



## Analisis Sentimen Terhadap Layanan Indihome Berdasarkan Twitter Dengan Metode Klasifikasi *Support Vector Machine* (SVM)

Rian Tineges, Agung Triayudi\*, Ira Diana Sholihati

Fakultas Teknologi Komunikasi dan Informatika, Sistem Informasi, Universitas Nasional, Jakarta, Indonesia

Email: <sup>1</sup>rian.tineges@gmail.com, <sup>2,\*</sup>agungtriayudi@civitas.unas.ac.id, <sup>3,\*</sup>iradiana2803@gmail.com

Email Penulis Korespondensi: [agungtriayudi@civitas.unas.ac.id](mailto:agungtriayudi@civitas.unas.ac.id)

**Abstrak**— Pada tahun 2018 sebesar 18,9% penduduk di Indonesia menyebutkan bahwa alasan utama mereka menggunakan internet adalah untuk media sosial. Salah satu media sosial dengan pengguna aktif sebesar 6,43 juta pengguna adalah Twitter. Berdasarkan lonjakan informasi yang dipublikasi melalui Twitter, maka memungkinkan bahwa informasi tersebut dapat mengandung opini pengguna terhadap suatu objek, objek tersebut dapat berupa kejadian disekitar masyarakat seperti suatu produk atau layanan. Hal ini membuat perusahaan menggunakan Twitter sebagai media untuk menyebarkan informasi. Contohnya adalah perusahaan *Internet Service Provider* (ISP) seperti Indihome. Melalui Twitter, pengguna dapat saling berdiskusi mengenai keluhan ataupun kepuasan mereka terhadap layanan Indihome. Dibutuhkan suatu metode yaitu analisis sentimen untuk memahami apakah data tekstual tersebut termasuk opini negatif atau opini positif. Maka, penulis menggunakan metode *Support Vector Machine* (SVM) dalam analisis sentimen tentang opini pengguna layanan Indihome pada Twitter, dengan tujuan untuk mendapatkan model klasifikasi sentimen menggunakan SVM, dan untuk mengetahui seberapa besar akurasi yang dihasilkan oleh metode SVM yang diterapkan pada analisis sentimen, serta untuk mengetahui seberapa puas pengguna layanan Indihome berdasarkan Twitter. Setelah dilakukan pengujian dengan metode SVM hasilnya adalah *accuracy* 87%, *precision* 86%, *recall* 95%, *error rate* 13%, dan *f1-score* 90%.

**Kata Kunci:** Analisis, Sentimen, Opini, Twitter, Indihome, SVM

**Abstract**—In the year 2018, 18.9% of the population in Indonesia mentioned that the main reason for their use of the Internet is social media. One of the social media with an active user of 6.43 million users is Twitter. Based on the surge of information published via Twitter, it is possible that such information may contain the user's opinions on an object, such objects may be events around the community such as a product or service. This makes the company use Twitter as a medium to disseminate information. An example is an Internet Service Provider (ISP) such as Indihome. Through Twitter, users can discuss each other's complaints or satisfaction with Indihome's services. It takes a method of sentiment analysis to understand whether the textual data includes negative opinions or positive opinions. Thus, the authors use the Support Vector Machine (SVM) method in sentiment analysis on the opinions of the Indihome service user on Twitter, with the aim of obtaining a sentiment classification model using SVM, and to know how much accuracy the SVM method generates, which is applied to sentiment analysis, and to see how satisfied the Indihome service users are based on Twitter. After testing with SVM method The result is *accuracy* 87%, *precision* 86%, *recall* 95%, *error rate* 13%, and *F1-score* 90%.

**Keywords:** Analysis, Sentiment, Opinions, Twitter, Indihome, SVM

### 1. PENDAHULUAN

Pada tahun 2018, 64,8% dari total populasi penduduk di Indonesia merupakan pengguna internet. Angka tersebut meningkat hampir 28 juta jiwa dari tahun 2017-2018. Dimana sebesar 18,9% diantaranya menyebutkan alasan utama mereka dalam menggunakan internet adalah untuk media sosial [1]. Salah satu media sosial dengan jumlah pengguna aktif mencapai 6,43 juta orang pada Januari 2019 adalah Twitter [2]. Twitter merupakan media sosial yang digunakan untuk mencari informasi tentang bisnis, hiburan, ekonomi, politik, dan lainnya. Dengan banyaknya jumlah *tweet* yang dipublikasikan melalui Twitter, *tweet* tersebut dapat mengandung opini pengguna mengenai sebuah objek, objek tersebut dapat berupa kejadian disekitar masyarakat seperti suatu produk atau layanan [3]. Hal ini membuat perusahaan menggunakan Twitter sebagai media untuk menyebarkan informasi. Contohnya adalah perusahaan jasa *Internet Service Provider* (ISP) seperti Indihome.

Indihome merupakan operator penyedia ISP yang paling banyak digunakan oleh masyarakat Indonesia, dimana sebanyak 8,7% memilih Indihome sebagai operator penyedia ISP yang mereka gunakan. Persentase tersebut lebih banyak dari pada persentase operator penyedia ISP lainnya [1]. Melalui Twitter, pengguna dapat saling berdiskusi mengenai kritik, saran ataupun kepuasan mereka terhadap layanan Indihome [3]. Perbedaan Twitter dari media sosial yang lain adalah Twitter memungkinkan tiap pengguna dapat melacak *tweet* pengguna lain tanpa persetujuan dari pengguna tersebut [4].

