

Analisis Sentimen Terhadap Layanan Tokopedia Berdasarkan Twitter dengan Metode Klasifikasi Support Vector Machine

Yumarlin MZ^{*1}, Jemmy Edwin Bororing², Sri Rahayu³, Jeffry Andhika F⁴

^{1,2,3,4}Program Studi Informatika Fakultas Teknik Universitas Janabadra

E-mail: ^{*1}yumarlin@janabadra.ac.id, ²jemmy@janabadra.ac.id, ³ayu.dj@janabadra.ac.id, ⁴jeffry@janabadra.ac.id

Abstrak

Pertumbuhan e-commerce di Indonesia begitu pesat dikarenakan perubahan perilaku masyarakat dalam berbelanja. Tokopedia menjadi salah satu e-commerce terkemuka yang ada di Indonesia saat ini. Tanggapan pengguna sangat penting bagi perusahaan untuk mengetahui layanan yang ditawarkan. Ulasan dari pengguna sendiri sangat mempengaruhi citra layanan perusahaan, terlebih di era digital saat ini. Akan tetapi, membangun dan menganalisis sentimen ulasan masyarakat bukan hal yang mudah dikarenakan jumlah nya yang begitu banyak, sehingga diperlukan metode untuk melakukan klasifikasi secara otomatis, yang meliputi ulasan positif atau negatif. Salah satu situs yang cukup banyak menampung opini masyarakat ialah sosial media, diantaranya Twitter. Data ulasan dari Twitter yang diperoleh dengan melakukan crawling data menggunakan library dari Python. Dan diberi pelabelan dan klasifikasi dengan menggunakan metode Support Vector Machine dimana klasifikasi dilakukan dengan membuat garis pembatas yang memisahkan antara kelas positif dan negatif. Untuk pengukuran tingkat akurasi dari metode ini tergantung dari kualitas data set yang digunakan. Hasil klasifikasi sentimen disajikan dalam bentuk grafik. Proses klasifikasi dimulai dengan melakukan cleaning data dan prapemrosesan data, yang selanjutnya data akan dilakukan pelabelan. Dari hasil pelabelan yang telah dilakukan kemudian dilakukan klasifikasi terhadap data uji untuk menemukan sentimen positif maupun negatif. Tingkat akurasi dari hasil penelitian yang dilakukan sebesar 91%. Selanjutnya dilakukan pengujian dengan metode K-Fold Cross Validation dan diperoleh rata-rata tingkat akurasi sebesar 94,5%.

Kata Kunci— Tokopedia, SVM (Support Vector Machine), K-Fold Cross Validation

1. PENDAHULUAN

Saat pandemi Coronavirus, pertumbuhan e-commerce di Indonesia begitu melesat. Hal ini disebabkan oleh perilaku masyarakat dalam berbelanja berubah drastis [1]. Kementerian Koperasi dan UKM mencatat terjadi peningkatan belanja online sebesar 26% atau mencapai 3.1 juta transaksi per hari selama pandemi [2].

Salah satu situs e-commerce terpopuler yang paling banyak dikunjungi di Indonesia ialah Tokopedia. Berdasarkan kunjungan ke situs Tokopedia tercatat ada sebanyak 32,04% jumlah traffic share dengan jumlah kunjungan bulanan sebanyak 129,1 juta. [3].

Twitter adalah layanan komunikasi bagi teman dan keluarga untuk tetap terhubung melalui pertukaran pesan yang cepat dan sering dengan mengunggah tweet yang dapat berisi foto, video, tautan dan teks [4]. Melalui Twitter, pengguna dapat saling berdiskusi mengenai kritik, saran ataupun kepuasan mereka terhadap layanan Tokopedia.

Ulasan dari pengguna sering digunakan sebagai alat yang efektif dan efisien dalam menemukan informasi terhadap suatu produk atau jasa. Penelitian baru-baru ini menemukan

hampir 50% dari pengguna internet bergantung pada rekomendasi word-of-mouth (opini) sebelum menggunakan suatu produk [5]. Analisis sentimen merupakan proses menganalisis pendapat seseorang yang biasanya berbentuk sebuah tulisan untuk permasalahan terkait suatu topik, layanan, produk, individu, organisasi atau kegiatan tertentu [6].

