

Analisis Sentimen Publik pada Media Sosial Twitter Terhadap Tiket.com Menggunakan Algoritma Klasifikasi

Zulmeida Silvana Anggraeni¹, Budiman^{2*}, Chairul Habibi³, Nur Alamsyah⁴

^{1,2,3,4} Fakultas Teknologi dan Informatika, Universitas Informatika dan Bisnis Indonesia
Jl. Soekarno Hatta No.643, Kota Bandung, Jawa Barat 40285, Indonesia

E-mail korespondensi: budiman@unibi.ac.id

Informasi Artikel Diterima: 26-09-2023 Direvisi: 15-12-2023 Disetujui: 03-01-2024

Abstrak

Analisis sentimen merupakan proses identifikasi emosional seseorang terhadap suatu objek yang akan menghasilkan sentimen positif, negatif dan netral. Kemajuan teknologi ini tentu memberikan pengaruh terhadap berbagai pelaku bisnis untuk saling mengintegrasikan sistem bisnisnya satu sama lain, salah satunya Tiket.com. Hal tersebut tentu menghasilkan sentimen dari masyarakat Indonesia yang diunggah pada platform media sosial Twitter, sehingga membantu individu maupun organisasi dalam mengambil keputusan. Penelitian ini dilakukan untuk mengetahui klasifikasi sentimen masyarakat Indonesia terhadap Tiket.com menggunakan algoritma *Naïve Bayes Classifier* (NBC), *K-Nearest Neighbor* (KNN), *Support Vector Machine* (SVM) dan *Random Forest* (RF). Berdasarkan perhitungan data sentimen terhadap Tiket.com terdapat 90.3% sentimen positif dan 9.7% sentimen negatif. Persentase tersebut menunjukkan bahwa Tiket.com cukup berpengaruh positif terhadap penggunaannya. Berdasarkan hasil pengujian algoritma klasifikasi, diketahui NBC memperoleh tingkat akurasi sebesar 88%, KNN dengan nilai $k = 11$ mendapatkan akurasi sebesar 91%, SVM menghasilkan tingkat akurasi sebesar 92%, dan tingkat akurasi RF mencapai 93% dengan $n_estimators = 100$. Kesimpulan pada penelitian ini, *Random Forest* merupakan algoritma yang memiliki tingkat akurasi paling tinggi dibanding dengan algoritma klasifikasi lain.

Kata Kunci: Analisis Sentimen, Tiket.com, Algoritma Klasifikasi

Abstract

Sentiment analysis is the process of identifying a person's emotions towards an object that will result in positive, negative and neutral sentiments. This technological advancement has certainly influenced various business actors to integrate their business systems with each other, one of which is Tiket.com. It certainly generates sentiments from the Indonesian people that are uploaded on the Twitter social media platform, thus helping individuals and organizations in making decisions. This research was conducted to determine the classification of Indonesian public sentiment towards Tiket.com using Naïve Bayes Classifier (NBC), K-Nearest Neighbor (KNN), Support Vector Machine (SVM) and Random Forest (RF) algorithms. Based on the calculation of sentiment data towards Tiket.com, there are 90.3% positive sentiments and 9.7% negative sentiments. These percentages show that Tiket.com has a positive impact on its users. Based on the classification algorithm test results, it is known that NBC obtained an accuracy rate of 88%, KNN with a value of $k = 11$ gets an accuracy of 91%, SVM produced an accuracy rate of 92%, and the RF accuracy rate reached 93% with $n_estimators = 100$. In conclusion, Random Forest is an algorithm that has the highest level of accuracy compared to other classification algorithms.

Keywords: Sentiment Analysis, Tiket.com, Classification Algorithm

1. Pendahuluan

Era revolusi industri 4.0 ini, teknologi memang memiliki peran penting terhadap kemajuan Indonesia, salah satunya yaitu teknologi data (Budiman, 2021). Cakupan pada teknologi data ini juga sangatlah luas seperti pengolahan data yang memiliki berbagai

langkah-langkah pengolahan. Pada pengolahan data, dilakukan perubahan data mentah untuk tahap pengujian sehingga menjadi sebuah informasi yang bermanfaat bagi pembaca. Dilakukannya pengolahan data ini guna meningkatkan efektivitas dan mengefisiensi setiap kegiatan operasional perusahaan,



meningkatkan inovasi dalam dunia bisnis, mengambil keputusan dan masih banyak lagi. Hal tersebut didapatkan dari hasil analisis tahapan pengolahan data menggunakan berbagai metode dan *tools* khusus pada bidang tersebut (Budiman et al., 2020). Ruang lingkup pengolahan data sangatlah luas seperti penambangan data (*data mining*) dan penambangan teks (*text mining*). Salah satunya *text mining* yang digunakan untuk menganalisis sentimen publik terhadap suatu produk maupun jasa guna mendeteksi masalah yang dapat terjadi di masa yang akan datang.

IBM mengutip dari Kwartler (2017) bahwa terdapat sekitar 80% data yang tidak terstruktur sehingga diperlukannya *tools* untuk *text mining* dan teknik *Natural Language Processing* (NLP) sebagai sarana mengekstraksi sekumpulan teks untuk mengubahnya menjadi sebuah informasi yang terstruktur dan berkualitas (Kwartler, 2017). Chauhan dan Hudaya pada tahun 2020 memaparkan bahwa berdasarkan data APJII 2017 pengguna internet di Indonesia berkembang sangat pesat, dengan peningkatan yang cukup signifikan dari tahun ke tahunnya sehingga memberikan dampak positif terhadap Indonesia. Gambar 1 merupakan grafik pertumbuhannya (Chauhan & Hudaya, 2020).