Pandangan masyarakat atau umpan balik selalu terbukti menjadi sumber daya paling penting dan berharga bagi perusahaan [5]. Sehingga berdasarkan *tweet* yang mereka publikasi, dapat dilihat gambaran mengenai opini pengguna melalui proses penggalian informasi lebih lanjut dari tiap *tweet* yang dipublikasi, akan tetapi jumlah data tersebut banyak sehingga dibutuhkan sebuah metode untuk dapat mengubah data tersebut menjadi sebuah informasi atau sebuah pengetahuan. Suatu metode yang dapat digunakan untuk mewujudkannya yaitu dengan menggunakan analisis sentimen [3][6]. Analisis sentimen ialah suatu metode dalam memahami, mengekstrasi data sentimen yang biasanya akan dikategorikan berdasarkan polaritasnya, apakah positif atau negatif dengan menggunakan teknik Natural Language Processing (NLP) [7][8][9]. Saat ini analisis sentimen menjadi salah satu topik panas yang banyak digunakan para peneliti, yang memiliki tujuan untuk menyediakan informasi dari sebuah



dataset yang tidak terstruktur [10]. Dari hasil tersebut akan menjadi sebuah informasi yang memiliki makna seperti penilaian suatu brand, pemberian rating, dan penyaringan opini untuk membantu perusahaan ataupun masyarakat.

Seperti pada penelitian analisis sentimen yang dilakukan oleh Pangestu tahun 2019 dengan objek opini tentang partai politik di Indonesia. Penelitian tersebut bertujuan untuk melakukan klasifikasi sikap politik terhadap partai politik Indonesia dengan metode *Support Vector Machine* (SVM) yang memiliki hasil *accuracy* sebesar 86% dan pengujian dengan *10-Fold Cross Validation* rata-rata *accuracy* sebesar 71% dengan *error rate* sebesar 29% [11].

Penelitian analisis sentimen dengan objek opini maskapai penerbangan pada Twitter yang dilakukan oleh Pravina tahun 2019. Penelitian tersebut bertujuan untuk mengklasifikasikan komentar masyarakat ke dalam kelas positif atau kelas negatif, agar dapat membantu pihak perusahaan maupun masyarakat, dengan metode *Support Vector Machine* penelitian tersebut memiliki hasil *accuracy* 40%, *precision* 40%, *recall* 100%, dan *f-measure* [12].

Penelitian analisis sentimen dengan objek konten radikal pada Twitter yang dilakukan oleh Alvianda tahun 2019. Penelitian tersebut bertujuan untuk mengklasifikasikan *tweet* yang termasuk kelas konten radikal positif dan kelas konten radikal negatif dengan metode *Support Vector Machine* yang mempunyai hasil akurasi tertingginya adalah 70% [13].

Penelitian analisis sentimen dengan objek penerapan sistem ganjil genap di Tol Bekasi yang dilakukan oleh Utama H.S, dkk tahun 2019. Penelitian tersebut bertujuan untuk memberikan informasi terhadap efektifitas sistem ganjil genap Tol Bekasi berdasarkan media sosial dengan menggunakan pengklasifikasian *Support Vector Machine*. Adapun hasil yang didapat dari penelitian tersebut yaitu *accuracy* 78,18%, *Precision* 74,03%, dan *Sensitivity* atau *Recall* 86,82% [14].

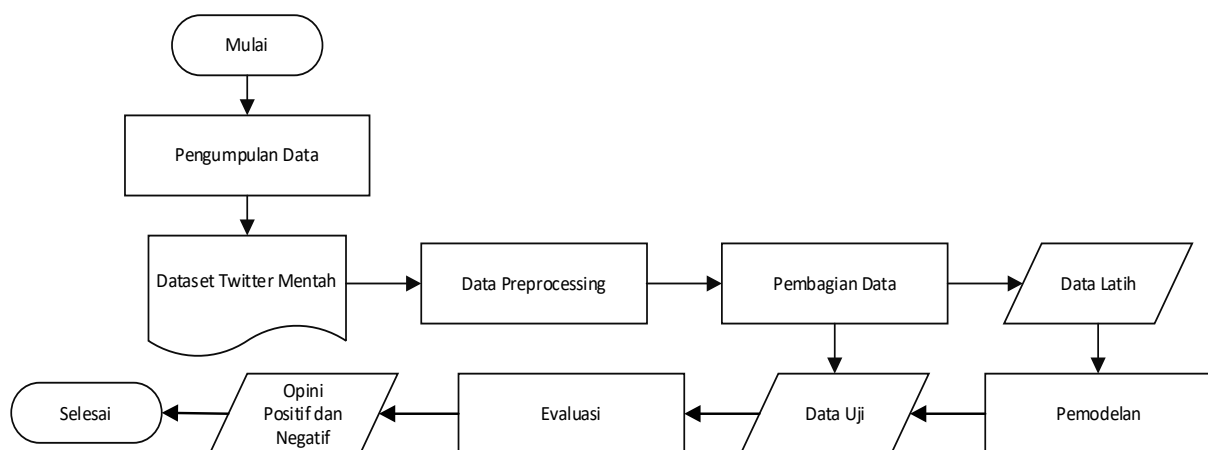
Penelitian lain mengenai analisis sentimen dengan objek opini pengguna layanan Telkom dan Biznet yang dilakukan oleh Haranto tahun 2019. Penelitian tersebut bertujuan untuk mengetahui persentase sentimen positif dan negatif dari Telkom dan Biznet dengan *Support Vector Machine*, memiliki hasil pengujian yaitu *accuracy* 79,6%, *precision* 76,5%, *recall* 72,8% , dan *F1-score* 74,6% untuk Telkom, serta *accuracy* 83,2%, *precision* 78,8%, *recall* 71,6%, dan *F1-score* 75% untuk Biznet [15].

