

# **Analisis Sentimen Terhadap Penerapan Sistem E-Tilang Pada Media Sosial Twitter Menggunakan Algoritma Support Vector Machine (SVM)**

**Dea Oktavia\*, Yudhi Raymond Ramadahan, Minarto**

Teknik Informatika, Sekolah Tinggi Teknologi Wastukencana, Purwakarta, Indonesia

Email: <sup>1,\*</sup>deaoctavia25@wastukencana.ac.id, <sup>2</sup>yudhi.raymond@wastukencana.ac.id, <sup>3</sup>minarto@stt-wastukencana.ac.id

Email Penulis Korespondensi: deaoctavia25@wastukencana.ac.id

**Abstrak**—Sistem e-Tilang menjadi solusi mendisiplinkan para pengendara kendaraan bermotor dari banyaknya melakukan pelanggaran berlalu-lintas. Keberadaan e-Tilang juga menjadi solusi mencegah kenakalan penegak hukum dari pungutan liar, istilah damai ditempat, hingga akuntabilitas uang denda. Efektif dan efisien sistem e-Tilang menimbulkan ragam komentar pada masyarakat. Belakangan ini populer sebuah teknik untuk menggali informasi dari tumpukan data khususnya pada *Twitter* yakni *Text Mining* atau sering juga disebut dengan analisis sentimen. *Twitter* merupakan jenis media sosial cukup populer dan diminati oleh seluruh masyarakat dunia termasuk Indonesia yang menyediakan berbagai informasi. Support Vector Machine (SVM) adalah seperangkat metode pembelajaran terbimbing yang menganalisis data dan mengenali pola, digunakan untuk klasifikasi dan analisis regresi dan dinilai merupakan metode yang relative baru. Tujuan penelitian ini untuk menganalisis sentimen pengguna twitter terhadap penerapan sistem e-Tilang dengan menggunakan algoritma *Support Vector Machine* (SVM) dengan menghitung nilai dari data tweet yang memiliki hasil sebesar 74.20%, *precision* sebesar 83.33% dan *recall* 5.28%. Hasil sentimen pada sosial media *Twitter* mengenai penerapan e-Tilang tergolong netral. Dari hasil penelitian menggunakan algoritma *Support Vector Machine* (SVM) hasil sentimen pada sosial media *Twitter* mengenai penerapan e-Tilang tergolong netral.

**Kata Kunci:** Text Mining; Analisis Sentimen; *Twitter*; Python; e-Tilang; SVM

**Abstract**—The e-Tilang system is a solution to disciplining motorized vehicle drivers from committing many traffic violations. The existence of e-Tilang is also a solution to preventing law enforcer delinquency from illegal levies, peaceful terms in place, to accountability for fines. The effectiveness and efficiency of the e-Tilang system raise various comments from the public. Lately, a technique has become popular for extracting information from piles of data, especially on *Twitter*, namely *Text Mining* or often also called sentiment analysis. *Twitter* is a type of social media that is quite popular and in demand by the whole world community, including Indonesia, which provides various information. Support Vector Machine (SVM) is a set of guided learning methods that analyze data and recognize patterns, is used for classification and regression analysis, and is considered a relatively new method. The purpose of this study was to analyze the sentiments of *Twitter* users regarding the implementation of the e-Tilang system using the Support Vector Machine (SVM) algorithm by calculating the value of the tweet data which has a yield of 74.20%, precision of 83.33% and recall of 5.28%. Sentiment results on social media *Twitter* regarding the implementation of e-tickets are classified as neutral. From the results of research using the Support Vector Machine (SVM) algorithm the results of sentiment on social media *Twitter* regarding the implementation of e-Tilang are classified as neutral.

**Keywords:** Text Mining; Sentiment Analysis; *Twitter*; Python; e-Tilang; SVM

## **1. PENDAHULUAN**

Pada era digital seperti sekarang data adalah sumber daya yang memenuhi kehidupan. Data yang banyak ini kemudian semakin menumpuk setiap harinya dan apabila dibiarkan tentunya hanya akan menjadi sampah data. Salah satunya adalah data tweets yang ada pada *Twitter*. Salah satu informasi yaitu opini pengguna twitter yang berisi tweet-tweet dan digunakan oleh masyarakat untuk sekedar meluapkan emosinya maupun membagikan suatu informasi [1]. Menurut data Statista pada bulan Juli 2021, negara Indonesia menempati posisi ke 6 pengguna twitter terbanyak dengan 15,7 juta pengguna [2]. Analisis sentimen merupakan proses memahami, mengekstrak dan mengolah data tekstual secara otomatis untuk mendapatkan informasi dengan tujuan untuk melihat kecenderungan opini terhadap sebuah objek, apakah cenderung beropini positif, negatif atau netral [3]. Opini masyarakat yang ada di twitter ini akan menjadi dataset untuk dilakukan analisis sentiment [4].

Peningkatan pelanggaran lalu lintas merupakan sebuah tantangan baru bagi para kepolisian untuk mampu menerapkan sanksi yang mendidik namun tetap memiliki efek jera. Salah satu cara untuk menekan pelanggaran adalah dengan melakukan sanksi administrative (tilang) yang dilakukan oleh pihak kepolisian. Namun yang terjadi selama ini sistem tilang sering disimpangkan oleh oknum sipil dan oknum anggota polisi untuk saling berkompromi agar kepentingan masing-masing bisa tercapai tanpa mengikuti prosedur yang berlaku. Bukan rahasia umum bila praktik suap-menyuap saat operasi lalu lintas terjadi. Itulah alasan yang mendasari Kepolisian Republik Indonesia menerapkan sistem baru bernama Electronic Traffic Law Enforcement (ETLE), atau masyarakat mengenal dengan istilah e-Tilang. Sistem ini launching serentak di Indonesia pada 6 Desember 2017. E-tilang diharapkan akan menjadi efek jera bagi para pelanggar lalu lintas sehingga diharapkan pengendara kendaraan bermotor akan menjadi tertib berkendara di jalan [5]. Penerapan sistem e-Tilang merupakan kebijakan untuk menggantikan sistem tilang manual yang menggunakan blanko/surat tilang [5]. Sistem e-Tilang menimbulkan ragam komentar pada masyarakat [6].

