

Penerapan Algoritma Multinomial Naïve Bayes dan Support Vector Machine (SVM) untuk Analisis Sentimen terhadap Ulasan Google Maps di Taman Mini Indonesia Indah (TMII)

Dwi Alfiyanti

Fakultas Teknologi Informasi, Program Studi Teknik Informatika
Universitas Budi Luhur
Email: 1911501011@student.budiluhur.ac.id

Indra

Fakultas Teknologi Informasi, Program Studi Teknik Informatika
Universitas Budi Luhur
Email: indra@budiluhur.ac.id

ABSTRAK

Taman Mini Indonesia Indah (TMII) adalah salah satu objek wisata di Jakarta yang sangat populer di kalangan wisatawan. Namun, TMII mengalami penurunan kunjungan yang signifikan pada tahun 2021. Oleh karena itu, pengelola ingin mengetahui pendapat pengunjung dari ulasan yang diberikan melalui situs *Google Maps* untuk meningkatkan daya tarik dan evaluasi layanan bagi pengelola. Untuk itu, dilakukan analisis sentimen terhadap data ulasan dengan menghasilkan dua kelas sentimen, yaitu kelas positif dan negatif. Analisis sentimen ini menggunakan metode *Multinomial Naive Bayes* dan *Support Vector Machine (SVM)*, yang merupakan salah satu metode dalam text mining yang digunakan dalam proses pengklasifikasian teks. Dalam pengujian yang dilakukan menggunakan metode *Multinomial Naive Bayes*, diperoleh hasil analisis dengan tingkat akurasi sebesar 87%, presisi sebesar 87% dan *recall* sebesar 100%. sedangkan metode SVM diperoleh tingkat akurasi sebesar 87%, presisi sebesar 88% dan *recall* sebesar 99%. Pengujian tersebut menggunakan 400 data *training* serta menggunakan 100 data *testing*. Berdasarkan hasil pengujian tersebut, dapat disimpulkan bahwa pilihan antara keduanya tergantung pada prioritas. SVM lebih cocok jika presisi prediksi positif lebih penting, sementara *Multinomial Naive Bayes* lebih baik jika mengidentifikasi semua *instance* dari kelas positif menjadi prioritas utama. Penelitian tentang sistem analisis sentimen terhadap data ulasan di *Google Maps* menunjukkan bahwa SVM merupakan algoritma yang lebih efektif dalam mengklasifikasikan sentimen ulasan *Google Maps*. Selain itu, penelitian ini memberikan gambaran umum tentang ulasan dari pengunjung mengenai kualitas dan layanan tempat wisata Taman Mini Indonesia Indah.

Kata kunci: *Multinomial Naive Bayes*, Taman Mini Indonesia Indah, Analisis Sentimen, Ulasan Google Maps

ABSTRACT

Taman Mini Indonesia Indah (TMII) is one of the tourist attractions in Jakarta that is very popular with tourists. However, TMII experienced a significant decline in visits in 2021. Therefore, the manager wants to know the opinions of visitors from the reviews given through the google maps site to increase attractiveness and service evaluation for the manager. For this reason, sentiment analysis of review data is carried out by producing two sentiment classes, namely positive and negative classes. This sentiment analysis uses the Multinomial Naive Bayes method and Support Vector Machine (SVM), which is one of the methods in text mining used in the text classification process. In tests conducted using the Multinomial Naive Bayes method, analysis results were obtained with an accuracy rate of 87%, precision of 87% and recall of 100%. while the SVM method obtained an accuracy rate of 87%, precision of 88% and recall of 99%. The test uses 400 training data and uses 100 testing data. Based on the test results, it can be concluded that the choice between the two depends on the priority: SVM is more suitable if the precision of positive predictions is more important, while Multinomial Naive Bayes is better if identifying all instances of the positive class is a top priority. Research on sentiment analysis systems for review data on Google Maps shows

that SVM is a more effective algorithm in classifying the sentiment of Google Maps reviews. In addition, this research provides an overview of reviews from visitors regarding the quality and services of Taman Mini Indonesia Indah tourist attractions.

Keywords: *Multinomial Naïve Bayes, Taman Mini Indonesia Indah, Analysis Sentiment, Google Maps Review.*

1. PENDAHULUAN

Pariwisata merupakan salah satu upaya yang dilakukan dengan skala besar untuk menjadi sumber utama devisa bagi setiap negara. Namun, tiga tahun pandemi Covid-19 telah memberikan dampak besar pada industri pariwisata dengan kebijakan tinggal di rumah yang menghambat mobilitas. Setelah penurunan kasus virus Covid-19, industri pariwisata mulai bangkit kembali dan destinasi pariwisata seperti Taman Mini Indonesia Indah (TMII) di Jakarta berusaha pulih dari dampak pandemi. Sebelum pandemi, TMII sangat populer di kalangan wisatawan, namun pada tahun 2021, jumlah pengunjung mengalami penurunan drastis. Setelah selesai menjalani proses renovasi pada pertengahan Oktober 2022, diperlukan pendapat dari pengunjung untuk meningkatkan daya tarik wisata TMII dan sebagai bahan evaluasi bagi pengelola taman. Oleh karena itu, penelitian ini menggunakan Google Review. Google Review adalah fitur yang disediakan oleh Google sebagai bagian dari layanan mereka untuk memberikan ulasan atau tinjauan pengguna terhadap berbagai bisnis, tempat, produk, atau layanan yang terdaftar di Google Maps atau Google My Business. Ulasan ini dapat mencakup berbagai aspek, seperti kualitas produk atau layanan, pelayanan pelanggan, suasana, harga, kebersihan, dan lain sebagainya [1]. Untuk mengumpulkan opini dan sentimen para pengunjung yang telah mengunjungi TMII pasca pandemi Covid-19. Namun, karena jumlah data yang cukup besar, diperlukan suatu metode untuk mengatasi hal tersebut, yaitu dengan menggunakan analisis sentimen. Analisis sentimen adalah cara yang dipakai untuk menentukan suatu opini, emosi dan sikap yang direfleksikan melalui sebuah teks, dan umumnya diklasifikasikan menjadi sebuah opini negatif dan positif [2]. Penelitian ini menggunakan metode *Multinomial Naïve Bayes* dalam analisis sentimen terhadap ulasan Google Maps tentang TMII.

Penelitian terkait sebelumnya telah dilakukan pada gerai minuman *Xing Fu Tang* di Indonesia dengan menggunakan *Multinomial Naïve Bayes*, dan hasilnya mencapai akurasi sebesar 78% dengan menggunakan 500 ulasan yang diambil dari situs Google Review dari beberapa gerai *Xing Fu Tang*. Namun, penting untuk dicatat bahwa hasil penelitian tersebut hanya berlaku untuk periode tertentu karena ulasan terus bertambah dan pendapat pelanggan dapat berubah seiring waktu berjalan [3]. Pada penelitian lainnya yang telah dilakukan uji coba analisis sentimen menggunakan metode *Multinomial Naïve Bayes* metode ini digunakan untuk mengklasifikasikan sentimen masyarakat terhadap penanganan Covid-19. Penelitian tersebut melibatkan analisis terhadap 2000 dataset teks yang berasal dari posting pengguna di media sosial Twitter. Data ini dibagi menjadi 3 kelas, yaitu positif, negatif, dan netral. Hasil penelitian menunjukkan bahwa metode tersebut mencapai nilai akurasi sebesar 74%, presisi sebesar 74%, dan *recall* sebesar 74% [4]. Pada penelitian berikutnya, dilakukan analisis sentimen terhadap ulasan aplikasi di Google Play. Metode yang digunakan mencakup *Naive Bayes* untuk menganalisis aplikasi yang dikembangkan, serta *Support Vector Machine* sebagai pendekatan pembandingan dalam penelitian ini. Hasil menunjukkan bahwa akurasi *Naive Bayes* mencapai 83.87%, sementara akurasi SVM mencapai 89.49%. Data training yang digunakan meliputi 9298 komentar, dengan 3104 komentar positif, 3092 komentar negatif, dan 3102 komentar *crash*. Data untuk pengujian diambil dari data *training* dengan total 9000 komentar, terdiri dari masing-masing 3000 komentar dengan sentimen positif, negatif, dan *crash* [5].

