

ISSN 2087-3336 (Print) | 2721-4729 (Online)

**TEKNOSAINS: Jurnal Sains, Teknologi dan Informatika**

Volume 10, Nomor 2, Juli 2023, hlm. 154-164

<http://jurnal.stmcileungsi.ac.id/index.php/tekno>

DOI: 10.37373

## **Analisis sentimen twitter terhadap steam menggunakan algoritma logistic regression dan support vector machine**

### ***Steam sentiment analysis using logistic regression algorithm and support vector machine***

**Edo Ridho Lidinillah, Tatang Rohana, Ayu Ratna Juwita\***

\* Teknik Informatika, Fakultas Ilmu Komputer, Universitas Buana Perjuangan Karawang, Indonesia, 41361

\*Jl. Hs. Ronggo Waluyo, Teluk Jambe Karawang, Jawa Barat, Indonesia

\*Koresponden Email: ayurj@ubpkarawang.ac.id

Artikel dikirim: 09/01/2023

Artikel direvisi: 19/01/2023

Artikel diterima: 20/01/2023

#### **ABSTRAK**

*Games online* yakni hal yang sudah menempel di masyarakat saat ini. Dalam beberapa tahun terakhir, ekspansi internet dan perangkat yang cepat telah mempercepat munculnya *games online*. Motivasi utama pemain untuk terus bermain *games online* adalah ketersediaan kemampuan *games online* multi-pengguna yang dapat diakses di mana saja[1][1][1]. Seiring perkembangan teknologi saat ini banyak *platform* penjualan *games online* seperti *Steam*, *Epic Games Store*, *Origin* dan sebagainya, opini masyarakat terkadang sulit dikomunikasikan secara eksklusif kepada pengelola khususnya developer *Steam* sebagai *platform* penjualan *games*, Ini mendorong individu untuk mengirimkan komentar, penilaian, dan konten serupa melalui media sosial, salah satu jejaring sosial yang paling terkenal saat ini ialah *Twitter*. Deretan *tweet* atau opini asal pengguna *Twitter* terkait *platform Steam* yang mana dapat dipergunakan menjadi analisis sentimen. Dalam penelitian ini data yang terkait menggunakan *platform Steam* dikumpulkan sampai 4363 data ulasan positif dan negatif terhadap *platform steam* dalam *twitter*, memakai *TextBlob Library* yang menyediakan *API* sederhana buat menyelam ke pada tugas *Natural Language Processing (NLP)* kemudian diproses menggunakan metode penambangan data (*data mining*), termasuk penambangan teks, *cleaning*, *case folding*, *tokenization*, *filtering stopword*, serta *wordcloud*. Untuk menghitung menggunakan metode *Confusion Matrix* dalam 2 algoritma yang berbeda buat perbandingan, algoritma yang digunakan adalah *Logistic Regression* dan *Support Vector Machine*. Dari percobaan perhitungan 2 metode itu diketahui bahwa algoritma *Super Vector Machine* mendapatkan nilai yang optimal dengan *accuracy* 0.81, *precision* 0.85 serta *recall* 0.77.

Kata Kunci: *twitter*; *data mining*; opini, analisis sentimen; algoritma.

#### **ABSTRACT**

*Online games are things that have stuck in today's society. In recent years, the rapid expansion of the Internet and devices has accelerated the emergence of online games. The main motivation for players to continue playing online games is the availability of multi-user online game capabilities that can be accessed anywhere. As technology develops, there are many online game sales platforms such as Steam, Epic Games Store, Origin and so on, public opinion is sometimes difficult to communicate exclusively to managers, especially Steam developers as a game sales platform. This encourages individuals to submit comments, ratings and similar content. through social media, one of the most popular social networks today is Twitter. A series of tweets or opinions from Twitter users regarding the Steam platform which can be used as sentiment analysis. In this study data related to the Steam platform was collected up to 4363 positive and negative review data on the Steam platform in Twitter, using the TextBlob Library which provides a simple API for diving into Natural Language Processing (NLP) tasks and then*



TEKNOSAINS: Jurnal Sains, Teknologi & Informatika is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License. ISSN 2087-3336 (Print) | 2721-4729 (Online)

processed using data mining methods (data mining), including text mining, cleaning, case folding, tokenization, stopword filtering, and wordcloud. Then it is calculated using the Confusion Matrix method in 2 different algorithms for comparison, the algorithms used are Logistic Regression and Support Vector Machine. From the experimental calculation of the 2 methods it is known that the Super Vector Machine algorithm obtains optimal values with an accuracy of 0.81, a precision of 0.85 and a recall of 0.77.

