bayes* untuk analisis sentimen.
2. Menjadi referensi tambahan bagi penelitian-penelitian lain yang berkaitan dengan analisis sentimen menggunakan kombinasi teknik seperti *Laplace smoothing* dan *SMOTE*.
3. Membantu pengembang atau peneliti lain dalam memahami pentingnya teknik *Laplace smoothing* dan *SMOTE* dalam meningkatkan kinerja algoritma analisis sentimen.