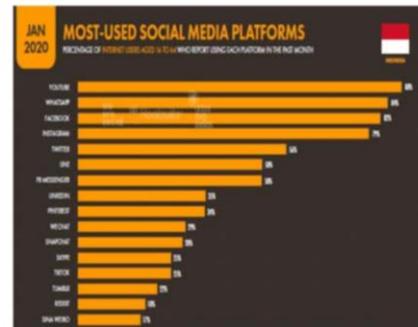
Sumber: Data APJII 2017 (Chauhan & Hudaya, 2020)

Gambar 1. Grafik Pertumbuhan Pengguna Internet di Indonesia

Grafik pada gambar 1 menunjukkan bahwa pada tahun 1998 hanya terdapat 0,5 juta penduduk yang menggunakan internet, lalu pada tahun 2017 meningkat pesat hingga mencapai 143,26 juta penduduk. Berdasarkan angka-angka tersebut dapat menjadi sebuah peluang bagi pelaku usaha dengan memanfaatkan untuk memperluas pangsa pasar (Chauhan & Hudaya, 2020).

Berkembangnya internet ini memang berpengaruh terhadap dunia karena adanya peningkatan pengguna media sosial. Menurut *Hootsuite (We Are Social) Indonesia* terdapat 160 juta orang pengguna aktif menggunakan media sosial pada tahun 2020. Berdasarkan grafik pada gambar 2, persentase media sosial yang banyak diakses oleh

masyarakat Indonesia adalah Youtube (88%) WhatsApp (84%), Facebook (82%), Instagram (79%) dan Twitter (56%). Melihat data tersebut, keberadaan media sosial ini dapat memberikan pengaruh kuat terhadap pola komunikasi dan interaksinya (Harahap & Adeni, 2020).



Sumber: *Hootsuite (We Are Social) Indonesian Digital Report 2022*

Gambar 2. Grafik Perkembangan Media Sosial di Indonesia

Kemajuan teknologi ini tentu dipengaruhi oleh tingkat kecepatan suatu data dan informasi yang diperoleh melalui berbagai media sosial. Dengan media sosial, penggunaannya dapat melakukan banyak kegiatan secara *online* seperti berkomunikasi jarak jauh, mengakses tren yang sedang terjadi, melakukan transaksi dan lain-lain. Twitter merupakan media sosial yang mudah untuk merepresentasikan opini-opini dalam mengakses informasi (Harahap & Adeni, 2020). Opini yang diberikan oleh pengguna Twitter dapat tertuju pada suatu objek seperti bisnis, pendidikan, hukum, politik, agama dan kesehatan. Kelas sentimen ini dapat berupa positif, negatif dan netral. Objek pada bidang bisnis seperti Tiket.com yang merupakan sebuah bisnis bergerak di bidang jasa pemesanan tiket *online*, seperti tiket pesawat, tiket hotel, *to do*, tiket kereta api, penyewaan mobil, tiket event, dan tiket+*Service* (Puspita et al., 2020). Opini yang diberikan biasanya berupa *review* dari pengguna mengenai layanan dan kualitas sistem yang didapatkan saat bertransaksi. Dengan sentimen, pelaku bisnis dapat menjadikan hal tersebut sebagai umpan balik (*feedback*) ke depannya, agar dapat meningkatkan performa dan kualitas bisnis perusahaan Tiket.com sesuai dengan kebutuhan pasar (Utomo, 2020). Hal ini juga berlaku bagi perusahaan diluar Tiket.com, karena dengan adanya opini publik dapat dijadikan sebagai sarana pendukung dalam meningkatkan kegiatan operasional suatu bisnis (Prastyo, 2019). Analisis sentimen juga dikenal sebagai *opinion mining* ini prosesnya terdiri dari memahami, mengekstraksi dan mengolah data

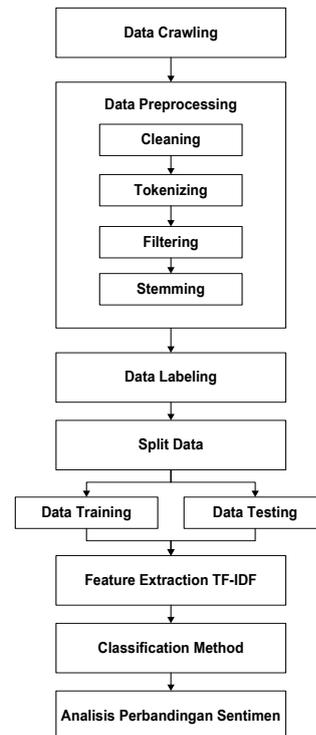
tekstual untuk mendapatkan informasi dari sekumpulan kalimat opini atau sentimen, hal tersebut tentunya menggunakan metode-metode sesuai dengan kebutuhan analisa (Anam et al., 2021). Penelitian terkait analisis sentimen publik menggunakan algoritma *Naïve Bayes*, *Random Forest* dan *Support Vector Machine* yang dilakukan oleh Evita dkk pada tahun 2020 mengenai aplikasi Ruangguru. Pengujian menggunakan 1629 data sentimen disertai *Cross Validation* dengan nilai $k = 10$. Diharapkan dari pengujian tersebut *Naïve Bayes* yang memperoleh tingkat akurasi sebesar 94.16% dan 0.999 nilai AUC, lalu *Random Forest* mendapatkan tingkat akurasi 97.16% dan 0.996 nilai AUC, serta *Support Vector Machine* menerima tingkat akurasi 96.01% dan 0.543 nilai AUC. Sehingga dibuktikan bahwa *Random Forest* memiliki tingkat akurasi yang lebih tinggi dari metode *Naïve Bayes* dan *Support Vector Machine* (Fitri, 2020). Penelitian lain dilakukan oleh Aida Indriani terkait analisis sentimen menggunakan *Naïve Bayes Classifier* (NBC) dan *K-Nearest Neighbor* (KNN) pada tahun 2020. Diperoleh hasil dari data forum yang diujikan bahwa KNN mendapatkan nilai akurasi sebesar 80% sedangkan data uji NBC mendapatkan nilai akurasi sebesar 73%, sehingga penelitian ini menyatakan bahwa hasil KNN lebih baik dibandingkan dengan metode NBC (Indriani, 2020). Penelitian lain terkait analisis sentimen dengan objek yaitu pengguna transportasi Jakarta telah dilakukan oleh Ismia Iwandini dkk pada Januari 2023 menggunakan metode *Naïve Bayes Classifier* (NBC) dan *K-Nearest Neighbor* (KNN) serta pembobotan TF-IDF dengan 4000 data tweet yang diujikan. Diperoleh dari metode NBC menghasilkan tingkat akurasi sebesar 61,1% dan KNN sebesar 75,7%. Kesimpulannya pada penelitian ini yaitu metode KNN jauh lebih baik dibandingkan metode NBC karena memiliki tingkat akurasi yang lebih tinggi (Iwandini et al., 2023).