**Keywords:** twitter; data mining; opinion; sentiment analysis; algorithms.

## 1. PENDAHULUAN

*Game online* yang sudah di kenal masyarakat umum ini dalam beberapa tahun terakhir telah mengalami peningkatan karena ekspansi internet dan perangkat yang cepat. Utamanya pemain termovitasi karena ketersediaan multi-pengguna dalam *game online* yang dapat di akses dimana saja [1]. Banyak platform penjualan *game online* seperti *Steam*, *Epic Game Store* dan yang lainnya seiring perkembangan teknologi [2]. Telah dilakukan penelitian bahwa seseorang yang mengaku sebagai *gamer* sejati akan cenderung bersikap ekstrem untuk membela komunitasnya. Hal itu menjadi pemicu adanya rasisme dan seksisme [3]. Hal ini mengakibatkan munculnya budaya komunikasi yang ditandai dengan penggunaan frase linguistik dengan konotasi sosial berupa makian, umpatan, komentar cabul, terminologi kasar, dan frase jahat. Kebiasaan komunikasi buruk berikutnya dapat muncul sebagai indikasi gangguan kebiasaan atau penyakit mental pada pengguna. Bentuk komunikasi ini sering disebut dan dianggap bahasa beracun karena sangat tidak sopan dan digunakan untuk alasan negatif seperti mempermalukan, menghina, mengganggu, dan menyakiti seseorang secara emosional [2].

*Twitter* ialah platform komunikasi yang basa digunakan masyarakat luas untuk berekspresi dan lebih diminati oleh masyarakat dunia. Pada *twitter* terdapat istilah *tweet* yaitu pengguna *twitter* dapat memberikan kabar terbaru, berekspresi, beraspirasi, dan beropini yang ditulis oleh pengguna *twitter* lainnya terutama topik atau hal yang sedang menjadi perbincangan utama [3]. Banyak orang suka menggunakan media sosial karena dapat digunakan untuk topik pribadi dan non-pribadi untuk memungkinkan beberapa orang berbagi pemikiran mereka tentang kejadian saat ini. Pandangan ini dapat diubah menjadi informasi melalui teknik yang dikenal sebagai analisis sentimen [4]. Analisis sentimen ialah proses mengekstraksi, mengolah dan memahami data berupa teks yang tidak terstruktur secara otomatis guna mengambil informasi sentimen yang terdapat pada sebuah kalimat pendapat atau opini [5]. Opini adalah ekspresi pendapat yang bertentangan terhadap suatu topik. Opini dapat disampaikan melalui komentar media sosial, artikel *website* berita, dan blog pribadi [6]. Penambangan data atau proses menggunakan alat dan teknik tertentu untuk mencari pola atau wawasan berharga dalam data dalam jumlah besar. Penambangan data (*data mining*) dapat dilakukan dengan menggunakan sejumlah pendekatan, metode, dan algoritme yang berbeda [7].

Pada penelitian terdahulu menggunakan lima *cross validation*, arsitektur percobaan dijalankan. Algoritma *KNN*, *SVM*, dan *CNN* dikonfigurasi dengan beberapa parameter. Menurut percobaan *CNN*, ia memiliki efisiensi tertinggi, dengan skor F1 0,942, akurasi 0,942, presisi (*precision*) 0,943, dan *recall* 0,942 [8]. Penelitian selanjutnya menganalisis dan mengembangkan model pandangan masyarakat terhadap kebijakan sekolah *online* yang diungkapkan melalui komentar media sosial *Twitter*. Penelitian ini dilakukan dalam bentuk klasifikasi kelas tunggal menggunakan model yang dibuat menggunakan teknik *Logistic Regression* dan *Supervised Machine Learning*. Dengan nilai akurasi 87%, nilai *precision* 0,88, nilai *recall* 0,87, dan nilai *f1-score* 0,88, *Logistic Regression* merupakan model yang optimal untuk analisis kategorisasi sentimen [9]. Penelitian ini memanfaatkan data sekunder berupa daftar lamaran beasiswa dan pemenang sebagai dataset dengan enam parameter penentu, meliputi semester, IPK, prestasi, kegiatan ekstrakurikuler, pendapatan orang tua, konsumsi tenaga, dan jumlah tanggungan. RapidMiner digunakan untuk pengujian validasi silang (*cross validation*) 10 kali lipat dan evaluasi kinerja model. Temuan menunjukkan bahwa saat membandingkan model algoritme C4.5 dengan model algoritma *Naive Bayes*, model C4.5 memiliki akurasi 96,40 poin persentase lebih besar, namun kedua algoritma tersebut membutuhkan waktu pemrosesan yang sama atau sebanding yaitu 0 detik. Hal ini karena jumlah kumpulan data tidak cukup untuk mendeteksi perbedaan waktu pemrosesan antara kedua model algoritma [10]. Studi berikutnya melakukan analisis sentimen terhadap “*CoronaVirus Disease- 2019 (covid19)*” yang saat ini menjadi topik populer di *Twitter*. Penularan virus