Support Vector Machine (SVM) merupakan salah satu metode klasifikasi dengan menggunakan metode machine learning yang memprediksi kelas berdasarkan pola dari hasil proses training, klasifikasi dilakukan dengan garis pembatas (hyperlane) yang memisahkan antara kelas opini positif dan opini negatif [7].

1.1. Support Vector Machine (SVM)

Support Vector Machine (SVM) merupakan salah satu metode klasifikasi dengan menggunakan metode machine learning yang memprediksi kelas berdasarkan pola dari hasil proses training yang diciptakan [8].

Tingkat akurasi pada model yang akan dihasilkan oleh proses pelatihan dengan SVM sangat bergantung terhadap fungsi kernel dan parameter yang digunakan. Berdasarkan dari karakteristiknya, metode SVM dibagi menjadi dua, yaitu SVM Linier dan SVM Non-Linier. SVM linier merupakan data yang dipisahkan secara linier, yaitu memisahkan kedua class pada hyperplane dengan soft margin. Sedangkan SVM Non-Linier menerapkan fungsi dari kernel trick terhadap ruang yang berdimensi tinggi [8]. SVM linear baik untuk memisahkan data yang bersifat Linearly separable data berarti data yang dapat dipisahkan secara linear. Misalnya terdapat dua kelas, positif dan negatif. Hyperplane pemisah terbaik antara kedua kelas dapat ditemukan dengan mengukur margin hyperplane tersebut dan mencari titik maksimalnya [8].

1.2. Cosine Similarity

Cosine similarity digunakan untuk melakukan perhitungan kesamaan dari sebuah dokumen [9]. Berikut adalah rumus untuk menghitung nilai dari cosine similarity:

$$\text{Cosine Similarity } (x_1 + x_2) = (V(x_1) \cdot V(x_2)) / (|V(x_1)| |V(x_2)|)$$

Keterangan:

$|V(x_1)|$ = Panjang vector x1

$|V(x_2)|$ = Panjang vector x2

$V(x_1) \cdot V(x_2)$ = perkalian titik x1 dan x2

1.3. Confusion Matrix

Confusion matrix merupakan sebuah tabel yang terdiri dari atas banyaknya baris data uji yang diprediksi benar dan tidak benar oleh model klasifikasi, tabel ini diperlukan untuk menentukan kinerja suatu model klasifikasi. Confusion Matrix digunakan untuk mencari nilai accuracy, precision, recall, dan F1- score, tabel confusion matrix dapat dilihat pada Gambar 1.

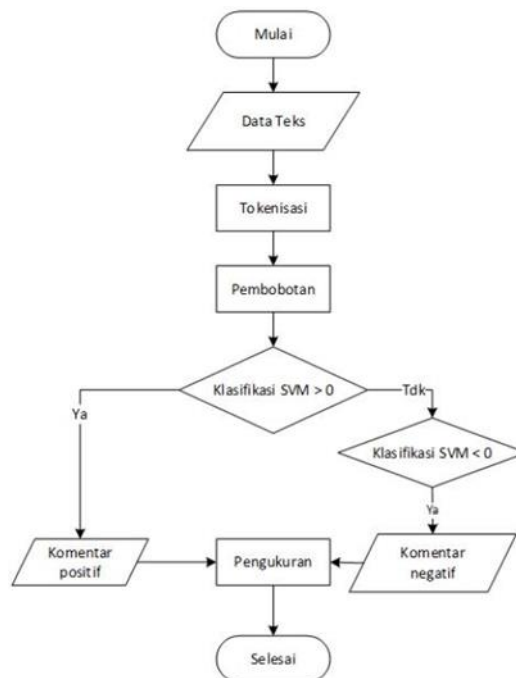
		Aktual	
Prediksi	Class	Positive	Negative
	Positive	True Positive (TP)	False Negative (FN)
	Negative	False Positive (FP)	True Negative (TN)

Gambar.1 Tabel Confusion Matrix

Dari Gambar 1, dijelaskan bahwa terdapat pengkategorian dokumen dalam suatu proses pencarian yaitu [10] : True Positive (TP), True Negative (TN), False Positive (FP), False Negative (FN). Berdasarkan tabel confusion matrix dapat dilakukan perhitungan untuk mendapatkan nilai accuracy, precision, recall dan F1- score.

