Jurnal Teknik Indonesia

Volume 4 Nomor 2 April 2025
https://jti.publicascientificsolution.com/index.php/rp

Analisis Sentimen Masyarakat Mengenai Relokasi Penduduk Rempang pada Media Sosial X Menggunakan Metode Naïve Bayes Classifier

[1]Muhammad Taufiq*, [2]Elin Haerani, [3]Fadhilah Syafria
[1,2,3,]Universitas Islam Negeri Sultan Syarif Kasim, Indonesia
Email : [1]11850114991@students.uin-suska.ac.id*, [2]elin.haerani@uin-suska.ac.id, [3]fadhilah.syafria@uin-suska.ac.id

Abstract

Social media platform X has become a primary space for the public to express opinions on various public issues, including the relocation policy of residents from Rempang Island as part of the National Strategic Project (PSN). The main problem is that public opinion is often unstructured, diverse, and widely dispersed, making it difficult to classify manually and objectively. Therefore, this study aims to develop an automated sentiment classification system for public opinion using a combination of lexical and machine learning approaches. A total of 1,000 relevant tweets were collected through a crawling process and filtered based on specific criteria. Sentiment labeling was conducted automatically using the InSet Lexicon, while text features were represented using the TF-IDF method. The classification model was built using the Naïve Bayes Classifier algorithm and evaluated through a confusion matrix, classification report, and 10-fold cross-validation. The results show that the model effectively classifies pro and contra sentiments, achieving a highest test accuracy of 81.00% (with a 90:10 split), and a highest cross-validation accuracy of 80.03% (with an 80:20 split). The highest precision was obtained in the pro class (up to 93%), while the highest recall was in the contra class (up to 89%). This approach has proven to be efficient and accurate for analyzing public opinion on social media, and it has strong potential for application to other relevant social issues.

Keywords: *Sentiment Analysis, Social Media X, Rempang Relocation, Naïve Bayes, TF-IDF, InSet Lexicon*

Abstrak

Media sosial X telah menjadi salah satu sarana utama bagi masyarakat dalam menyampaikan opini terhadap isu publik, termasuk kebijakan relokasi penduduk Pulau Rempang sebagai bagian dari Proyek Strategis Nasional (PSN). Permasalahan yang muncul adalah opini publik yang bersifat tidak terstruktur, beragam, dan tersebar luas sulit untuk diklasifikasikan secara manual dan objektif. Oleh karena itu, penelitian ini bertujuan untuk mengembangkan sistem klasifikasi sentimen otomatis terhadap opini masyarakat dengan pendekatan kombinasi leksikal dan pembelajaran mesin. Sebanyak 1.000 tweet relevan dikumpulkan melalui proses crawling dan disaring menggunakan kriteria tertentu. Pelabelan sentimen dilakukan secara otomatis menggunakan InSet Lexicon, sedangkan representasi fitur teks dilakukan dengan metode TF-IDF. Algoritma Naïve Bayes Classifier digunakan sebagai model klasifikasi dan dievaluasi menggunakan confusion matrix, classification report, dan 10-fold cross-validation. Hasil evaluasi menunjukkan bahwa model mampu mengklasifikasikan sentimen pro dan kontra secara efektif, dengan akurasi tertinggi pada data uji sebesar 81,00% (rasio 90:10), dan akurasi validasi silang tertinggi sebesar 80,03% (rasio 80:20). Precision tertinggi diperoleh pada

Jurnal Teknik Indonesia
E-ISSN: 2963-2293 | P-ISSN: 2964-8092
DOI: 10.58860/jti.v4i2.711

kelas pro (hingga 93%), sedangkan recall tertinggi pada kelas kontra (hingga 89%). Pendekatan ini terbukti efisien dan akurat untuk menganalisis opini publik berbasis media sosial, serta memiliki potensi untuk diterapkan pada isu-isu sosial lainnya yang relevan.

Kata kunci: Analisis Sentimen, Media Sosial X, Relokasi Rempang, Naïve Bayes, TF-IDF, InSet Lexicon

Corresponding Author;
E-mail: 11850114991@students.uin-suska.ac.id

Pendahuluan

Perkembangan teknologi informasi dan komunikasi telah mengubah pola interaksi masyarakat secara signifikan. Media sosial kini menjadi salah satu sarana utama dalam menyampaikan opini, kritik, dan dukungan terhadap berbagai isu. Media sosial juga digunakan pengguna dalam rangka mempresentasikan dirinya, berinteraksi, bekerja sama, berbagi informasi, maupun berinteraksi dengan pengguna lainnya (Krisdiyanto, 2021). Di antara berbagai platform, *X* menempati posisi penting karena sifatnya yang real-time, terbuka, dan berbasis teks, sehingga menjadikannya sumber data yang sangat potensial untuk dianalisis (Nandaresta & Warman, 2023).

Salah satu isu nasional yang menjadi sorotan dan menimbulkan berbagai respons masyarakat adalah kebijakan relokasi penduduk Pulau Rempang, yang dilakukan pemerintah sebagai bagian dari Proyek Strategis Nasional (PSN) dalam pembangunan kawasan Rempang Eco-City (Walangare & Syaiful, 2023). Proyek ini digagas sebagai kawasan industri dan pariwisata terintegrasi yang bertujuan meningkatkan daya saing ekonomi Indonesia di kawasan Asia Tenggara, terutama untuk menyaingi Singapura dan Malaysia. Namun, sejak diumumkan secara luas pada 2023, proyek ini menuai penolakan dari masyarakat adat Rempang yang menganggap relokasi mengabaikan hak historis dan sosial mereka atas tanah (Saly et al., 2023).