Penelitian selanjutnya mengenai analisis sentimen masyarakat di Twitter terhadap tokoh publik, memanfaatkan tweet yang mencakup kata kunci seperti "nadiem makarim", "kemendikbud", dan "pak nadiem". Menggunakan metode *Naive Bayes* dan *Support Vector Machine* (SVM). Dari percobaan yang dilakukan, *Naive Bayes* menunjukkan hasil yang lebih baik dibandingkan SVM, dengan akurasi mencapai 91.48%. Presisi terukur sebesar 89.28%, dan *recall* sebesar 91.58%. Sementara itu, SVM mencapai akurasi 85.47%, presisi 90.95%, dan *recall* 76.18%. Total data training yang digunakan adalah 2200, sedangkan data

testing berjumlah 300 [6]. Berdasarkan perbedaan dengan penelitian sebelumnya, usulan penelitian ini memiliki tujuan untuk:

- Mengidentifikasi perbandingan tingkat akurasi dari metode *Multinomial Naïve Bayes* dan *Support Vector Machine (SVM)* dalam analisis sentimen terhadap ulasan pengunjung TMII di Google Maps di Taman Mini Indonesia Indah,
- Menjelaskan gambaran umum tentang ulasan dari pengunjung mengenai kualitas dan layanan tempat wisata Taman Mini Indonesia Indah.

2. METODOLOGI PENELITIAN

2.1. Data Penelitian

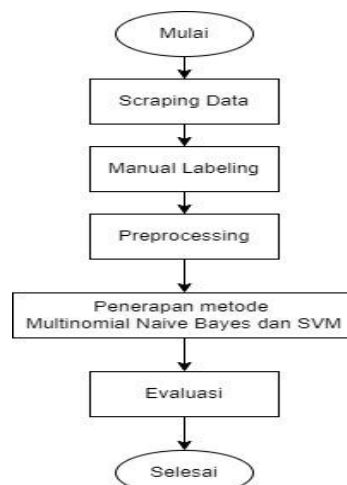
Dalam penelitian ini, penulis menggunakan *dataset* yang terdiri dari 500 ulasan mengenai tempat wisata Taman Mini Indonesia Indah yang berasal dari periode satu tahun terakhir, yaitu dari tanggal 11 Juni 2022 hingga tanggal 7 Juni 2023. Data tersebut diperoleh melalui proses *scraping* menggunakan *SerpApi* dan *Google Colab*, dengan dukungan *package google-search-results* dan *library GoogleSearch* dalam bahasa pemrograman *python*. Data yang berhasil diambil dari Google Maps mencakup informasi seperti nama pengunjung, rating dan ulasan. Tabel 1 contoh data ulasan Google Maps hasil dari pengumpulan data melalui tahapan *scraping*.

Tabel 1. Contoh Data Ulasan

<i>Nama Pengunjung</i>	<i>Rating</i>	<i>Ulasan</i>
<i>Krisna Widya Putra</i>	<i>1</i>	<i>Sistemnya ribet. Ga boleh lagi mobil pribadi keliling. Waktu Habis buat nunggu wara Wiri.</i>
<i>Heri Abdullah</i>	<i>2</i>	<i>Kecewa motor g boleh masuk..mau keliling jadi susah..</i>
<i>Eko Prasetyo</i>	<i>3</i>	<i>Antrian kliling anjungan membludak tapi armada mobil listriknya tidak memadai..Parah antrinya</i>

2.2. Penerapan Metode

Dalam penerapan metode ini dilakukan langkah-langkah untuk menganalisis opini masyarakat tentang wisata Taman Mini Indonesia Indah. Berikut penerapan metode seperti pada Gambar 1.



Gambar 1. Penerapan Metode

2.2.1. Analisis Sentimen

Kata 'analisis' diadaptasi dari Bahasa Inggris 'analysis', yang secara etimologis berasal dari Bahasa Yunani Kuno 'Analisis'. Kata ini terdiri dari dua suku kata, yaitu 'ana' yang berarti kembali, dan 'luein' yang berarti melepas atau mengurai. Bila digabungkan, memiliki arti menguraikan kembali. Secara umum, analisis adalah aktivitas yang terdiri dari serangkaian kegiatan seperti mengurai, membedakan, dan memilah sesuatu untuk dikelompokkan kembali menurut kriteria tertentu. Kemudian, keitannya dicari dan maknanya ditafsirkan [7].

Analisis sentimen melibatkan pengolahan Bahasa alami, analisis teks, dan komputasi linguistik untuk mengenali sentimen dari sebuah dokumen informasi teks. Secara umum, analisis sentimen bisa dibedakan menjadi informasi fakta dan opini. Tugas utama dalam analisis sentimen adalah mengelompokkan polaritas teks dalam dokumen, yaitu apakah pendapat yang diungkapkan dalam dokumen tersebut bersifat positif, negatif, atau netral [8].

2.2.2. Pengumpulan Data

Pada penelitian ini proses pengumpulan data dilakukan melalui *scraping*. *Web scraping* adalah istilah yang digunakan untuk menggambarkan proses pengambilan dokumen semi-terstruktur dari internet. Dokumen ini biasanya berbentuk halaman web yang ditulis dalam bahasa markup seperti HTML atau XHTML. Setelah berhasil diambil, dokumen tersebut kemudian dianalisis untuk mengambil data tertentu dari halaman web tersebut. Data ini bisa berupa teks, gambar, tautan, tabel, atau informasi lain yang relevan [9]. *Scraping* dilakukan dengan mengambil data yang terdapat dalam fitur ulasan di Google Maps dengan cara *web scraping* yang dilakukan menggunakan bahasa pemrograman *python* dengan bantuan *API* yang bernama *SerpAPI* yang nantinya *API Key SerpAPI* dipakai untuk *scraping* data dari ulasan Google Maps. *SerpAPI* adalah sebuah layanan atau *API* yang menyediakan akses ke hasil pencari melalui permintaan program secara otomatis. *Dataset* yang diambil sebanyak 500 ulasan.

2.2.3. Labeling Manual

Labeling manual adalah proses pemberian label atau kategori pada data oleh pakar secara manual (Dosen Komunikasi Pariwisata, Fakultas Komunikasi dan Desain Kreatif Unviersitas Budi Luhur Jakarta). Dalam konteks pembelajaran mesin, data yang telah diberi label ini kemudian digunakan sebagai contoh untuk melatih algoritma atau model pembelajaran mesin agar dapat memahami pola dan mengenali pola yang serupa pada data baru yang belum diberi label.

