

ANALISIS TOLERANSI BERAGAMA DI KOTA SALATIGA MENGGUNAKAN SUPPORT VECTOR MACHIE NAIVE BAYES DAN RANDOM FOREST

ANALYSIS OF RELIGIOUS TOLERANCE IN SALATIGA CITY USING SUPPORT VECTOR MACHINE NIVE BAYES AND RANDOM FOREST

Yoel Agustian¹, Magdalena A. Ineke Pakereng²

Teknologi dan Informasi, Teknik Informatika, Universitas Kristen Satya Wacana, Salatiga, Indonesia¹

Teknologi dan Informasi, Teknik Informatika, Universitas Kristen Satya Wacana, Salatiga, Indonesia²

672022326@student.uksw.edu¹

ABSTRACT

The rapid development of social media provides opportunities for communities to express opinions on various social issues, including religious tolerance. This study aims to analyze public perceptions of religious tolerance in Salatiga City using a sentiment analysis approach based on machine learning. Research data were obtained from the Twitter platform through a crawling process, resulting in approximately 670 comments. The dataset was subsequently processed through text pre-processing stages, including cleaning, case folding, normalization, tokenization, and stemming, before feature extraction was performed using the TF-IDF method. Sentiment classification was conducted by comparing three algorithms, namely Support Vector Machine (SVM), Naive Bayes, and Random Forest. Model performance evaluation employed several metrics, including accuracy, precision, recall, and F1-score. The results indicate that public opinion is predominantly dominated by positive sentiment, reflecting Salatiga City's image as an area with a high level of religious tolerance. The Naive Bayes algorithm achieved the highest accuracy rate of 92%, while Random Forest demonstrated more stable classification performance in handling data imbalance. Furthermore, this study confirms that the application of text mining techniques can generate relevant digital data-based information to support an objective understanding of social conditions within society.

Keywords: Sentiment Analysis, Religious Tolerance, Machine Learning, Twitter, Text Mining.

ABSTRAK

Perkembangan media sosial memberikan ruang bagi masyarakat untuk menyampaikan opini terkait berbagai isu sosial, termasuk toleransi beragama. Penelitian ini bertujuan menganalisis persepsi masyarakat terhadap toleransi beragama di Kota Salatiga menggunakan pendekatan analisis sentimen berbasis *machine learning*. Data penelitian diperoleh dari platform Twitter melalui proses *crawling* dengan jumlah sekitar 670 komentar. *Dataset* kemudian diproses melalui tahapan *text pre-processing* yang meliputi *cleaning*, *case folding*, normalisasi, tokenisasi, dan *stemming* sebelum dilakukan ekstraksi fitur menggunakan metode *TF-IDF*. Proses klasifikasi sentimen dilakukan dengan membandingkan tiga algoritma, yaitu *Support Vector Machine (SVM)*, *Naive Bayes*, dan *Random Forest*. Evaluasi performa model menggunakan metrik akurasi, *precision*, *recall*, dan *F1-score*. Hasil penelitian menunjukkan bahwa opini masyarakat didominasi oleh sentimen positif yang merepresentasikan citra Kota Salatiga sebagai wilayah dengan tingkat toleransi beragama yang tinggi. *Naive Bayes* memperoleh nilai akurasi tertinggi sebesar 92%, namun *Random Forest* menunjukkan kemampuan klasifikasi yang lebih stabil dalam menangani ketidakseimbangan data. Penelitian ini menunjukkan bahwa penerapan teknik *text mining* mampu menghasilkan informasi berbasis data digital yang relevan untuk mendukung pemahaman kondisi sosial masyarakat secara objektif

Kata kunci: Analisis Sentimen, Toleransi Beragama, *Machine Learning*, Twitter.

PENDAHULUAN

Indonesia merupakan negara dengan tingkat keberagaman suku, agama, ras, dan antargolongan yang sangat tinggi. Terdapat enam agama yang diakui secara resmi, yaitu Islam, Kristen, Katolik, Hindu, Buddha, dan Kong Hu Cu. Setiap agama mengajarkan nilai-nilai kebaikan, termasuk

sikap saling menghormati, menghargai perbedaan, serta membangun toleransi antarumat beragama. Toleransi beragama merupakan bentuk interaksi sosial yang mencerminkan sikap saling menghormati dan menerima keberagaman keyakinan dalam kehidupan bermasyarakat (Fitriani, 2020). Implementasi toleransi dapat terlihat

dalam kehidupan sehari-hari, seperti menghargai pelaksanaan ibadah agama lain maupun melalui sikap bijak dalam menyikapi isu-isu SARA yang berkembang di media sosial.

Kota Salatiga dikenal sebagai salah satu kota yang memiliki tingkat keberagaman masyarakat yang tinggi dan sering memperoleh predikat sebagai kota dengan tingkat toleransi terbaik di Indonesia. Predikat tersebut didasarkan pada hasil pengukuran Indeks Kota Toleran (IKT) yang dilakukan oleh SETARA Institute serta berbagai penelitian yang menunjukkan bahwa masyarakat Salatiga mampu menjaga hubungan sosial yang harmonis di tengah perbedaan agama dan budaya (Sofyan & Badi'ati, 2022). Meskipun demikian, penelitian mengenai toleransi di Salatiga juga menunjukkan bahwa dinamika sosial dan potensi konflik tetap dapat muncul dalam kehidupan masyarakat sehingga kondisi toleransi perlu terus dijaga dan dievaluasi secara berkelanjutan.