Situasi tersebut memicu perdebatan intens di berbagai kanal media sosial, khususnya *X*. Cuitan-cuitan dari masyarakat menggambarkan berbagai ekspresi: mulai dari dukungan terhadap pembangunan hingga kecaman terhadap kebijakan relokasi (Silalahi, 2020). Hal ini menjadikan *X* sebagai cerminan persepsi publik yang layak dianalisis secara sistematis.

Untuk memahami opini masyarakat secara kuantitatif, dapat dilakukan pendekatan analisis sentimen**,** yaitu proses mengidentifikasi dan mengklasifikasikan sikap atau emosi dalam suatu teks. Dalam konteks penelitian ini, sentimen dibagi menjadi dua kelas utama, yaitu *pro* terhadap relokasi dan *kontra* terhadap relokasi (Puspita & Widodo, 2021). Analisis ini penting bagi pemerintah dan pemangku kepentingan untuk mengetahui kecenderungan publik dan merumuskan kebijakan berbasis aspirasi masyarakat.

Untuk proses klasifikasi, digunakan algoritma *Naïve Bayes Classifier (NBC)* (Muslimin & Lusiana, 2023)**,** yang merupakan metode statistik berbasis probabilistik yang telah terbukti efektif dalam klasifikasi teks. NBC memiliki keunggulan dalam kecepatan proses pelatihan dan kemampuannya mengolah data besar dengan tingkat akurasi yang kompetitif (Tanggraeni & Sitokdana, 2022).

Seperti penelitian yang dilakukan oleh Rahmat Yasmin dengan judul "Analisis Sentimen Masyarakat Mengenai Vaksin Covid-19 Menggunakan Metode Naive Bayes Classifier Pada Media Sosial *X*" mendapatkan hasil akurasi tertinggi diperoleh dengan rasio perbandingan 90:10% dengan nilai akurasi sebesar 84%, kemudian nilai precision tertinggi yaitu sebesar 86%, dan nilai recall tertinggi yaitu sebesar 82,69%.

Sebelum dilakukan klasifikasi, data perlu diolah melalui proses *preprocessing teks,* termasuk *case folding*, *tokenizing*, *stopword removal*, dan *stemming* (Rahayu et al., 2022). Kemudian *dataset* diberi sentiment berupa Pro dan Kontra dan dilakukan penerapan metode TF-IDF (Term Frequency–Inverse Document Frequency) untuk membobot kata-kata dalam teks, sehingga memberikan bobot lebih besar pada kata-kata yang penting namun jarang muncul di seluruh dokumen (Ilmar Rifaldi et al., 2023).

Untuk pelabelan data sentimen secara otomatis, digunakan pendekatan berbasis kamus (*lexicon-based*) dengan menggunakan *InSet Lexicon***.** pada penelitian yang dilakukan oleh (Koto & Rahmaningtyas, 2018) mereka berhasil membangun sebuah sentiment leksikon dalam bahasa Indonesia yang dinamai dengan InSet (Indonesian Sentiment) Lexicon yang memiliki polarity score berkisar antara -5 dan +5. Pada penelitiannya juga terbukti bahwa InSet Lexicon memiliki akurasi tertinggi untuk di setiap algoritma classifier seperti Naive Bayes, Linear Regression, SVM, dan KNN dibandingkan dengan baseline lainnya. Mencapai 65,78% sebagai yang tertinggi dan lebih baik dari Vania Lexicon dengan akurasi 61,48% (Fauzan, 2022). *InSet Lexicon* dirancang khusus untuk teks media sosial yang berbahasa indoensia dan memiliki akurasi tinggi dalam mendeteksi polaritas emosional dalam kalimat-kalimat pendek dan tidak formal yang khas di *X*.

Akurasi model klasifikasi kemudian dievaluasi menggunakan confusion matrix**,** classification report (precision, recall, F1-score), serta metode k-fold cross validation guna memastikan model yang digunakan andal dan tidak overfitting (Ulya, 2022).

Penelitian ini tidak hanya bertujuan untuk memahami persepsi masyarakat terhadap relokasi Rempang, tetapi juga memberikan kontribusi terhadap pengembangan metode analisis sentimen dengan kombinasi pendekatan *lexicon-based* dan *machine learning*. Selain itu, hasil analisis ini diharapkan dapat menjadi masukan kebijakan publik yang lebih inklusif dan responsif terhadap kebutuhan masyarakat.

Penelitian ini bertujuan untuk melakukan pelabelan sentimen secara otomatis terhadap data opini masyarakat mengenai relokasi penduduk Rempang menggunakan metode Lexicon InSet, yang mampu mengidentifikasi polaritas sentimen berdasarkan skor kata dalam Bahasa Indonesia. Selanjutnya, penelitian ini mengimplementasikan algoritma Naïve Bayes Classifier untuk mengklasifikasikan sentimen menjadi kategori pro dan kontra, dengan menggunakan metode pembobotan fitur TF-IDF untuk merepresentasikan data teks secara numerik. Terakhir, performa model klasifikasi dievaluasi menggunakan metrik confusion matrix, classification report (meliputi akurasi, precision, recall, dan F1-score), serta k-fold cross validation untuk menguji konsistensi dan keandalan model.