2.2.4. Pre-processing

Pre-processing adalah proses untuk mengubah bentuk teks data yang belum terstruktur atau sembarang menjadi data yang terstruktur, maka dari itu data menjadi siap untuk dilakukan analisis. Pada proses *text-processing* terdapat beberapa tahapan diantaranya sebagai berikut:

- a. *Case Folding*
Case Folding merupakan metode awal dalam pengolahan teks yang sederhana dan efisien, namun sering terlupakan. Proses *case folding* mengonversi semua huruf menjadi huruf kecil, menggantikan karakter "A" dan "Z" dalam data dengan karakter "a" dan "z" [10].
- b. *Cleansing*
Cleansing merupakan proses evaluasi terhadap kualitas data dengan melakukan perubahan, modifikasi, atau menghilangkan tanda baca, karakter selain huruf, menghilangkan *mention*, email, url dengan tujuan menghasilkan data yang berkualitas tinggi. Proses ini juga sering disebut sebagai pembersihan data atau pembersihan data [11].
- c. *Tokenizing*
Tokenisasi merupakan proses di mana sekumpulan kata yang tersusun dalam sebuah kalimat akan dipisahkan menjadi bagian-bagian kata tunggal atau dalam bentuk token [12].
- d. *Slangword*

Slangword merupakan suatu metode untuk mengumpulkan informasi dari sebuah teks secara optimal. Istilah-istilah tidak resmi seperti *slang*, singkatan dan lain sebagainya akan dikumpulkan dalam kamus *slangword*. Selanjutnya, istilah-istilah tersebut akan diganti agar sesuai dengan tata bahasa Indonesia yang standar sesuai dengan EYD.

- e. *Stopword*
Stopword merupakan kata-kata yang tidak deskriptif dan bukan merupakan kata penting dari suatu dokumen sehingga dapat dibuat. Contoh *stopword* adalah “yang”, “dan”, “dari” dan seterusnya [13]. Proses *stopword* bertujuan untuk menghilangkan kata-kata yang tidak memberikan makna yang signifikan dalam suatu kalimat. Kata-kata tersebut termasuk dalam daftar *stopword* pada pustaka *Sastrawi* [14].
- f. *Stemming*
Stemming merupakan tahap untuk mengubah kata-kata yang masih memiliki imbuhan menjadi kata dasar dengan cara menghilangkan imbuhan yang berada di awal dan akhir kata. Proses ini menggunakan modul *library Sastrawi* [14].

2.2.5. Visualisasi WordCloud

Word Cloud adalah teknik yang digunakan untuk memvisualisasikan data teks dengan mudah terlihat. Karena kemudahannya, teknik ini menjadi favorit dalam penambangan teks. Metode word cloud mengidentifikasi kata-kata yang paling sering muncul dalam dokumen dan menghasilkan visualisasi yang menarik, mencerminkan frekuensi kata-kata tersebut sambil memberikan informasi yang berguna [15].

2.2.6. Multinomial Naïve Bayes

Algoritme Multinomial Naïve Bayes adalah metode pembelajaran probabilistik yang salah satu dasarnya menggunakan teorema Bayes, digunakan dalam *Natural Language Processing* (NLP). Algoritma ini beroperasi berdasarkan pada konsep *term frequency*, yaitu seberapa sering sebuah kata muncul dalam suatu dokumen. Model ini menjelaskan dua fakta, yaitu apakah kata tersebut muncul dalam dokumen atau tidak, serta frekuensi kemunculannya dalam dokumen [4]. Proses pembelajaran diawali dengan memasukkan data latih yang digunakan dalam pelatihan. Setelah itu, dilakukan langkah pembentukan daftar kata-kata yang terdiri dari sekumpulan kata unik yang berasal dari data latih. Selanjutnya, untuk menghitung prioritas atau peluang kemunculan suatu kategori pada seluruh dokumen latih, dapat dilakukan menggunakan persamaan (1) berikut [16]:

$$P(c) = \frac{N_c}{N_{doc}} \quad (1)$$

Catatan:

- c : Kategori atau kelas
- doc : Dokumen
- P(c) : Probabilitas kategori c
- Nc : Banyaknya kategori c pada dokumen latih
- Ndoc : Banyaknya keseluruhan dokumen latih yang digunakan

Langkah selanjutnya adalah melakukan perhitungan peluang bahwa kata "i" termasuk dalam suatu kategori atau kelas tertentu. Proses ini dapat dilakukan dengan menerapkan persamaan (2) berikut:

$$P(w_i|c) = \frac{\text{count}(w_i|c)+1}{\text{count}(w|c)+|V|} \quad (2)$$

Catatan:

- P(w_i|c) : Probabilitas kata tertentu yang muncul dalam suatu kategori atau kelas
- w_i : Kata ke-i dalam seluruh dokumen yang berkategori c
- count(w_i|c) : Jumlah kata tertentu yang muncul dalam suatu kategori atau kelas
- count(w|c) : Jumlah seluruh kata pada kelas
- |V| : Jumlah seluruh kata unik pada kelas

Pada rumus *Multinomial Naive Bayes*, kelas dokumen tidak hanya ditentukan oleh kemunculan kata-kata, melainkan juga oleh jumlahnya. Perhitungan dilakukan dengan menggunakan persamaan (3) berikut [17]:

$$P(c|d) = \frac{N_c}{N_{doc}} x P(w_1|c) x \dots x P(w_i|c) \quad (3)$$

Catatan:

P(c d)	:	Probabilitas suatu dokumen termasuk kelas c
Nc	:	Banyaknya kategori c pada dokumen latih
Ndoc	:	Banyaknya keseluruhan dokumen latih yang digunakan
wi	:	Kata ke-i dalam seluruh dokumen yang berkategori c
P(wi c)	:	Probabilitas kata tertentu yang muncul dalam suatu kategori atau kelas

2.2.7. Support Vector Machine (SVM)

Support Vector Machine (SVM) memiliki tujuan utama dalam menemukan fungsi pemisah optimal diantara berbagai fungsi yang tersedia, dengan maksud untuk mengartikan antara dua jenis objek yang berbeda [18]. Suatu *dataset* memiliki variabel x_i untuk data dan variabel y_i untuk kelasnya. SVM membagi *dataset* menjadi dua kelas, dengan *hyperplane* memisahkan kelas pertama yang memiliki nilai 1 dari kelas lain yang memiliki nilai -1.

$$X_i \cdot W + b \geq \text{untuk } Y_i = 1 \quad (4)$$

$$X_i \cdot W + b \leq \text{untuk } Y_i = -1 \quad (5)$$

Keterangan:

X_i	:	data ke - i
W	:	nilai bobot <i>support vector</i> yang tegak lurus dengan <i>hyperplane</i>
b	:	nilai bias
Y_i	:	kelas data ke -i

Bobot vektor (W) merupakan vektor yang tegak lurus terhadap garis *hyperplane* dan melewati titik pusat koordinat. Sedangkan, bias (b) adalah koordinat relatif garis terhadap titik koordinat. Persamaan (6) digunakan untuk menentukan nilai b , sementara persamaan (7) digunakan untuk menghitung nilai W .

$$b = -\frac{1}{2}(W \cdot X^+ + W \cdot X^-) \quad (6)$$

$$w = \sum_{i=1}^n a_i y_i x_i \quad (7)$$

Keterangan:

b	:	nilai bias
$W \cdot X^+$:	nilai bobot untuk kelas data positif
$W \cdot X^-$:	nilai bobot untuk kelas data negatif
W	:	bobot vektor
a_i	:	nilai bobot data ke-i
y_i	:	kelas data ke-i
x_i	:	data ke-i