Menurut pemaparan Dedi dkk yang mengutip dari A. Samad dkk, Twitter merupakan salah satu media sosial yang banyak diminati oleh masyarakat di dunia. Tahun 2016, Twitter mendapati pengguna aktif sekitar 313 juta per bulannya (Darwis et al., 2021). Berdasarkan fenomena yang terjadi, maka akan dilakukan analisis sentimen menggunakan algoritma klasifikasi yaitu *Naïve Bayes Classifier* (NBC), *K-Nearest Neighbor* (KNN), *Support Vector Machine* (SVM) dan *Random Forest* (RF) dari pengguna media sosial Twitter terhadap Tiket.com di Indonesia.

2. Metode Penelitian

Text mining dimulai dengan proses pengumpulan dokumen dari berbagai sumber.

Text Mining akan mengambil dokumen tertentu dan memprosesnya terlebih dahulu dengan memeriksa format dan rangkaian karakter, Kemudian dokumen akan melalui tahap analisis teks. Analisis teks adalah analisis semantik untuk memperoleh informasi berkualitas tinggi dari teks (VijayGaikwad et al., 2014). seperti yang telah dijelaskan pada Pendahuluan, penelitian ini akan menggunakan empat algoritma klasifikasi.



Sumber: Proses Penelitian

Gambar 3. Tahapan Penelitian

Gambar 3 merupakan tahapan alur penelitian analisis sentimen publik terhadap Tiket.com, berikut ini merupakan tahapan yang dilakukan dalam proses penelitian:

1. *Data crawling* merupakan proses otomatisasi dalam mengumpulkan data *mention* Twitter untuk dilakukan pengujian. Tools yang digunakan pada tahap ini adalah RapidMiner 10.3 yang merupakan software independen untuk menganalisis data dengan mengintegrasikan bahasa pemrograman, RapidMiner didedikasikan untuk pengambilan data dengan skala besar (Prasetyo et al., 2021).
2. *Data preprocessing* merupakan tahap penghapusan kata yang tidak terstruktur (*unsupervised*) (Mustofa & Novita, 2022). Penelitian ini akan melalui empat tahap *preprocessing*. *Cleaning* merupakan proses menghilangkan dan mengubah karakter pada teks untuk menghasilkan kata asli (Dwi & Surya, 2021). *Tokenizing* proses pemisahan

kalimat deskripsi menjadi sebuah kata yang dipotong sesuai dengan karakter spasi (Gunawan et al., 2018). *Filtering* proses menghilangkan beberapa kata yang tidak penting agar lebih mudah ketika diolah (Mustofa & Novita, 2022). *Stemming* proses pencarian kata kasar dari hasil *filtering* yaitu *stopword removing*. *Stemming* memiliki aturan dalam pengolahan seperti pendekatan kamus (Anam et al., 2021).

3. *Data labeling* adalah tahap pelabelan teks hasil *preprocessing*, pada penelitian ini akan menggunakan dua klasifikasi sentimen positif dan negatif.
4. *Split data* yaitu tahap untuk membagi *dataset* ke dalam dua bagian yaitu *data training* dan *data testing*. Pada pengklasifikasian data biasanya *dataset* dibagi ke dalam *data training* dan *data testing* untuk diselesaikan oleh metode *text mining* yang digunakan (Indriani, 2020).
5. *Feature extraction* TF-IDF adalah teknik pengambilan ciri teks atau *feature* yang akan dilakukan ekstraksi ke dalam bentuk numerik dan dilakukan analisis (Religia, 2019).
6. *Classification method* adalah tahap klasifikasi sentimen publik menggunakan empat algoritma, NBC, KNN, SVM dan RF. Algoritma NBC yaitu teknik klasifikasi yang mengasumsikan bahwa fitur tidak memiliki keterkaitan dengan fitur lain walaupun berada di kelas yang sama yang didasarkan pada probabilitas Teorema Bayes. Lalu, KNN merupakan algoritma bagian dari *instance based learning* (Batta, 2018). KNN ini algoritma yang membutuhkan pelatihan data dan nilai *k* yang telah ditentukan untuk mencari *k* terdekat berdasarkan jarak komputasi (Sholeha et al., 2022). Kemudian, SVM adalah teknik yang mengklasifikasikan teks yang mencari performa yang baik untuk suatu objek. Cara kerjanya dengan mencari data yang jaraknya paling dekat dengan *hyperplane* diantara dua kelas yang berbeda agar dapat memaksimalkan *output*. Sedangkan RF adalah metode klasifikasi dan regresi yang menginterpretasikan pohon keputusan sebagai *base classifier* yang dikombinasikan (Fitri, 2020).
7. Analisis perbandingan sentimen adalah tahap evaluasi dari setiap hasil pengujian di masing-masing algoritma, dengan memunculkan *Confusion Matrix*, *Classification Report* dan *Area Under Curve* (Budiman et al., 2023).