2. METODE PENELITIAN

Penelitian ini dilakukan melalui beberapa tahapan, meliputi: pengumpulan data, pra-pemrosesan data, tokenisasi, pembobotan dan pengklasifikasian data menggunakan algoritma Support Vector Machine. Gambar 2 menunjukkan diagram alur proses penelitian dalam penelitian ini.



Gambar 2. Diagram Alir Proses Penelitian

2.1. Pengumpulan Data

Jenis data yang digunakan dalam penelitian adalah data primer. Data tersebut diperoleh dengan proses crawling dari halaman situs twitter. Data yang diperoleh merupakan data dari database situs twitter, yaitu data yang berupa ulasan pengguna Tokopedia yang berbahasa Indonesia. Dari data twitter yang telah dikumpulkan sejumlah 1.000 data, kemudian dikelompokkan menjadi dua kelas yaitu data ulasan positif dan ulasan negatif.

2.2. Pra Pemrosesan Data

Data ulasan yang telah diperoleh belum sepenuhnya siap digunakan untuk proses klasifikasi, hal itu dikarenakan data masih tidak terstruktur dengan baik dan masih terdapat banyak noise. Data masih memuat angka, tanda baca, serta kata-kata yang tidak diperlukan. Data perlu dilakukan preprocessing yang bertujuan untuk menjadikan data lebih terstruktur. Dalam proses preprocessing ini menggunakan bantuan library “nltk” yang sudah tersedia dalam bahasa pemrograman python.

2.3. Pembobotan TF-IDF

TF-IDF merupakan proses pembobotan terhadap kata yang nantinya akan digunakan untuk proses klasifikasi data dan berbentuk binary. Dengan menggunakan metode TF-IDF atau

Frequency-inverse document frequency, diharapkan akan meningkatkan akurasi dari proses klasifikasi. Nilai dari TF didapat dengan menggunakan persamaan:

$$TF(t) = ft, d / \sum t, d \quad (2)$$

Dimana ft, d merupakan frekuensi sebuah kata (t) yang muncul pada dokumen d , sedangkan $\sum t, d$ merupakan total keseluruhan kata yang terdapat pada dokumen d [11].

Untuk menghitung nilai IDF (Inverse Document Frequency) dari sebuah kata dalam kumpulan dokumen dapat menggunakan persamaan:

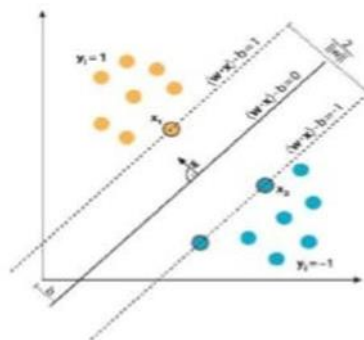
$$IDF(t) = \log (|D| / ft, D) \quad (3)$$

$|D|$ merupakan jumlah dokumen yang ada di dalam koleksi, sedangkan ft, D merupakan jumlah dokumen dimana t muncul dalam D . Langkah terakhir menghitung nilai TF-IDF dengan menggunakan rumus:

$$TF - IDF(t) = TF(t) * IDF(t) \quad (4)$$

2.4. Klasifikasi SVM

SVM (Support Vector Machine) merupakan salah satu metode klasifikasi dengan menggunakan metode machine learning (supervised learning) yang memprediksi kelas berdasarkan pola dari hasil proses training yang diciptakan. Klasifikasi dilakukan dengan garis pembatas (hyperlane) yang memisahkan antara kelas yang ditentukan [8], dapat dilihat pada Gambar 3.



Gambar 3. Hyperlane SVM

Berdasarkan Gambar 3, menunjukkan beberapa pattern dari dua kelas yaitu negatif (-1) dan positif (+1). Hyperplane pemisah kedua kelas terbaik dapat ditemukan dengan memaksimalkan margin. Margin merupakan jarak antara hyperplane dengan pattern terdekat dari masing-masing kelas. Pattern terdekat tersebut disebut support vector [12]. Secara matematika, formulasi untuk SVM untuk kasus klasifikasi linear adalah:

$$y_i [w^T x_i + b] \geq 1 \quad (5)$$

Dimana, x_i adalah data input; y_i adalah keluaran dari data x_i ; w dan b adalah parameter yang dicari nilainya. Untuk mencari Hyperplane pemisah terbaik antara kedua kelas dapat ditemukan dengan mengukur margin hyperplane tersebut dan mencari titik maksimalnya.