H_1 adalah *hyperplane* pendukung dari kelas +1 yang memiliki fungsi $wx_1 + b = +1$.

$$\text{Margin} = |dH_1 dH_2| = \frac{2}{\|w\|} \quad (8)$$

Keterangan:

dH_1	:	jarak <i>hyperplane</i> pendukung kelas +1
dH_2	:	jarak <i>hyperplane</i> pendukung kelas -1

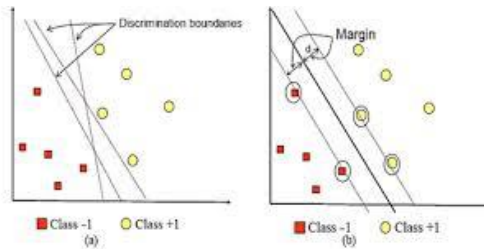
Selanjutnya, untuk menemukan *hyperplane* optimal yang memisahkan kedua kelas, digunakan persamaan berikut:

$$\text{Minimize } \int 1 [W] = \frac{1}{2} \|W\|^2 \quad (9)$$

$$y_i(X_i \cdot W + b) - 1 \geq 0, i = 1, \dots, n \quad (10)$$

Pada Gambar 2, terlihat beberapa pola yang termasuk anggota dari dua kelas: +1 dan -1. Pola yang termasuk ke dalam kelas -1 ditandai dengan simbol kotak merah, sedangkan pola pada kelas +1 ditandai dengan simbol lingkaran kuning. Masalah dalam klasifikasi ini timbul ketika mencari

garis *hyperplane* yang dapat memisahkan kedua kelompok tersebut. Garis solid pada Gambar 2 menunjukkan *hyperplane* optimal, yaitu yang berada di tengah-tengah kedua kelas, sedangkan titik merah dan kuning yang berada dalam lingkaran hitam adalah support vector [19].



Gambar 2. Hyperplane Terbaik yang Memisahkan Kedua Class Negatif dan Positif

2.3. Evaluasi

Dalam penelitian ini, dilakukan pengujian menggunakan metode *Multinomial Naive Bayes* untuk mengukur tingkat akurasi, presisi, dan *recall* dari metode yang dihasilkan. Pengujian tersebut menggunakan *confusion matrix*, sebuah tabel yang digunakan dalam *machine learning* untuk mengevaluasi kinerja model klasifikasi. *Confusion matrix* membandingkan nilai aktual atau nilai sebenarnya dengan nilai hasil prediksi model, dan terdiri dari empat kemungkinan hasil proses klasifikasi, yaitu TP, TN, FP, dan FN. Lebih lanjut, tabel ini dijelaskan dalam Tabel 2.

Tabel 2. Confusion Matrix

Predicted Values	Actual Values	
	Positive	Negative
	Positive	TP (True Positive)
Negative	FN (True Negative)	TN (True Negative)

Catatan:

- TP (True Positive) : Jumlah data yang ada pada nilai positive dan diprediksi benar berada pada nilai *positive*.
- FP (False Positive) : Jumlah data yang ada pada nilai positive dan diprediksi seharusnya berada pada nilai *negative*.
- FN (False Negative) : Jumlah data yang ada pada nilai negative dan diprediksi seharusnya berada pada nilai *positive*.
- TN (True Negative) : Jumlah data yang ada pada nilai negative dan diprediksi benar berada pada nilai *negative*.

Setelah jumlah data yang diperoleh dari tabel *confusion matrix*, dari itu dapat dihitung nilai akurasi, presisi dan *recall* sebagai berikut:

2.3.1. Akurasi

Akurasi adalah nilai akurasi dihitung dengan membagi jumlah total data dalam *dataset* dengan jumlah data positif yang diprediksi positif dan data nilai negatif yang diprediksi negatif. Adapun rumus perhitungan dari akurasi dapat dilihat pada persamaan berikut:

$$\text{Akurasi} = \frac{TP+TN}{TP+FP+TN+FN} \times 100\% \quad (11)$$

2.3.2. Presisi

Presisi adalah rasio jumlah data kategori positif yang diklasifikasikan dengan benar dibagi dengan total yang diklasifikasikan positif [20]. Adapun rumus perhitungan dari presisi dapat dilihat pada persamaan berikut:

$$\text{Presisi} = \frac{TP}{TP+FP} \times 100\% \quad (12)$$

2.3.3. Recall

Recall adalah menggambarkan persentase data kategori positif yang terklasifikasi dengan benar oleh sistem [20]. Adapun rumus perhitungan dari *recall* dapat dilihat pada persamaan berikut:

$$\text{Recall} = \frac{TP}{TP+FN} \times 100\% \quad (13)$$

3. HASIL DAN PEMBAHASAN

3.1. Implementasi Metode

Implementasi metode membahas tentang hasil penelitian yang terkait dengan metode yang digunakan sesuai dengan metodologi penelitian yang telah disusun sebelumnya. Pada tahap ini, penelitian mengungkapkan hasil yang diperoleh melalui penerapan metode yang telah dirancang.

3.1.1. Scraping

Pada penelitian ini proses pengumpulan data dilakukan melalui *scraping* data dari Google Maps dengan menggunakan *library GoogleSearch* dan *API Key*. Penggunaan package *google-search-results* dan *library GoogleSearch* memerlukan *API Key* untuk mengakses data dari ulasan *Google Maps* yang akan dikumpulkan. *API Key* tersebut dapat diperoleh melalui akun *SerpAPI* (<https://serpapi.com>) dan berupa *private API Key*. Dengan menggunakan *library GoogleSearch*, program akan mengirimkan permintaan pencarian ke mesin pencari Google dan mendapatkan hasilnya dalam bentuk data yang dapat diolah oleh program *python*. Data ulasan dari *Google Maps* yang berhasil dikumpulkan akan disimpan dalam bentuk file *csv*.

3.1.2. Manual Labeling

Manual labeling adalah tahapan dalam melakukan label pada *dataset* dengan cara manual. Setiap *dataset* akan diklasifikasikan berlandaskan kelas sentimen positif, dan negatif. Teknik manual labeling *dataset* ini dilakukan menggunakan pakar dari Fakultas Komunikasi Desain Kreatif (FKDK) Universitas Budi Luhur untuk mevalidasi data yaitu Nexen Alexandre Pinontoan, S.I.Kom., M.I.Kom adalah seorang dosen Ilmu Komunikasi dengan spesialisasi dalam komunikasi visual pariwisata. Dengan pengalaman mengajar dan keahliannya dalam analisis komunikasi visual. Dalam konteks ulasan *Google Maps* di TMII, beliau adalah pakar yang memahami bagaimana ulasan tersebut memengaruhi citra dan reputasi destinasi pariwisata. Dengan pengalaman akademis dan pemahaman yang mendalam, Nexen Alexandre Pinontoan memberikan wawasan berharga tentang bagaimana mengelola ulasan *online* secara efektif untuk memperkuat daya tarik TMII di era digital. Melalui proses pelabelan manual pada data *training*. Pada label yang bersifat positif umumnya mencakup kalimat-kalimat yang memuji, memberikan saran, menyampaikan masukan, serta merenungkan tentang perasaan-perasaan positif seperti kegembiraan, kesenangan, dan kebahagiaan. Di sisi lain, label-label yang bersifat negatif mungkin berisi keluhan, ekspresi sindiran, kritik, dan refleksi atas emosi negatif seperti frustrasi, ketidakpuasan, dan kekecewaan. Proses labeling dilakukan melalui *crowdsourcing*, dengan dua orang *annotator* (Fakultas Komunikasi Desain Kreatif) yang melabeli setiap data ulasan [21]. Setiap orang akan melabelkan 400 ulasan data, dan label akhir untuk setiap ulasan data akan divalidasi oleh pakar. Contoh anotasi label dapat ditemukan dalam Tabel 3. Dalam proses labeling tersebut, beberapa ulasan data yang memiliki label positif dan negatif.