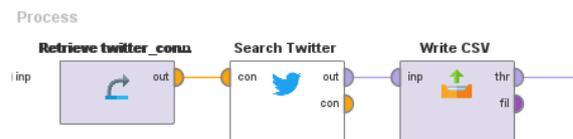
3. Hasil dan Pembahasan

Bagian tiga merupakan proses pendokumentasian dari setiap tahapan penelitian yang dijalankan. Penelitian ini menggunakan dua tools, pertama RapidMiner 10.3 yang digunakan

untuk proses crawling data. Kedua, bahasa pemrograman Python yang digunakan untuk melaksanakan tahapan penelitian mulai dari data preprocessing sampai dengan pemodelan klasifikasi.

3.1 Data Crawling

Teknik pengumpulan data menggunakan RapidMiner 10.3 dengan memakai beberapa *query* yang berhubungan dengan Tiket.com seperti "tiket.com", "tiketdotcom", "tiket/.com", "tiket .com", "tiket com", "tiketcom", "tikom", yang diselaraskan ke dalam satu *query* "tiketdotcom" untuk mempermudah pengujian.



Sumber: Proses Penelitian

Gambar 4. Data Crawling

Pengumpulan dataset diperoleh dari mention tweet masyarakat Indonesia melalui beberapa *query* yang berhubungan dengan Tiket.com, seperti "tiket.com", "tiketdotcom", "tiketcom", "sistem tiket.com", "qn tiket.com", "antri tiket.com", "murah tiket.com", "promo tiket.com" pada media sosial Twitter. Data yang diperoleh merupakan dataset dari mention tweet pada rentang waktu satu bulan yaitu Juni 2023 sampai dengan Juli 2023. Hasil *data crawling* sebanyak 1278 data *tweet* dengan menggunakan RapidMiner 10.3. Data tersebut kemudian dilakukan penghapusan duplikasi data sehingga menghasilkan 769 data sentimen yang akan diolah pada penelitian.

3.2 Data Preprocessing

Tahap *preprocessing* ini adalah tahap pembersihan dan merapikan *dataset* dengan bantuan bahasa pemrograman Python lengkap dengan *library* pada setiap tahapan. *Cleaning*, *tokenizing* dan *filtering* di sini menerapkan *library Regular Expression*, sedangkan *stemming* menggunakan *library Stemmer*.

Tabel 1 merupakan tahapan preprocessing dapat dilihat terjadi beberapa perubahan pada setiap tahapan sesuai dengan *function* yang terdapat pada Python.

3.3 Data Labeling

Proses pemberian label dilakukan secara manual menggunakan data hasil *preprocessing* dengan menerapkan dua klasifikasi sentimen yaitu positif (+) dan negatif (-).

Tabel 1. Data Preprocessing

Create d-At	From-User	To-User	Langu age	Text	Cleaning	Tokenizin g	Filtering	Stemmi ng
12/06/2023 12:50	WINANGS UONG LAKA WINANGS UENG	bonekb oys	in	@bonekb oys Mulai apa ya? Emang mulai dibuka un...	bonekbo ys mulai apa ya emang mulai dibuka un...	[bonekbo ys, mulai, apa, ya, emang, mulai, dibu...	bonekbo ys mulai apa iya emang mulai dibuka unt...	bonekb oys mulai apa iya emang mulai buka untuk...
15/06/2023 12:51	GrnSyndct 27	yusfidw s	in	@yusfidw s di tiketdotco m keterangan annya udah ha...	yusfidws di tiketdotc om keterangan annya udah ha...	[yusfidws , di, tiketdotco m, keterangan annya, uda...	yusfidws di tiketdotc om keterangan annya sudah ha...	yusfidw s di tiketdotc om terang sudah habis pes
17/06/2023 6:19	GrnSyndct 27	bengben gcklt	in	@bengben gcklt kan yang asli belinya di tiketdo...	bengben gcklt kan yang asli belinya di tiketdo...	[bengben gcklt, kan, yang, asli, belinya, di, t...	bengben gcklt kan yang asli belinya di tiketdot...	bengben gcklt kan yang asli beli di tiketdotc om...
28/06/2023 10:47	JASTIP Barang TAEYEON , Le Sserafim	NaN	in	group hybe biasanya dibawa ime dan pake tiketd...	group hybe biasanya dibawa ime dan pake tiketd...	[group, hybe, biasanya, dibawa, ime, dan, pake...	group hybe biasanya dibawa ime dan pakai tiket...	group hybe biasa bawa ime dan pakai tiketdotc o...
28/06/2023 12:08	ran	xxzlingli ng	in	@xxzlingl ing Kemarin lg ada promo gledek dari ...	xxzlinglin g kemarin lg ada promo gledek dari ...	[xxzlinglin g, kemarin, lg, ada, promo, gledek,...	xxzlinglin g kemarin lagi ada promo gledek dari...	xxzlingli ng kemarin lagi ada promo gledek dari...

Sumber: Proses Penelitian

Tabel 2. Data Labeling

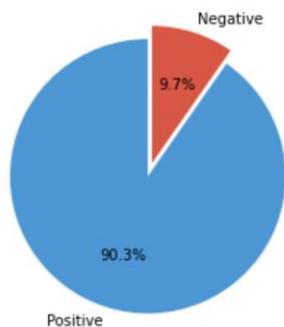
TextPreprocessing	Label
bonekboys mulai apa iya emang mulai buka untuk kalang batas belum ada di tiketdotcom bahkan baya lama error dan hancur	Positif
yusfidws di tiketdotcom terang sudah habis pes	Positif
bengbengcklt kan yang asli beli di tiketdotcom itu ciri ciri	Positif