Berdasarkan latar belakang diatas, serta berpedoman pada penelitian sebelumnya, maka didapat rumusan masalah, yaitu bagaimana mendapatkan model klasifikasi sentimen menggunakan *Support Vector Machine* (SVM). Selain itu, ruang lingkup dari penelitian ini yaitu *tweet* yang dianalisis diambil dari 16 Maret 2020 sampai 22 Maret 2020, data yang digunakan hanya yang menyebutkan/*mention* akun Indihome (@IndiHome), *tweet* yang digunakan berbahasa Indonesia dan metode klasifikasi yang diterapkan adalah *Support Vector Machine* (SVM). Adapun tujuan dari penelitian ini adalah untuk dapat pemodelan klasifikasi sentimen dengan *Support Vector Machine* (SVM), untuk mengetahui seberapa besar akurasi yang dihasilkan oleh metode SVM, serta untuk mengetahui seberapa puas pengguna layanan Indihome berdasarkan Twitter.

## 2. METODOLOGI PENELITIAN

### 2.1 Diagram Alur Proses Penelitian

Penelitian berupa kumpulan fakta atau data yang didapat dari suatu objek penelitian. Dalam hal ini objek penelitian yang dimaksud adalah data *tweet* yang menyebut akun Twitter @IndiHome yang kemudian datanya akan dijadikan sebagai sumber data. Dalam melakukan penelitian ini diperlukan alur proses penelitian, agar dalam proses penelitian berjalan sesuai dengan rencana. Gambar 1 menunjukkan diagram alur proses penelitian dalam penelitian ini.



Gambar 1. Diagram Alur Proses Penelitian

Pada Gambar 1. Terdapat beberapa tahap dalam penelitian kali ini diantaranya tahap pengumpulan data, *text preprocessing*, vektorisasi, pembagian data, pemodelan metode klasifikasi, dan terakhir tahap evaluasi.

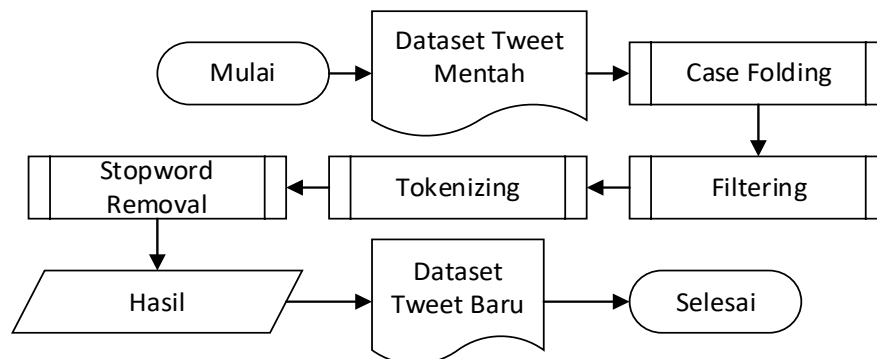


**2.2 Pengumpulan Data**

Tahap pengumpulan data dengan melakukan *crawling* melalui API Twitter dengan menggunakan modul *tweepy* pada bahasa pemrograman *Python*. Data yang *dicrawl* adalah berdasarkan kumpulan *tweet* yang ditujukan kepada akun Twitter IndiHome berdasarkan kata kunci yang menyebutkan *username* @IndiHome. Setelah itu dataset *tweet* yang masih mentah disimpan dalam bentuk CSV (*Comma Separated Values*) yang nantinya akan di proses lebih lanjut.

**2.3 Text Preprocessing**

Setelah proses pengumpulan data dikarenakan dataset tersebut termasuk dalam *unstructured data* (data tidak terstruktur). Sebelum dilakukan analisis lebih lanjut, terlebih dahulu dataset melalui proses *text preprocessing* untuk menghilangkan serta mengatasi *noisy* data agar hasil perhitungan optimal. Adapun alur proses *text preprocessing* antara lain, *case folding*, *filtering*, *tokenizing*, dan *stopword removal*. Setelah data bersih maka dilakukan proses pelabelan secara manual. Berikut ini adalah diagram alur *text preprocessing* yang ditunjukkan pada Gambar 2.



**Gambar 2.** Diagram Alur Proses *Text Preprocessing*

**2.4 Vektorisasi**

Penelitian ini menerapkan algoritma pembobotan kata atau TF-IDF (*Term Frequency-Inverse Document Frequency*) yang berguna untuk melakukan pembobotan sebagai cara untuk memberi nilai pada kata (*term*) berdasarkan suatu dokumen/*tweet*. Adapun rumus algoritma TF-IDF seperti pada persamaan (1) sebagai berikut:

$$W_{if} = tf_{if} + \left( \log \frac{D}{df_i} \right) \quad (1)$$

Keterangan:

- D = jumlah dokumen/*tweet*
- $df_i$  = kemunculan kata dari D
- $W_{if}$  = bobot kata  $t_j$  terhadap dokumen/*tweet*  $d_i$
- $tf_{if}$  = jumlah kemunculan kata  $t_j$  dokumen/*tweet*  $d_i$

**2.5 Support Vector Machine (SVM)**

Pada tahap ini terlebih dahulu dilakukan proses pembagian data menjadi 2 yaitu 80% data latih dan 20% data uji. Data latih digunakan untuk melatih algoritma dan pengembangan pemodelan, sedangkan data uji digunakan sebagai acuan dan untuk mengetahui performa dari algoritma atau model yang sudah dilatih sebelumnya [16]. Selanjutnya barulah proses pemodelan dengan metode SVM.