Perkembangan teknologi informasi dan media sosial telah mengubah cara masyarakat menyampaikan opini terhadap berbagai isu sosial, termasuk isu keagamaan. Twitter (X) menjadi salah satu platform yang banyak digunakan masyarakat untuk mengekspresikan pandangan, pengalaman, maupun respons terhadap suatu peristiwa secara terbuka. Berdasarkan laporan Digital Indonesia 2023, jumlah pengguna media sosial aktif di Indonesia mencapai sekitar 167 juta pengguna (Simon Kemp, 2023). Besarnya jumlah pengguna media sosial menjadikan Twitter sebagai sumber data yang potensial untuk memahami persepsi masyarakat terhadap isu toleransi beragama. Selain itu, media sosial juga menjadi ruang interaksi yang berperan dalam membentuk dan menyebarkan narasi toleransi di masyarakat digital.

Meskipun berbagai survei dan penelitian telah menunjukkan bahwa Kota Salatiga memiliki tingkat toleransi yang tinggi, sebagian besar penilaian tersebut

diperoleh melalui survei lapangan maupun indikator kelembagaan (Sofyan & Badi'ati, 2022). Hingga saat ini masih terbatas penelitian yang mengkaji bagaimana persepsi masyarakat terhadap toleransi beragama di Kota Salatiga berdasarkan opini yang disampaikan secara langsung melalui media sosial (Riyan Hidayat & El-Feyza², n.d.). Di sisi lain, data yang dihasilkan dari media sosial memiliki karakteristik tidak terstruktur dan jumlah yang sangat besar sehingga sulit dianalisis secara manual (Panjaitan et al., 2025). Kondisi tersebut menyebabkan diperlukan suatu metode yang mampu mengolah data secara otomatis untuk mengidentifikasi kecenderungan opini masyarakat terhadap isu toleransi beragama secara lebih efektif dan berbasis data (Panjaitan et al., 2025) (Basri et al., 2024).

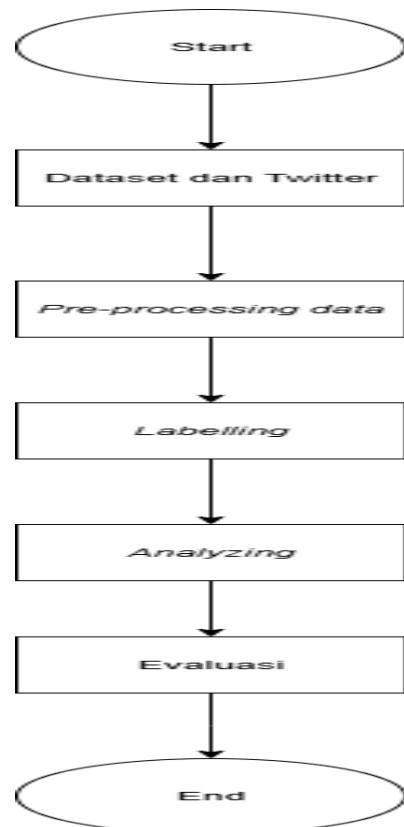
Salah satu pendekatan yang dapat digunakan untuk mengatasi permasalahan tersebut adalah analisis sentimen. Analisis sentimen merupakan teknik dalam bidang *text mining* dan *Natural Language Processing* (NLP) yang digunakan untuk mengidentifikasi dan mengklasifikasikan opini publik ke dalam kategori positif, negatif, maupun netral (Panjaitan et al., 2025). Melalui analisis sentimen, data teks yang berasal dari media sosial dapat diolah menjadi informasi yang lebih mudah dipahami dan digunakan sebagai dasar pengambilan keputusan.

Berbagai algoritma *machine learning* dapat digunakan dalam proses klasifikasi sentimen, di antaranya *Support Vector Machine* (SVM), *Random Forest* (RF), dan *Naive Bayes* (NB) (Panjaitan et al., 2025). Ketiga metode tersebut memiliki karakteristik yang berbeda. SVM dikenal efektif dalam menangani data berdimensi tinggi dan menghasilkan pemisahan kelas yang optimal, *Random Forest* mampu mengurangi risiko *overfitting* melalui kombinasi banyak pohon keputusan, sedangkan *Naive Bayes* memiliki keunggulan dalam kecepatan proses dan performa yang baik pada data teks (Basri et al., 2024). Oleh karena itu, penelitian ini

membandingkan ketiga algoritma tersebut untuk mengetahui metode yang paling efektif dalam mengklasifikasikan sentimen masyarakat terhadap toleransi beragama di Kota Salatiga berdasarkan data yang diperoleh dari Twitter.

Penelitian ini bertujuan untuk menganalisis persepsi masyarakat mengenai toleransi beragama di Kota Salatiga melalui pendekatan analisis sentimen serta membandingkan performa algoritma Support Vector Machine, Random Forest, dan Naive Bayes dalam proses klasifikasi sentimen. Hasil penelitian diharapkan dapat memberikan gambaran mengenai kondisi persepsi masyarakat terhadap toleransi beragama berdasarkan data digital yang dihasilkan secara langsung oleh pengguna media sosial. Selain itu, informasi yang diperoleh dapat dimanfaatkan oleh pemerintah daerah, akademisi, maupun organisasi sosial sebagai bahan evaluasi dan pendukung pengambilan kebijakan dalam menjaga kerukunan serta memperkuat toleransi antarumat beragama di era digital (Fitriani, 2020).