TextPreprocessing	Label
group hybe biasa bawa ime dan pakai tiketdotcom teman teman kita bisa bantu presale gensale yuk kita sudah alam dan dapat tiket group hybe jastip jasa titip konser le sserafim murah trusted tiketdotcom	Positif
hblfess ime sbneryna itungannya lumayan bagus dia sediaiin exchange tiket tuh oke service oke yang ga oke sistem tiketing di tiketdotcom yabg tidak jelas sold engga nya terus calo super banyak sksks	Negatif
rifkypalmy rubyjxnee maaf nimbrung agoda memang kurang aman iya aku lagi booking hotel bingung antara tiketdotcom atau agoda di agoda harga jatuh murah banget ini benar gak iya tertawa	Negatif
umanggg lwat mana emg war nya kak umang lwat link apa tiketdotcom	Positif
pakai kode promo tiketdotcom inap di hotel swiss belcourt bogor jadi makin hemat tiketdotcom tiketdotcom	Positif
ohlalaloey iyes betul kakk sama kayak kalau kita veli di tiketdotcom juga kokk bisa wakil asal tiket bukan yang package soundcheck kakk	Positif
yang udh nuker tiket numberseat di tiket sesuai sama yang di tiketdotcom ga	Positif
fairiyutopiaa asli tiketdotcom emamg ngeselin aku dapat presale tapi buat teman sedang diri aku sendiri waktu itu dapat gensale tapi jahat kelamin jastip	Negatif
rt bankbtn siapa yang bisa nolak coba pergi libur tapi bisa hemat sahabat btn buru cek deh promo tarik dari tiketdotcom dan jangan	Positif
teman teman kalau misal saat beli di tiketdotcom kurang lengkap isi nama kira jadi masalah ga nanti saat tukar tiket tapi kalau no nik sih sudah sesuai dengan ktp	Negatif
sampai bandara check masuk sudah beres semua iseng tanya ke kakak kak terbang dimajuiin yah kakanya dengan santai jawab tidak kak sesuai jadwal kok jam emang niat ngeprank ini tiketdotcom	Negatif
meski ada drama dari tiketdotcom dapat email jam yang mana sudah tidur bangun jam siapa yang gak kaget kalau terbang dimajuiin jadi jam tiketdotcom	Negatif

Sumber: Proses Penelitian

Langkah berikutnya yaitu melihat jumlah data sentimen positif dan negatif serta persentasenya dengan bantuan *library Matplotlib*.

Percentage of Positive and Negative Sentiments of Tiket.com



Sumber: Proses Penelitian

Gambar 6. Visualisasi *Data Labeling*

Gambar 6 menunjukkan visualisasi data *labeling* yang dapat dilihat bahwa terdapat 719 data sentimen positif dan 77 data sentimen negatif terhadap Tiket.com, dengan persentase yang diperoleh yaitu 90.3% sentimen positif dan 9.7% sentimen negatif.

3.4 Split Data

Tahapan yang dilakukan sebelum pembobotan TF-IDF ini yaitu dengan membagi

dataset ke dalam *data training* dan *data testing*. Persentase pembagian data ini digunakan 80% sebagai data *training* dan 20% sebagai data *testing*. Setelah melalui tahap split data dari 796 sentimen, maka dihasilkan bahwa 636 data latih dan 160 data uji.

3.5 Feature Extraction TF-IDF

Feature extraction merupakan tahap pembobotan TF-IDF dengan memanfaatkan *library sklearn* dengan fungsi *TfidfVectorizer()*, yang akan memunculkan frekuensi kemunculan *term* pada suatu dokumen.

Tabel 3. Pembobotan TF-IDF

Term	TF-IDF
(0, 1918)	0.10531175368477318
(0, 1826)	0.04723219618389431
(0, 1164)	0.5217871519069835
(0, 1002)	0.23342001362012255
(0, 875)	0.276964579679738
(0, 777)	0.13837710340746245
(0, 672)	0.276964579679738
(0, 513)	0.24949101734636878
(0, 499)	0.21317294343124882
(0, 420)	0.07902934526301438

Sumber: Proses Penelitian

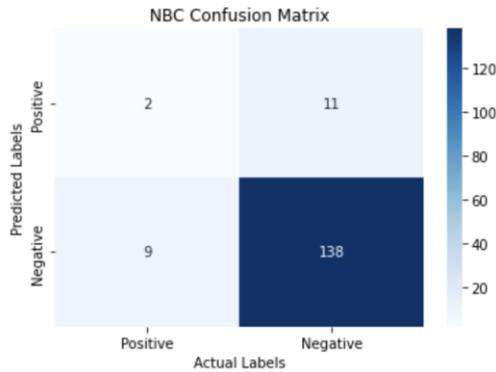
3.6 Classification Method

Proses pengklasifikasian dengan menerapkan algoritma pada setiap metode sebagai perbandingan. Algoritma yang digunakan pada metode NBC adalah BernoulliNB, lalu KNN yaitu algoritma *euclidean distance* dengan nilai $k = 11$, kemudian SVM dan terakhir yaitu RF dengan nilai estimasi $n = 100$.

3.7 Analisis Perbandingan Sentimen

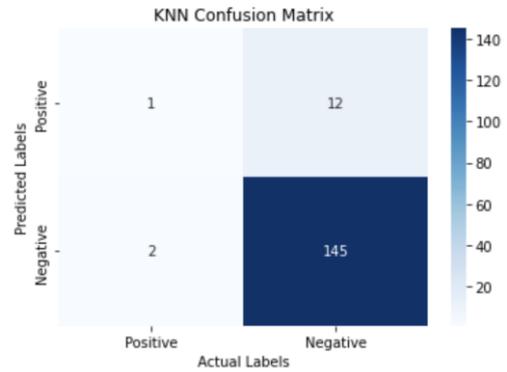
Tahap ini akan menginterpretasikan hasil dari pengujian dari model evaluation yang dilakukan yaitu *confusion matrix*, *classification report* dan *area under curve*.