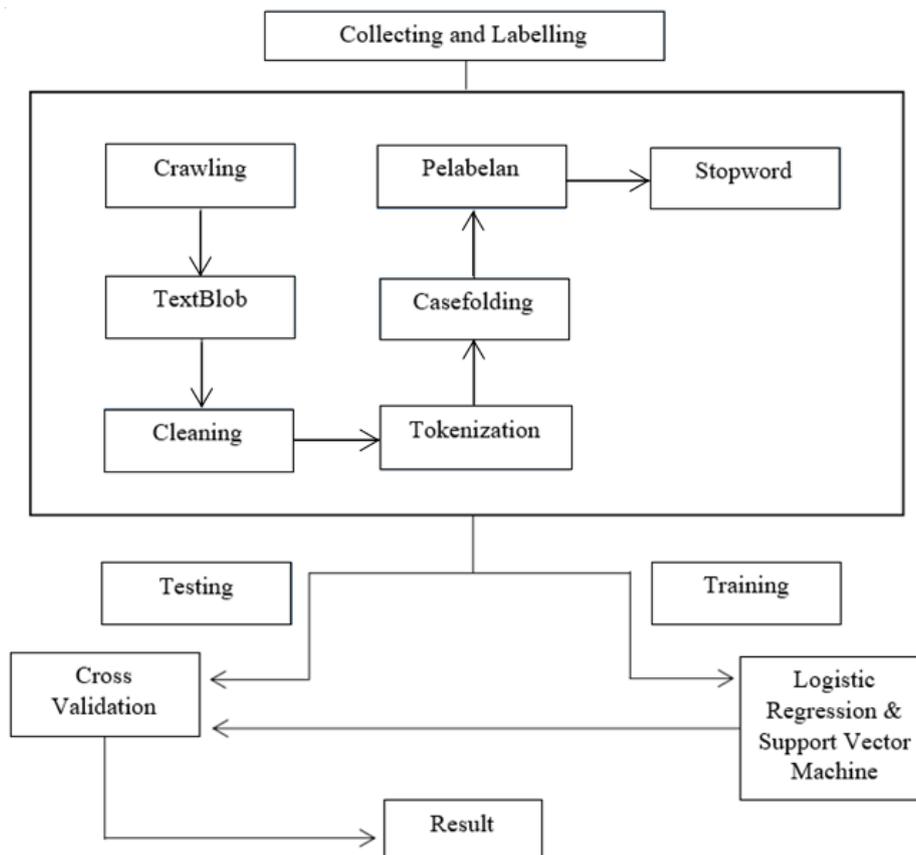
ini pun langsung diperdebatkan oleh banyak pengguna *Twitter*, dan virus corona saat itu tengah menjadi perhatian dunia. Banyak perspektif berkembang tentang virus corona akibat jumlah korban yang sangat banyak dan penyebaran virus yang cepat. Polaritas pandangan ini kemudian ditentukan oleh analisis sentimen. *Logistic Regression* dan *Support Vector Machine* (SVM) digunakan, dengan SVM mencapai akurasi 91,15% pada data uji dan *Logistic Regression* mencapai akurasi 87,68% pada data uji [11].

Beberapa penelitian yang telah disampaikan diatas menunjukkan bahwa penelitian perbandingan algoritma klasifikasi telah banyak dilakukan. Penelitian ini kemudian mengklasifikasikan pandangan ke dalam statistik opini positif dan negatif. Studi ini menggunakan dua algoritma berbeda, *Logistic Regression* dan *Support Vector Machine*, untuk mengkategorikan data dan mengekstrak ide atau perasaan individu dari *tweet* mereka. Tujuan penelitian ini adalah untuk mengevaluasi kinerja dari dua algoritma *classifier* yaitu *Logistic Regression* dan *Support Vector Machine* dalam hal akurasi, *precision*, dan *recall* untuk menentukan *classifier* mana yang memberikan hasil terbaik dalam mencari tahu tingkat ulasan positif dan negatif terhadap platform steam dalam *twitter*. Agar penelitian ini dapat menjadi landasan bagi penelitian selanjutnya dengan menggunakan metodologi yang paling sesuai.

## 2. METODE

### 2.1. Analisis sentimen

Penelitian tentang analisis sentimen *tweet* pada platform steam di *Twitter*. Analisis sentimen atau dikenal juga dengan *opinion mining* adalah sub bidang dari *data mining* yang berupaya mengevaluasi, memahami, dan mengolah data tekstual berupa pandangan terhadap hal-hal seperti barang, jasa, organisasi, individu, dan tema tertentu [12]. Informasi yang dijelaskan untuk menunjukkan bahwa studi sebelumnya telah dilakukan pada analisis sentimen *Twitter*. Masalah klasifikasi harus diprioritaskan dalam analisis sentimen *Twitter*. Klasifikasi adalah proses mencari pemahaman komprehensif tentang suatu fungsi atau himpunan dengan maksud membangun model untuk menentukan kategori item yang tidak diketahui [7].



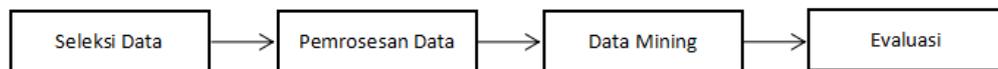
Gambar 1. Diagram klasifikasi analisis sentimen

Dua metode klasifikasi, termasuk *Logistic Regression* dan *Support Vector Machine*, telah digunakan untuk menentukan hasil yang optimal. **Gambar 1** mengilustrasikan proses, dimulai dengan pengumpulan data dan pelabelan dan diakhiri dengan prediksi label, untuk mendapatkan hasil pengujian yang sebaik mungkin. Ketika melakukan analisis sentimen, ada tahapan yang perlu dilakukan buat mendapatkan hasil tes yang bagus. Adapun **Gambar 2** menunjukkan langkah-langkah yang dipergunakan diawali dari Collecting and Labelling hingga mendapatkan hasil.