Langkah pertama dalam algoritma SVM adalah mendefinisikan persamaan suatu hyperplane ditulis dengan [12] :

$$W \cdot X + b = 0 \quad (6)$$

W merupakan bobot suatu vektor, yaitu $W = \{W_1, W_2, \dots, W_n\}$; n adalah banyaknya atribut, sedangkan b adalah suatu skalar yang disebut sebagai bias. Jika menggunakan penamaan atribut A_1, A_2 dengan pemisalan tupel pelatihan $X = (x_1, x_2)$, x_1 dan x_2 merupakan nilai dari atribut A_1 dan A_2 . Sehingga persamaan yang memiliki 2 atribut dapat ditulis menjadi [12]:

1. Hyperplane (H_0)

$$w_1 x_1 + w_2 x_2 + b = 0 \quad (7)$$

2. Hyperplane pendukung dari kelas positif (H1)

$$w_1 x_1 + w_2 x_2 + b \geq 1 \text{ untuk } y = +1 \quad (8)$$

3. Hyperplane pendukung dari kelas negatif (H2)

$$w_1 x_1 + w_2 x_2 + b \leq -1 \text{ untuk } y = -1 \quad (9)$$

2.5. Evaluasi

Pada tahapan ini dilakukan evaluasi terhadap klasifikasi data yang sudah diperoleh, Ukuran yang dapat digunakan untuk menilai atau mengevaluasi model klasifikasi, diantaranya adalah: accuracy, precision, recall dan F1- Score. Hasil dari nilai akurasi, presisi, recall dan F-1 Score akan ditampilkan dalam bentuk persentase.

3. HASIL DAN PEMBAHASAN

3.1. Persiapan Data

Data yang diperoleh merupakan data dari database situs twitter, yaitu data yang berupa ulasan pengguna Tokopedia berbahasa Indonesia dengan waktu crawling selama bulan Februari dan mendapatkan data sebanyak 20.000 data. Dari data tersebut dilakukan pemilihan kata dengan rata-rata 500 data ulasan positif dan 500 data ulasan negatif.

Data yang diperoleh merupakan data mentah yang masih berformat (.json), selanjutnya dilakukan proses labeling data dimana label “1” untuk data positif dan “0” untuk data negatif. Data yang telah di labeling dapat dilihat pada Gambar 4.

No	Jenis Data	Data
1	1	halo min gaada kendala min, tokped bagus banget, mamaku kalo beli apa-apa selalu di tokped 🤗
2	1	Nanya-nanya tu selalu dijawab. Mana jawabannya tu friendly banget jadi gaada sungkan.
3	1	Selama pakai tokped untuk belanja ga pernah ada kendala, fiturnya user friendly
4	1	Nah akhirnya barang pesanan ku sampai di aku dengan kondisi yang baik. Tokopedia Care selalu ada solusi
5	1	Baik terimakasih banyak kak 🥰
6	1	Terimakasih Mimin
7	1	Pokoknya Tokopedia Care selalu ada solusi disetiap aku punya permasalahan dengan aplikasinya. Terimakasih
8	1	Terimakasih banyak mintoo🥰 tokopedia hadiah GA nya sudah datang, huhu keren banget, sekali lagi Terimakasih
9	1	Terimakasih respon & bantuannya Minto :)
10	1	My Bestie Tokopedia memang terbaik 🥰 Selalu Ada Solusi Sahabat Terbaik Toppers

Gambar 4. Data labeling

3.2. Prapemrosesan Data

Tahapan preprocessing sendiri terdiri dari cleaning data/remove punctuation, case folding, tokenisasi, stopwords removal dan stemming. Proses preprocessing sendiri dilakukan dengan perintah dan library “nltk” dalam Python.

3.3. Klasifikasi SVM

Dalam proses pembobotan dan klasifikasi data ini digunakan library “sklearn” untuk menjalankan proses pembobotan dan klasifikasi secara sistem. Dimana data training yang sudah jadi akan masuk ke dalam proses klasifikasi support vector machine dan menghasilkan beberapa output berupa data yang terstruktur, nilai accuracy, precision, recall dan f-score.