NBC Confusion Matrix:
 $\begin{bmatrix} 2 & 11 \\ 9 & 138 \end{bmatrix}$



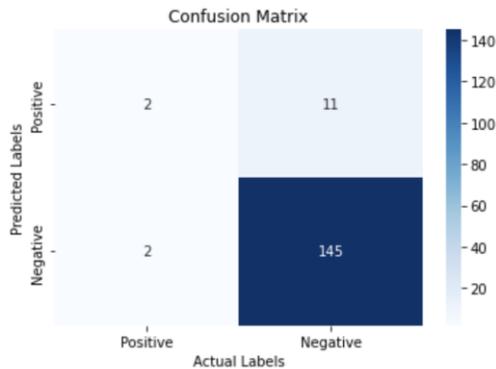
(a)

KNN Confusion Matrix:
 $\begin{bmatrix} 1 & 12 \\ 2 & 145 \end{bmatrix}$



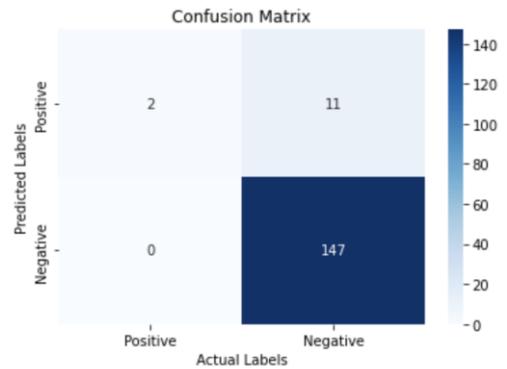
(b)

SVM Confusion Matrix:
 $\begin{bmatrix} 2 & 11 \\ 2 & 145 \end{bmatrix}$



(c)

Random Forest Confusion Matrix:
 $\begin{bmatrix} 2 & 11 \\ 0 & 147 \end{bmatrix}$



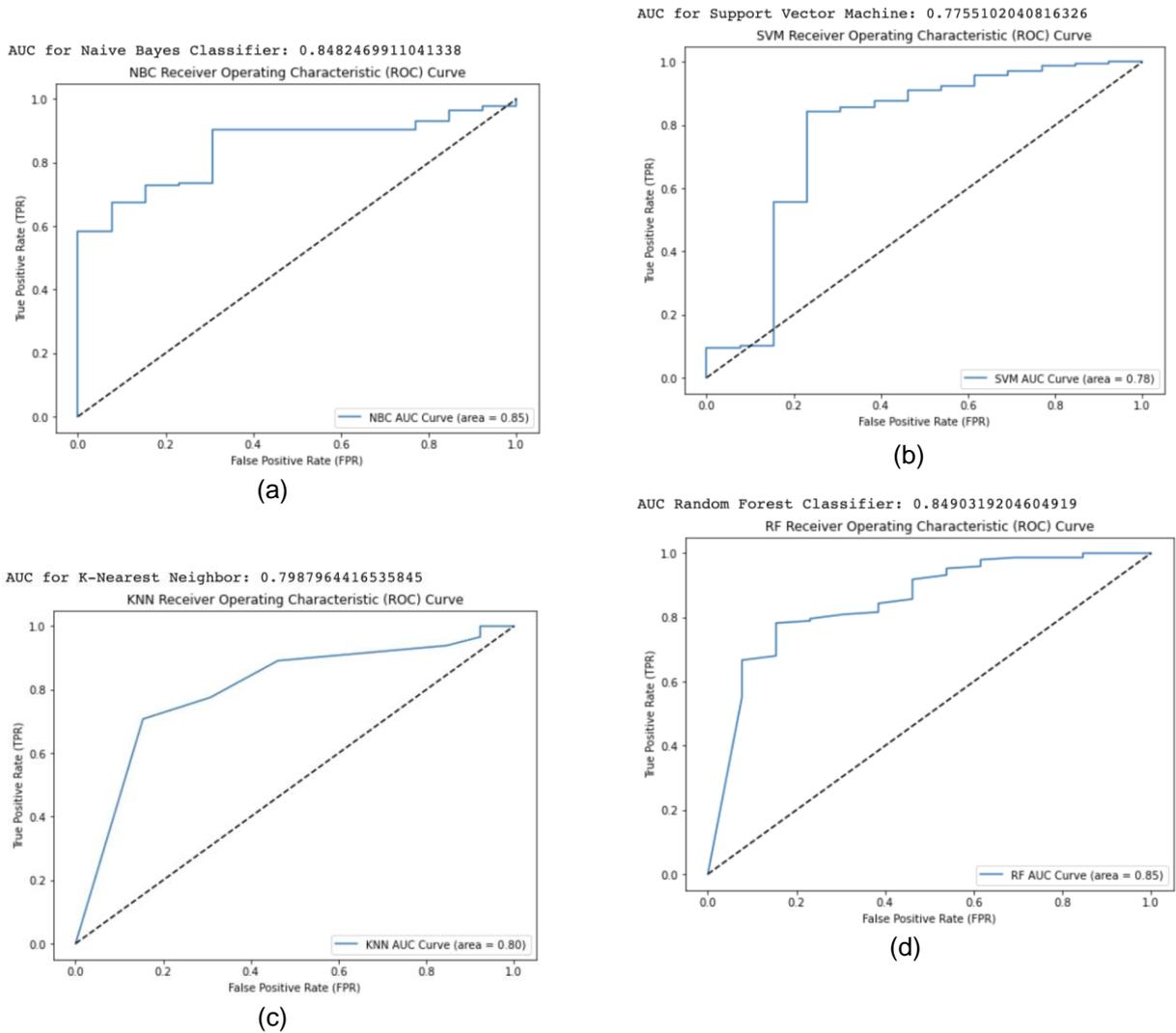
(d)

Sumber: Proses Penelitian

Gambar 8. Confusion Matrix: (a) Naïve Bayes Classifier Confusion Matrix (b) K-Nearest Neighbor Confusion Matrix (c) Support Vector Machine Confusion Matrix (d) Random Forest Confusion Matrix

Gambar 8 menunjukkan *confusion matrix* dari empat algoritma klasifikasi sebagai sarana menganalisis kualitas data dari kelas positif dan negatif terhadap nilai *actual* dan *predicted*. *Confusion matrix* merupakan tahapan sebelum memperoleh nilai *area under curve* dan *classification report* dari setiap algoritma yang digunakan. Langkah berikutnya yaitu menghitung nilai AUC, *Area Under Curve* (AUC) dapat diartikan sebagai rata-rata sensitivitas pada probabilitas (Irawan & Samopa, 2018).