Data mining merupakan suatu proses penemuan informasi baru dengan cara mencari pola-pola atau aturan tertentu dari sejumlah data yang sangat besar merupakan pengertian data mining secara sederhana [7]. Text mining merupakan bagian dari data mining, yaitu proses untuk memperoleh suatu pengetahuan menggunakan seperangkat alat analisis

dimana pengguna berinteraksi dengan sekumpulan dokumen dari waktu ke waktu [8]. Analisis sentimen merupakan sebuah pemrosesan bahasa alami yang mampu mendeteksi emosi dan opini dari seseorang terhadap suatu topik tertentu. Analisis sentimen diawali dengan tahap pengumpulan data atau yang biasa disebut dengan data crawling [9]. Twitter ialah salah satu media sosial yang memungkinkan para penggunanya untuk membagikan informasi [10]. Support Vector Machine (SVM) adalah salah satu Teknik dalam machine learning yang berdasarkan teori struktural pembelajaran statistik. SVM mempunyai performa yang lebih baik apabila dibandingkan dengan metode machine learning lain [11]. Python merupakan bahasa pemrograman tinggi yang bisa melakukan eksekusi sejumlah instruksi multi guna secara langsung (interpretatif) dengan metode Object Oriented Programming dan juga menggunakan semantik dinamis untuk memberikan tingkat keterbacaan syntax [12]. Snsrape adalah library python yang dapat scraping data untuk beberapa media sosial, seperti Facebook, Twitter, Instagram [13].

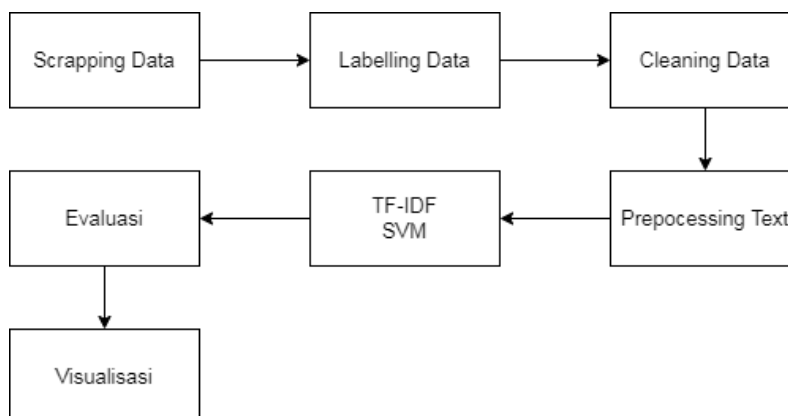
Setelah peneliti melakukan literature review, terdapat beberapa artikel yang berkaitan dengan penelitian ini, seperti yang dilakukan oleh Rakhmi Khalida, Siti Setiawati dengan judul “Analisis Sentimen Sistem E-Tilang Menggunakan Algoritma Naive Bayes Dengan Optimalisasi Information Gain”. Opini pada twitter dikumpulkan, lalu dilakukan tahapan preprocessing, selanjutnya dengan seleksi fitur information gain membantu mengurangi noise yang disebabkan oleh label-label yang tidak relevan, tahap selanjutnya adalah klasifikasi sentimen dengan algoritma Naive Bayes dan terakhir sentimen polarity. Penelitian ini menghasilkan accuracy 41,82%, presisi 50,51% dan recall 45,45% [6]. Artikel tersebut berkaitan dengan e-Tilang sesuai dengan penelitian ini.

Artikel lainnya dengan judul “Analisis Sentimen Pinjaman Online Di Twitter Menggunakan Algoritma Support Vector Machine (SVM)”. Pada penelitian ini data menunjukkan bahwa tingkat keakurasian untuk pinjaman online yaitu 62.00%. Hasil akurasi pada penelitian dinilai cukup baik. Analisis sentimen dengan algoritma Support Vector Machine (SVM) berhasil mengklasifikasikan sentimen masyarakat di twitter tentang pinjaman online [14].

Selanjutnya artikel dengan judul “Analisis Sentimen Twitter Terhadap Bitcoin dan Cryptocurrency Berbasis Python TextBlob”, Pada penelitian ini, seluruh data tweets yang berjumlah 3433 data telah diklasifikasikan menggunakan library python textblob dan didapatkan hasil dimana 41,3% tweets adalah positif, 44,9% adalah netral, dan 13,7% adalah negative [3]. Penelitian ini berhasil mengklasifikasikan 3 kelas sentimen.

Berdasarkan beberapa artikel yang berkaitan dengan penelitian ini, menunjukkan bahwa algoritma Support Vector Machine (SVM) metode untuk memprediksi dengan tingkatan akurasi nilai yang cukup baik.. Tujuan penelitian ini untuk membahas bagaimana cara memanfaatkan pendapat pengguna Twitter terhadap penerapan sistem e-Tilang menjadi analisis sentimen, dan kemudian membagi pendapat tersebut menjadi sentimen positif, negatif dan netral. Pengklasifikasian kumpulan data tweet tersebut memakai salah satu tata cara klasifikasi Support Vector Machine (SVM).

## 2. METODOLOGI PENELITIAN



Gambar 1. Kerangka Penelitian

Penelitian ini menggunakan metode Knowledge Discovery in Database (KDD). Metode KDD ini dapat berguna menemukan suatu teknik dari database yang ada [15]. Knowledge Discovery in Database merupakan sebuah proses pengetahuan dalam database. Knowledge Discovery in Database (KDD) adalah penerapan metode saintifik pada data mining. Data Mining (DM) adalah inti dari proses KDD, melibatkan kesimpulan dari algoritma yang mengeksplorasi data, mengembangkan model dan menemukan pola yang sebelumnya tidak diketahui [16]. Penggalan informasi tersembunyi dari sebuah basis data yang besar merupakan tujuan dari KDD dan data mining [17]. Pada penelitian ini menggunakan algoritma Support Vector Machine (SVM) sebagai model dalam klasifikasi yang kemudian dilakukan proses evaluasi. Tahapan-tahapan dari metode penelitian ini yaitu, Pengumpulan Data (Scrapping Data), Labelling Data, Cleaning Data, Preprocessing Text, Klasifikasi (TF-IDF dan SVM), Evaluasi, dan Visualisasi. Berikut penjelasan secara detail mengenai tahapan metode penelitian yang digunakan :

### 2.1 Pengumpulan Data

Scraping adalah mendapatkan data untuk kemudian melakukan ekstraksi informasi yang dimiliki oleh data tersebut. Cara kerja web scraping adalah dengan mengakses halaman web, memilih elemen data yang ada dalam halaman tersebut, melakukan ekstraksi dan transformasi bila diperlukan, dan terakhir menyimpan data tersebut menjadi dataset terstruktur [18].