Metode SVM memungkinkan perhitungan untuk masalah linear dengan menerapkan transformasi matematis untuk ruang belajar menggunakan fungsi kernel [17]. Metode SVM memiliki konsep sentral dalam mengklasifikasikan data, yaitu menentukan *hyperplane* terbaik untuk memberi jarak atau pemisah antara dua kelas yang telah ditentukan [18]. Cara kerja SVM yaitu dengan dengan menemukan *hyperplane* optimal yang memberi jarak atau pemisah antar dua kelas. *hyperplane* yang mempunyai margin maksimum. Jarak antar titik data terdekat dengan *hyperplane* diklaim sebagai margin. Adapun vector pendukung sebagai titik terdekat dengan *hyperplane* [19].

Tahapan dalam metode *Support Vector Machine* adalah sebagai berikut:

- a. Menentukan kata yang sering muncul dari tiap dokumen atau *tweet* yang digunakan.
- b. Menentukan inialisasi awal untuk nilai  $\alpha=0.5, C=1, \lambda=0.5, \text{gamma}=0.5$  dan  $\text{epsilon}=0.001$ .
- c. Menghitung matriks dengan rumus (2):

$$D_{ij} = y_i y_j (K(\vec{x}_i \cdot \vec{x}_j) + \lambda^2) \quad (2)$$



Keterangan:

- $D_{ij}$  = elemen matriks data ke- $ij$
- $y_i$  = kelas atau label data ke- $i$
- $y_j$  = kelas atau label data ke- $j$
- $\lambda$  = turunan batas teoritis
- $K(\vec{x}_i \cdot \vec{x}_j)$  = fungsi kernel

d. Untuk data ke  $n= 1,2,3,\dots n$  gunakan persamaan (3),(4),(5) berikut ini:

$$E_i = \sum_{j=1}^n \alpha_i D_{ij} \quad (3)$$

$$\delta \alpha_i = \min\{\max[\gamma(1 - E_i), -\alpha_i], C - \alpha_i\} \quad (4)$$

$$\alpha_i = \alpha_i + \delta \alpha_i \quad (5)$$

Keterangan:

- $E_i$  = nilai error data ke- $i$
- $\gamma$  = tingkat pembelajaran
- $\max_{(i)} D_{ij}$  = nilai maksimum diagonal matriks hessian

e. Mencari nilai bias (b) dengan menggunakan persamaan (6)

$$b = -\frac{1}{2} [w \cdot x^+ + w \cdot x^-] \quad (6)$$

f. Pengujian pada dokumen yang diuji

g. Perhitungan keputusan

$$h(x) = \begin{cases} +1, & \text{if } w \cdot x + b \geq 0 \\ -1, & \text{if } w \cdot x + b < 0 \end{cases}$$

Jika hasil perhitungan keputusan lebih dari sama dengan 0 maka nilai  $sign h(x)$  adalah +1, maka termasuk kelas positif dan jika hasil perhitungan keputusan kurang dari 0 maka nilai  $sign h(x)$  nilai -1, maka termasuk kelas negatif. Perhitungan keputusan dengan menggunakan persamaan (7).

$$h(x) = w \cdot x + b \quad (7)$$

atau

$$h(x) = \sum_{i=1}^m \alpha_i y_i K(x, x_i) + b \quad (7)$$

## 2.6 Evaluasi

Tahap evaluasi dilakukan untuk mengetahui keakuratan dari pemodelan yang telah diterapkan pada data latih. Kemudian membandingkan hasil dari dua dataset yang berbeda dengan menerapkan *confusion matrix* untuk menghitung *precision*, *recall*, *f1-score*, dan *accuracy*.

## 3. HASIL DAN PEMBAHASAN

Pada bab ini berisi tentang hasil serta pembahasan dari topik penelitian, yang sudah dipaparkan pada bab metodologi penelitian. Bab ini juga merepresentasikan penjelasan berupa narasi, gambar, dan juga table.

### 3.1 Persiapan Data

Penelitian ini mengambil data dari Twitter, dengan cara *crawling data*. Data yang diambil berdasarkan kumpulan *tweet* yang ditujukan kepada akun twitter IndiHome, yakni, dengan menggunakan kata kunci yang menyebut *username* @IndiHome. Data yang berhasil diambil adalah sebanyak 10.000 data *tweet* yang dicrawl saat awal-awal pemberitahuan *work from home* dan *social distancing* yaitu tanggal 16 – 23 Maret 2020. Setelah itu data disimpan dalam bentuk file CSV (*Comma Separated Values*) untuk diolah lebih lanjut. Hasil dari tahap ini adalah data *username*, *timestamp*, dan *tweet* yang dapat dilihat pada Tabel 1.

**Tabel 1.** Data Tweet Indihome

Username	Timestamp	Tweet
xxxxxxxxxxxx	3/20/2020 12:11:00 PM	'@IndiHome Gimana bisa mendukung WFH kalau internet sejak beberapa hari ini sering terputus dan sama sekali tidak bisa digunakan.'