Data *training* adalah kumpulan data dengan karakteristik klasifikasi/label yang diterapkan pada karakteristik kumpulan data oleh mesin untuk tujuan membangun pola/model data. Data uji bukan hanya sekumpulan angka; itu juga memiliki label dan kategori yang dapat digunakan untuk menilai kebenaran dan tingkat akurasi pola/model klasifikasi, tidak diuji kebenaran pola/modelnya [13].

### 2.2. Alir penelitian

Ketika menganalisa sentimen dan mengetahui tingkat akurasinya, ada beberapa tahapan untuk mendapatkan nilai hasil yang terbaik. Tahapan yang dilakukan sesuai gambar 2.



**Gambar 2.** Alir metodologi penelitian.

Pada tahap seleksi data dilakukan proses pencarian data memakai *TextBlob Library* yang menyediakan *API* sederhana buat menyelam ke pada tugas *Natural Language Processing (NLP)*, kemudian masuk ke tahap pemrosesan data dengan tujuan menghilangkan noise atau kata tidak sesuai, selanjutnya dilakukan proses *data mining* dimana pengolahan data dengan menerapkan algoritma, pada tahap akhir data akan di uji dengan tujuan mengevaluasi model klasifikasi yang telah digunakan.

### 2.3. Pengumpulan dan pelabelan data

Pengumpulan data adalah langkah pertama dalam proses analisis sentimen. Menggunakan *TextBlob*, permintaan pencarian di *Twitter* untuk "*Steam*" menghasilkan sebanyak 4363 item.

**Tabel 1.** Dataset

No	Text	Username	Label	Output
0	<i>Surprised that Soul Hackers 2 isn't on switch. Thankfully I can grab on steam and play in the deck cuz that seems another perfect game on the go</i>	MNOLuffy	1	Positif
1	<i>Just Added: £12.85 SnowRunner - Year 2 Pass PC Download #PC DIGITAL #FocusHomeInteractive: Expand your SnowRunner experience with the Year 2 Pass! This gives you access to four phases of content, including new vehicles,...</i> <a href="https://t.co/U3g0IL6kWE">https://t.co/U3g0IL6kWE</a> -DIGITAL #Steam #steamkeys #games <a href="https://t.co/Yu5wXwOrpX">https://t.co/Yu5wXwOrpX</a>	shoptonet	1	Positif
2	<i>After more testing, I think I'm gonna give Madden until Wednesday and if there isn't another update then we will start the Franchise then. Tonight we will be trying out a new game that is in Early Access I found scrolling through queues on Steam called Neophyte. Stream @ 10pm EST</i>	GoncherGaming	1	Positif
...	...	...	...	...
436	<i>@Optimus4587 Everyone in Smash are actually alright to be in Smash at all These are literally the only ones that don't really have a good reason at all - Corrin was only added in Smash 4 to promote Fates, which ended up being a mediocre Fire Emblem game - ARMS lost steam</i> <a href="https://t.co/DQILRnYFVQ">https://t.co/DQILRnYFVQ</a>	PSI_4ce	-1	Negatif
436	<i>Someday I'd like to have a game published on Steam</i>	adamc0dez	-1	Negatif

436	Game	🚀	On itch: <a href="https://t.co/6dywhy3Cs2">https://t.co/6dywhy3Cs2</a>	🚀	On Steam: <a href="https://t.co/MklXpSo1hO">https://t.co/MklXpSo1hO</a>	AppasPapas	-1	Negatif
2								

Have fun!

👍👍👍👍👍👍👍👍 Daniel

**Tabel 1** enunjukkan hasil dari seleksi data yang kemudian akan masuk pada tahap pelabelan data, langkah selanjutnya adalah memberi label pada informasi yang memerlukan penempatan data ke dalam kelas berbeda dikarenakan klasifikasi terdiri dari fakta-fakta baik dan buruk, yang nantinya akan digunakan. Penandaan ini dimaksudkan untuk membedakan antara data positif dan kumpulan data negatif. Kumpulan data positif berisi kalimat yang baik, sedangkan kumpulan data negatif berisi kalimat yang buruk. Contoh data yang sudah diberi label ditunjukkan pada **Tabel 2**.

**Tabel 2.** Contoh label *dataset*

Text	Sentiment
Surprised that Soul Hackers 2 isn't on switch. Thankfully I can grab on steam and play in the deck cuz that seems another perfect game on the go	POSITIVE
Hello. I've tried to play this on steam deck but can't even start the game. Are you aware of this problem?	NEGATIVE

**Tabel 2** menunjukkan contoh ulasan sentimen yang telah diberi label positif (baik) dan negatif (buruk), kemudian akan masuk pada tahap pra-pemrosesan.