## 2.2 Labelling Data

Pada sisi proses pelabelan data, diputuskan untuk mengategorikan data menjadi tiga kategori, yaitu positif, negatif, dan netral. Kategori positif digunakan untuk emosi seperti senang, ceria, rileks, santai, dan emosi yang secara garis besar berhubungan dengan kebajikan. Sedangkan kategori negatif digunakan untuk untuk emosi seperti sedih, marah, tertekan, takut, dan emosi-emosi lain yang mengakibatkan penderitaan. Lalu, tweet-tweet yang memenuhi kriteria penyaringan tetapi tidak dapat dikategorikan ke dalam dua kategori ini diberi label sebagai kategori netral. Proses pengkategorian ini dilakukan secara manual dan subjektif [19].

## 2.3 Cleaning Data

Cleaning data merupakan kegiatan penghapusan karakter-karakter yang tidak sesuai ketentuan yang dibuat seperti huruf atau karakter diluar alphabet a-z (termasuk tanda baca), menghapus link atau URL, hashtag, username [20]. Tahap cleaning ini peneliti menambahkan penghapusan duplikat pada data tweet.

## 2.4 Preprocessing Text

Dalam melakukan text mining, text dokumen harus dilakukan persiapan terlebih dahulu sehingga dapat digunakan untuk proses utama. Proses mempersiapkan text atau data mentah ini disebut text preprocessing. Text preprocessing berfungsi mengubah data yang tidak terstruktur menjadi terstruktur [21]. Tahap Preprocessing text ini terdapat 3 tahapan lanjutan yaitu, transformation, tokenization dan filtering.

## 2.5 Klasifikasi

Klasifikasi pada penelitian ini terdapat beberapa tahap yaitu TF-IDF dan melakukan klasifikasi Support Vector Machine (SVM). Berikut tahapan-tahapannya adalah:

### a. Term Frequency – Inverse Document Frequency TF-IDF

Term Frequency – Inverse Document Frequency (TF-IDF) merupakan pemodelan yang dimanfaatkan untuk menemukan hubungan antara term terhadap data dengan cara memberikan bobot pada setiap kata atau term [2]. Perhitungan nilai TF secara manual dilakukan dengan persamaan (1).

$$TF(k) = \sum F_{f,d} \quad (1)$$

Keterangan :

F = frekuensi kemunculan

k = kata

d = dokumen

Dilanjutkan dengan mencari nilai IDF dengan persamaan (2).

$$IDF(k) = \text{Log} \frac{D}{Df(k)} \quad (2)$$

Keterangan :

D = total jumlah dokumen

DF(k) = total dokumen yang mengandung frekuensi kemunculan kata k

Setelah kedua nilai didapatkan, dilakukan perhitungan nilai TF-IDF dengan persamaan (3).

$$TF - IDF(k) = TF(k) \times IDF(k) \quad (3)$$

### b. Support Vector Machine (SVM)

Support Vector Machine (SVM) adalah suatu teknik yang relatif baru untuk melakukan prediksi, baik dalam kasus klasifikasi maupun regresi [22].

$$f(x) = \sum_{i=1}^{\infty} a_i y_i K(x, x') + b \quad (4)$$

$a_i$  = alfa ke-i

$y_i$  = kelas data latih ke-i

M = jumlah data

$K(x, x')$  = fungsi kernel yang dipakai, dengan

x = data uji

$x_i$  = data latih ke-i

b = bias

## 2.6 Evaluasi

Pada tahap evaluasi dilakukan perhitungan accuracy, precision, recall dengan confusion matrix [19]. Evaluasi pada penelitian ini menggunakan Confusion Matrix (pengukuran performa untuk masalah klasifikasi Machine Learning dimana keluaran dapat berupa dua kelas atau lebih) yang dilakukan untuk mengetahui nilai performansi dari sistem yang sudah dibuat berdasarkan hasil dari klasifikasi. Proses evaluasi dilakukan pada python dan menggunakan model SVM.

## 2.7 Visualisasi

Visualisasi dilakukan untuk menyederhanakan data dan memudahkan pemahaman bagi pembaca. Hasil dari visualisasi ini berupa wordcloud yang dilakukan dari python dan visualisasi berupa dashboard menggunakan tools Microsoft Power BI.

# 3. HASIL DAN PEMBAHASAN

## 3.1 Pengumpulan Data

Data yang dikumpulkan dalam penelitian ini diperoleh dari media sosial twitter dengan cara meng-scraping tweet berbahasa Indonesia menggunakan snsrape Python dengan kata kunci e-Tilang. Diambil pada bulan Maret 2023 dengan rentang waktu dari 1 Januari 2019 sampai 28 Februari 2023. Ditunjukkan pada gambar 1.

```
#Scraping Data

query="(E-Tilang) lang:id until:2023-2-28 since:2019-1-1 -filter:links"
tweets = []
limits = 3000
# columns = ['Date', 'User', 'Tweet']
for tweet in sntwitter.TwitterSearchScrapper(query).get_items():
    # print(vars(tweet))

    if len(tweets) == limits:
        break
    else :
        tweets.append([tweet.content, tweet.date])

df = pd.DataFrame(tweets, columns=['Tweet', 'tanggal'])
print(df)
```