**3.2 Text Preprocessing**

Setelah proses pengumpulan data Twitter selesai, terlebih dahulu dilakukan *text preprocessing*. Hal ini dilakukan dikarenakan dataset yang digunakan termasuk dalam data *unstructured data* (data tidak terstruktur). Adapun beberapa tahap *text preprocessing* yang dilakukan pada penelitian kali ini sebagai berikut:

- Case Folding* adalah tahap pengubahan semua huruf campuran baik itu *uppercase*, atau *lowercase*, menjadi *lowercase* (huruf kecil) semua.
- Filtering* adalah tahap pembersihan data dari tanda baca, simbol atau elemen yang tidak dibutuhkan seperti URL.
- Tokenizing* adalah tahap pemecahan teks menjadi kata-kata, seperti “saya pergi ke sekolah” menjadi “saya”, “pergi”, “ke”, “sekolah”.
- Stopword Removal* adalah tahap penghapusan kata sambung seperti “ke”, “di”, “dan”, “dia”, “kami”, “aku”, “saya”.

Hasil dari *text preprocessing* seperti yang ada pada Tabel 2. dimana setelah melalui *text preprocessing* data yang dilabeli adalah 1400 data (542 *tweet* positif dan 858 *tweet* negatif).

**Tabel 2.** Hasil *Text Preprocessing*

IndiHome	
<b>Input</b>	'@IndiHome Gimana bisa mendukung WFH kalau internet sejak beberapa hari ini sering terputus dan sama sekali tidak bisa digunakan.'
<b>Output</b>	gimana mendukung wfh internet hari terputus digunakan

**3.3 Vektorisasi**

Manfaat dari proses vektorisasi adalah untuk memberi nilai pada kata dalam suatu dokumen atau *tweet* agar dapat diolah dan dicari polanya ketika masuk dalam tahap klasifikasi. Proses vektorisasi dengan menggunakan algoritma perhitungan pembobotan kata atau TF-IDF yaitu dengan menganalisis sampel data *tweet* kemudian melakukan perhitungan berdasarkan rumus (1). Berikut ini merupakan hasil perhitungan TF-IDF terhadap 3 sampel data *tweet* yang dapat dilihat pada Tabel 3.

**Contoh:**

**Input teks:**

D1 = “kembali normal terimakasih atas perhatian”

D2 = “terimakasih indihome atas gangguannya hari selesai selesai katanya hari selesai”

D3 = “kembali normal”

**Output :**

**Tabel 3.** Sampel Data yang di Vektorisasi

Tweet	TFIDF										y
	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10	
D1	0.42	0	0	0	0	0.42	0.42	0.55	0	0.42	1
D2	0.18	0.24	0.48	0.24	0.24	0	0	0	0.72	0.18	-1
D3	0	0	0	0	0	0.71	0.71	0	0	0	1

**3.4 Proses Klasifikasi SVM**

Proses klasifikasi dilakukan pada data latih untuk membuat model yang nantinya model ini akan digunakan untuk data uji. Sebelum melakukan proses klasifikasi terlebih dahulu melakukan pembagian data latih 80% (1120 *tweet*) dan data uji 20% (280 *tweet*).

Pada Tabel 3 terdapat 3 sampel *tweet* dengan variable X sebagai data term inputan dan Y adalah label, dimana -1 sebagai label negatif dan 1 sebagai label positif. Data latih dihitung dengan menggunakan metode SVM. Hasil data latih adalah sebuah pembelajaran di dalam sistem yang akan menjadi acuan bagi sistem. Sedangkan hasil data uji adalah sebuah pengetahuan untuk menentukan suatu *tweet* merupakan kelas positif atau kelas negatif.

Pada tahapan fungsi kernel, setiap data akan dibandingkan dengan data itu sendiri dan data lainnya, maka dari itu apabila data uji berjumlah 3 maka dapat dilihat pada Tabel 4.

**Tabel 4.** Tabel Fungsi Kernel

	D1	D2	D3
D1	K(D1,D1)	K(D1,D2)	K(D1,D3)
D2	K(D2,D1)	K(D2,D2)	K(D2,D3)
D3	K(D3,D1)	K(D3,D2)	K(D3,D3)

Selanjutnya contoh perhitungan pada data D1, dan D1:

$$\begin{aligned}
 K(D1,D1) &= ((0.42 \times 0.42) + (0 \times 0) + (0 \times 0) + (0 \times 0) + (0 \times 0) + (0.42 \times 0.42) + (0.42 \times 0.42) + (0.55 \times 0.55) + (0 \times 0) + (0.42 \times 0.42)) \\
 &= 0.1764 + 0.1764 + 0.1764 + 0.3025 + 0.1764 \\
 &= 1.0081
 \end{aligned}$$



Lakukan cara yang sama hingga mendapatkan matriks 3x3 yang hasilnya dipaparkan dalam Tabel 5 sebagai berikut:

**Tabel 5.** Tabel Hasil Perhitungan Fungsi Kernel

	D1	D2	D3
D1	1.0081	0.1512	0.5964
D2	0.1512	0.9864	0
D3	0.5964	0	0.5041

Berikutnya menghitung matriks dengan memasukkan hasil perhitungan ke dalam rumus (2)

$$D_{ij} = y_i y_j (K(\bar{x}_i \cdot \bar{x}_j) + \lambda^2)$$

Sampel perhitungan untuk data D1, D1:

$$D_{ij} = (1)(1)(1.0081)+0.5^2 = 1.2581$$

Untuk hasil data lainnya terlihat pada Tabel 6 sebagai berikut:

**Tabel 6.** Tabel Matriks

	D1	D2	D3
D1	1.2581	-0.0378	0.8464
D2	-0.0378	0.2466	-0.25
D3	0.8464	0.25	0.7541

Setelah mendapatkan matriks maka kita akan mencari nilai error dengan mensubstitusi matriks tersebut ke dalam persamaan (3), yang hasilnya disajikan dalam Tabel 7.