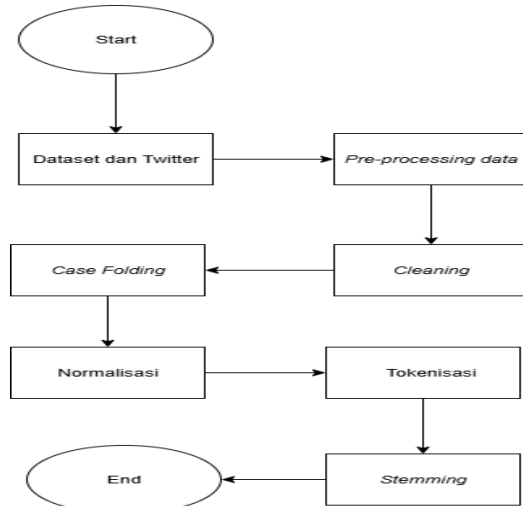
METODE



Gambar 1. Tahapan Penelitian

Tahap pertama adalah pengumpulan dataset, yaitu memperoleh data melalui proses crawling komentar pada platform Twitter (X) yang berkaitan dengan toleransi beragama di Kota Salatiga. Data yang berhasil dikumpulkan berjumlah sekitar 670 komentar dan disimpan dalam format CSV. Tahap kedua adalah pre-processing data, yaitu membersihkan dan mempersiapkan data teks melalui beberapa proses, meliputi cleaning, case folding, normalisasi, tokenisasi, dan stemming sehingga menghasilkan data yang lebih bersih dan siap digunakan dalam proses klasifikasi (Syahira & Kurniawan, 2024). Tahap ketiga adalah pelabelan data (labelling), di mana setiap tweet yang telah diproses diberikan label sentimen sesuai kategorinya, yaitu positif, negatif, atau netral. Tahap keempat adalah analisis data (analyzing), yaitu merepresentasikan data teks menggunakan metode TF-IDF (Term Frequency-Inverse Document Frequency), kemudian melakukan klasifikasi sentimen dengan algoritma Support Vector Machine (SVM), Naive Bayes (NB), dan Random Forest (RF) (Sriani et al., 2024). Tahap

terakhir adalah evaluasi, yang dilakukan untuk mengukur dan membandingkan kinerja masing-masing algoritma menggunakan metrik accuracy, precision, recall, dan F1-score, sehingga dapat ditentukan model yang memiliki performa terbaik dalam melakukan klasifikasi sentimen.



Gambar 2. Flowchart Pre-Processing

Tahap pre-processing data dilakukan untuk membersihkan dan mempersiapkan data teks agar dapat digunakan secara optimal dalam proses analisis sentimen. Proses ini diawali dengan cleaning, yaitu membersihkan data mentah dari elemen-elemen yang tidak diperlukan seperti karakter khusus, URL, angka, dan simbol yang tidak relevan. Selanjutnya dilakukan case folding, yaitu mengubah seluruh teks menjadi huruf kecil untuk mengurangi variasi kata yang disebabkan oleh perbedaan penggunaan huruf kapital (Shinta Angreni et al., 2025). Tahap berikutnya adalah normalisasi, yang bertujuan mengubah kata-kata slang, singkatan, atau bahasa tidak baku menjadi bentuk kata baku yang umum digunakan (Shinta Angreni et al., 2025). Setelah itu dilakukan tokenisasi, yaitu proses memisahkan teks menjadi kata-kata individual (token) yang dapat digunakan dalam analisis lebih lanjut. Tahap terakhir adalah stemming, yaitu mengubah setiap kata ke dalam bentuk dasarnya (stem word), misalnya kata "bermain" diubah menjadi "main", sehingga dapat mengurangi variasi

kata yang memiliki makna yang sama (Savitri et al., 2024).

2.1 Support Vector Machine (SVM)

Support Vector Machine (SVM) adalah algoritma pembelajaran mesin yang digunakan secara luas dalam penerapan analisis sentimen. SVM bekerja dengan mencari Hyperplane optimal yang memisahkan dua kelas data (Positif dan negatif) Dengan margin terbesar. Pada SVM, hyperplane terbaik adalah yang memaksimalkan jarak antara data dari kedua kelas (Anggraini & Alita, 2024).

Rumus sederhana model SVM :

$$w \cdot x_i + b = 0 \text{ for } y_i = 0 \text{ (hyperline)}$$

$$w \cdot x_i + b \geq 0 \text{ for } y_i = +1 \text{ (positive class)}$$

$$w \cdot x_i \leq 0 \text{ for } y_i = -1 \text{ (negative class)}$$

Kemudian penentuan nilai margin terbesar dari kedua grup dengan notasi berikut:

$$\frac{1 - b - (1 - b)}{w} = \frac{1}{|w|}$$

Keterangan:

w = berat, berbentuk vector dari kemiringan hyperplane suatu training model

x = data(input), titik data yang dipakai

b = bias, penggeser hyperplane layaknya pembiasan.

Model SVM ini akan menentukan nilai w dan b yang memberikan margin maksimal antara dua kelas data SVM berhasil digunakan dalam analisis sentimen pemilihan presiden 2024 pada aplikasi Twitter dengan tingkat akurasi mencapai 95,43% menggunakan kernel linier. Hasil serupa diperoleh pada penelitian yang melaporkan SVM mampu mengklasifikasikan opini publik dengan tingkat akurasi 94% dan F1-score 96%. Hal ini menunjukkan bahwa SVM efektif dalam menangani data berdimensi tinggi dan memberikan hasil klasifikasi yang stabil (Anggraini & Alita, 2024).

2.2 Naive Bayes (NB)

Naive Bayes merupakan metode pengklasifikasi probabilitas dan statistik yang dikembangkan berdasarkan Teorema Bayes. Metode ini bekerja untuk memprediksi kemungkinan suatu data termasuk kedalam kelas tertentu berdasarkan pengalaman atau pola dari data sebelumnya, Naive Bayes menghitung peluang kemunculan kata terhadap kategori sentimen tertentu untuk menentukan apakah sebuah teks bersifat positif negatif atau netral(Purnomo et al., n.d.). Berikut rumus dasar Naive Bayes sebagai berikut :

$$P(c|d) = \frac{P(c) \prod_{i=1}^n P(x_i|c)}{P(d)}$$

Keterangan :

$P(c|d)$: probabilitas dokumen d termasuk kelas c

$P(c)$: Probabilitas awal kelas c

$P(x_i|c)$: Probabilitas kelas ke-i Muncul pada kelas c

$P(d)$: probabilitas dokumen d.