Gambar 2. Source Code Scrapping

Data yang didapat setelah proses scrapping di atas berjumlah 2.999 tweet, dengan kolom tweet dan tanggal. Seperti pada gambar 2.

```
<ipython-input-10-e34a07161bc0>:11: DeprecateWarning: content is deprecated, use rawContent instead
tweets.append([tweet.content, tweet.date])

                                Tweet \
0    Oiya. Traffic Light di Surabaya juga setertib ...
1    @catuaries Kapan ada bounty buat laporin e-tilang
2    Sementara itu Kapolda DIY Irjen Pol Suwondo Na...
3    Panewu Wates mengusulkan agar pembayaran denda...
4    @ZoneVenomWar Sekarang udah jarang distop di j...
...
2995 @bdgfess E-Tilang terapkan\nRekam cctv&gt;telu...
2996 @Anissa9889 @bacuulll @bintangmhyh Waduh, enak...
2997 Please @PolresBogorKota uji coba E-tilang dibi...
2998 @Rajadewa6660 Jd, harusnya penerapan e-tilang ...
2999 Opo grgr saiki e-tilang duite bfc gak entek en...

                                tanggal
0    2023-02-27 02:12:57+00:00
1    2023-02-26 14:04:32+00:00
2    2023-02-24 06:38:51+00:00
3    2023-02-24 06:38:39+00:00
4    2023-02-24 04:40:03+00:00
...
2995 2020-02-08 14:34:01+00:00
2996 2020-02-08 13:07:14+00:00
2997 2020-02-08 08:15:00+00:00
2998 2020-02-08 03:30:50+00:00
2999 2020-02-08 02:03:06+00:00

[3000 rows x 2 columns]
```

Gambar 3. Hasil Scrapping

## 3.2 Labelling Data

Pelabelan dilakukan secara manual dengan membaca tweet satu per satu dan melihat kata-kata. Jika ada kata-kata yang mengandung emosi maka diberi label negative, sedangkan kata-kata yang mengandung pujian diberi label positif dan yang tidak mengandung keduanya diberi label netral. Contoh labelling ditunjukkan pada tabel 1.

**Tabel 1.** Contoh Labelling Data

Tweet	Tanggal	Label
Kapan ada bounty buat laporin e-tilang	2/26/2023	0
Akibat e-tilang belum sempurna	12/23/2022	1
Hati-hati banyak kamera E-Tilang di Jakarta sekarang, silent lagi	2/17/2023	2

### 3.3 Cleaning Data

Pada tahap cleaning pertama, dilakukan penghapusan duplikat data secara manual pada excel, berikut merupakan contoh data tweet duplikat netral, contoh data duplikat ditunjukkan pada tabel 2.

**Tabel 2.** Contoh Data Duplikat

Tweet	Tanggal	Label
@TMCPoldaMetro izin nanya pak, kalau nakes kena e-tilang gage gimana konfirmasinya? Harusnya ga kena kan?	10/30/2023	0
@TMCPoldaMetro izin nanya pak, kalau nakes kena e-tilang gage gimana konfirmasinya? Harusnya ga kena kan?	10/30/2023	0

Tahap cleaning selanjutnya dilakukan penghapusan username. Tahap ini dilakukan untuk menghapus username dan simbol lainnya menggunakan python. Source code cleaning data ditunjukkan pada gambar 3.

```
[ ] # Cleaning Data

def remove_pattern(input_txt, pattern):
    r = re.findall(pattern, input_txt)
    for i in r:
        input_txt = re.sub(i, '', input_txt)
    return input_txt
data['remove_user'] = np.vectorize(remove_pattern)(data['Tweet'], "@[\w]*")
```

**Gambar 4.** Source Code Cleaning Data

Hasil data tweet setelah melewati proses cleaning penghapusan username. Hasil cleaning ditunjukkan pada gambar 4.

Tweet	tanggal	label	remove_user
Oiya. Traffic Light di Surabaya juga setertib ...	2023-02-27	1	Oiya. Traffic Light di Surabaya juga setertib ...
@catuaries Kapan ada bounty buat laporin e-tilang	2023-02-26	0	Kapan ada bounty buat laporin e-tilang
Sementara itu Kapolda DIY Irjen Pol Suwondo Na...	2023-02-24	2	Sementara itu Kapolda DIY Irjen Pol Suwondo Na...
Panewu Wates mengusulkan agar pembayaran denda...	2023-02-24	2	Panewu Wates mengusulkan agar pembayaran denda...
@ZoneVenomWar Sekarang udah jarang distop di j...	2023-02-24	1	Sekarang udah jarang distop di jalan\nE-tilan...

**Gambar 5.** Hasil Cleaning Hapus Username

### 3.4 Preprocessing Text

Proses preprocessing text terdiri dari transformation, tokenize dan filtering menggunakan Google Colab. Berikut adalah langkah-langkah text preprocessing :

#### a. Transformation

Tahap transformation terdiri dari case folding yang merupakan proses mengubah huruf besar menjadi huruf kecil. Contoh case folding ditunjukkan pada tabel 3.

**Tabel 3.** Case Folding

Sebelum	Sesudah
Hati-hati banyak kamera E-Tilang di Jakarta sekarang silent lagi	hati-hati banyak kamera e-tilang di jakarta sekarang silent lagi

#### b. Tokenization

Setelah dilakukan transformation selanjutnya dilakukan tokenisasi, yaitu pemecahan kalimat menjadi beberapa bagian atau beberapa kata. Contoh tokenization ditunjukkan pada tabel 4.

**Tabel 4.** Tokenization

Sebelum	Sesudah
---------	---------

e-tilang ini kalo dikumpulin bisa buat bayar pajak mobil gila	“e-tilang” “ini” “kalo” “dikumpulin” “bisa” “buat” “bayar” “pajak” “mobil” “gila”
---	---

**c. Filtering**

Tahap filtering dilakukan untuk penyaringan kalimat tweet atau pengurangan kata. Pada proses filtering ini peneliti menggunakan filtering stopword dan stemming.

**1. Stopword**

Filter stopword biasanya disebut sebagai stopword removal, tahap ini dilakukan untuk menghilangkan kata-kata yang tidak relevan tetapi tidak berpengaruh pada suatu kalimat seperti “ada” “yang”. Contoh stopword ditunjukkan pada tabel 5.