**Tabel 7.** Hasil Perhitungan Nilai Error

D1	1.03335
D2	-0.0206
D3	0.92525

Kemudian menghitung nilai delta alpha dengan menggunakan persamaan (4), hasilnya seperti pada Tabel 8 berikut ini:

**Tabel 8.** Hasil Perhitungan Delta Alpha

D1	-0.016675
D2	0.5
D3	0.037375

Berikutnya menghitung nilai alpha baru dengan menggunakan persamaan (5), maka didapatkan hasil pada Tabel 9 sebagai berikut:

**Tabel 9.** Nilai Alpha Baru

D1	0.483325
D2	1
D3	0.537375

Langkah selanjutnya menentukan nilai bias dengan persamaan (6), tetapi sebelum mencari nilai bias terlebih dahulu menghitung nilai  $w$ . Dimana  $w_1^+$  merupakan bobot dot product data di kelas positif sedangkan  $w_1^-$  sebagai bobot dot product data di kelas negatif, berikut ini adalah perhitungannya:

$$w \cdot x^+ \text{ (kelas positif)} = (0.483325 \times 1 \times 0.5964) + (1 \times -1 \times 0) + (0.537375 \times 1 \times 0.5041) = 0.5591457675$$

$$w \cdot x^- \text{ (kelas negatif)} = (0.483325 \times 1 \times 0.1512) + (1 \times -1 \times 0.9864) + (0.537375 \times 1 \times 0) = -0.91332126$$

Jika nilai  $w \cdot x^+$  (kelas positif) dan  $w \cdot x^-$  (kelas negatif) sudah diketahui, maka nilai  $b$  dapat dicari dengan rumus persamaan (6):

$$b = -1/2 (w \cdot x^+ + w \cdot x^-) = 0.1770877463$$

Setelah nilai  $\alpha$ ,  $w$  dan  $b$  sudah ditemukan barulah bisa melakukan pengujian pada sampel data uji pada Tabel 10 berikut ini:

D1 = “terimakasih kembali normal”

**Tabel 10.** Tabel Data Uji

TF-IDF	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	Y
D1	0	0	0	0	0	0.57735	0.57735	0	0	0.57735	?



Langkahnya adalah dengan menghitung *dot product* dengan mengalikan data uji dengan semua data latih yang kemudian dimasukkan ke fungsi kernel.

$$K(x_i, x) = (0 \times 0) + (0 \times 0) + (0 \times 0) + (0 \times 0) + (0 \times 0) + (0.57735 \times 0.71) + (0.57735 \times 0.71) + (0 \times 0) + (0 \times 0) + (0.57735 \times 0.71) = 0.819837$$

lakukan hal yang sama antara data uji dengan semua data latih, hingga mendapatkan hasil pada Tabel 11 berikut:

**Tabel 11.** Hasil *Dot Product* Data Uji Dengan Data Latih

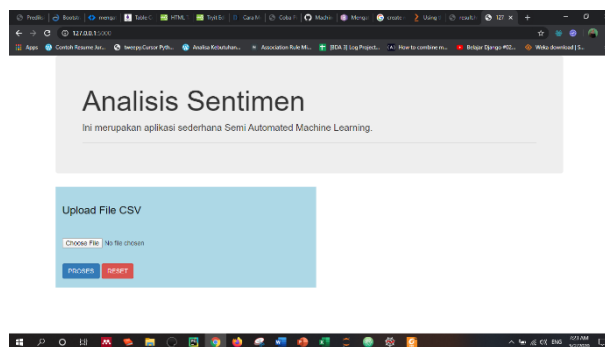
D1	0.819837
D2	0.103923
D3	0.727461

Setelah mendapatkan *dot product* antara data uji ,barulah dilakukan perhitungan fungsi keputusan dengan menggunakan rumus (7).

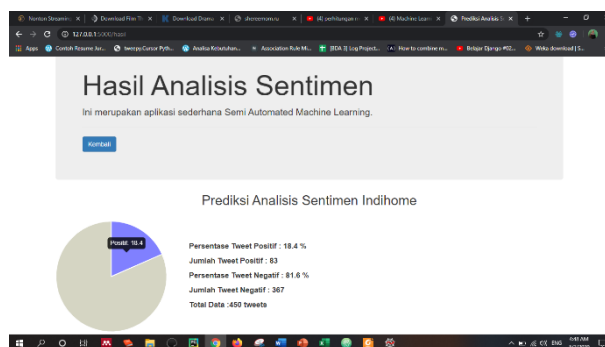
$$h(x) = \text{sign}((0.483325 \times 1 \times 0.819827) - 0.1770877463 + (1 \times -1 \times 0.103923) - 0.1770877463 + (0.537375 \times 1 \times 0.727461) - 0.1770877463) = 0.0073752 = \text{sign}(0.0073752) = +1$$

Berdasarkan hasil perhitungan fungsi keputusan terhadap sampel data uji, maka data uji tersebut tersebut adalah kelas **positif**.