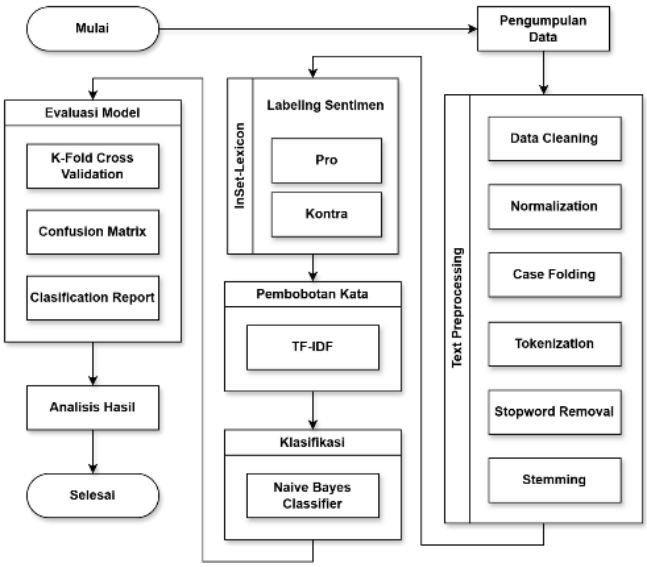
Penelitian ini diharapkan dapat memberikan sejumlah manfaat. Dari sisi akademik, penelitian ini berkontribusi dalam pengembangan studi analisis sentimen, khususnya dalam konteks Bahasa Indonesia, dengan menggabungkan pendekatan lexicon-based dan machine learning secara efektif. Dari aspek praktis, hasil penelitian

E-ISSN: 2963-2293 | P-ISSN: 2964-8092
DOI: 10.58860/jti.v4i2.711

ini dapat memberikan wawasan kepada pemerintah dan pemangku kebijakan mengenai kecenderungan opini publik terhadap proyek Rempang Eco-City, yang dapat dijadikan sebagai masukan dalam perumusan kebijakan maupun strategi komunikasi publik. Sementara dari sisi teknologis, penelitian ini menunjukkan bahwa pendekatan otomatis berbasis Natural Language Processing (NLP) dan pembelajaran mesin mampu digunakan secara efisien untuk menganalisis opini masyarakat dari data yang bersumber dari media sosial.

Metode Penelitian

Penelitian ini dilaksanakan melalui serangkaian tahapan sistematis untuk menganalisis sentimen masyarakat terhadap relokasi penduduk Rempang berdasarkan data dari media sosial *X*. Sentimen diklasifikasikan ke dalam dua kategori utama, yaitu **pro** relokasi dan **kontra** relokasi. Model klasifikasi dibangun menggunakan pendekatan **Naïve Bayes Classifier** dan juga dibandingkan dengan pendekatan **lexicon-based** menggunakan ***InSet (Indonesia Sentimen)***. Berikut merupakan flowchart dari kerangka kerja:



Gambar 1. Flowchard Kerangka Kerja

Pengumpulan Data

Pengumpulan data dalam penelitian ini dilakukan melalui metode web crawling pada media sosial X, yang dipilih karena perannya sebagai platform publik utama dalam menyuarakan opini, termasuk terkait isu relokasi penduduk Rempang. Proses crawling menggunakan tools *tweet-harvest* berbasis Node.js versi 20, dengan dukungan pustaka seperti *pandas* untuk pengolahan data.

Data dikumpulkan berdasarkan kata kunci relevan seperti “Pulau Rempang”, “Konflik Rempang”, “Relokasi Rempang”, “Tolak Relokasi”, dan “Rempang Eco City”, terbatas pada tweet berbahasa Indonesia yang dipublikasikan antara Januari hingga Oktober 2023. Jenis data yang diambil meliputi teks tweet, tanggal, dan ID tweet, dan disimpan dalam format CSV untuk kebutuhan preprocessing dan analisis.

Dari total 6.079 tweet yang terkumpul, dilakukan proses seleksi untuk menjaga relevansi, dan sebanyak 1.000 tweet yang sesuai digunakan sebagai dataset akhir dalam penelitian.

Text Preprocessing

Tahap preprocessing merupakan bagian krusial dalam analisis teks, karena memengaruhi kualitas data masukan yang akan digunakan oleh model klasifikasi. Tujuan utama dari tahapan ini adalah untuk membersihkan dan menormalkan data mentah agar memiliki struktur yang konsisten serta informatif, sehingga dapat diolah secara optimal oleh algoritma pembelajaran mesin.

Dalam penelitian ini, preprocessing dilakukan melalui enam tahap utama. Pertama, *data cleaning* dilakukan dengan menghapus elemen-elemen yang tidak relevan seperti URL, mention (@), hashtag (#), emoji, angka, tanda baca, dan karakter khusus lainnya. Kedua, dilakukan *normalisasi* untuk mengganti kata-kata tidak baku atau slang menjadi bentuk baku, seperti "gk" menjadi "tidak" dan "dgn" menjadi "dengan". Tahap ketiga adalah *case folding*, yaitu mengubah seluruh huruf menjadi huruf kecil guna menjaga konsistensi representasi teks.