**Tabel 5.** Contoh Stopword

Sebelum	Sesudah
Kapan ada bounty buat laporin e-tilang	bounty lapor etilang

**2. Stemming**

Proses aktual digunakan untuk menyaring kata-kata yang mengandung konjungsi, pronomina, dan preposisi menjadi kata dasar dengan menghilangkan awalan atau akhiran. Contoh stemming ditunjukkan pada tabel 6.

**Tabel 6.** Contoh Stemming

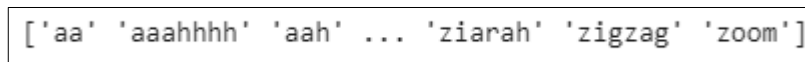
Sebelum	Sesudah
iya masih, jadi e-tilang buat yang kalo lolos dari polisi	iya etilang lolos polisi

**3.5 Klasifikasi**

Setelah tahap preprocessing dilakukan tahap klasifikasi yang terdiri dari tahap TF-IDF dan SVM Model, berikut tahapan – tahapan dari klasifikasi :

**a. Pembobotan Kata TF-IDF**

Pada topik penelitian e-Tilang ini terdapat 5590 kata. Contoh kata ditunjukkan pada gambar 5.



**Gambar 6.** Contoh Kata

Selanjutnya tahap pembobotan kata (TF) yang telah dilakukan, hasil ditunjukkan pada gambar 6.

	Nilai	Fitur
0	0.306198	aa
1	0.210604	aaahhhh
2	0.139479	aah
3	0.200745	aamanilah
4	0.406268	aamiin
...	...	...
5585	2.289985	zaman
5586	0.038178	zebra
5587	0.148101	ziarah
5588	1.166820	zigzag
5589	1.352945	zoom

**Gambar 7.** Hasil TF

Setelah mendapatkan bobot (nilai) pada dokumen tahap selanjutnya adalah menghitung document frequency (DF) atau banyak kata yang ada dalam dokumen, yang kemudian total dari seluruh dokumen tersebut dibagi dengan document frequency (DF) untuk menghasilkan dokumen/DF. Hasil ditunjukkan pada gambar 7.

	Nilai	Fitur
	151.774294	pungli
	104.089416	polri
	76.540975	satu
	71.047980	superappskanaksimu
	70.056093	aplikasi

**Gambar 8.** Hasil IDF

Untuk mendapatkan nilai TF-IDF yaitu dengan cara mengalikan bobot tiap kata dari seluruh dokumen, dengan nilai dari idf kata tersebut. Hasil ditunjukkan pada gambar 8.

	aamiin	abang	abdulrahman	abi	about	abraham	abram	abrar	absurd	abu	...	yekan	yha	yiya	yogyakarta	yok	yol	yoo	zaman	zigzag	zoom
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0
2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0
3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.259262	0.0	0.0	0.0	0.0	0.0	0.0
4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0

Gambar 9. Hasil TF-IDF

Perhitungan nilai TF secara manual dilakukan dengan persamaan (1).

$$TF(k) = \sum F_{f,d} \quad (1)$$

Dilanjutkan dengan mencari nilai IDF dengan persamaan (2).

$$IDF(k) = \text{Log} \frac{D}{df(k)} \quad (2)$$

Setelah kedua nilai didapatkan, dilakukan perhitungan nilai TF-IDF dengan persamaan (3).

$$TF - IDF(k) = TF(k) \times IDF(k) \quad (3)$$

### b. SVM Model

Setelah melakukan pembobotan (TF-IDF) tahap selanjutnya pada penelitian ini adalah implementasi model Support Vector Machine (SVM). Tahap pertama yaitu import SVM model yang ditunjukkan pada gambar 9.

```
#Import Library
import random
from sklearn.model_selection import train_test_split

#Algoritme
#Support Vector Machine
from sklearn import svm
```

Gambar 10. Source Code Import Model SVM

Tahap kedua yaitu memisahkan data training dan data testing dengan perbandingan 80 data training dan 20 data testing. Ditunjukkan pada gambar 10.

```
x = selected_x
y = data.label
#Memisahkan data training dan data testing dengan perbandingan 80:20
x_train, x_test, y_train, y_test = train_test_split(x,y,test_size=0.2, random_state=0)
```

Gambar 11. Source Code Memisahkan Data

Tahap ketiga yaitu menampilkan berapa banyak data training dan data testing. Ditunjukkan pada gambar 11.

```
#print
print('Banyak data x_train :',len(x_train))
print('Banyak data x_test :',len(x_test))
print('Banyak data y_train :',len(y_train))
print('Banyak data y_test :',len(y_test))
```

Gambar 12. Source Code Menampilkan Data

Berikut adalah hasil data yang dapat ditampilkan. Ditunjukkan pada gambar 12.

```
Banyak data x_train : 2261
Banyak data x_test : 566
Banyak data y_train : 2261
Banyak data y_test : 566
```

Gambar 13. Jumlah Data Training dan Testing

Berdasarkan gambar 13, perbandingan Data yang digunakan yaitu data training 80% dan data testing 20%. Data training sebanyak 2261 data dan data testing sebanyak 566 data.

Tahap keempat yaitu memasukkan training model yang digunakan yaitu Support Vector Machine (SVM). Source code dapat dilihat pada gambar 13.

```
#Training Model

from datetime import datetime
start_time = datetime.now()
from joblib import dump
#algoritme fitting

#text_algorithm = MultinomialNB()
text_algorithm = svm.SVC(kernel="rbf", C=1.0)

model = text_algorithm.fit(x_train, y_train)

# save the model to disk
dump(model, filename="model_sentiment_naive.joblib")

end_time = datetime.now()
result_time = end_time-start_time
print("Duration:",result_time)

#y_train
```

Gambar 14. Source Code Training Model

Tahap kelima yaitu memasukkan model prediksi yang berisi data testing. Source code dapat dilihat pada gambar 14.

```
#Prediksi
predicted = model.predict(x_test)

#Hasil ini akan dibandingkan dengan nilai y_test (labeling) / y_train
predicted
```