Dalam penerapan nya, nilai $P(d)$ dianggap konstan, sehingga prediksi kelas diperoleh dengan mencari nilai maksimum dari :

$$\hat{c} = \arg \max_c P(c) \prod_{i=1}^n P(x_i|c)$$

Algoritma ini dikenal dengan efisiensinya pada data set besar dan menghasilkan performa yang kompetitif dengan waktu untuk melatih algoritma yang relatif cepat. Menurut penelitian Supian et al. (2024), Naive Bayes dapat mencapai akurasi lebih dari 80% dalam analisis. Karena kemudahan dan kecepatan prosesnya, algoritma ini banyak dijadikan pembandingan dalam penelitian analisis sentimen(Kinerja Naïve Bayes et al., n.d.).

2.3 Random Forest

Random Forest merupakan algoritma yang *ensemble learning* yang dimana algoritma ini menggabungkan banyak *decision tree* untuk meningkatkan akurasi dan mengurangi resiko overfitting. Setiap pohon keputusan dibangun dari subset acak data dan subset acak fitur. Hasil akhir dari klasifikasi ditentukan melalui mekanisme

majority voting atau pemilihan suara terbanyak dari seluruh pohon keputusan. rumus prediksi akhir random forest(Kuyoro et al., 2022).

$$\hat{y} = \text{mode}\{h_1(x), h_2(x), \dots, h_k(x)\}$$

Dengan :

$h_i(x)$: hasil prediksi dari decision tree ke i.

Mode : kelas yang paling sering muncul dari seluruh hasil pohon.

Selain untuk memilih pemisahan terbaik pada setiap node, random forest menggunakan ukuran ketidak murnian seperti *Gini Index* yang dirumuskan sebagai berikut:

$$Gini(D) = 1 - \sum_{i=1}^m p_i^2$$

Dengan P_i adalah proporsi data pada kelas ke-i dalam dataset D. Semakin kecil nilai yang dihasilkan Gini semakin baik pemisahan yang diperoleh(Kuyoro et al., 2022).

Menurut penelitian Falah Penerapan algoritma Random forest pada sentimen Tweet tentang isu ginjal akut menghasilkan akurasi yang tinggi sebesar 94,47% dimana algoritma ini mampu mengolah data teks(Falah et al., 2024), sementara penelitian Ningrum menemukan bahwa performa algoritma Random Forest mampu menganalisis opini dimana kedua penelitian ini mampu menegaskan bahwa algoritma Random Forest mampu menghasilkan tingkat prediksi yang stabil, dan tingkat kinerja yang konsisten dengan menggunakan data yang besar(Ningrum, 2024).

Setelah berbagai proses dibutuhkan beberapa nilai untuk keperluan evaluasi model pada langkah akhir dengan rumus sebagai berikut:

$$\text{Accuracy} = \frac{(TP+TN)}{(TP+FP+TN+FN)}$$

$$\text{Precision} = \frac{TP}{(TP+FP)}$$

$$\text{Recall} = \frac{TP}{(TP+FN)}$$

$$F1 \text{ Score} = \frac{2 \times (\text{Recall} \times \text{Precision})}{(\text{Recall} + \text{Precision})}$$

Keterangan:

TP = True Positive,
jumlah prediksi positif yang benar
TN = True Negative,
jumlah prediksi negatif yang benar
FP = False Positive,
jumlah prediksi positif yang salah
FN = False Negative,
jumlah prediksi negatif yang salah (Falah et al., 2024).

HASIL DAN PEMBAHASAN

3.1 Dataset

Pengambilan dataset dilakukan dengan metode crawling terhadap aplikasi twitter dengan total data yang didapatkan sebanyak ± 670 data.

Bentuk sederhana dari crawling itu sendiri dengan memasukan auth-token dari twitter yang sedang dibuka, kemudian menggunakan node.js untuk membiarkan program masuk kehalaman utama twitter, lalu diberi kata kunci (toleransi salatiga), dan program sederhananya akan melakukan scrolling layar sampai komentar terbawah hingga habis.

Tabel 1. Tabel Hasil Crawling

conversation_id	created_at	favorite_count	full_text	id_str	image_url	in_reply_to_lang	location	quote_count	reply_count	retweet_count	tweet_url	user_id_str	username
#####	Sun Jun 08	300%	@jawafess #####	#####	#####	in		0%	200%	0%	https://x.cc/#####	#####	
#####	Wed Jun 04	3200%	Toleransi b #####	#####	#####	https://pbs.twimg.com/in		200%	100%	1300%	https://x.cc/#####	#####	
#####	Mon Jun 02	500%	Kota Salatiga #####	#####	#####	https://pbs.twimg.com/in		0%	100%	200%	https://x.cc/#####	#####	
#####	Thu May 21	13800%	Selamat ur #####	#####	#####	https://pbs.twimg.com/in		0%	400%	3200%	https://x.cc/#####	#####	
#####	Thu May 21	100%	@twotwof #####	#####	#####	twotwof/in		0%	0%	0%	https://x.cc/#####	#####	
#####	Wed May 2	1800%	Salatiga dr #####	#####	#####	https://pbs.twimg.com/in		400%	0%	700%	https://x.cc/#####	#####	
#####	Wed May 2	0%	[SALATIGA] #####	#####	#####	https://pbs.twimg.com/in		0%	0%	0%	https://x.cc/#####	#####	
#####	Wed May 2	0%	Selamat! Si #####	#####	#####	https://pbs.twimg.com/in		0%	0%	0%	https://x.cc/#####	#####	

3.2 Pre-processing data

Tahap *pre-processing* dilakukan untuk meningkatkan kualitas data teks sebelum digunakan dalam proses klasifikasi sentimen. Data hasil *crawling* masih mengandung berbagai elemen yang tidak relevan seperti URL, simbol, angka, dan variasi penulisan kata yang dapat memengaruhi performa model. Oleh karena itu dilakukan beberapa tahapan, yaitu *cleaning* dan *case folding*, normalisasi, serta tokenisasi dan *stemming*.