Tabel 3. Sample Manual Labeling

Ulasan	Label
<i>Seru banget, bisa keliling indonesia sambil main wahana dan nambah wawasan. Buat yang lagi cari tempat liburan yang asik, menarik, dan beredukasi ini cocok banget buat dikunjungi</i>	<i>Positive</i>
<i>Parkiran jauh, odong2 dan bus shuttle nya sedikit sehingga terjadi penumpukan antrian, tolong dibenahi management transportasi nya bos ku, KITA MAU LIBURAN BUKAN MAU OLAH RAGA JALAN KAKI Toilet pria dan wanita rusak :(, anak terpaksa</i>	<i>Negative</i>

pipis di taman this place very not recomended for lansia , ibu hamil dan yang bawa balita karena harus berjalan jauh

3.1.3. Pre-processing

Pada tahapan *pre-processing* ini digunakan untuk menghapus istilah yang tidak berarti dan tidak perlu dari data. Tahapan ini meliputi *case folding, cleansing, tokenizing, slangword, stopword* dan *stemming*.

Tabel 4. Sample Penerapan Pre-processing

<i>Proses</i>	<i>Ulasan Awal</i>	<i>Ulasan Hasil Pre-processing</i>
<i>Case Folding</i>	<i>Tempat wisata di Jakarta sekaligus edukasi buat keluarga, tiketnya murah dan mudah untuk mengakses tempatnya</i>	<i>tempat wisata di jakarta sekaligus edukasi buat keluarga, tiketnya murah dan mudah untuk mengakses tempatnya</i>
<i>Cleansing</i>	<i>Tempat wisata di Jakarta sekaligus edukasi buat keluarga, tiketnya murah dan mudah untuk mengakses tempatnya</i>	<i>tempat wisata di jakarta sekaligus edukasi buat keluarga tiketnya murah dan mudah untuk mengakses tempatnya</i>
<i>Tokenizing</i>	<i>Tempat wisata di Jakarta sekaligus edukasi buat keluarga, tiketnya murah dan mudah untuk mengakses tempatnya</i>	<i>tempat, wisata, di, jakarta, sekaligus, edukasi, buat, keluarga, tiketnya, murah, dan, mudah, untuk, mengakses, tempatnya</i>
<i>Slangword</i>	<i>Tempat wisata di Jakarta sekaligus edukasi buat keluarga, tiketnya murah dan mudah untuk mengakses tempatnya</i>	<i>tempat wisata di jakarta sekaligus edukasi buat keluarga tiketnya murah dan mudah untuk mengakses tempatnya</i>
<i>Stopword</i>	<i>Tempat wisata di Jakarta sekaligus edukasi buat keluarga, tiketnya murah dan mudah untuk mengakses tempatnya</i>	<i>wisata jakarta edukasi keluarga tiketnya murah mudah mengakses tempatnya</i>
<i>Stemming</i>	<i>Tempat wisata di Jakarta sekaligus edukasi buat keluarga, tiketnya murah dan mudah untuk mengakses tempatnya</i>	<i>wisata jakarta edukasi keluarga tiket murah mudah akses tempat</i>

3.1.4. Visualisasi WordCloud Positive

Pada Gambar 3, menunjukkan bahwa *WordCloud* Positive merepresentasikan sentiment positif yang mendominasi ulasan mengenai kepuasan pelanggan terhadap pemesanan tiket via *online* (tiket masuk), kebersihan, fasilitas, keramahan staf, keindahan alam, hingga keanekaragaman budaya yang ditampilkan (tertulis "anjungan"). Hal ini membenarkan ketertarikan dan kelebihan yang dimiliki TMII melalui ulasan data.



Gambar 3. WordCloud Positive

3.1.5. Visualisasi WordCloud Negative

Pada Gambar 4, menunjukkan bahwa WordCloud Negatif merepresentasikan sentiment negative berkaitan dengan keluhan terkait kondisi fasilitas, harga, layanan, kepadatan, lahan parkir terbatas (tertulis di Gambar 4 "parkir"), keterbatasan armada (tertulis "Mobil"). Dari beberapa keluhan yang disampaikan mungkin hanya merupakan sebagian kecil dari keseluruhan ulasan yang cenderung positif.



Gambar 4. WordCloud Negative

3.1.6. Multinomial Naive Bayes

Pada proses klasifikasi menggunakan metode *Multinomial Naive Bayes*, dilakukan untuk merubah teks ulasan dengan menggunakan sampel data *training* dan data *testing* yang terdapat pada Tabel 5 dan Tabel 6 menjadi angka agar dapat diolah lebih lanjut.

Tabel 5. Sample Ulasan Bersih Training

No	Ulasan	Label
1	sistem ribet tidak boleh mobil pribadi keliling habis tunggu wara wiri	Negative
2	kecewa motor tidak boleh masuk keliling susah	Negative
3	antri keliling anjung membludak armada mobil listrik tidak ada parah antri	Negative
4	bagus lengkap asri indah nyaman suka area papu bersih keren banget pokok	Positive
5	wisata jakarta edukasi keluarga tiket murah mudah akses tempat	Positive
6	rekreasi nyaman harga tiket jangkau tersedia anjung daerah bersih rapih	Positive

Tabel 6. Sample Ulasan Bersih Testing

No	Ulasan	Label
1	nyaman libur keluarga kesini	Positive

Setelah data dipisahkan menjadi data *training* dan data *testing*, langkah berikutnya menghitung jumlah kemunculan *term* data *training* yang terdapat pada Tabel 7.

Tabel 7. Pembobotan TF

<i>No</i>	<i>Kata</i>	<i>TF Negatif</i>	<i>TF Positif</i>
1	<i>sistem</i>	1	0
2	<i>ribet</i>	1	0
3	<i>tidak</i>	3	0
4	<i>boleh</i>	2	0
5	<i>mobil</i>	2	0
6	<i>pribadi</i>	1	0
7	<i>keliling</i>	3	0
8	<i>habis</i>	1	0
9	<i>tunggu</i>	1	0
10	<i>wara</i>	1	0
11	<i>wiri</i>	1	0
12	<i>kecewa</i>	1	0
13	<i>motor</i>	1	0
14	<i>masuk</i>	1	0
15	<i>susah</i>	1	0
16	<i>antri</i>	2	0
17	<i>anjung</i>	1	1
18	<i>membludak</i>	1	0
19	<i>armada</i>	1	0
20	<i>listrik</i>	1	0
21	<i>ada</i>	1	0
22	<i>parah</i>	1	0
23	<i>bagus</i>	1	0
24	<i>lengkap</i>	0	1
25	<i>asri</i>	0	1
26	<i>indah</i>	0	1
27	<i>suka</i>	0	1
28	<i>area</i>	0	1
29	<i>papua</i>	0	1
30	<i>keren</i>	0	1
31	<i>banget</i>	0	1
32	<i>pokok</i>	1	1
33	<i>rekreasi</i>	1	1
34	<i>wisata</i>	0	1
35	<i>jakarta</i>	0	1
36	<i>edukasi</i>	0	1
37	<i>keluarga</i>	0	1
38	<i>tiket</i>	0	1
39	<i>murah</i>	0	2
40	<i>mudah</i>	0	1
41	<i>akses</i>	0	1
42	<i>tempat</i>	0	1
43	<i>reaksi</i>	0	2
44	<i>nyaman</i>	0	3
45	<i>harga</i>	0	1
46	<i>jangkau</i>	0	1
47	<i>tersedia</i>	0	1