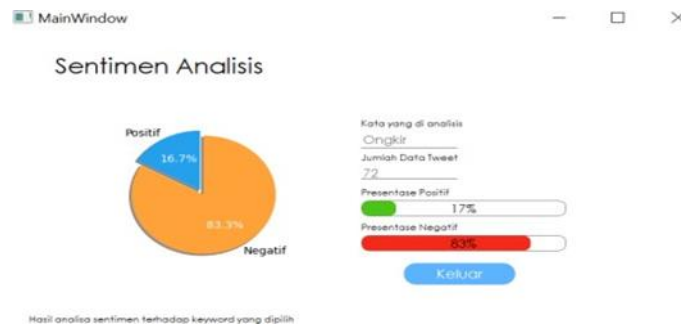
3.4. Implementasi Hasil Program

Penelitian ini mengembangkan fitur untuk melakukan analisis sentimen secara realtime. Proses ini akan menggunakan library Twitter- Sentiment-Classifier dimana sistem akan berjalan otomatis untuk melakukan crawling data yang diambil dari situs twitter secara realtime, dapat dilihat pada Gambar 5 berikut ini.



Gambar 5. Tampilan awal program

Dalam program akan terintegrasi otomatis dimana data yang akan di analisis bergantung dengan keyword yang dimasukan. Setelah menambahkan keyword, maka sistem akan otomatis melakukan crawling data dari situs twitter yang selanjutnya dilakukan proses preprocessing dan klasifikasi secara otomatis. Pengguna diharuskan memasukkan kata kunci yang berkaitan dengan Tokopedia, dicontohkan dengan kata “ongkir” Setelah memasukkan keyword yang berkaitan dengan Tokopedia, pengguna akan diarahkan untuk menekan tombol mulai untuk lanjut ke proses klasifikasi dan reset untuk memasukkan ulang. Proses klasifikasi akan menghasilkan grafik berupa besaran data sentimen positif maupun negatif dan banyaknya jumlah data yang diperoleh. Hasil klasifikasi dapat dilihat pada Gambar 6.



Gambar 6. Tampilan hasil Klasifikasi

Berdasarkan Gambar 6 dapat dijelaskan bahwa data yang berkaitan dengan kata “ongkir” diperoleh sebanyak 72 data. Dari 72 data tersebut menghasilkan presentase bahwa kata ongkir bermakna positif sebesar 17% dan 83% kata yang mengandung kata ongkir bermakna negatif, menunjukkan bahwa kata ongkir berdasarkan ulasan pengguna masih banyak yang mengeluhkan mengenai ongkir.

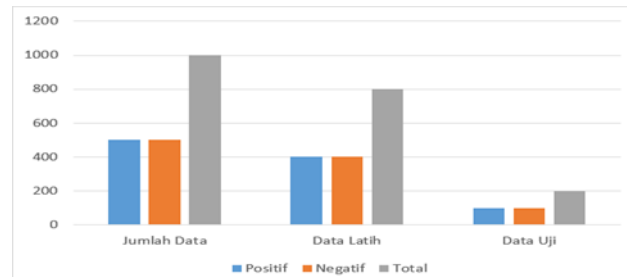
3.5. Nilai Akurasi

Penelitian ini menggunakan metode 10- Fold Validation dalam proses evaluasi. Evaluasi tersebut mencakup nilai, Accuracy, precision, recall dan F1-Score. Data yang digunakan untuk data latih dan data uji adalah data yang telah memiliki label kelas, dengan perbandingan 80%:20%. Perbandingan data latih dan data uji dapat dilihat pada Tabel 1.

Tabel 1. Perbandingan data latih & uji

Klasifikasi	Jumlah	Data Latih (80%)	Data Uji (20%)
Positif	500	400	100
Negatif	500	400	100
Total	1.000	800	200

Berdasarkan Tabel 1, dengan perbandingan 80:20 dari data total 1.000 ulasan berbahasa Indonesia, digunakan sebanyak 800 ulasan sebagai data latih dan 200 ulasan sebagai data uji. Untuk lebih jelasnya dapat dilihat pada Gambar 7 berikut.