#### 2.4. Logistic Regression

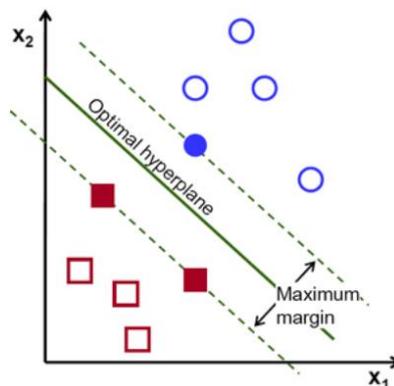
Probabilitas merupakan suatu peristiwa bakal terjadi dari banyaknya peristiwa dalam suatu percobaan. Probabilitas selalu berkisar antara 0 dan 1. *Odds* dijelaskan sebagai probabilitas bahwa peristiwa akan terjadi dibagi dengan probabilitas bahwa peristiwa tidak akan terjadi [14].

$$O(Y) = \frac{P(Y)}{1-P(Y)} \quad (1)$$

*Logistic regression* bekerja untuk mengetahui apakah ada hubungan antara variabel dependen (Y) yang terkait dengan kejadian positif (YA) atau negatif (TIDAK) dari satu peristiwa (tipe dikotomis) dengan satu atau lebih banyak variabel independen yang bertipe *categorical* atau *continuos* [14].

#### 2.5. Support Vector Machine

Tujuan dari algoritma *Support Vector Machine* adalah untuk mencari *hyperplane* yang secara *linear* memisahkan kumpulan data menjadi dua kelas. *Hyperplan* adalah istilah yang dibuat umum untuk semua dimensi. Algoritma *Support Vector Machine* mencoba mencari *hyperplane* yang paling optimal [15].



**Gambar 3.** Margin *hyperplane*

**Gambar 3** menunjukkan *hyperplane* (batas keputusan) pemisah terbaik antara kedua kelas yaitu positif (YA) atau negatif (TIDAK) dengan mengukur margin *hyperplane* dan mencari titik maksimalnya. Margin merupakan jarak antara *hyperplane* tersebut dengan data terdekat dari setiap kelas. Data yang paling dekat ini disebut *support vector*. Garis tebal pada **Gambar 3** menunjukkan *hyperplane* terbaik yaitu yang terletak tepat di tengah kedua kelas, sedangkan data lingkaran dan persegi yang dilalui oleh garis tepi (garis putus-putus) merupakan *support vector*. Menemukan lokasi *hyperplane* ini adalah inti dari proses pelatihan *Support Vector Machine*.

### 3. HASIL DAN PEMBAHASAN

#### 3.1 Pra-pemrosesan

Label data yang telah diberikan sebelumnya pada data *testing* dan data *training* dengan sentimen positif dan negatif, langkah selanjutnya adalah pra-pemrosesan. Pada langkah ini, data diubah menjadi informasi yang dapat dianalisis. Beberapa fase terdiri dari *preprocessing*, termasuk *cleaning*, *case folding*, *tokenization*, *filtering stopword* dan hasil *preprocessing* akan dikumpulkan untuk memperoleh frekuensi istilah yang paling banyak muncul kemudian kata-kata tersebut akan divisualisasikan menjadi *wordcloud* positif dan negatif.

##### 3.1.1. *Cleaning*

Langkah prosedur ini menghilangkan karakteristik dari data input yang tidak diperlukan, seperti simbol dan tanda baca. Bekerja untuk meminimalkan *noise* dalam dataset. Contoh karakter yang dihilangkan antara lain titik (.), koma (,), dan tanda baca lainnya. Berikut ini adalah contoh frasa yang terkait dengan pembersihan data:

**Tabel 3.** Contoh kalimat pembersihan data

<i>Input</i>	<i>Output</i>
<i>Surprised that Soul Hackers 2 isn't on switch. Thankfully I can grab on steam and play in the deck cuz that seems another perfect game on the go</i>	<i>Surprised that Soul Hackers 2 isn t on switch Thankfully I can grab on steam and play in the deck cuz that seems another perfect game on the go</i>

**Tabel 3** menunjukkan hasil dari proses pembersihan data dimana telah dihilangkan antara lain titik (.), koma (,), dan tanda baca lainnya.

##### 3.1.2. *Case folding*

Seringkali ada banyak bentuk huruf yang digunakan saat men-tweet, ini adalah tata cara untuk memastikan keseragaman huruf tentang penggunaan huruf kapital. Berikut adalah contoh *case folding*:

**Tabel 4.** Contoh kalimat *case folding*.

<i>Input</i>	<i>Output</i>
<i>Surprised that Soul Hackers 2 isn't on switch. Thankfully I can grab on steam and play in the deck cuz that seems another perfect game on the go</i>	<i>surprised that soul hackers 2 isn t on switch thankfully i can grab on steam and play in the deck cuz that seems another perfect game on the go</i>

**Tabel 4** menunjukkan hasil dari proses *case folding* dimana telah dilakukan keseragaman huruf tentang penggunaan huruf kapital.