48	<i>daerah</i>	0	1
49	<i>bersih</i>	0	2
50	<i>rapih</i>	0	1
Jumlah		29	33

Setelah menghitung jumlah kemunculan *term*, langkah selanjutnya adalah menghitung probabilitas untuk setiap label menggunakan Persamaan 1.

$$P(c) = \frac{N_c}{N_{doc}} \quad P(\text{positive}) = \frac{3}{6} \quad P(\text{negative}) = \frac{3}{6}$$

Setelah melalui tahap *pre-processing*, data testing yang telah dibersihkan akan digunakan untuk menghitung probabilitas kata-kata terhadap label dengan menerapkan persamaan 2.

Data ulasan : nyaman libur keluarga kesini

Perhitungan kelas positive

Perhitungan kelas negative

$$P(\text{nyaman, positive}) = \frac{3 + 1}{33 + 50} = \frac{4}{83}$$

$$P(\text{nyaman, negative}) = \frac{0 + 1}{29 + 50} = \frac{1}{79}$$

$$P(\text{libur, positive}) = \frac{0 + 1}{33 + 50} = \frac{1}{83}$$

$$P(\text{libur, negative}) = \frac{0 + 1}{29 + 50} = \frac{1}{79}$$

$$P(\text{keluarga, positive}) = \frac{1 + 1}{33 + 50} = \frac{2}{83}$$

$$P(\text{keluarga, negative}) = \frac{0 + 1}{29 + 50} = \frac{1}{79}$$

$$P(\text{kesini, positive}) = \frac{0 + 1}{33 + 50} = \frac{1}{83}$$

$$P(\text{kesini, negative}) = \frac{0 + 1}{29 + 50} = \frac{1}{79}$$

Setelah selesai menghitung probabilitas dari tiap label sentimen dan probabilitas kata-kata pada data testing terhadap label sentimen, langkah berikutnya adalah menghitung kemunculan frekuensi kata pada data testing yang terdapat pada Tabel 8.

Tabel 8. Kemunculan Frekuensi Testing

No	Kata	Frekuensi
1	<i>nyaman</i>	1
2	<i>libur</i>	1
3	<i>keluarga</i>	1
4	<i>kesini</i>	1

Setelah selesai menghitung kemunculan frekuensi kata pada *testing*, langkah selanjutnya adalah melakukan perhitungan probabilitas dokumen dari data *testing* terhadap label sentimen menggunakan Persamaan 3.

$$P(\text{positive, d1}) = \frac{3}{6} X \left(\frac{4}{83}\right)^1 X \left(\frac{1}{83}\right)^1 X \left(\frac{2}{83}\right)^1 X \left(\frac{1}{83}\right)^1$$

$$= 0.5 \times 0.0481928 \times 0.0120482 \times 0.0240964 \times 0.0120482$$

$$= 0.0000000842847$$

$$P(\text{negative, d1}) = \frac{3}{6} X \left(\frac{1}{79}\right)^1 X \left(\frac{1}{79}\right)^1 X \left(\frac{1}{79}\right)^1 X \left(\frac{1}{79}\right)^1$$

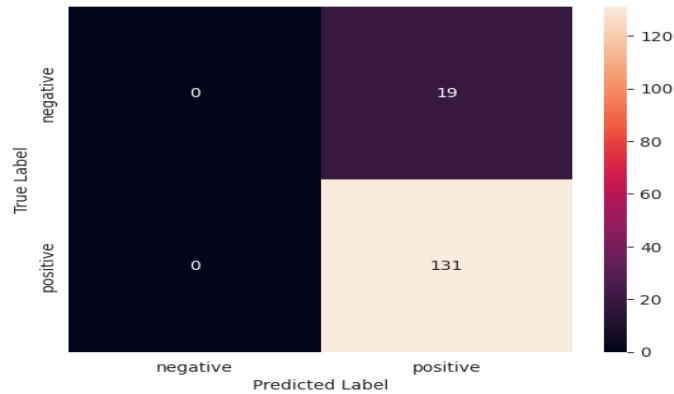
$$= 0.5 \times 0.0126582 \times 0.0126582 \times 0.0126582 \times 0.0126582$$

$$= 0.0000000128368$$

Hasil perhitungan menunjukkan bahwa probabilitas dokumen terhadap label positif lebih besar dibandingkan dengan probabilitas dokumen terhadap label negatif. Ini mengindikasikan bahwa sampel data uji diklasifikasikan sebagai sentiment positif.

3.2. Pengujian

Pada Gambar 5, menunjukkan hasil *confusion matrix* pada algoritma *Multinomial Naive Bayes* dengan diperoleh nilai TP adalah 131, nilai TN adalah 0, nilai FP adalah 0, dan nilai FN adalah 19. Hasil yang diperoleh mendapatkan nilai *Accuracy* 87%, *Recall* 87%, dan *Precision* 100%.



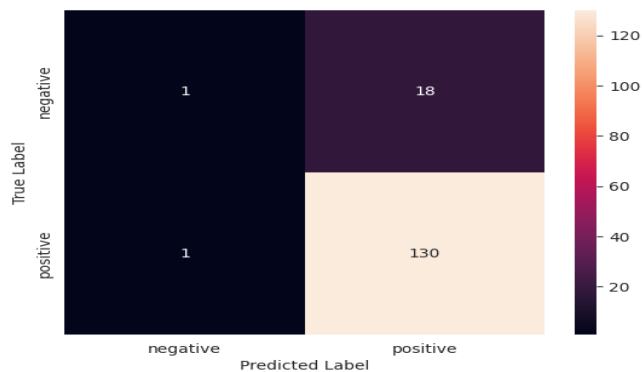
Gambar 5. Confusion Matrix Multinomial Naïve Bayes

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% = \frac{131 + 0}{131 + 0 + 0 + 19} = \frac{131}{150} \times 100\% = 87\%$$

$$Recall = \frac{TP}{TP + FN} \times 100\% = \frac{131}{131 + 19} = \frac{131}{150} \times 100\% = 87\%$$

$$Precision = \frac{TP}{TP + FP} \times 100\% = \frac{131}{131 + 0} = \frac{131}{131} \times 100\% = 100\%$$

Pada Gambar 6, menunjukkan hasil *confusion matrix* pada algoritma *Support Vector Machine* (SVM) dengan diperoleh nilai TP adalah 131, nilai TN adalah 1, nilai FP adalah 1, dan nilai FN adalah 18. Hasil yang diperoleh mendapatkan nilai *Accuracy* 87%, *Recall* 88%, dan *Precision* 99%.