Selanjutnya, dilakukan *tokenisasi* untuk memecah kalimat menjadi unit kata (token). Setelah itu, *stopword removal* diterapkan untuk menghapus kata-kata umum yang tidak memiliki kontribusi signifikan dalam analisis, dengan mengacu pada daftar stopword Bahasa Indonesia. Terakhir, tahap *stemming* menggunakan algoritma Sastrawi dilakukan untuk mengembalikan kata ke bentuk dasarnya, seperti "menolak" menjadi "tolak".

Dengan melalui tahapan tersebut, data teks menjadi lebih bersih, ringkas, dan terstruktur, sehingga siap digunakan dalam proses pelabelan dan klasifikasi. Proses ini juga berkontribusi dalam pengurangan dimensi fitur yang berlebihan dan meningkatkan kinerja model klasifikasi secara keseluruhan.

Labeling Sentimen

Labeling sentimen merupakan proses penentuan kategori sentimen dari setiap tweet berdasarkan opini yang terkandung di dalamnya. Dalam penelitian ini, proses pelabelan dilakukan menggunakan pendekatan ***lexicon-based*** dengan memanfaatkan ***InSet Lexicon,*** sebuah metode analisis sentimen berbasis kamus yang dirancang khusus untuk bahasa Indonesia.

Pembobotan Kata

Proses pembobotan kata merupakan langkah penting dalam mengubah data teks hasil preprocessing menjadi representasi numerik yang dapat diproses oleh algoritma pembelajaran mesin. Dalam penelitian ini, metode yang digunakan adalah *Term Frequency–Inverse Document Frequency* (TF-IDF), yang memberikan bobot tinggi pada kata-kata yang sering muncul dalam dokumen tertentu namun jarang ditemukan di dokumen lain, sehingga dianggap lebih representatif terhadap isi dokumen tersebut.

TF-IDF efektif dalam mengurangi pengaruh kata-kata umum yang bersifat generik, serta memperkuat kontribusi kata-kata yang relevan terhadap sentimen. Implementasi dilakukan menggunakan pustaka *scikit-learn* dalam bahasa pemrograman

E-ISSN: 2963-2293 | P-ISSN: 2964-8092
DOI: 10.58860/jti.v4i2.711

Python, yang secara otomatis mengubah teks menjadi vektor numerik berdimensi tetap. Metode ini dipilih karena efisien dalam menangani data teks pendek seperti tweet, serta mampu mengolah data dalam jumlah besar secara optimal.

Klasifikasi

Algoritma *Naïve Bayes Classifier* (NBC) digunakan sebagai metode klasifikasi dalam penelitian ini. NBC merupakan algoritma berbasis probabilistik yang mengasumsikan independensi antar fitur, menjadikannya efisien dan akurat untuk klasifikasi teks, terutama pada data pendek seperti tweet.

NBC dipilih karena kemampuannya dalam menangani data skala besar, proses komputasi yang ringan, serta kemudahan implementasi dan interpretasi. Dalam penelitian ini, NBC digunakan untuk mengklasifikasikan tweet ke dalam dua kelas sentimen: Pro Relokasi dan Kontra Relokasi terhadap proyek Rempang Eco City.

Proses klasifikasi mencakup tiga tahap: pelatihan model menggunakan data yang telah dipreproses dan diberi bobot dengan TF-IDF, pengujian terhadap data uji, serta evaluasi performa menggunakan metrik akurasi, precision, recall, F1-score, dan *confusion matrix*. Implementasi dilakukan dengan library *Scikit-learn* di Python, menggunakan *TfidfVectorizer* dan *MultinomialNB*, dengan data yang telah dilabeli otomatis menggunakan *InSet Lexicon*.

Evaluasi Model

Evaluasi model bertujuan untuk mengukur performa algoritma klasifikasi dalam mengelompokkan sentimen tweet secara akurat. Evaluasi dilakukan setelah pelatihan model, menggunakan data uji yang telah dipisahkan sebelumnya. Tiga pendekatan utama digunakan, yaitu *confusion matrix*, *classification report*, dan *k-fold cross validation*.

Confusion matrix digunakan untuk menampilkan jumlah prediksi benar dan salah pada masing-masing kelas (Pro dan Kontra), terdiri dari empat komponen: *true positive*, *true negative*, *false positive*, dan *false negative*. Sementara itu, *classification report* menyajikan metrik evaluasi secara lebih rinci, mencakup *accuracy*, *precision*, *recall*, dan *F1-score*, yang dihitung menggunakan pustaka sklearn.metrics.

Untuk memastikan konsistensi model dan menghindari bias akibat pembagian data, digunakan metode *10-fold cross validation*, yang membagi dataset menjadi 10 bagian dan melakukan pelatihan serta pengujian secara bergilir. Hasil dari setiap iterasi dirata-rata untuk memperoleh estimasi performa yang lebih stabil dan andal, serta mengurangi risiko *overfitting*.

Hasil dan Pembahasan

Deskripsi Dataset

Dataset dalam penelitian ini diperoleh melalui metode *web crawling* menggunakan X API (sebelumnya Twitter API), dengan memanfaatkan endpoint *Recent Search*. Platform X dipilih karena perannya sebagai ruang publik digital yang aktif dalam menyuarakan opini terhadap isu-isu sosial, termasuk relokasi penduduk Rempang. Pengambilan data dilakukan dengan kata kunci relevan seperti #PulauRempang, #KonflikRempang, #RelokasiRempang, #TolakRelokasi, dan #RempangEcoCity,

terbatas pada tweet berbahasa Indonesia yang dipublikasikan antara Januari hingga Oktober 2023.