Gambar 15. Source Code Model Prediksi

### 3.6 Evaluasi

Pengujian dilakukan menggunakan metode cross validation kemudian diukur menggunakan Confusion Matrix melalui library SVM, untuk mengetahui hasil akurasi, presisi dan recall. Ditunjukkan pada gambar 15.

```
[ ] #Library evaluasi
from sklearn.metrics import confusion_matrix
from sklearn.metrics import classification_report
from sklearn.metrics import classification_report
from sklearn.metrics import confusion_matrix
import seaborn as sns

#Compute performance manually
NewprediksiBenar = (predicted == y_test).sum()
NewprediksiSalah = (predicted != y_test).sum()

print("prediksi benar: ", NewprediksiBenar, " data")
print("prediksi salah: ", NewprediksiSalah, " data")
print("Akurasi Algoritme: ", NewprediksiBenar/(NewprediksiBenar+NewprediksiSalah)*100,"%")

CM = confusion_matrix(y_test,predicted)

TN = CM[0][0]
FN = CM[1][0]
TP = CM[1][1]
FP = CM[0][1]
precision = TP/(TP+FP)
recall = TP/(TP+FN)
print ("TRUE NEGATIVE (TN):",TN)
print ("FALSE NEGATIVE (FN):",FN)
print ("TRUE POSITIVE (TP):",TP)
print ("FALSE POSITIVE (FP):",FP)
print ("PRECISION:",precision*100,"%")
print ("RECALL:",recall*100,"%")
print(classification_report(y_test,predicted))
```

Gambar 16. Source Code Pengujian

Setelah dilakukan tahap evaluasi confusion matrix menggunakan python dengan menggunakan algoritma SVM pada data tweet dengan data training 80% dan data testing 20%. Menghasilkan nilai akurasi sebesar 74%, precision 83%, recall 5.2%. Hasil pengujian ditunjukkan pada gambar 16.



```

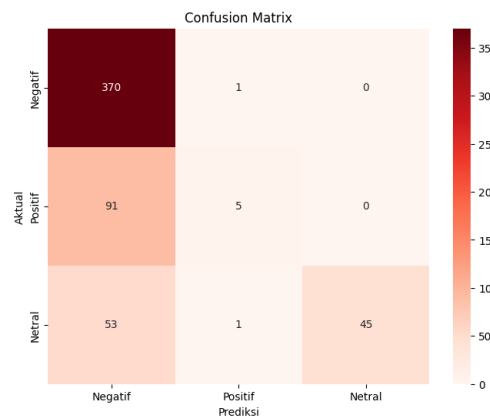
prediksi benar: 420 data
prediksi salah: 146 data
Akurasi Algoritme: 74.20494699646643 %
TRUE NEGATIVE (TN): 370
FALSE NEGATIVE (FN): 91
TRUE POSITIVE (TP): 5
FALSE POSITIVE (FP): 1
PRECISION: 83.33333333333334 %
RECALL: 5.208333333333334 %

```

	precision	recall	f1-score	support
0	0.72	1.00	0.84	371
1	0.71	0.05	0.10	96
2	1.00	0.45	0.62	99
accuracy			0.74	566
macro avg	0.81	0.50	0.52	566
weighted avg	0.77	0.74	0.67	566

Gambar 17. Hasil Pengujian

Berdasarkan gambar 17, hasil pengujian menunjukkan prediksi benar sebanyak 420 data, dan prediksi salah sebanyak 146 data. Nilai True Negative (TN) 370, Flase Negative (FN) 91, True Positive (TP) 5, False Positif 1. Nilai Precision Netral (0) 0.72, recall 1.00, f1-score 0.84, support 371. Nilai Precision Positif (1) 0.71, recall 0.05, f1-score 0.10, support 96, dan nilai precision positif (2) 1.00, recall 0.45, fi-score 0.62 support 99.



Gambar 18. Heatmap Confusion Matix

Heatmap confusion matrix berdasarkan gambar 18, divisualisasikan agar mempermudah dalam menentukan sentiment aktual dan sentimen prediksi.

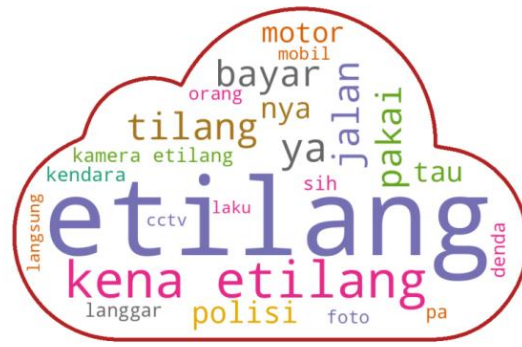
### 3.7 Visualisasi

Berdasarkan hasil dari tahapan-tahapan yang telah dilakukan, hasil penelitian dapat disajikan dalam bentuk visualisasi. Hasil visualisasi disajikan dalam bentuk word cloud dan dashboard. Visualisasi word cloud keseluruhan sentimen dapat ditunjukkan pada gambar 18 untuk sentiment positif, gambar 19 sentimen netral dan gambar 20 sentimen negatif.



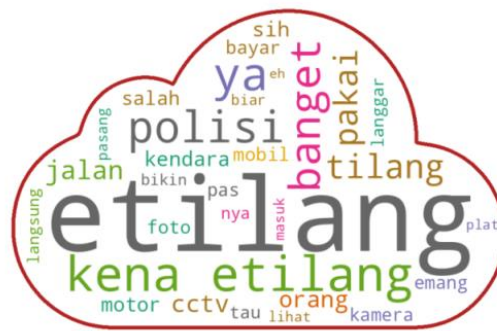
Gambar 19. Word Cloud Sentimen Positif Keseluruhan

Pada gambar 19 word cloud menunjukkan kata yang sering muncul, pada sentimen positif yang berjumlah 426 tweet.