Gambar 9 merupakan hasil pengujian AUC dari metode NBC, KNN, SVM dan RF. Pada AUC ini akan menggunakan *classification level* (Putri et al., 2020). Nilai AUC dari NBC diperoleh sebesar 0.84824 atau *good classification*, lalu KNN sebesar 0.7987 yang *level*-nya termasuk ke dalam *fair classification*, kemudian SVM mendapatkan nilai 0.7755 yaitu masuk ke dalam *fair classification* dan RF yaitu sebesar 0.8490 yang termasuk ke dalam *good classification*.



Sumber: Proses Penelitian

Gambar 9. Area Under Curve: (a) Area Under Curve untuk Naïve Bayes Classifier (b) Area Under Curve untuk Support Vector Machine (c) Area Under Curve untuk K-Nearest Neighbor (d) Area Under Curve untuk Random Forest Classifier

Tabel 4. Analisis Perbandingan Sentimen

	NBC	KNN	SVM	RF
Precision	0.55	0.63	0.71	0.91
Recall	0.55	0.53	0.57	0.58
F1-Score	0.55	0.54	0.60	0.62
AUC	0.8482	0.7987	0.7755	0.8490
Accuracy	88%	91%	92%	93%

Sumber: Proses Penelitian

Diketahui dari tabel 4, bahwa NBC memperoleh tingkat akurasi sebesar 88%, lalu KNN sebesar 91%, kemudian SVM sebesar 92% dan RF sebesar 93%. Hal ini menunjukkan algoritma yang memiliki performa paling baik dalam analisis sentimen terhadap Tiket.com ini adalah *Random Forest*, karena memiliki tingkat akurasi yang lebih unggul dari algoritma

klasifikasi lain yang digunakan pada penelitian ini.

4. Kesimpulan

Penelitian analisis sentimen terhadap Tiket.com menggunakan algoritma klasifikasi ini membuktikan bahwa dari ke empat algoritma yang dilakukan pengujian bahwa Random Forest memperoleh nilai akurasi yang paling tinggi yaitu sebesar 93%. Dalam artian, *Random Forest* (RF) berhasil menganalisis sentimen masyarakat di Indonesia yang menggunakan layanan dari Tiket.com. Berdasarkan hasil pengujian, menunjukkan banyak *customer* yang merasa puas atas layanan yang diterima, ditandai dengan banyaknya sentimen positif dari media sosial. Karena dari sisi promosi dan pemasaran Tiket.com selalu aktif dalam memberikan penawaran khusus kepada *user*-nya dengan

promo terbaik dalam melakukan transaksi. Satu hal yang perlu ditingkatkan ke depannya adalah dari operasional bisnis khususnya *customer care* guna menangani permasalahan yang dihadapi oleh *customer* dan *user*. Dengan meningkatkan layanan operasional bisnis, tentunya Tiket.com akan mendapatkan *customer loyalty* berupa *feedback* dari kepuasan *customer*, sehingga memberikan peluang besar terhadap perusahaan untuk selalu mengembangkan inovasi dan strategi agar semakin dapat bersaing dengan para kompetitor Tiket.com.

Berdasarkan penelitian yang telah dilaksanakan maka terdapat beberapa saran yang direkomendasikan untuk penelitian ke depannya yaitu melakukan pelabelan data dengan tiga kelas data yaitu sentimen positif, negatif dan netral agar data yang digunakan lebih seimbang. Lalu, algoritma klasifikasi yang digunakan sebagai pembanding dengan *Random Forest* adalah *Logistic Regression*, *XGBoost* dan *Decision Tree* dengan parameter *random_state*. Teknik pengujian lain yang dapat diimplementasikan adalah *Cross Validation*, dan teknik pengambilan data sentimen dapat dari objek lain seperti data yang bersumber dari *review* Google Playstore sebagai sarana pengoptimalan hasil analisis sentimen terhadap Tiket.com