Berikut ini dilakukan uji coba melalui antarmuka dengan menggunakan data baru sebanyak 450 data *tweet* yang sudah dibersihkan terlebih dahulu. Gambar 3 merupakan halaman antarmuka metode SVM untuk klasifikasi analisis sentimen terhadap layanan Indihome dengan cara mengupload file CSV data *tweet* baru, dan ketika klik proses maka akan tampil Gambar 4. Gambar 4 adalah hasil dari analisis sentimen terhadap layanan Indihome berupa pie chart beserta keterangan persentase komentar positif dan komentar negatif, total data keseluruhan, serta banyaknya jumlah komentar positif dan banyaknya jumlah komentar negatif yang ada disampingnya.



**Gambar 3.** Halaman Upload File CSV



**Gambar 4.** Halaman Hasil Analisis Sentimen

Berdasarkan hasil uji dengan menggunakan data baru menghasilkan persentase 18,4% untuk sentimen positif dan 81,6% untuk sentimen negatif. Dengan jumlah sentimen positif sebanyak 83 *tweet* dan sentimen negatif sebanyak 367 *tweet*. Untuk persentase positif dan negatif bisa berubah sesuai dengan opini yang masyarakat publikasikan melalui Twitter kepada Indihome.

### 3.5 Evaluasi

Setelah melakukan pemodelan dengan metode SVM, pemodelan dari hasil data latih kemudian digunakan untuk data uji dengan jumlah data uji sebanyak 280 *tweet* (104 *tweet* positif dan 176 *tweet* negatif). Berikut adalah hasil prediksi ditunjukkan pada Gambar 5.



Sebenarnya=0, Prediksi=1  
 Sebenarnya=0, Prediksi=0  
 Sebenarnya=0, Prediksi=1  
 Sebenarnya=0, Prediksi=1  
 Sebenarnya=0, Prediksi=0  
 Sebenarnya=0, Prediksi=0  
 Sebenarnya=0, Prediksi=0  
 Sebenarnya=0, Prediksi=0  
 Sebenarnya=1, Prediksi=1  
 Sebenarnya=0, Prediksi=1  
 Sebenarnya=0, Prediksi=0  
 Sebenarnya=0, Prediksi=0

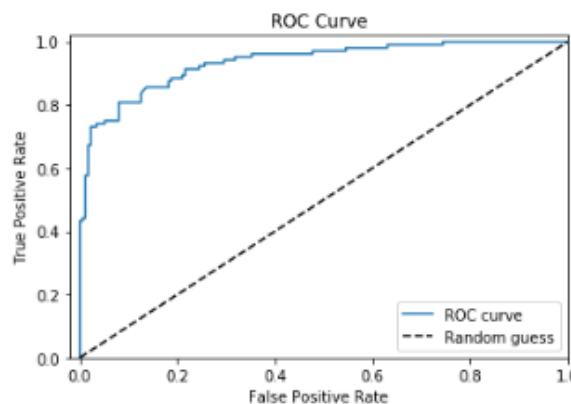
**Gambar 5.** Sampel Hasil Data Uji

Berdasarkan gambar diatas artinya 0 sebagai kelas negatif dan 1 sebagai kelas positif. Adapun hasil evaluasi terhadap data uji digambarkan dalam *confussion matrix* pada Tabel 12.

**Tabel 12.** *Confussion Matrix*

	Prediksi Negatif	Prediksi Positif	Jumlah
Sebenarnya Negatif	167	9	176
Sebenarnya Positif	27	77	104
Jumlah	194	86	280

Pada Tabel 12 terdapat 194 data yang di prediksi negatif , dan 86 data yang di prediksi positif. Namun dari 194 data yang diprediksi negatif memiliki kesalahan prediksi sebanyak 27 data, yang seharusnya berada pada kelas positif, sedangkan dari 86 data di prediksi positif terdapat kesalahan prediksi sebanyak 9 data yang seharusnya berada di kelas negatif. Berdasarkan *confussion matrix* tersebut didapatkan hasil perhitungan nilai akurasi sebesar 87%, ketepatan antara hasil prediksi dengan data sebenarnya (*precision*) sebesar 86%, tingkat keberhasilan sistem dalam memprediksi sebuah data (*recall*) sebesar 95%, tingkat kesalahan semua data yang diprediksi (*error rate*) sebesar 13%, sedangkan untuk nilai perbandingan rata-rata *precision* dan *recall* (*f1-score*) adalah sebesar 90%. Visualisasi Kurva ROC pada Gambar 6 dengan nilai AUC model klasifikasi dengan *Support Vector Machine* sebesar 0.93 atau 93%.



**Gambar 6.** Kurva ROC Metode *Support Vector Machine*

#### 4. KESIMPULAN

Berdasarkan hasil dan pembahasan yang telah diuraikan di atas, maka dapat ditarik kesimpulan mengenai analisis sentimen terhadap layanan Indihome berdasarkan Twitter menggunakan metode *Support Vector Machine* sebagai berikut:

- Dengan menerapkan metode *Support Vector Machine* untuk analisis sentimen terhadap layanan Indihome berdasarkan data *tweet* berdasarkan Tabel 12 hasil evaluasi menggunakan *Confussion Matrix*, didapat akurasi sebesar 87% dengan ketepatan antara hasil prediksi dengan data sebenarnya (*precision*) sebesar 86%, tingkat keberhasilan sistem dalam memprediksi sebuah data (*recall*) sebesar 95%, tingkat kesalahan semua data yang diprediksi (*error rate*) sebesar 13%, sedangkan untuk nilai perbandingan rata-rata *precision* dan *recall* (*f1-score*) adalah sebesar 90%.
- Hasil sentimen layanan Indihome berdasarkan data baru, dengan metode *Support Vector Machine* mendapatkan hasil nilai positif sebesar 18,4 % dan hasil nilai negatif sebesar 81,6%. Berdasarkan hasil tersebut dapat disimpulkan bahwa tingkat kepuasan pengguna layanan Indihome cukup rendah.
- Berdasarkan hasil dan pembahasan menunjukkan bahwa metode SVM dapat diterapkan untuk analisis sentimen data Twitter berbahasa Indonesia.