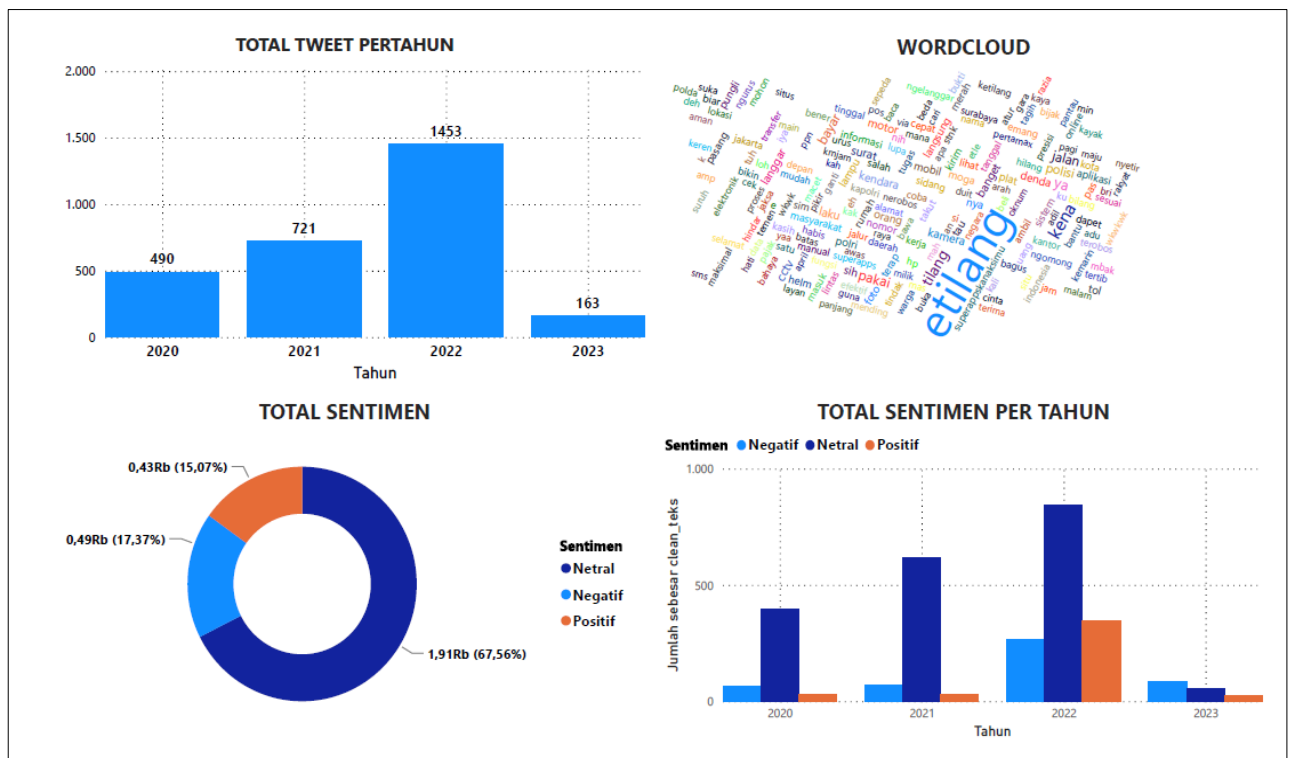
Gambar 20. Word Cloud Sentimen Netral Keseluruhan

Pada gambar 20 wordcloud menunjukkan kata yang sering muncul, pada sentiment netral yang berjumlah 1910 tweet.



Gambar 21. Word Cloud Sentimen Negatif Keseluruhan

Pada gambar 21 word cloud menunjukkan kata yang sering muncul pada sentiment positif yang berjumlah 491 tweet.



Gambar 22. Visualisasi Dashboard

Visualisasi dashboard ditunjukkan pada gambar 18, menunjukkan total tweet pertahun, wordcloud keseluruhan sentiment total sentiment dan total sentiment pertahun.

#### 4. KESIMPULAN

Analisis sentimen menggunakan algoritma Support Vector Machine (SVM) berhasil mengklasifikasikan sentimen masyarakat pada sosial media Twitter mengenai penerapan e-Tilang. Berdasarkan proses pengumpulan data diperoleh 2999 data, setelah melakukan proses Text Mining diperoleh sebanyak 2827 data. Terdapat 426 data opini positif, 491 negatif dan 1910 netral. Dapat disimpulkan bahwa hasil sentimen masyarakat pada sosial media Twitter mengenai penerapan e-Tilang tergolong netral. Implementasi algoritma Support Vector Machine (SVM) pada penelitian ini menunjukkan tingkat keakurasian sebesar 74.20%, precision sebesar 83.33% dan recall 5.28%.