3.2.1. Cleaning dan Case Folding

Tahap *cleaning* dan *case folding* diawali dengan proses identifikasi karakter yang tidak diperlukan seperti URL, *hashtag*, *mention*, angka, tanda baca, emoji,

dan karakter khusus menggunakan pola *Regular Expression* (Regex). Seluruh karakter yang sesuai dengan pola tersebut kemudian dihapus dari dokumen. Setelah proses pembersihan selesai, sistem mengubah seluruh huruf menjadi huruf kecil (*lowercase*) menggunakan fungsi *lower()* sehingga kata yang memiliki makna sama tetapi berbeda penggunaan huruf kapital dapat dianggap sebagai satu representasi yang sama. Sebagai contoh, kata "Salatiga", "SALATIGA", dan "salatiga" akan diubah menjadi "salatiga".

Tabel 2. Tabel Hasil Cleaning dan Case Folding

Text	Cleaning dan Case Folding
Toleransi berperan penting dalam harmoni inklusivitas menjadi faktor yang tak bisa diabaikan dari kesuksesan pembangunan daerah. Dari sederet kota di Jawa Salatiga saat ini memimpin sebagai kota paling toleran berdasarkan Indeks Kota Toleran 2024. Kota ini juga https://t.co/dznPkAe0kL	toleransi berperan penting dalam menjaga harmoni dan inklusivitas menjadi faktor yang tak bisa diabaikan dari kesuksesan pembangunan daerah dari sederet kota di jawa salatiga saat ini memimpin sebagai kota paling toleran berdasarkan indeks kota toleran kota ini juga

3.2.2. Normalisasi

Proses normalisasi dilakukan dengan membandingkan setiap kata hasil *cleaning* terhadap kamus normalisasi yang berisi pasangan kata tidak baku dan kata baku. Jika ditemukan kata yang terdapat dalam

kamus, sistem akan mengganti kata tersebut dengan bentuk bakunya. Sebagai contoh, kata "gk" akan diubah menjadi "tidak", sedangkan kata "yg" akan diubah menjadi "yang". Tahapan ini bertujuan mengurangi variasi penulisan kata yang memiliki makna sama.

Tabel 3. Tabel Hasil Normalisasi

Cleaning dan Case Folding	Normalisasi
toleransi berperan penting dalam menjaga harmoni dan inklusivitas menjadi faktor yang tak bisa diabaikan dari kesuksesan pembangunan daerah dari sederet kota di jawa salatiga saat ini memimpin sebagai kota paling toleran berdasarkan indeks kota toleran kota ini juga	toleransi berperan penting dalam menjaga harmoni dan inklusivitas menjadi faktor yang tak bisa diabaikan dari kesuksesan pembangunan daerah dari sedekat kota di jawa salatiga saat ini memimpin sebagai kota paling toleransi berdasarkan indeks kota toleransi kota ini juga

3.2.3. Tokenisasi dan Stemming

Tahap tokenisasi dilakukan dengan memecah setiap kalimat menjadi unit-unit kata berdasarkan spasi (*whitespace tokenization*). Sebagai contoh, kalimat "salatiga kota toleran" akan diubah menjadi token ["salatiga", "kota", "toleran"]. Selanjutnya, setiap token diproses menggunakan algoritma Nazief-Adriani melalui pustaka Sastrawi untuk menghilangkan imbuhan dan memperoleh kata dasar. Algoritma ini bekerja dengan menghapus prefiks, sufiks, konfiks, serta partikel sesuai aturan morfologi Bahasa Indonesia. Sebagai contoh, kata "menjalankan", "dijalankan", dan "berjalan" akan direduksi menjadi kata dasar "jalan". Hasil dari proses ini berupa kumpulan kata dasar yang lebih konsisten

untuk digunakan pada tahap ekstraksi fitur.

Tabel 4. Tabel Hasil Tokenisasi dan Stemming

Normalisasi	Tokenisasi	Stemming
toleransi berperan penting dalam menjaga harmoni dan inklusivitas menjadi faktor yang tak bisa diabaikan n dari kesuksesan pembangunan daerah dari sedekat kota di jawa salatiga saat ini memimpin sebagai kota paling toleransi berdasar kan indeks kota toleransi kota ini juga	toleransi,ber peran,menjag a, harmoni,inkl usivitas,fakto r, diabaikan,ke suksesan, pembanguna n,daerah, sedekat,kota, jawa,salatiga, memimpin,k ota,toleransi, berdasarkan,i ndeks,kota, toleransi,kota	toleransi peran jaga harmoni inklusivitas faktor abai sukses bangun daerah dekat kota jawa salatiga pimpin kota toleransi dasar indeks kota toleransi kota

3.3. Labelling

Setelah data dibersihkan, dilakukan proses pelabelan (*labelling*) untuk mengategorikan setiap opini ke dalam tiga kelas sentimen: Positif, Negatif, dan Netral. Penentuan label didasarkan pada keberadaan kata kunci yang

merepresentasikan nilai toleransi dalam teks tersebut.