##### 3.1.3. *Tokenization*

Tokenisasi merupakan proses menyortir dan membagi frase menjadi kata-kata individu, yang dikenal sebagai tokenisasi. Ada beberapa model tokenisasi yang tersedia, seperti *unigram*, *bigram*, *trigram*, dan *ngram*. Kalimat *input* di bawah ini adalah contoh tokenisasi “*surprised that soul hackers 2 isn t on switch thankfully i can grab on steam and play in the deck cuz that seems another perfect game on the go*”, *output* “*surprised, that, soul, hackers, 2, isn, t, on, switch, thankfully, I, can, grab, on, steam, and, play, in, the, deck, cuz, that, seems, another, perfect, game, on, the, go*”.

##### 3.1.4. *Filtering*



$$Precision = \frac{TP}{TP+FP} \tag{3}$$

3. Recall

Recall adalah bagian dari dokumen teks yang relevan di bawah kendali relatif terhadap jumlah total dokumen teks yang relevan dalam koleksi. Rumus untuk recall dapat dinyatakan sebagai.

$$Recall = \frac{TP}{TP+FN} \tag{4}$$

Kriteria kualitas model berdasarkan kategorisasi data uji masing-masing metode dirinci dalam tabel 6.

Tabel 6. Confusion matrix data tes algoritma logistic regression

Class	Positive	Negative	Support
Positive	340	82	422
Negative	83	368	451
accuracy			873

Dari confusion matrix di Tabel 6 menjelaskan mengapa model tersebut dengan tepat memberi label 368 observasi sebagai negatif dan 340 observasi sebagai positif. Selain itu, model salah memprediksi 82 contoh data negatif yang seharusnya positif (false negative) dan 83 contoh data positif yang seharusnya negatif (false positive).

Tabel 7. Laporan klasifikasi algoritma logistic regression

	precision	recall	f1-score	support
0	0.80	0.81	0.80	422
1	0.82	0.82	0.82	451
Accuracy			0.81	873
macro avg	0.81	0.81	0.81	873
weighted avg	0.81	0.81	0.81	873

Tabel 7 laporan menunjukkan tingkat akurasi precision, recall, f1-score, dan support pada data sentimen positif (1) dan negatif (0). Hasil akurasi yang didapatkan pada data sentimen positif yaitu 0.82 (precision), 0.82 (recall), 0.82 (f1-score), 451 (support). Sedangkan hasil akurasi yang didapatkan pada data sentimen negatif yaitu 0.80 (precision), 0.81 (recall), 0.80 (f1-score), 422 (support).

Tabel 8. Confusion matrix data tes algoritma support vector machine

Class	Positive	Negative	Support
Positive	361	61	422
Negative	102	349	451
accuracy			873

Dari confusion matrix di Tabel 8 Dapat dijelaskan mengapa model memberi label 349 observasi sebagai negatif dan 361 observasi sebagai positif. Selain itu, model salah memprediksi 82 contoh data negatif yang seharusnya positif (false negative) dan 102 contoh data positif yang seharusnya negatif (false positive).

Tabel 9. Laporan klasifikasi algoritma support vector machine

	precision	recall	f1-score	support
0	0.78	0.86	0.82	422
1	0.85	0.77	0.81	451
accuracy			0.81	873
macro avg	0.82	0.81	0.81	873
weighted avg	0.82	0.81	0.81	873

**Tabel 9** laporan menunjukkan tingkat akurasi *precision*, *recall*, *f1-score*, dan *support* pada data sentimen positif (1) dan negatif (0). Hasil akurasi yang didapatkan pada data sentimen positif yaitu 0.85 (*precision*), 0.77 (*recall*), 0.81 (*f1-score*), 451 (*support*). Sedangkan hasil akurasi yang didapatkan pada data sentimen negatif yaitu 0.78 (*precision*), 0.86 (*recall*), 0.82 (*f1-score*), 422 (*support*).

### 3.3 Evaluasi

Pada tahap ini menggunakan *Confusion Matrix* untuk mengetahui performa model klasifikasi pada jumlah hasil prediksi data label yang dijadikan sebagai data *testing*. Prediksi kelas positif yang benar dinamakan (*TP*) *True Positive*, sementara yang salah dinamakan (*FP*) *False Positive*. Untuk kelas negatif yang diprediksi benar dinamakan (*TN*) *True Negative*, sedangkan yang salah dinamakan (*FN*) *False Negative*. Hasil nya dapat dilihat pada tabel 12 berikut.

**Tabel 10.** Nilai *accuracy*, *precision*, dan *recall* menggunakan formula

Metode	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>
<i>Logistic Regression</i>	0.81	0.80	0.80
<i>Super Vector Machine</i>	0.81	0.85	0.77

**Tabel 10** menunjukan nilai *Accuracy*, *Precision*, dan *Recall* yang di dapat dari percobaan perhitungan manual memakai metode *confusion matrix*.