#### REFERENCES

- [1] N. Legiawati, T. I. Hermanto, and Y. R. Ramadhan, "Analisis Sentimen Opini Pengguna Twitter Terhadap Perusahaan Jasa Ekspedisi Menggunakan Algoritma Naïve Bayes Berbasis PSO," vol. 9, no. 4, pp. 930–937, 2022, doi: 10.30865/jurikom.v9i4.4629.
- [2] A. D. Pangestu, I. E. S. Kom, M. Si, N. C. S. Kom, and M. Kom, "Analisis Sentimen Terhadap Ppk Darurat Pada Media Sosial Twitter Menggunakan Metode Naïve Bayes Dengan Seleksi Fitur Information Gain," pp. 825–834, 2022.
- [3] R. Parlita, S. I. Pradika, A. M. Hakim, and K. R. N M, "Analisis Sentimen Twitter Terhadap Bitcoin Dan Cryptocurrency Berbasis Python Textblob," *J. Ilm. Teknol. Inf. dan Robot.*, vol. 2, no. 2, pp. 33–37, 2020, doi: 10.33005/jifti.v2i2.22.
- [4] R. A. Raharjo, I. Made, G. Sunarya, D. Gede, and H. Divayana, "Perbandingan Metode Naïve Bayes Classifier Dan Support Vector Machine Pada Kasus Analisis Sentimen Terhadap Data Vaksin Covid-19 Di Twitter," *J. Ilm. Elektron. Dan Komput.*, vol. 15, no. 2, pp. 456–464, 2022, [Online]. Available: <http://journal.stekom.ac.id/index.php/elkom/page456>
- [5] F. A. Abdullah and F. Windiyastuti, "Electronic Traffic Law Enforcement (ETLE) sebagai Digitalisasi Proses Tilang," *J. Kewarganegaraan*, vol. 6, no. 2, pp. 3004–3008, 2022.
- [6] R. Khalida and S. Setiawati, "Analisis Sentimen Sistem E-Tilang Menggunakan Algoritma Naive Bayes Dengan Optimalisasi Information Gain," *J. Inform. Inf. Secur.*, vol. 1, no. 1, pp. 19–26, 2020, doi: 10.31599/jiforty.v1i1.137.
- [7] I. P. Ninditama, I. P. Ninditama, W. Cholil, M. Akbar, and D. Antoni, "Klasifikasi Keluarga Sejahtera Study Kasus : Kecamatan Kota Palembang," vol. 15, no. 2, pp. 37–49, 2020.
- [8] Y. Rahman, "Penerapan machine learning dalam e-commerce perusahaan bukalapak," no. 1118093000049, pp. 1–9, 2018.
- [9] A. Hasiholan, I. Cholissodin, and N. Yudistira, "Analisis Sentimen Tweet Covid-19 Varian Omicron pada Platform Media Sosial Twitter menggunakan Metode LSTM berbasis Multi Fungsi Aktivasi dan GLOVE," vol. 6, no. 10, pp. 4653–4661, 2022, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [10] S. M. Tambunan, Y. Nataliani, and E. S. Lestari, "Perbandingan Klasifikasi dengan Pendekatan Pembelajaran Mesin untuk Mengidentifikasi Tweet Hoaks di Media Sosial Twitter," *J. Edukasi dan Penelit. Inform.*, vol. 7, no. 2, p. 112, 2021, doi: 10.26418/jp.v7i2.47232.
- [11] A. C. Najib, A. Irsyad, G. A. Qandi, and N. A. Rakhmawati, "Perbandingan Metode Lexicon-based dan SVM untuk Analisis Sentimen Berbasis Ontologi pada Kampanye Pilpres Indonesia Tahun 2019 di Twitter," *Fountain Informatics J.*, vol. 4, no. 2, p. 41, 2019, doi: 10.21111/fij.v4i2.3573.
- [12] M. K. Rahmadhika and A. M. Thantawi, "Rancang Bangun Aplikasi Face Recognition Pada Pendekatan CRM Menggunakan Opencv Dan Algoritma Haarcascade," *IKRA-ITH Inform. J. Komput. dan Inform.*, vol. 5, no. 1, pp. 109–118, 2021.
- [13] I. Kurniawan, "Klasifikasi Opini Masyarakat Di Twitter Tentang Kebocoran Data Yang Terjadi Di Indonesia," *J. Inform. Kaputama*, vol. 7, no. 1, pp. 1–7, 2023.
- [14] D. S. Utami and A. Erfina, "Analisis Sentimen Pinjaman Online di Twitter Menggunakan Algoritma Support Vector Machine (SVM)," *SISMATIK (Seminar Nas. Sist. Inf. dan Manaj. Inform.)*, vol. 1, no. 1, pp. 299–305, 2021.
- [15] I. A. Ropikoh, R. Abdulhakim, U. Enri, and N. Sulistyowati, "Penerapan Algoritma Support Vector Machine (SVM) untuk Klasifikasi Berita Hoax Covid-19," *J. Appl. Informatics Comput.*, vol. 5, no. 1, pp. 64–73, 2021, doi: 10.30871/jaic.v5i1.3167.
- [16] M. R. Andryan, M. Fajri, and N. Sulistyowati, "Komparasi Kinerja Algoritma Xgboost Dan Algoritma Support Vector Machine (Svm) Untuk Diagnosis Penyakit Kanker Payudara," *JIKO (Jurnal Inform. dan Komputer)*, vol. 6, no. 1, p. 1, 2022, doi: 10.26798/jiko.v6i1.500.
- [17] A. Asroni, H. Fitri, and E. Prasetyo, "Penerapan Metode Clustering dengan Algoritma K-Means pada Pengelompokan Data Calon Mahasiswa Baru di Universitas Muhammadiyah Yogyakarta (Studi Kasus: Fakultas Kedokteran dan Ilmu Kesehatan, dan Fakultas Ilmu Sosial dan Ilmu Politik)," *Semesta Tek.*, vol. 21, no. 1, pp. 60–64, 2018, doi: 10.18196/st.21i2.11.
- [18] E. Suryati and A. A. Aldino, "Analisis Sentimen Transportasi Online Menggunakan Ekstraksi Fitur Model Word2vec Text Embedding Dan Algoritma Support Vector Machine (SVM)," vol. 4, no. 1, pp. 96–106, 2023.
- [19] R. Ferdiana, F. Jatmiko, D. D. Purwanti, A. S. T. Ayu, and W. F. Dicka, "Dataset Indonesia untuk Analisis Sentimen," *J. Nas. Tek. Elektro dan Teknol. Inf.*, vol. 8, no. 4, p. 334, 2019, doi: 10.22146/jnteti.v8i4.533.
- [20] Y. Nurtikasari, Syariful Alam, and Teguh Iman Hermanto, "Analisis Sentimen Opini Masyarakat Terhadap Film Pada Platform Twitter Menggunakan Algoritma Naive Bayes," *INSOLOGI J. Sains dan Teknol.*, vol. 1, no. 4, pp. 411–423, 2022, doi: 10.55123/insologi.v1i4.770.
- [21] A. Rahman Isnain, A. Indra Sakti, D. Alita, and N. Satya Marga, "Sentimen Analisis Publik Terhadap Kebijakan Lockdown Pemerintah Jakarta Menggunakan Algoritma Svm," *Jdmsi*, vol. 2, no. 1, pp. 31–37, 2021, [Online]. Available: <https://t.co/NfhfMjtXw>
- [22] R. Mahendrajaya, G. A. Buntoro, and M. B. Setyawan, "Analisis Sentimen Pengguna Gopay Menggunakan Metode Lexicon Based Dan Support Vector Machine," *Komputek*, vol. 3, no. 2, p. 52, 2019, doi: 10.24269/jkt.v3i2.270.