Dari 6.079 tweet yang terkumpul, dilakukan proses filtrasi berdasarkan beberapa kriteria: menghapus tweet berbahasa asing, kosong, terlalu pendek (<25 suku kata), duplikat, dan yang tidak menyebutkan "Rempang". Hasilnya, sebanyak 1.000 tweet dipertahankan sebagai dataset akhir.

Tweet tersebut kemudian diklasifikasikan ke dalam dua kategori sentimen, yaitu Pro Relokasi (mendukung program pemerintah) dan Kontra Relokasi (menolak proyek dan mendukung hak masyarakat adat). Setiap entri dalam dataset memuat isi teks dan label sentimen.

Pengumpulan data dilakukan menggunakan *tweet-harvest*, tool berbasis Node.js, dengan lingkungan kerja yang dikonfigurasi menggunakan Node.js versi 20 dan pustaka *pandas* untuk pengolahan data. Untuk mengatasi batasan *rate limit* dari API, digunakan beberapa akun X tambahan guna memperluas jangkauan data.

Perlu dicatat bahwa seluruh data bersumber dari satu platform, sehingga hasil analisis tidak dimaksudkan untuk merepresentasikan keseluruhan opini publik Indonesia secara menyeluruh.

Preprocessing

Dalam preprocessing, tweet akan melalui serangkaian proses, agar mudah dalam pelabelan data. Berikut merupakan 2 contoh data yang di dapatkan dengan Teknik Crawling pada media sosial tweeter:

Tabel 1. Contoh Data Komentar Tweeter

NO	DATASET
1	@IND0_Update @MataNajwa @Metro_TV @tvOneNews Yakin dehhh klo proyek Rempang Eco City nih bakal berhasil krn Bahlil Kawal Investasi #HIPMIRempang,
…	…
1000	"Hipmi berkomitmen untuk mendukung investasi Rempang Eco City dengan membantu meningkatkan keterampilan warga. Kolaborasi kami dengan pemerintah adalah kunci kesuksesan, membuka jalan bagi kemajuan bersama. Bahlil Kawal Investasi #HIPMIRempang https://t.co/eZlG0ZVwPR",

1. Cleaning Data

Menghapus karakter tidak penting seperti URL, mention (@username), hashtag (#tag), simbol, angka, serta karakter non-alfabet yang tidak relevan untuk analisis teks.

Tabel 2. Proses Cleaning

NO	Sebelum *Cleaning*	Hasil *Cleaning*
1	@IND0_Update @MataNajwa @Metro_TV @tvOneNews Yakin dehhh klo proyek Rempang Eco City	Yakin dehhh klo proyek Rempang Eco City nih bakal berhasil krn Bahlil Kawal Investasi

E-ISSN: 2963-2293 | P-ISSN: 2964-8092
DOI: 10.58860/jti.v4i2.711

	nih bakal berhasil krn Bahlil Kawal Investasi #HIPMIRempang,

2. Case Folding

Mengubah seluruh huruf menjadi huruf kecil agar tidak terjadi duplikasi token karena perbedaan kapitalisasi.

Tabel 3 Proses Case Folding

NO	Sebelum *Case Folding*	Setelah *Case Folding*
1	Yakin dehhh klo proyek Rempang Eco City nih bakal berhasil krn Bahlil Kawal Investasi	yakin dehhh klo proyek rempang eco city nih bakal berhasil krn bahlil kawal investasi

3. Tokenization

Memecah kalimat atau paragraf menjadi kata-kata (token) tunggal.

Tabel 4 Proses Tokenization

NO	Sebelum *Tokenization*	Setelah *Tokenization*
1	yakin dehhh klo proyek rempang eco city nih bakal berhasil krn bahlil kawal investasi	['yakin', 'dehhh', 'klo', 'proyek', 'rempang', 'eco', 'city', 'nih', 'bakal', 'berhasil', 'krn', 'bahlil', 'kawal', 'investasi']

4. Normalization

Menormalkan teks seperti menyamakan kata-kata tidak baku atau singkatan menjadi bentuk baku (contoh: “gk” → “tidak”, “tdk” → “tidak”).

Tabel 5 Proses Normalization

NO	Sebelum *Normalization*	Setelah *Normalization*
1	['yakin', 'dehhh', 'klo', 'proyek', 'rempang', 'eco', 'city', 'nih', 'bakal', 'berhasil', 'krn', 'bahlil', 'kawal', 'investasi']	['yakin', 'deh', 'kalau', 'proyek', 'rempang', 'eco', 'city', 'ini', 'bakal', 'berhasil', 'karena', 'bahlil', 'kawal', 'investasi']

5. Stopword Removal

Menghapus kata-kata umum yang tidak memiliki makna penting dalam klasifikasi seperti “yang”, “dan”, “dari”, dsb.

Tabel 6 Proses Stopword Removal

No	Sebelum *Stopword Removal*	Setelah *Stopword Removal*
1	['yakin', 'deh', 'kalau', 'proyek', 'rempang', 'eco', 'city', 'ini', 'bakal', 'berhasil', 'karena', 'bahlil', 'kawal', 'investasi']	['yakin', 'proyek', 'rempang', 'eco', 'city', 'bakal', 'berhasil', 'bahlil', 'kawal', 'investasi']

6. Stemming

Mengubah kata ke bentuk dasar menggunakan algoritma stemming Bahasa Indonesia, seperti Sastrawi.