Tabel 5. Tabel Hasil Labeling

Negatif	Netral	Positif
Salatiga langganan peringkat 1 atau 2.	Salatiga Tengah menjadi tujuan	Jawa terpilih kota program
Percayalah kami orang baik meski kadang disalah mengerti	Peace Indonesia yg digelar akhir pekan kemarin. Apa yang dipelajari dalam program #PeaceTrain? Bagaimana kota Salatiga merawat kebhinekaan? Kami bahas di#WhatsTrending	Toleransi berperan penting dalam menjaga harmoni dan inklusivitas menjadi faktor yang tak bisa diabaikan dari kesuksesan pembangunan daerah. Dari sederet kota di Jawa Salatiga saat ini memimpin sebagai kota paling toleran berdasarkan Indeks Kota Toleran 2024. Kota ini juga

3.4. Analysis

Dengan data yang sudah berlabel, tahap processing dilakukan melalui dua langkah utama, yaitu ekstraksi fitur menggunakan metode TF-IDF dan klasifikasi sentimen menggunakan tiga algoritma machine learning. Berikut penjelasan langkah kerja masing-masing tahapan.

3.4.1. Ekstraksi Fitur (TF-IDF)

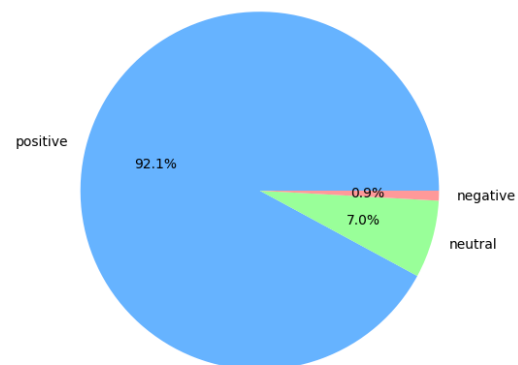
Data teks hasil pre-processing direpresentasikan ke dalam bentuk vektor numerik menggunakan metode TF-IDF. Pada tahap ini, setiap dokumen dihitung nilai TF-nya berdasarkan frekuensi kemunculan kata, kemudian dikombinasikan dengan nilai IDF yang mengukur seberapa jarang kata tersebut muncul di seluruh koleksi dokumen. Kata-kata yang terlalu umum secara otomatis mendapat bobot lebih rendah, sehingga fitur yang dihasilkan lebih representatif. Hasil perkalian TF dan IDF membentuk matriks fitur numerik yang selanjutnya digunakan sebagai masukan algoritma klasifikasi. Dataset kemudian dibagi menjadi data latih dan data uji dengan rasio

80:20 menggunakan train_test_split dari pustaka Scikit-learn.

3.4.2. Klasifikasi Sentimen

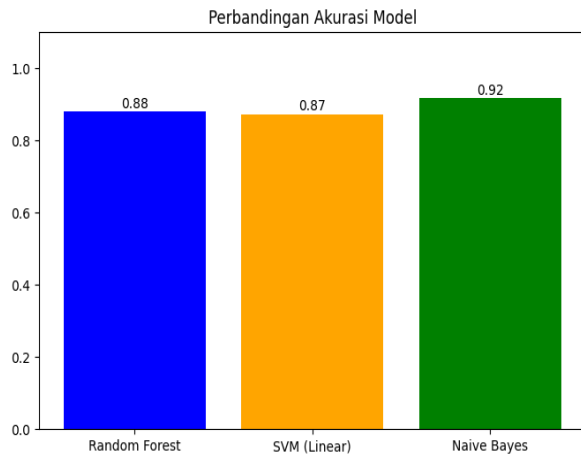
Klasifikasi sentimen dilakukan menggunakan tiga algoritma dengan langkah kerja sebagai berikut. Pertama, Support Vector Machine (SVM): matriks TF-IDF dari data latih dimasukkan ke model SVM dengan kernel linear; algoritma mencari hyperplane optimal yang memisahkan kelas sentimen positif, negatif, dan netral dengan margin terbesar; kemudian model yang telah dilatih digunakan untuk memprediksi label sentimen pada data uji berdasarkan posisi titik data terhadap hyperplane yang ditemukan. Kedua, Naive Bayes: model menghitung probabilitas prior $P(c)$ untuk setiap kelas sentimen berdasarkan data latih, sedangkan untuk setiap kata dalam data uji dihitung probabilitas kondisional $P(x_i|c)$, serta prediksi kelas diperoleh dengan memilih kelas yang menghasilkan nilai posterior $P(c|d)$ tertinggi. Ketiga, Random Forest: sejumlah decision tree dibangun secara paralel menggunakan subset acak data latih dan subset acak fitur TF-IDF; setiap pohon memberikan prediksi kelas sentimen secara independen; prediksi akhir ditentukan melalui mekanisme majority voting dari seluruh pohon. Seluruh proses klasifikasi diimplementasikan menggunakan pustaka Scikit-learn dalam bahasa pemrograman Python.

Distribusi Sentimen (Final)



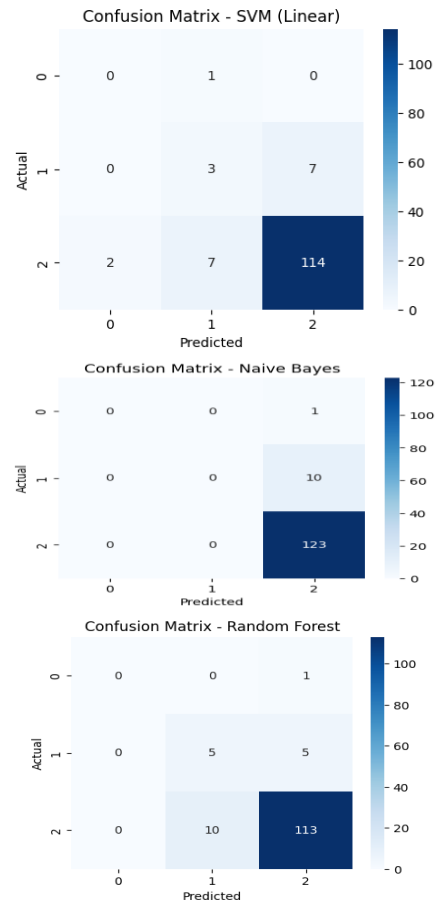
Gambar 3. Pie Chart dari Distribusi Sentimen