#### 3.3.1 *Logistic regression*

$$\text{Accuracy} = \frac{340+368}{(340+83+368+82)} \times 100\% = \frac{708}{873} \times \frac{100}{100} = 0.81 \quad (5)$$

$$\text{Precision} = \frac{340}{(340+83)} \times 100\% = \frac{340}{423} \times \frac{100}{100} = 0.80 \quad (6)$$

$$\text{Recall} = \frac{340}{(340+82)} \times 100\% = \frac{340}{422} \times \frac{100}{100} = 0.80 \quad (7)$$

#### 3.3.2 *Support Vector Machine*

$$\text{Accuracy} = \frac{361+349}{(361+61+349+102)} \times 100\% = \frac{710}{873} \times \frac{100}{100} = 0.81 \quad (8)$$

$$\text{Precision} = \frac{361}{(361+61)} \times 100\% = \frac{361}{422} \times \frac{100}{100} = 0.85 \quad (9)$$

$$\text{Recall} = \frac{361}{(361+102)} \times 100\% = \frac{361}{463} \times \frac{100}{100} = 0.77 \quad (10)$$

Hasil ini menunjukkan *accuracy* dari *Logistic Regression* dan *Support Vector Machine* sebesar 0.81, dan 0.81. Hasil untuk *precision* dari *Logistic Regression* dan *Support Vector Machine* sebesar 0.80, dan 0.85. Sementara hasil untuk *Recall* dari *Logistic Regression* dan *Support Vector Machine* adalah 0.80, dan 0.77. Jadi dapat dikatakan bahwa pengklasifikasi *Support Vector Machine* adalah pengklasifikasi yang unggul untuk digunakan pada kumpulan data media sosial karena akurasi prediktifnya yang lebih tinggi.

Kemudian dilakukan proses pengujian tahapan ini memakai teknik *5 K-FOLD Cross Validation*. Teknik yang bekerja dengan cara memisahkan K data menjadi data uji dan data latih. Kemudian membuat prediksi menggunakan algoritma regresi logistik dan mesin vektor pendukung. Nilai akurasi dan presisi didapatkan dari setiap metode deteksi yang digunakan. Tahapan *K-fold crossing* adalah:

Pertama bagi data dengan faktor k. Kemudian pada percobaan pertama, jadikan bagian pertama sebagai data uji dan bagian kedua sebagai data latih. Pada percobaan kedua, alihkan sebagian ke data uji dan sebagian ke data pelatihan. Tes ketiga mengubah bagian ketiga menjadi data uji dan sisanya menjadi data pelatihan. Sepertiga dari hasil ini, *confusion matrix* digunakan untuk menyimpan perkiraan kinerja model. Kemudian tentukan rata-rata untuk setiap percobaan. Keempat, eksperimen dilakukan

dengan menggunakan model algoritma terpilih yang dapat dijadikan referensi. Berikut tabel hasil *k-fold* 5 yang dapat dilihat pada **Tabel 11**.

**Tabel 11.** Hasil persilangan k-fold 5

K-Fold	1.	2.	3.	4.	5.
<i>LogisticRegression</i>	0.79	0.82	0.81	0.81	0.79
<i>SVM</i>	0.61	0.79	0.78	0.81	0.79

**Tabel 11** memperlihatkan hasil dari metode persilangan *K-Fold Cross Validation* yang digunakan untuk mengevaluasi kinerja model. **Tabel 11** hasil dari bagian *Logistic Regression* jika dirata-rata kinerja algoritma ini lebih baik dibandingkan *SVM*.

#### 4. SIMPULAN

Penelitian yang dimaksud melibatkan upaya menentukan *classifier* mana yang memberikan hasil terbaik dalam mencari tahu tingkat ulasan positif dan negatif terhadap platform *steam* dalam *twitter*. Data yang terkait menggunakan platform *Steam* dikumpulkan sampai 4363 data ulasan positif dan negatif terhadap platform *steam* dalam *twitter* memakai *TextBlob Library* yang menyediakan *API* sederhana buat menyelam ke pada tugas *Natural Language Processing (NLP)* lalu diproses menggunakan metode penambangan data (*data mining*), termasuk penambangan teks, *cleaning*, *case folding*, *tokenization*, *filtering stopword*, serta *wordcloud*. lalu dihitung dalam 2 algoritma yang berbeda untuk dijadikan perbandingan, algoritma yg digunakan adalah *Logistic Regression* dan *Support Vector Machine*. Dari percobaan perhitungan menggunakan metode *Confusion Matrix* dapat disimpulkan bahwa algoritma *Support Vector Machine* mendapatkan tingkat akurasi serta presisi yang lebih baik dibanding dari algoritma *Logistic Regression*. oleh karena itu metode *Support Vector Machine* adalah pengklasifikasi terbaik untuk membuat prediksi platform *Steam* yang lebih akurat dan *precision*.