Tabel 7 Proses Stemming

No	Sebelum *Stemming*	Setelah *Stemming*
1	['yakin', 'proyek', 'rempang', 'eco', 'city', 'bakal', 'berhasil', 'bahlil', 'kawal', 'investasi']	['yakin', 'proyek', 'rempang', 'eco', 'city', 'bakal', 'hasil', 'bahlil', 'kawal', 'investasi']

Pelabelan

Proses pelabelan sentimen dalam penelitian ini dilakukan secara otomatis menggunakan pendekatan **lexicon-based** dengan memanfaatkan **InSet Lexicon**, yaitu kamus sentimen Bahasa Indonesia yang berisi daftar kata-kata positif dan negatif beserta bobot polaritasnya. Pendekatan ini memungkinkan pemberian label sentimen tanpa keterlibatan anotator manusia, sehingga efisien untuk skala data yang besar.

Mekanisme pelabelan dilakukan dengan membandingkan setiap kata dalam tweet hasil preprocessing terhadap daftar kata yang terdapat dalam InSet Lexicon. Jika suatu kata termasuk dalam daftar kata positif, maka akan diberikan bobot antara **+1 hingga +5**. Sebaliknya, jika kata tersebut termasuk dalam daftar kata negatif, maka diberikan bobot antara **–1 hingga –5**. Kata-kata yang tidak ditemukan dalam kamus diberi bobot **0**. Total skor sentimen suatu tweet dihitung dengan menjumlahkan seluruh bobot kata-kata yang ada dalam tweet tersebut.

Berdasarkan total skor, label sentimen ditentukan dengan ketentuan sebagai berikut: jika skor total lebih dari 0, maka tweet dikategorikan sebagai **Pro Relokasi** karena merepresentasikan opini positif terhadap kebijakan pembangunan. Jika skor total kurang dari 0, maka tweet diberi label **Kontra Relokasi**, karena mengandung opini negatif terhadap proyek tersebut. Sementara itu, jika skor total sama dengan 0, tweet dianggap **Netral**, dan dalam konteks penelitian ini, data netral tidak digunakan dalam pelatihan dan pengujian model.

Penerapan metode ini pada 5 tweet hasil preprocessing ditunjukkan dalam tabel berikut:

Tabel 8. Hasil Pelabelan InSet-Lexicon

No	Data	Jumlah
1	yakin proyek rempang eco city bakal hasil bahlil kawal investasi	+8
…	…	…
1000	hipmi komit dukung investasi rempang eco city bantu tingkat keterampil warga kolaborasi pemerintah kunci sukses buka jalan maju sama bahlil kawal investasi	+18

E-ISSN: 2963-2293 | P-ISSN: 2964-8092
DOI: 10.58860/jti.v4i2.711

Dengan metode ini, labeling terhadap seluruh dataset dapat dilakukan secara lebih objektif dan konsisten. Untuk eksperimen manual, 4 data pertama digunakan sebagai data latih, dan 1 data terakhir digunakan sebagai data uji pada tahapan klasifikasi.

Hasil Pembobotan dan Klasifikasi

Sebelum dilakukan proses klasifikasi, terlebih dahulu dilakukan tahap pembobotan kata, yaitu proses pemberian nilai bobot pada setiap kata yang muncul dalam dokumen. Pembobotan ini dilakukan menggunakan metode TF-IDF (Term Frequency-Inverse Document Frequency). Empat data yang telah melalui tahap preprocessing diasumsikan sebagai D1, D2, D3, dan D4.

Naïve Bayes Classifier merupakan metode yang digunakan pada penelitian ini untuk melakukan proses klasifikasi sentimen. Tahap klasifikasi menggunakan metode Naïve Bayes Classifier dengan keseluruhan dataset dibagi 2 tahapan yaitu training (latih) dan tahapan testing (uji). Berikut dijelaskan tahapan training dan testing.

1. *Training* (Latih)

Dalam membentuk model klasifikasi data yang didapatkan bobotnya pada tahapan training digunakan untuk menjadi referensi. Selanjutnya, akan dilakukan pencarian nilai probabilitas pada tiap kelas terhadap dokumen dan probabilitas kata masing-masing kelas yang berasal dari data latih. Berikut penjelasan perhitungan nya:

2. *Testing* (Uji)

Pengujian dilakukan pada tahapan testing (uji) dengan menempatkan data uji di dalam model klasifikasi yang telah dibentuk pada tahapan training (latih). Berikut tabel dari contoh data uji yang digunakan.

Mengikuti data uji (testing) diatas, nilai probabilitas terhadap kelas akan dihitung dengan menggunakan metode Naive Bayes Classifier. Perhitungan nilai probabilitas kata terhadap kelas pada dokumen ini dilakukan dengan perkalian beruntun.