Hasil pada Gambar 3 menunjukkan dominasi sentimen Positif sebesar 92,1%. Hal ini mengindikasikan bahwa mayoritas opini publik di media sosial merefleksikan citra Kota Salatiga sebagai kota dengan tingkat toleransi yang sangat tinggi



Gambar 4. Perbandingan Tingkat Akurasi Model

Gambar 4 menyajikan perbandingan tingkat akurasi dari tiga algoritma klasifikasi yang diimplementasikan dalam penelitian ini. Berdasarkan hasil pengujian, Naive Bayes menunjukkan performa yang paling unggul dengan capaian akurasi tertinggi sebesar 0.92. Hal ini mengindikasikan bahwa asumsi independensi fitur pada *Naive Bayes* bekerja sangat efektif dalam mengenali pola kata pada dataset sentimen toleransi beragama di Salatiga. Sementara itu, Random Forest menempati posisi kedua dengan skor akurasi 0.88, yang menunjukkan kemampuan metode *ensemble* dalam menjaga stabilitas prediksi tetap baik. Di sisi lain, *Support Vector Machine* (SVM) mencatatkan nilai akurasi terendah dibandingkan kedua model lainnya, yaitu sebesar 0.87. Secara keseluruhan, visualisasi ini menegaskan bahwa untuk karakteristik data teks yang digunakan dalam penelitian ini, algoritma Naive Bayes merupakan model yang paling optimal dan akurat dalam mengklasifikasikan sentimen masyarakat.



Gambar 5 Confusion Matrix

Berdasarkan hasil evaluasi pada Gambar 5, terlihat bahwa ketidakseimbangan data sangat memengaruhi cara kerja ketiga model dalam mengklasifikasikan sentimen. Naive Bayes memang meraih akurasi tertinggi (92%), namun hasil ini kurang akurat secara fungsi karena model mengalami "bias" dengan hanya menebak kategori positif untuk seluruh data (123 data). Akibatnya, Naive Bayes gagal total dalam mengenali sentimen negatif maupun netral.

Di sisi lain, SVM menunjukkan kelebihan dalam mengenali kelompok kecil dengan mendeteksi 3 data netral secara benar, meskipun ia tetap gagal pada kategori negatif. Sementara itu, Random Forest terbukti sebagai model yang paling cerdas dan stabil dalam penelitian ini. Model ini mampu memprediksi 113 data positif dan 5 data netral dengan tepat, menjadikannya pilihan yang paling realistis. Meski semua model masih kesulitan mendeteksi sentimen negatif yang jumlahnya sangat sedikit, Random Forest

tetap menjadi algoritma paling optimal karena memberikan hasil yang paling seimbang dan mendekati kondisi asli data di lapangan.

4.5. Evaluasi

Pada tahap ini, dilakukan evaluasi mendalam terhadap model-model yang telah dilatih untuk menentukan algoritma dengan performa paling optimal dalam mengklasifikasikan sentimen terkait tingkatan toleransi beragama

Tabel 6. Hasil Evaluasi Model

Peringkat	Model	Akurasi test	Rata rata cross val	F1_Score
1	Random forest	0.88	0.92	0.88
2	SVM	0.87	0.89	0.87
3	Naive Bayes	0.91	0.92	0.87

"Data yang disajikan pada Tabel 6 mengonfirmasi bahwa *Naive Bayes* mencapai performa optimal dengan akurasi pengujian sebesar 91% dan rata-rata *Cross-Validation* sebesar 92%, mengungguli algoritma SVM dan *Random Forest*. Oleh karena itu, *Naive Bayes* terbukti sebagai metode yang paling efektif dalam mengklasifikasikan sentimen terkait toleransi beragama dalam ekosistem digital."

SIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa analisis sentimen terhadap opini masyarakat di media sosial Twitter mampu memberikan gambaran mengenai persepsi publik terhadap tingkat toleransi beragama di Kota Salatiga. Hasil analisis menunjukkan bahwa sebagian besar komentar masyarakat didominasi oleh sentimen positif, yang menandakan bahwa Salatiga masih dipandang sebagai kota dengan kehidupan sosial yang harmonis serta memiliki tingkat penerimaan antarumat beragama yang baik. Informasi ini relevan dan dibutuhkan oleh pemerintah daerah Kota Salatiga, akademisi, serta organisasi kemasyarakatan sebagai dasar evaluasi dan pengambilan kebijakan dalam

mempertahankan predikat kota toleran. Urgensi dari temuan ini terletak pada kebutuhan akan data berbasis opini publik digital yang bersifat real-time dan objektif, mengingat penilaian toleransi selama ini masih didominasi oleh survei konvensional dan indikator kelembagaan yang belum sepenuhnya mencerminkan dinamika persepsi masyarakat di ruang digital. Penerapan tiga algoritma *machine learning*, yaitu *Support Vector Machine* (SVM), *Naive Bayes*, dan *Random Forest*, memperlihatkan perbedaan kinerja dalam proses klasifikasi sentimen. *Naive Bayes* memperoleh nilai akurasi tertinggi dalam pengujian model, namun performanya kurang optimal dalam mengenali variasi kelas sentimen akibat ketidakseimbangan distribusi data. Sementara itu, SVM menunjukkan kemampuan klasifikasi yang cukup baik meskipun belum mampu mengidentifikasi seluruh kategori sentimen secara maksimal. *Random Forest* menunjukkan performa yang lebih stabil karena mampu menghasilkan prediksi yang lebih seimbang pada beberapa kategori sentimen, sehingga dinilai lebih representatif dalam menggambarkan kondisi data yang sebenarnya. Secara keseluruhan, penelitian ini membuktikan bahwa pendekatan text mining dan *machine learning* efektif digunakan untuk menganalisis opini masyarakat berbasis data digital, serta dapat dimanfaatkan sebagai sumber informasi dalam mendukung upaya menjaga dan meningkatkan nilai toleransi beragama di lingkungan masyarakat modern.