**REFERENCES**

- [1] Asosiasi Penyelenggara Jasa Internet Indonesia [APJII], “Penetrasi & Profil Perilaku Pengguna Internet Indonesia: Asosiasi Penyelenggara Jasa Internet Indonesia,” 2018.
- [2] H. – W. A. Social, “Social Media Advertising Audiances.” 2019.
- [3] I. M. B. S. Darma, R. S. Perdana, and Indriati, “Penerapan Sentimen Analisis Acara Televisi Pada Twitter Menggunakan Support Vector Machine dan Algoritma Genetika sebagai Metode Seleksi Fitur,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 3, pp. 998–1007, 2018, [Online]. Available: <http://j-ptiik.ub.ac.id>.
- [4] I. P. Windasari, F. N. Uzzi, and K. I. Satoto, “Sentiment Analysis on Twitter Posts: An analysis of Positive or Negative Opinion on GoJek,” *Int. Conf. Inf. Tech, Comput. Electr. Eng.*, pp. 266–269, 2017.
- [5] M. Ahmad, S. Aflab, and I. Ali, “Sentiment Analysis of Tweets using SVM,” vol. 177, no. 5, pp. 25–29, 2017.
- [6] I. D. Sholihati, Irmawati, and D. Glory, “Aplikasi Data Mining Berbasis Web Menggunakan Algoritma Apriori untuk Data Penjualan di Apotek,” *Semin. Nas. Teknol. Informasi, Komun. dan Apl.*, vol. 4, no. 2, pp. 121–126, 2019.
- [7] M. Cindo, D. P. Rini, and Ermatita, “Studi Komparatif Metode Ekstraksi Fitur pada Analisis Sentimen Maskapai Penerbangan Menggunakan Support Vector Machine dan Maximum Entropy,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 3, no. 3, pp. 402–407, 2019.
- [8] U. Rofiqoh, R. S. Perdana, and M. A. Fauzi, “Analisis Sentimen Tingkat Kepuasan Pengguna Penyedia Layanan Telekomunikasi Seluler Indonesia Pada Twitter Dengan Metode Support Vector Machine dan Lexicon Based Features,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 1, no. 12, pp. 1725–1732, 2017.
- [9] A. Triayudi, “Convolutional Neural Network For Text Classification On Twitter,” *J. Softw. Eng. Intellident Syst.*, vol. 4, no. 3, pp. 123–131, 2019.
- [10] I. Santoso, W. Gata, and A. B. Paryanti, “Penggunaan Feature Selection di Algoritma Support Vector Machine untuk Sentimen Analisis Komisi Pemilihan Umum,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 1, no. 1, pp. 364–370, 2019.
- [11] S. Y. Pangestu, Y. Astuti, and L. D. Farida, “Algoritma Support Vector Machine Untuk Klasifikasi Sikap Politik Terhadap Partai Politik Indonesia,” *J. Mantik Penusa*, vol. 3, no. 1, pp. 236–241, 2019.
- [12] A. M. Pravina, I. Cholissodin, and P. P. Adikara, “Analisis Sentimen Tentang Opini Maskapai Penerbangan pada Dokumen Twitter Menggunakan Algoritma Support Vector Machine (SVM),” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 3, pp. 2789–2797, 2019.
- [13] F. Alvianda and P. P. Adikara, “Analisis Sentimen Konten Radikal Di Media Sosial Twitter Menggunakan Metode Support Vector Machine ( SVM ),” *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. 3, no. 1, pp. 241–246, 2019.
- [14] H. S. Utama, D. Rosiyadi, B. S. Prakoso, and D. Ariadarma, “Analisis Sentimen Sistem Ganjil Genap di Tol Bekasi Menggunakan,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 1, no. 10, pp. 2–8, 2019.
- [15] B. W. Sari and F. F. Haranto, “Implementasi Support Vector Machine Untuk Analisis Sentimen Pengguna Twitter Terhadap Pelayanan Telkom Dan Biznet,” *J. Pilar Nusa Mandiri*, vol. 15, no. 2, pp. 171–176, 2019, doi: 10.33480/pilar.v15i2.699.
- [16] A. Rahmansyah, O. Dewi, P. Andini, T. Hastuti, P. Ningrum, and M. E. Suryana, “Membandingkan Pengaruh Feature Selection Terhadap Algoritma Naïve Bayes dan Support Vector Machine,” *Semin. Nas. Apl. Teknol. Inf.*, pp. 1–7, 2018.
- [17] Y. Al-Amrani, M. Lazaar, and K. E. El Kadiri, “Sentiment Analysis Using Hybrid Method Of Support Vector Machine And Decision Tree,” *J. Theor. adn Appl. Inf. Technol.*, vol. 96, no. 7, pp. 1886–1895, 2018.
- [18] K. A. *Support Vector Machines Succinctly*. 2017.
- [19] S. Rani and J. Singh, “Sentiment Analysis Of Tweets Using Support Vector Machine,” *Int. J. Comput. Sci. Mob. Appl.*, vol. 5, no. 10, pp. 83–91, 2017.