Gambar 7. Perbandingan data latih & uji

Setelah mendapatkan besaran data uji maka akan dilakukan perhitungan akurasi dengan confusion matrix. Confusion matrix sendiri digunakan untuk memudahkan dalam proses menghitung nilai akurasi dengan mengetahui jumlah data uji yang terklasifikasi benar maupun salah. Berikut adalah tampilan pengujian Confusion matrix oleh sistem dapat dilihat pada Gambar 8.

```

IDLE Shell 3.9.4
File Edit Shell Debug Options Window Help
Python 3.9.4 (tags/v3.9.4:1ff2e308, Apr 6 2021, 13:40:21) [MSC v.1928 64 bit (AMD64)] on win32
Type "help()", "copyright()", "credits()" or "license()" for more information.
>>> = RESTART: C:\Users\User\Music\Source Code-Analysis Sentimen Tokopedia\training.PY
precision    recall  f1-score   support
0           0.89      0.90      0.89         87
1           0.92      0.91      0.92        113

accuracy          0.90      0.90      0.91        200
macro avg          0.90      0.90      0.90        200
weighted avg          0.91      0.91      0.91        200

The length of the string is : 200

```

Gambar 8. Hasil pengujian

Dari Gambar 8, dapat dijelaskan bahwa terdapat data uji sebanyak 200 data yang mana terdapat 87 data yang terkonfirmasi negatif dan 113 data yang terkonfirmasi data positif. Confusion matrix dari data uji yang telah diperoleh dapat dilihat pada Tabel 2 berikut.

Tabel 2. Tabel Confusion matrix

		Nilai Sebenarnya		Jumlah
		Positif	Negatif	
Nilai Prediksi	Positif	102	8	110
	Negatif	11	79	90
Jumlah		113	87	200

Dari Tabel 2, dapat dijabarkan bahwa terdapat 113 sentimen positif dan 87 sentimen negatif yang dihasilkan oleh sistem namun terdapat 8 data positif yang diprediksi sebagai data negatif dan 11 data negatif yang diprediksi sebagai data positif. Berdasarkan Tabel 2 dapat dihitung nilai akurasi dan pengujian cross validation seperti dijelaskan pada sub bab berikut:

1. Perhitungan Nilai Akurasi

a. Precision

Nilai Precision merupakan rasio prediksi tepat terhadap keseluruhan data prediksi positif.

$$\begin{aligned}
 \text{Precision} &= \text{TP}/(\text{TP}+\text{FP}) \\
 &= 102/(102+8) \\
 &= 0.92 \times 100\% = 92\%
 \end{aligned}$$

Ket:

TP : True Positif

TN : True Negatif

FP : False Positif

FN : False Negatif

b. Recall

Recall adalah proporsi jumlah dokumen yang relevan diantara semua dokumen teks relevan yang ada pada klasifikasi.

$$\begin{aligned}\text{Recall} &= \text{TP}/(\text{TP}+\text{FN}) \\ &= 102/(102+11) \\ &= 0.91 \times 100\% \\ &= 91\%\end{aligned}$$

c. F- Score

Merupakan perbandingan rata-rata antara precision dengan recall

$$\begin{aligned}\text{F-Score} &= (2 \times \text{Precision} \times \text{Recall})/(\text{Precision} + \text{Recall}) \\ &= (2 \times 0.92 \times 0.91)/(0.92 + 0.91) \\ &= 0.91 \times 100\% \\ &= 91\%\end{aligned}$$

d. Accuracy

Merupakan jumlah proporsi prediksi yang tepat

$$\begin{aligned}\text{Accuracy} &= \text{TP}+\text{TN}/(\text{TP}+\text{TN}+\text{FP}+\text{FN}) \\ &= 102+79/(102+79+8+11) \\ &= 0.91 \times 100\% \\ &= 91\%\end{aligned}$$

4. Pengujian Cross Validation

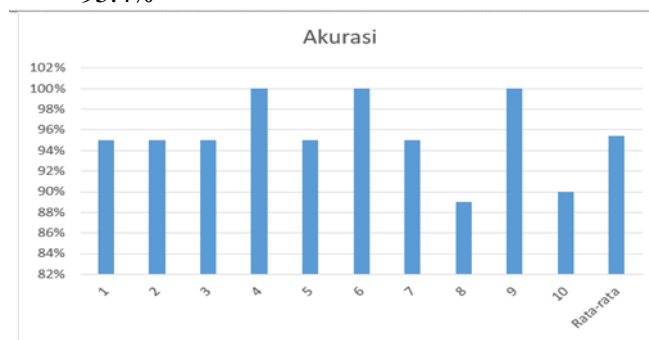
Untuk menguji performa machine learning dalam melakukan klasifikasi akan dilakukan pengujian cross validation dengan 10 Fold dan rasio data 80% untuk data training dan 20% untuk data testing yang selanjutnya akan dihitung untuk nilai rata-rata akurasi setiap fold. Perbandingan cross validation dapat dilihat pada Tabel 3 berikut:

Tabel 3. Perbandingan Cross Validation

FOLD	ASPEK	SVM
Fold 1	Accuracy (%)	95%
	Precision (%)	96%
	Recall (%)	95%
	F-1 Score	95%
	Jumlah Data	100
Fold 2	Accuracy (%)	95%
	Precision (%)	95%
	Recall (%)	95%
	F-1 Score	96%
	Jumlah Data	100
Fold 3	Accuracy (%)	95%
	Precision (%)	96%
	Recall (%)	95%
	F-1 Score	95%
	Jumlah Data	100
Fold 4	Accuracy (%)	100%
	Precision (%)	100%
	Recall (%)	100%
	F-1 Score	100%
	Jumlah Data	100
Fold 5	Accuracy (%)	95%
	Precision (%)	96%
	Recall (%)	95%
	F-1 Score	95%
	Jumlah Data	100
Fold 6	Accuracy (%)	100%
	Precision (%)	100%
	Recall (%)	100%
	F-1 Score	100%
	Jumlah Data	100
Fold 7	Accuracy (%)	95%
	Precision (%)	96%
	Recall (%)	95%
	F-1 Score	95%
	Jumlah Data	100
Fold 8	Accuracy (%)	89%
	Precision (%)	91%
	Recall (%)	89%
	F-1 Score	89%
	Jumlah Data	100
Fold 9	Accuracy (%)	100%
	Precision (%)	100%
	Recall (%)	100%
	F-1 Score	100%
	Jumlah Data	100
Fold 10	Accuracy (%)	90%
	Precision (%)	92%
	Recall (%)	90%

Dari data Tabel 3, dapat disimpulkan bahwa pada fold 4, fold 6 dan fold 9 memiliki nilai akurasi terbesar dengan nilai 100% sedangkan pada fold 8 memiliki nilai akurasi paling kecil dengan nilai 89%. Berdasarkan data pada Tabel 3, dapat dilakukan proses perhitungan nilai rata-rata akurasi dari setiap fold dengan rumus, dapat di lihat Gambar 9 berikut.

$$\begin{aligned}
 \text{Rata-rata} &= \frac{\text{Jumlah Nilai}}{\text{Banyak data}} \\
 &= \frac{95+95+95+100+95+100+95+89+100+90}{10} \\
 &= 95.4 \\
 \text{Rata-rata} &= 95.4 \times 100\% \\
 &= 95.4\%
 \end{aligned}$$



Gambar 9. Perbandingan nilai akurasi

Berdasarkan Gambar 9, dapat dijabarkan rata-rata nilai akurasi yang dihasilkan oleh sistem berdasarkan data uji ke-10 fold menghasilkan nilai rata-rata akurasi sebesar 95.4%. Ketepatan klasifikasi dapat dilihat dari akurasi klasifikasi. Akurasi klasifikasi menunjukkan ketepatan dan performansi model klasifikasi secara keseluruhan, dimana semakin tinggi akurasi klasifikasi semakin baik ketepatan dan performansi model klasifikasi [13].

4. KESIMPULAN

Berdasarkan hasil analisis dan pembahasan dapat menarik kesimpulan sebagai berikut:

1. Berdasarkan penelitian yang telah dilakukan, proses crawling dapat dilakukan dengan library tweepy dan untuk melakukan klasifikasi data ulasan dapat dilakukan dengan library SVM.
2. Proses membangun dan analisis sentimen terhadap layanan Tokopedia menggunakan data ulasan yang diperoleh dari hasil crawling data ulasan Tokopedia pada twitter dengan menggunakan library tweepy pada python yang selanjutnya data mentah tersebut dilakukan prapemrosesan data yang bertujuan memperoleh data yang terstruktur. Kemudian data yang terstruktur tadi diklasifikasikan menggunakan metode support vector machine untuk mendapatkan nilai akurasi bahwa ulasan tersebut masuk kedalam ulasan positif atau negatif. Data ulasan yang telah melalui proses klasifikasi ditampilkan dalam bentuk diagram persentase.
3. Dengan menggunakan perbandingan data latih dan data uji sebesar 80:20% dengan metode support vector machine diperoleh nilai akurasi analisis sentimen sebesar 91% yang artinya dari 200 data ulasan yang diujikan, terdapat 113 ulasan yang benar klasifikasinya. Sedangkan setelah dilakukan evaluasi dengan metode K-Fold Cross Validation terhadap 10-Fold yang diujikan, diperoleh rata-rata data akurasi sebesar 94.5%, yang artinya dari 1.000 data ulasan yang diujikan terdapat 945 data ulasan yang benar pengklasifikasiannya.