#### REFERENSI

- [1] H. Simorangkir and K. M. Lhaksana, "Analisis Sentimen pada Twitter untuk Games Online Mobile Legends dan Arena of Valor dengan Metode Naïve Bayes Classifier," *e-proceeding of Englineering*, vol. 5, no. 3, pp. 8131–8140, 2018, [Online]. Available: [https://openlibrary.telkomuniversity.ac.id/pustaka/files/144621/jurnal\\_eproc/analisis-sentimen-pada-twitter-untuk-games-online-mobile-legends-dan-arena-of-valor-dengan-metode-na-ve-bayes-classifier.pdf](https://openlibrary.telkomuniversity.ac.id/pustaka/files/144621/jurnal_eproc/analisis-sentimen-pada-twitter-untuk-games-online-mobile-legends-dan-arena-of-valor-dengan-metode-na-ve-bayes-classifier.pdf)
- [2] Shintia Dwi Alike, A. P. Dewi, I. R. Anggara, R. Hayyu, Shabrany, and S. Y. Madhasatya, "Urgensi Penggunaan Tata Bahasa yang Baik dalam Berkomentar di Aplikasi Media Sosial Tiktok Terhadap Kesehatan Mental dan Pembentukan Karakter pada Siswa SMP dan SMA," *ikanJurnal Paedagogy J. Penelit. dan Pengemb. Pendid.*, vol. 7, no. 4, pp. 281–288, 2020, [Online]. Available: <https://e-journal.undikma.ac.id/index.php/pedagogy>
- [3] D. Darwis, E. S. Pratiwi, and A. F. O. Pasaribu, "Penerapan Algoritma Svm Untuk Analisis Sentimen Pada Data Twitter Komisi Pemberantasan Korupsi Republik Indonesia," *Eduitic - Sci. J. Informatics Educ.*, vol. 7, no. 1, pp. 1–11, 2020, doi: 10.21107/edutic.v7i1.8779.
- [4] D. Abimanyu, E. Budianita, E. P. Cynthia, F. Yanto, and Yusra, "Analisis sentimen akun twitter apex legends menggunakan vader tugas akhir," 2022.
- [5] P. Arsi and R. Waluyo, "Analisis Sentimen Wacana Pemindahan Ibu Kota Indonesia Menggunakan Algoritma Support Vector Machine (SVM)," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 8, no. 1, p. 147, 2021, doi: 10.25126/jtiik.0813944.
- [6] N. M. S. Hadna, P. I. Santosa, and W. W. Winarno, "Studi Literatur Tentang Perbandingan Metode Untuk Proses Analisis Sentimen di Twitter," no. March, 2016.
- [7] M. Syarifuddin, "Analisis Sentimen Opini Publik Terhadap Efek Psbb Pada Twitter Dengan Algoritma Decision Tree, Knn, Dan Naïve Bayes," *INTI Nusa Mandiri*, vol. 15, no. 1, pp. 87–94, 2020, doi: 10.33480/inti.v15i1.1433.
- [8] M. F. Naufal, "ANALISIS PERBANDINGAN ALGORITMA SVM, KNN, DAN CNN UNTUK KLASIFIKASI CITRA CUACA," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 8, no. 2, pp. 311–318, 2021, doi: 10.25126/jtiik.202184553.

- [9] N. L. P. C. Savitri, R. A. Rahman, R. Venyutzky, and N. A. Rakhmawati, "Analisis Klasifikasi Sentimen Terhadap Sekolah Daring pada Twitter Menggunakan Supervised Machine Learning," *J. Tek. Inform. dan Sist. Inf.*, vol. 7, no. 1, pp. 47–58, 2021, doi: 10.28932/jutisi.v7i1.3216.
- [10] C. Anam and H. B. Santoso, "Perbandingan Kinerja Algoritma C4.5 dan Naive Bayes untuk Klasifikasi Penerima Beasiswa," *Energy - J. Ilm. Ilmu-Ilmu Tek.*, vol. 8, no. 1, pp. 13–19, 2018, [Online]. Available: <https://ejournal.upm.ac.id/index.php/energy/article/view/111>
- [11] K. Kelvin, J. Banjarnahor, E. I. -, and M. NK Nababan, "Analisis perbandingan sentimen Corona Virus Disease-2019 (Covid19) pada Twitter Menggunakan Metode Logistic Regression Dan Support Vector Machine (SVM)," *J. Sist. Inf. dan Ilmu Komput. Prima(JUSIKOM PRIMA)*, vol. 5, no. 2, pp. 47–52, 2022, doi: 10.34012/jurnalsisteminformasidanilmukomputer.v5i2.2365.
- [12] A. Novantirani, M. K. Sabariah, and V. Effendy, "Analisis Sentimen pada Twitter untuk Mengenai Penggunaan Transportasi Umum Darat Dalam Kota dengan Metode Support Vector Machine," *e-Proceeding Eng.*, vol. 2, no. 1, pp. 1177–1183, 2015.
- [13] W. Musu *et al.*, "Pengaruh Komposisi Data Training dan Testing terhadap Akurasi Algoritma C4 . 5," vol. X, no. 1, pp. 186–195.
- [14] M. I. Gunawan, D. Sugiarto, and I. Mardianto, "Peningkatan Kinerja Akurasi Prediksi Penyakit Diabetes Mellitus Menggunakan Metode Grid Search pada Algoritma Logistic Regression," *J. Edukasi dan Penelit. Inform.*, vol. 6, no. 3, p. 280, 2020, doi: 10.26418/jp.v6i3.40718.
- [15] C. Chazar and B. Erawan, "Machine Learning Diagnosis Kanker Payudara Menggunakan Algoritma Support Vector Machine," *Inf. (Jurnal Inform. dan Sist. Informasi)*, vol. 12, no. 1, pp. 67–80, 2020, doi: 10.37424/informasi.v12i1.48.