Gambar 6. Confusion Matrix SVM

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% = \frac{130 + 1}{131 + 1 + 1 + 18} = \frac{131}{151} \times 100\% = 87\%$$

$$Recall = \frac{TP}{TP + FN} \times 100\% = \frac{131}{131 + 18} = \frac{131}{149} \times 100\% = 88\%$$

$$Precision = \frac{TP}{TP + FP} \times 100\% = \frac{131}{131 + 1} = \frac{131}{132} \times 100\% = 99\%$$

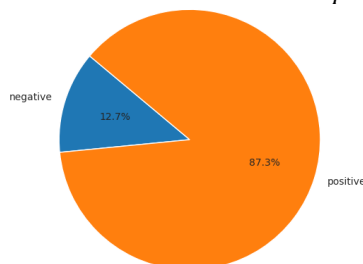
Pada Tabel 9, terdapat hasil *confusion matrix* dari metode *Multinomial Naïve Bayes* dan SVM yang digunakan dalam evaluasi model. Sehingga diperoleh hasil perhitungan nilai akurasi, presisi, dan *recall* dari setiap metode.

Tabel 9. Tabel Hasil Perbandingan akurasi, presisi, dan recall

Metode	Akurasi	Presisi	Recal
<i>Multinomial Naive Bayes</i>	87%	87%	100%
<i>Support Vector Machine (SVM)</i>	87%	88%	99%

Hasil evaluasi kinerja kedua model, yaitu *Multinomial Naïve Bayes* dan *Support Vector Machine (SVM)*, memiliki tingkat akurasi yang sama, karena keduanya berhasil memprediksi sebagian besar data dengan benar dari keseluruhan data yang ada. Penerapan tahapan *pre-processing* mempengaruhi hasil akurasi. Hal ini dibuktikan, penggunaan kamus *slangword* (kamus NLP https://github.com/louisowen6/NLP_bahasa_resources), kamus *stopword* (kamus Tala *Stopword* https://github.com/nltk/nltk_data) dan kamus *stemming* (cari kamus *Stemming* Nazief Adriani) mempengaruhi hasil akurasi dari hasil penelitian ini. Disisi lain, tipikal *dataset* yang dikumpulkan dalam durasi panjang (satu tahun) dari 11 Juni 2022 sampai dengan 7 Juni 2023 berisi sebaran *term* yang cukup kompleks dan komentar yang beragam sehingga menghasilkan temuan-temuan informasi yang beragam. Hal ini menjadi penentu faktor tingginya akurasi dari metode *Multinomial Naïve Bayes* dan SVM dengan nilai 87%. Presisi, yang mengukur proporsi prediksi positif yang benar dari semua prediksi positif, sedikit berbeda antara keduanya karena metode SVM memiliki lebih sedikit kesalahan dalam mengidentifikasi kelas positif, sehingga presisinya sedikit lebih tinggi, yaitu 88%, dibandingkan dengan 87% pada *Multinomial Naïve Bayes*. Namun, dalam hal *recall*, yang mengukur proporsi kelas positif yang berhasil diidentifikasi dengan benar dari keseluruhan kelas positif yang sebenarnya, *Multinomial Naïve Bayes* mencapai 100%, sementara SVM mencapai 99%.

Hal ini menunjukkan bahwa *Multinomial Naïve Bayes* mungkin lebih cenderung untuk mengidentifikasi semua *instance* dari kelas positif dibandingkan dengan SVM. Dengan demikian, berdasarkan evaluasi ini, dapat disimpulkan bahwa pilihan antara keduanya tergantung pada prioritas: SVM lebih cocok jika presisi prediksi positif lebih penting, sementara *Multinomial Naïve Bayes* lebih baik jika mengidentifikasi semua *instance* dari kelas positif menjadi prioritas utama. Hal ini memperkuat penelitian [7], yang menghasilkan akurasi metode Naïve Bayes diatas atau sama dengan 85% dengan *dataset* ulasan pantai di Kabupaten Karawang. Namun, hasil penelitian ini berbeda dengan penelitian [22] dibuktikan dengan nilai akurasi SVM lebih unggul dibandingkan dengan *Naïve Bayes* yaitu 100% dan 80,95% dengan data ulasan sebanyak 2143 komentar. Sedangkan pada penelitian [23] menunjukkan bahwa kinerja dari *Naïve Bayes* lebih baik dan akurat daripada SVM untuk permasalahan klasifikasi sentimen ulasan *Spotify*.



Gambar 7. Visualisasi Analisis Sentiment

Selain itu, pada Gambar 7 terlihat bahwa sentimen positif memiliki presentase yang lebih besar daripada sentiment negatif. Menunjukkan bahwa sentimen pengunjung TMII melalui ulasan Google Maps cenderung kearah ulasan positif.

4. KESIMPULAN

Dari hasil penelitian yang dilakukan menunjukkan bahwa sentimen pengunjung Taman Mini Indonesia Indah (TMII) melalui ulasan Google Maps cenderung positif. Dalam penelitian ini, data dianalisis menggunakan dua metode klasifikasi berbeda, yaitu *Multinomial Naive Bayes* mencapai akurasi sebesar 87%, presisi sebesar 88%, dan recall sebesar 99% dan *Support Vector Machine* (SVM) akurasi sebesar 87%, presisi sebesar 87%, dan recall sebesar 100%. dengan penggunaan 500 data ulasan yang terdiri dari 400 data *training* dan 100 data *testing* dengan label. Dengan demikian, dari hasil evaluasi kinerja, Berdasarkan hasil pengujian tersebut, dapat disimpulkan bahwa memiliki tingkat akurasi yang sama, karena keduanya berhasil memprediksi sebagian besar data dengan benar dari keseluruhan data yang ada. pilihan antara keduanya tergantung pada prioritas: SVM lebih cocok jika presisi prediksi positif lebih penting, sementara *Multinomial Naive Bayes* lebih baik jika mengidentifikasi semua *instance* dari kelas positif menjadi prioritas utama, masih ada beberapa kekurangan yang perlu dipertimbangkan. Pertama, SVM cenderung sensitif terhadap skala data dan memerlukan normalisasi atau penskalaan fitur sebelum penerapan. Selain itu, pelatihan SVM dapat menjadi komputasi yang mahal dan memakan waktu, terutama pada *dataset* besar. Di sisi lain, *Multinomial Naive Bayes* berasumsi bahwa setiap fitur independen satu sama lain, yang tidak selalu realistis dan dapat mengakibatkan kinerja yang buruk jika fitur-fitur tersebut tergantung satu sama lain. Peluang penelitian yang dikembangkan dimasa yang akan datang pada variasi jumlah dataset berdasarkan periode pengumpulan dan jumlah *record* yang berbeda. Kemudian, diperlukan penelitian lebih lanjut menggunakan *dataset* dengan objek wisata yang berbeda. Selanjutnya, diperlukan penelitian lanjut menggunakan perbandingan beberapa metode *deep learning* dan *machine learning* tradisional untuk menghasilkan metode yang paling akurat. Selain itu, diperlukan variasi dari tahapan *pre-processing* seperti variasi penggunaan *stemming*, *non-stemming* atau tahapan variasi penggunaan kamus *slangword* atau *stopword* untuk melihat pengaruhnya terhadap hasil akurasi, presisi dan recall.