DAFTAR PUSTAKA

- [1] R.Pratama,E.Prasetyo, A.Rinaldi, and I.Rusdi, "Aplikasi Startup rekomendasiin.com dengan Fitur Perekomendasi Berbasis SPK dengan Metode SAW," Smart Comp., vol. 10, no. 3, 2021, ISSN :2089-676X.
- [2] Riana, "Pandemi, Penjualan di e- Commerce Tembus 3,1 Juta Transaksi per Hari," [www.pontos.id](https://pontos.id/2021/04/15/pandemi-penjualan-di-e-commerce-tembus-31-juta-transaksi-per-hari/), 2021 <https://pontos.id/2021/04/15/pandemi-penjualan-di-e-commerce-tembus-31-juta-transaksi-per-hari/> (accessed Sep. 21, 2021).
- [3] E.Catrina, "5 E-commerce yang paling banyak dikunjungi di Indonesia," [www.kompas.com](https://money.kompas.com/read/2021/02/24/072440626/5-e-commerce-yang-paling-banyak-dikunjungi-di-indonesia?page=all), 2021. <https://money.kompas.com/read/2021/02/24/072440626/5-e-commerce-yang-paling-banyak-dikunjungi-di-indonesia?page=all> (accessed Sep. 21, 2021).
- [4] D. A. Agustina, S. Subanti, and E. Zukhronah, "Implementasi Text Mining Pada Analisis Sentimen Pengguna Twitter Terhadap Marketplace di Indonesia Menggunakan Algoritma Support Vector Machine," Indones. J. Appl. Stat., vol. 3, no. 2, p. 109, 2021, doi: 10.13057/ijas.v3i2.44337.
- [5] F. Fanani, "Klasifikasi Review Software Pada Google Play Menggunakan Pendekatan Analisis Sentimen," 2017.
- [6] Y. Lin, X. Wang, and A. Zhou, "Opinion spam detection," Opin. Anal. Online Rev., no. May, pp. 79–94, 2016, doi: 10.1142/9789813100459_0007.

- [7] R. S. Rofiqoh, Umi 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 Twitter event detection View project Human Detection and Tracking View project,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 1(12), no. October, pp. 1725–1732, 2017, [Online]. Available: <https://www.researchgate.net/publication/320234928>.
- [8] D. Yosmita Praptiwi, “Analisis Sentimen Online Review Pengguna E-Commerce Menggunakan Metode Support Vector Machine Dan Maximum Entropy,” 2018.
- [9] Wahyuni, Rizki Tri, Dhidik Prastiyanto, and Eko Suprptono. 2017. “Penerapan Algoritma Cosine Similarity Dan Pembobotan TF-IDF Pada Sistem Klasifikasi Dokumen Skripsi.” *Jurnal Teknik Elektro Universitas Negeri Semarang* 9(1):18–23.
- [10] Monica Pravina, Arsyia Cholissodin, Imam Pandu Adikara, Putra. 2019. “Tampilan Analisis Sentimen Tentang Opini Maskapai Penerbangan Pada Dokumen Twitter Menggunakan Algoritme Support Vector Machine (SVM).Pdf.”
- [11] H. Syahputra, L. K. Basyar, and A. A. S. Tamba, “Setiment Analysis of Public Opinion on the Go-Jek Indonesia Through Twitter Using Algorithm Support Vector Machine,” *J. Phys. Conf. Ser.*, vol. 1462, no. 1, 2020, doi: 10.1088/1742-6596/1462/1/012063.
- [12] E. F. U. Latifah, “Perbandingan Kinerja Machine Learning Berbasis Algoritma Support Vector Machine dan Naive Bayes (Studi Kasus: Data Tanggapan Mengenai Traveloka Melalui Media Sosial Twitter),” 2018, [Online]. Available: <https://dspace.uui.ac.id/handle/123456789/7996>.
- [13] E. Prasetyo, *Data mining : konsep dan aplikasi menggunakan MATLAB*. CV Andi Offset, 2012.