Referensi

- Anam, M. K., Pikir, B. N., & Firdaus, M. B. (2021). Penerapan Naïve Bayes Classifier, K-Nearest Neighbor (KNN) dan Decision Tree untuk Menganalisis Sentimen pada Interaksi Netizen danPemerintah. *MATRIK: Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, 21(1), 139–150. <https://doi.org/10.30812/matrik.v21i1.1092>
- Batta, M. (2018). Machine Learning Algorithms— A Review. *International Journal of Science and Research (IJSR)*, 18(8), 381–386. <https://doi.org/10.21275/ART20203995>
- Budiman, B. (2021). Perbandingan Algoritma Klasifikasi Data Mining untuk Penelusuran Minat Calon Mahasiswa Baru. *NUANSA INFORMATIKA*, 15(2), 37–52. <https://doi.org/10.25134/nuansa.v15i2.4162>
- Budiman, B., Habibi, C., & Wulandari, W. (2023). Comparative Analysis of Community Sentiment Against the Implementation of Booster Vaccination in Indonesia Using the K-Nearest Neighbor and Naïve Bayes Classifier Methods. *International Journal of Ethno-Sciences and Education Research (IJEER)*, 3(3).
- Budiman, B., Nursyanti, R., Alamsyah, R. Y. R., & Akbar, I. (2020). *Data Mining Implementation Using Naïve Bayes Algorithm and Decision Tree J48 In Determining Concentration Selection*. 1(3).
- Chauhan, R., & Hudaya, A. (2020). *April 2020 Revised: 20*. 1(4), 525–534. <https://doi.org/10.31933/DIJEMSS>
- Darwis, D., Siskawati, N., & Abidin, Z. (2021). Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Review Data Twitter Bmkg Nasional. *Jurnal Tekno Kompak*, 15(1), 131. <https://doi.org/10.33365/jtk.v15i1.744>
- Fitri, E. (2020). Analisis Sentimen Terhadap Aplikasi Ruangguru Menggunakan Algoritma Naive Bayes, Random Forest Dan Support Vector Machine. *Jurnal Transformatika*, 18(1), 71. <https://doi.org/10.26623/transformatika.v18i1.2317>
- Gunawan, B., Pratiwi, H. S., & Pratama, E. E. (2018). Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode Naive Bayes. *Jurnal Edukasi Dan Penelitian Informatika (JEPIN)*, 4(2), 113. <https://doi.org/10.26418/jp.v4i2.27526>
- Harahap, M. A., & Adeni, S. (2020). Tren Penggunaan Media Sosial Selama Pandemi Di Indonesia. *Jurnal Professional FIS UNIVED*, 7(2), 13–23.
- Indriani, A. (2020). Analisa Perbandingan Metode Naïve Bayes Classifier Dan K-Nearest Neighbor Terhadap Klasifikasi Data. *Sebatik*, 24(1), 1–7. <https://doi.org/10.46984/sebatik.v24i1.909>
- Irawan, F., & Samopa, F. (2018). A Comparative Assessment of the Random Forest and SVM Algorithms Using Combination of Principal Component Analysis and SMOTE For Accounts Receivable Seamless Prediction case study company X in Surabaya. *Proceedings of the 2nd International Seminar Of Contemporary ONBUSINESS Research & Management*, July, 2–14.
- Iwandini, I., Triayudi, A., & Soepriyono, G. (2023). Analisa Sentimen Pengguna Transportasi Jakarta Terhadap Transjakarta Menggunakan Metode Naives Bayes dan K-Nearest Neighbor. *Journal of Information System Research (JOSH)*, 4(2), 543–550. <https://doi.org/10.47065/josh.v4i2.2937>

- Kwartler, T. (2017). *What is Text Mining?* John Wiley & Sons, Ltd.
<https://dx.doi.org/10.1002/9781119282105.ch1>
- Mustofa, A., & Novita, R. (2022). Klasifikasi Sentimen Masyarakat Terhadap Pemberlakuan Pembatasan Kegiatan Masyarakat Menggunakan Text Mining Pada Twitter. *Building of Informatics, Technology and Science (BITS)*, 4(1), 200–208.
<https://doi.org/10.47065/bits.v4i1.1628>
- Prasetyo, V. R., Lazuardi, H., Mulyono, A. A., & Lauw, C. (2021). Penerapan Aplikasi RapidMiner Untuk Prediksi Nilai Tukar Rupiah Terhadap US Dollar Dengan Metode Linear Regression. *Jurnal Nasional Teknologi Dan Sistem Informasi*, 7(1), 8–17.
<https://doi.org/10.25077/teknosi.v7i1.2021.8-17>
- Prastyo, M. (2019). *Sentimen Analisis Untuk Membandingkan Kinerja Pada Traveloka Dan Tiket.Com* [Other, Universitas Amikom Purwokerto].
<https://doi.org/10/LAMPIRAN.pdf>
- Puspita, B. H., Muhajir, M., & Aliady, H. (2020). Topic Modeling Using Latent Dirichlet Allocation (LDA) and Sentiment Analysis for Marketing Planning Tiket.com: *Proceedings of the 2nd International Seminar on Science and Technology (ISSTEC 2019)*. The 2nd International Seminar on Science and Technology (ISSTEC 2019), Yogyakarta, Indonesia.
<https://doi.org/10.2991/assehr.k.201010.004>
- Putri, D. A., Kristiyanti, D. A., Indrayuni, E., Nurhadi, A., & Hadinata, D. R. (2020). Comparison of Naive Bayes Algorithm and Support Vector Machine using PSO Feature Selection for Sentiment Analysis on E-Wallet Review. *Journal of Physics: Conference Series*, 1641(1), 012085.
<https://doi.org/10.1088/1742-6596/1641/1/012085>
- Religia, Y. (2019). Feature Extraction Untuk Klasifikasi Pengenalan Wajah Menggunakan Support Vector Machine Dan K-Nearest Neighbor. *Pelita Teknologi: Jurnal Ilmiah Informatika, Arsitektur Dan Lingkungan*, 14(2), 85–92.
- Sholeha, E. W., Yunita, S., Hammad, R., Hardita, V. C., & Kaharuddin, K. (2022). Analisis Sentimen Pada Agen Perjalanan Online Menggunakan Naïve Bayes dan K-Nearest Neighbor. *JTIM : Jurnal Teknologi Informasi Dan Multimedia*, 3(4), 203–208.
<https://doi.org/10.35746/jtim.v3i4.178>
- Utomo, M. H. (2020). *Analisis Komparasi E-Service Quality Pada Aplikasi Online Travel Agent Dengan Menggunakan Analisis Sentimen (Traveloka, Tiket.Com, Dan Pegipegi)*. Universitas Telkom, S1 Manajemen (Manajemen Bisnis Telekomunikasi dan Informatika).
- VijayGaikwad, S., Chaugule, A., & Patil, P. (2014). Text Mining Methods and Techniques. *International Journal of Computer Applications*, 85(17), 42–45.
<https://doi.org/10.5120/14937-3507>