DAFTAR PUSTAKA

- [1] 2005 Rumondang, R., Batubara, J. P., & Simbolon K, “Membangun Sistem Navigasi di Surabaya Menggunakan Google Maps API,” *Membangun Sist. Navig. di Surabaya Menggunakan Google Maps API*, vol. 15, no. 1, pp. 165–175, 2016, [Online]. Available: <https://core.ac.uk/download/pdf/196255896.pdf>
- [2] L. F. S. Coletta, N. F. F. De Silva, E. R. Hruschka, and E. R. Hruschka, “Combining classification and clustering for tweet sentiment analysis,” *Proc. - 2014 Brazilian Conf. Intell. Syst. BRACIS 2014*, pp. 210–215, 2014, doi: 10.1109/BRACIS.2014.46.
- [3] A. J. Fadhillah, “Analisis Sentimen Terhadap Gerai Minuman Xing Fu Tang Indonesia Menggunakan Data Ulasan Google,” vol. 8, no. 1, pp. 886–894, 2022.
- [4] Yuyun, Nurul Hidayah, and Supriadi Sahibu, “Algoritma Multinomial Naive Bayes Untuk Klasifikasi Sentimen Pemerintah Terhadap Penanganan Covid-19 Menggunakan Data Twitter,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 4, pp. 820–826, 2021, doi: 10.29207/resti.v5i4.3146.
- [5] L. B. Ilmawan and E. Winarko, “Aplikasi Mobile untuk Analisis Sentimen pada Google Play,” *IJCCS (Indonesian J. Comput. Cybern. Syst.)*, vol. 9, no. 1, p. 53, 2015, doi: 10.22146/ijccs.6640.
- [6] T. T. Widowati and M. Sadikin, “Analisis Sentimen Twitter terhadap Tokoh Publik dengan Algoritma Naive Bayes dan Support Vector Machine,” *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 11, no. 2, pp. 626–636, 2021, doi: 10.24176/simet.v11i2.4568.
- [7] W. Khofifah, D. N. Rahayu, and A. M. Yusuf, “Analisis Sentimen Menggunakan Naive

- Bayes Untuk Melihat Review Masyarakat Terhadap Tempat Wisata Pantai Di Kabupaten Karawang Pada Ulasan Google Maps,” *J. Interkom J. Publ. Ilm. Bid. Teknol. Inf. dan Komun.*, vol. 16, no. 4, pp. 28–38, 2022, doi: 10.35969/interkom.v16i4.192.
- [8] D. N. I. Huda and C. Prianto, “Analisis Sentimen Layanan Jasa Pengiriman Pada Ulasan Play Store: Systematic Literature Review,” *J. Inform. dan Teknol. Komput.*, vol. 4, no. 2, pp. 87–98, 2023, [Online]. Available: <https://ejournalunsam.id/index.php/jicom/>
- [9] M. Turland, *Php-Architect’s Guide to Web Scraping*. 2010.
- [10] W. A. Muhammad Asjad Adna Jihad, Adiwijaya, “Analisis Sentimen Terhadap Ulasan Film Menggunakan Algoritma Random Forest,” *e-Proceeding Eng.*, vol. 8, no. 4, pp. 4136–4144, 2021.
- [11] D. Darwis, N. Siskawati, and Z. Abidin, “Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Review Data Twitter Bmkg Nasional,” *J. Tekno Kompak*, vol. 15, no. 1, p. 131, 2021, doi: 10.33365/jtk.v15i1.744.
- [12] H. C. Husada and A. S. Paramita, “Analisis Sentimen Pada Maskapai Penerbangan di Platform Twitter Menggunakan Algoritma Support Vector Machine (SVM),” *Teknika*, vol. 10, no. 1, pp. 18–26, 2021, doi: 10.34148/teknika.v10i1.311.
- [13] A. Alwasi’a, “Analisis Sentimen Pada Review Aplikasi Berita Online Menggunakan Metode Maximum Entropy (Studi Kasus: Review Detikcom pada Google Play 2019),” *Skripsi*, 2020.
- [14] W. Parasati, F. Abdurrachman Bachtiar, and N. Y. Setiawan, “Analisis Sentimen Berbasis Aspek pada Ulasan Pelanggan Restoran Bakso President Malang dengan Metode Naive Bayes Classifier,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 4, no. 4, pp. 1090–1099, 2020, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [15] R. Yunita and M. Kamayani, “Perbandingan Algoritma SVM Dan Naive Bayes Pada Analisis Sentimen Penghapusan Kewajiban Skripsi,” *Indones. J. Comput. Sci.*, vol. 12, no. 5, pp. 2879–2890, 2023, doi: 10.33022/ijcs.v12i5.3415.
- [16] L. Lesmana, Mukrodin, and F. Nabyla, “Analisis Sentimen Pengguna Twitter PPDB Menggunakan Algoritma Multinomial Naive Bayes,” *J. Sist. Inf. dan Teknol. Perad.*, vol. 1, no. 1, 2020, [Online]. Available: <https://journal.peradaban.ac.id/index.php/jsitp/article/view/604>
- [17] F. A. Ramadhan, S. H. Sitorus, and T. Rismawan, “Penerapan Metode Multinomial Naive Bayes untuk Klasifikasi Judul Berita Clickbait dengan Term Frequency - Inverse Document Frequency Application of Multinomial Naive Bayes News Clickbait with Term Frequency - Inverse Document Frequency,” vol. 11, no. 1, pp. 70–76, 2023, doi: 10.26418/justin.v11i1.57452.
- [18] N. Hendrastuty et al., “Analisis Sentimen Masyarakat Terhadap Program Kartu Prakerja Pada Twitter Dengan Metode Support Vector Machine,” *J. Inform. J. Pengemb. IT*, vol. 6, no. 3, pp. 150–155, 2021, [Online]. Available: <http://situs.com>

- [19] I. Monika Parapat and M. Tanzil Furqon, “Penerapan Metode Support Vector Machine (SVM) Pada Klasifikasi Penyimpangan Tumbuh Kembang Anak,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 10, pp. 3163–3169, 2018, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [20] B. P. Pratiwi, A. S. Handayani, and S. Sarjana, “Pengukuran Kinerja Sistem Kualitas Udara Dengan Teknologi Wsn Menggunakan Confusion Matrix,” *J. Inform. Upgris*, vol. 6, no. 2, pp. 66–75, 2021, doi: 10.26877/jiu.v6i2.6552.
- [21] M. Ihsan, Benny Sukma Negara, and Surya Agustian, “LSTM (Long Short Term Memory) for Sentiment COVID-19 Vaccine Classification on Twitter,” *Digit. Zo. J. Teknol. Inf. dan Komun.*, vol. 13, no. 1, pp. 79–89, May 2022, doi: 10.31849/digitalzone.v13i1.9950.
- [22] H. Herlawati, R. T. Handayanto, P. D. Atika, F. N. Khasanah, A. Y. P. Yusuf, and D. Y. Septia, “Analisis Sentimen Pada Situs Google Review dengan Naïve Bayes dan Support Vector Machine,” *J. Komtika (Komputasi dan Inform.)*, vol. 5, no. 2, pp. 153–163, 2021, doi: 10.31603/komtika.v5i2.6280.
- [23] A. S. Rahayu, A. Fauzi, and R. Rahmat, “Komparasi Algoritma Naïve Bayes Dan Support Vector Machine (SVM) Pada Analisis Sentimen Spotify,” *J. Sist. Komput. dan Inform.*, vol. 4, no. 2, p. 349, 2022, doi: 10.30865/json.v4i2.5398.