Setelah itu perhitugan probabilitas pada data uji (testing) dilakukan dengan menggunakan hasil probabilitas yang didapatkan pada data latih yang sudah melewati tahap pre-processing. Dan melakukan perhitungan probabilitas kata pada setiap kelas menggunakan persamaan rumus 2 berikut:

$$P(w_k, pro|kontra) = \frac{1 + n_k}{n + |kata|}$$

Selanjutnya, tahapan setelah dilakukan perhitungan probabilitas pada setiap kelas yaitu semua nilai probabilitas beruntun untuk menentukan kelas data uji tersebut sebagai berikut:

$$\begin{aligned} = 0.5 \times 0.018377292 \times\ & 0.017857143 \times 0.017857143 \times 0.018475029 \\ & \times 0.018181818 \times 0.018377292 \times 0.018377292 \times 0.01837729 \\ & \times 0.017857143 \times 0.017857143 \times 0.017857143 \times 0.017857143 \\ & \times 0.017857143 \times 0.017857143 \times 0.018377292 \times 0.017857143 \\ & \times 0.017857143 \times 0.018181818 \times 0.017857143 \times 0.01826294 \\ & \times 0.018475029 \end{aligned}$$

$= 1.27153E-37$

$= 0.5 \times 0.015384615 \times 0.015151515 \times 0.015151515 \times 0.015384615 \times 0.015384615 \times 0.015384615 \times 0.015384615 \times 0.015151515 \times 0.015151515 \times 0.015151515 \times 0.015151515 \times 0.015151515 \times 0.015151515 \times 0.015384615 \times 0.015151515 \times 0.015151515 \times 0.015524956 \times 0.015151515 \times 0.015413739 \times 0.015384615$

$= 3.62759E-39$

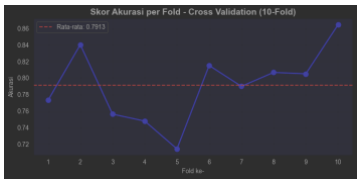
Tabel 9. Hasil Data Uji (Testing)

Data Uji	Probabilitas Pro	Probabilitas Kontra	Label
hipmi komit dukung investasi rempang eco city bantu tingkat terampil warga kolaborasi perintah kunci sukses buka jalan maju sama bahlil kawal investasi	$1.27153E-37$	$3.62759E-39$	Pro

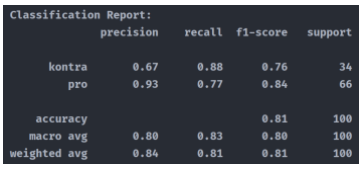
Hasil klasifikasi dengan Naive Bayes menunjukkan bahwa data uji memiliki probabilitas lebih tinggi pada kelas Pro, sehingga model memutuskan label Pro. Ini menunjukkan bahwa prediksi bergantung pada distribusi kata dalam data latih, meskipun secara konteks kalimat bernada mendukung masyarakat Relokasi.

Hasil Evaluasi Model

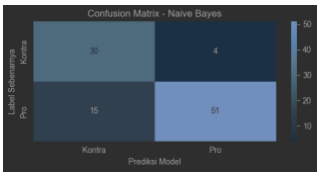
Evaluasi performa model dilakukan dengan tiga pendekatan utama. Pertama, data dibagi ke dalam tiga skenario rasio pelatihan dan pengujian: 90:10, 80:20, dan 70:30, untuk mengamati pengaruh proporsi data latih terhadap performa model. Kedua, digunakan *10-fold cross validation* guna memperoleh estimasi performa yang lebih stabil dan mengurangi bias dari satu pembagian data tunggal. Ketiga, evaluasi dilakukan menggunakan *confusion matrix* dan *classification report* yang mencakup metrik presisi, recall, dan F1-score untuk masing-masing kelas sentimen.



Gambar 2 K-Fold Cross Validation (90:10)



Gambar 3 Classification Report (90:10)



Gambar 4 Confusion Matrix (90:10)

Model klasifikasi opini menggunakan algoritma Naïve Bayes dievaluasi melalui tiga skenario pembagian data, yaitu 90:10, 80:20, dan 70:30, menggunakan metrik

E-ISSN: 2963-2293 | P-ISSN: 2964-8092
DOI: 10.58860/jti.v4i2.711

akurasi, precision, recall, F1-score, serta 10-fold cross validation. Hasil evaluasi menunjukkan performa model yang stabil, dengan perbedaan kinerja bergantung pada proporsi data latih.

Skenario 90:10 memberikan hasil terbaik, dengan akurasi 81% dan F1-score makro 80%. Precision tertinggi dicapai pada kelas Pro (0.93), sementara recall tertinggi terdapat pada kelas Kontra (0.88). Evaluasi cross validation juga mendukung stabilitas model dengan rata-rata akurasi 79.13%. Pada rasio 80:20 dan 70:30, akurasi turun menjadi masing-masing 76% dan 76.33%, dengan F1-score makro 75% dan 76%. Meski precision pada kelas Pro tetap tinggi di seluruh skenario, recall lebih dominan pada kelas Kontra.

Secara keseluruhan, model menunjukkan akurasi lebih tinggi dengan proporsi data latih yang lebih besar. Nilai F1-score makro yang konsisten dalam kisaran 75–80% menunjukkan kinerja model yang seimbang dan andal dalam mengklasifikasikan opini Pro dan Kontra terhadap relokasi Rempang di media sosial.