DAFTAR PUSTAKA

- Anggraini, J., & Alita, D. (2024). Implementasi Metode SVM Pada Sentimen Analisis Terhadap Pemilihan Presiden (Pilpres) 2024 Di Twitter. *Jurnal Informatika: Jurnal Pengembangan IT*, 9(2), 102–111. <https://doi.org/10.30591/jpit.v9i2.6560>
- Basri, H., Junianto, M. B. S., & Kusyad, I. (2024). Enhancing Usability Testing

- Through Sentiment Analysis: A Comparative Study Using SVM, Naive Bayes, Decision Trees and Random Forest. *Jurnal Teknologi Sistem Informasi Dan Aplikasi*, 7(4), 1603–1610.
<https://doi.org/10.32493/jtsi.v7i4.45117>
- Falah, H. M., Jamil, M. R., Taufik, A., Botha, M., & Agustina, N. (2024). Analysis Sentiment Terhadap Ginjal Akut pada Twitter Menggunakan Algoritma Random Forest. *Jurnal Ilmu Komputer Dan Informatika*, 3(2), 99–106.
<https://doi.org/10.54082/jiki.65>
- Fitriani, S. (2020). Analisis: Jurnal Studi Keislaman Keberagaman dan Toleransi Antar Umat Beragama. *Jurnal Studi Keislaman*, 20(2), 179–192. <https://doi.org/10.24042/ajsk>
- Kinerja Naïve Bayes, P., Supian, A., Tri Revaldo, B., Marhadi, N., & Efrizoni, L. (n.d.). *Acuan Supian Perbandingan Kinerja Naïve Bayes dan SVM pada Analisis Sentimen Twitter Ibukota Nusantara*. Retrieved <https://github.com/syenirasheila/Sentiment-Analysis-IKN->
- Kuyoro, A. O., Ogunyolu, O. A., Ayanwola, T. G., & Ayankoya, F. Y. (2022). Dynamic Effectiveness of Random Forest Algorithm in Financial Credit Risk Management for Improving Output Accuracy and Loan Classification Prediction. *Ingenierie Des Systemes d'Information*, 27(5), 815–821.
<https://doi.org/10.18280/isi.270515>
- Ningrum, A. F. (2024). Sentiment Analysis of Public Opinion on Handling Stunting in Indonesia using Random Forest. *Jurnal Statistika Dan Aplikasinya*, 8(1), 31–40.
<https://doi.org/10.21009/jsa.08103>
- Panjaitan, F., Ce, W., Oktafiandy, H., Kanugrahan, G., Ramdhani, Y., Hafizh, V., Putra, C., & Permai, A. (2025). STUDI KOMPARATIF ALGORITMA MACHINE LEARNING PADA ANALISIS SENTIMEN MEDIA SOSIAL. In *Jurnal Mahasiswa Teknik Informatika* (Vol. 9, Number 2).
- Purnomo, H., Arifin, O., Arpan, A., Permata, R., & Handoko, D. (n.d.). EVALUASI KINERJA ALGORITMA NAÏVE BAYES, KNN, DAN SVM DALAM ANALISIS SENTIMEN MEDIA SOSIAL. In *Atika Arpan* (Vol. 9, Number 2). Dwi Handoko.
- Riyan Hidayat, M., & El-Feyza², M. (n.d.). *From Pixels to Peace: Unraveling the Multifaceted Narrative of Religious Tolerance in Social Media*. <https://doi.org/10.15575/jw.v7i2.22629>
- Savitri, P. R. A., Suarjaya, I. M. A. D., & Vihikan, W. O. (2024). Sentiment Analysis of X (Twitter) Comments on The Influence of South Korean Culture in Indonesia. *Journal of Information Systems and Informatics*, 6(2), 979–991.
<https://doi.org/10.51519/journalisi.v6i2.749>
- Shinta Angreni, D., Khaerawati Nur, S., & Luvi Jayanto, D. (2025). Sentiment Analysis for the 2024 DKI Jakarta Gubernatorial Election Using a Support Vector Machine Approach. In *Journal of Applied Informatics and Computing (JAIC)* (Vol. 9, Number 2).
<http://jurnal.polibatam.ac.id/index.php/JAIC>
- Simon kemp. (2023, February 9). *Digital 2023: Indonesia*. DATAREPORTAL.
<https://datareportal.com/reports/digital-2023-indonesia>
- Sofyan, M. A., & Badi'ati, A. Q. (2022). Reexamining Salatiga as a Tolerant City in Java: An Exposure of Lesser-Known Facts. *Walisongo: Jurnal Penelitian Sosial Keagamaan*, 30(2), 241–262.
<https://doi.org/10.21580/ws.30.2.15010>

- Sriani, Lubis, A. H., & Lubis, L. P. A. (2024). Sentiment analysis on twitter about the death penalty using the support vector machine method. *TEKNOSAINS: Jurnal Sains, Teknologi Dan Informatika*, 11(2), 312–321.
<https://doi.org/10.37373/tekno.v11i2.1096>
- Syahira, M. A., & Kurniawan, R. (2024). Analisis Sentimen Cyberbullying Pada Media Sosial X Menggunakan Metode Support Vector Machine. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 8(3), 1724.
<https://doi.org/10.30865/mib.v8i3.7926>