Kesimpulan

Penelitian ini berhasil mengembangkan sistem klasifikasi sentimen terhadap isu relokasi penduduk Rempang dengan memanfaatkan opini masyarakat dari media sosial X. Sistem ini dibangun melalui tahapan pelabelan data secara otomatis menggunakan pendekatan InSet Lexicon, serta pengembangan model klasifikasi menggunakan algoritma Naïve Bayes Classifier (NBC) dengan representasi fitur TF-IDF. Hasil penelitian menunjukkan bahwa metode InSet Lexicon efektif dan efisien dalam mengelompokkan komentar menjadi dua kelas utama, yaitu *Pro* dan *Kontra*, sekaligus menggantikan proses anotasi manual yang memakan waktu.

Model klasifikasi diuji pada tiga skenario pembagian data. Pada rasio 90:10, model mencapai akurasi tertinggi sebesar 81%, yang menunjukkan hasil optimal ketika data latih lebih besar. Pada rasio 80:20, diperoleh akurasi validasi silang tertinggi sebesar 80,03%, mencerminkan stabilitas performa model. Sementara pada rasio 70:30, akurasi uji sebesar 76,33% tetap menunjukkan kinerja yang layak meskipun data latih lebih sedikit.

Evaluasi model berdasarkan confusion matrix, classification report, dan 10-fold cross-validation memperlihatkan bahwa precision tertinggi secara konsisten berada pada kelas *Pro* dengan rentang 91%–93%, sedangkan recall tertinggi terdapat pada kelas *Kontra* dengan nilai 87%–89%. Nilai F1-score makro berkisar antara 75%–80%, mencerminkan keseimbangan kinerja model dalam mengklasifikasikan kedua kelas.

Secara keseluruhan, model Naïve Bayes yang dikembangkan mampu mengklasifikasikan opini publik terhadap relokasi Rempang secara akurat dan konsisten, serta memiliki potensi untuk diaplikasikan pada isu-isu sosial lainnya yang melibatkan analisis opini masyarakat di media sosial.

Daftar Pustaka

Fauzan, M. A. (2022). Penerapan sentimen leksikon Indonesia pada analisis sentimen mengenai opini Masyarakat di twitter terhadap kebijakan PPKM darurat menggunakan algoritma …. In *Repository.Uinjkt.Ac.Id*.

Ilmar Rifaldi, M., Raymond Ramadhan, Y., & Jaelani, I. (2023). Analisis Sentimen Terhadap Aplikasi Chatgpt Pada Twitter Menggunakan Algoritma Naïve Bayes. *Jurnal Sains Komputer & Informatika (J-SAKTI*, *7*(2), 802–814.

Krisdiyanto, T. (2021). Analisis Sentimen Opini Masyarakat Indonesia Terhadap Kebijakan PPKM pada Media Sosial Twitter Menggunakan Naïve Bayes Clasifiers. *Jurnal CoreIT: Jurnal Hasil Penelitian Ilmu Komputer Dan Teknologi Informasi*, *7*(1), 32. https://doi.org/10.24014/coreit.v7i1.12945

Muslimin, M., & Lusiana, V. (2023). Analisis Sentimen Terhadap Kenaikan Harga Bahan Pokok Menggunakan Metode Naive Bayes Classifier. *Jurnal Media Informatika Budidarma*, *7*(3), 1200. https://doi.org/10.30865/mib.v7i3.6418

Nandaresta, S. C., & Warman, C. (2023). Analisis Sentimen Tanggapan Masyarakat Terhadap Tiktok Shop Dan Shopee Di Twitter Menggunakan Metode Naïve Bayes Dan Knn (K- Nearest Neighbor. *Sismatik*, *12*(1), 1–9.

Puspita, R., & Widodo, A. (2021). Perbandingan Metode KNN, Decision Tree, dan Naïve Bayes Terhadap Analisis Sentimen Pengguna Layanan BPJS. *Jurnal Informatika Universitas Pamulang*, *5*(4), 646. https://doi.org/10.32493/informatika.v5i4.7622

Rahayu, I. P., Fauzi, A., & Indra, J. (2022). Analisis Sentimen Terhadap Program Kampus Merdeka Menggunakan Naive Bayes Dan Support Vector Machine. *Jurnal Sistem Komputer Dan Informatika (JSON)*, *4*(2), 296. https://doi.org/10.30865/json.v4i2.5381

Saly, J. N., Ekalia, E., & Tarumanagara, U. (2023). Status Perlindungan Hukum Kepada Masyarakat Setempat Terkait Relokasi Pulau Rempang. *Jurnal Kewarganegaraan*, *7*(2), 1668–1676.

Silalahi, R. C. (2020). Faktor-Faktor yang Menyebabkan Permasalahan Relokasi Bantaran Sungai. *Jurnal Muara Ilmu Sosial, Humaniora, Dan Seni*, *1*(2), 488–499.

Tanggraeni, A. I., & Sitokdana, M. N. N. (2022). Analisis Sentimen Aplikasi E-Government pada Google Play Menggunakan Algoritma Naïve Bayes. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, *9*(2), 785–795. https://doi.org/10.35957/jatisi.v9i2.1835

Ulya, S. Z. (2022). *Analisa Sentimen Masyarakat Mengenai Ppkm Di Kota Pekanbaru Pada Instagram Menggunakan Metode Naïve Bayes Classifier*. ii–123.

Walangare, S. W., & Syaiful, B. (2023). Kontestasi Kepentingan Pro-Growth Coalition dan Anti-Growth Coalition dalam Konflik Pembangunan Rempang Eco-City Tahun 2023. *Madani: Jurnal Politik Dan Sosial Kemasyarakatan*, *15*(2), 381–403.