

Sentiment Analysis of OYO App Reviews Using the *Support Vector Machine* Algorithm

Analisis Sentimen terhadap Ulasan Aplikasi OYO menggunakan Algoritma Support Vector Machine

Zaenal, Ika Ratna Indra Astutik
{zaenal167@gmail.com, ikaratna@umsida.ac.id}

Program Studi Informatika, Fakultas Sains dan Teknologi, Universitas Muhammadiyah Sidoarjo

Abstract. *The rapidly growing tourism industry causes the need for hotels to increase. This has led to innovation in the form of virtual hotel operators, one of which is OYO. OYO is a VHO application with the highest rating found on the Google playstore. OYO is one of the applications with millions of users. Of course, this cannot be separated from the ratings and reviews from users. The reviews contained in the playstore itself can contain positive, neutral or even negative opinions. Given the importance of user reviews to application development, this research classifies reviews on the OYO application to determine user sentiment. In this study, the data used is 2,000 data which will be classified into positive, neutral and negative sentiments. The Support vector machine algorithm was chosen because it is capable of producing high accuracy. Based on testing, the Radial basis function kernel is able to produce the highest accuracy among other kernels and by using a dataset division ratio of 80:20 the accuracy obtained is 78.98%. While testing using the Confusion matrix produces an accuracy of 80.36%.*

Keywords – OYO; Playstore; RBF; Support vector machine

Abstrak. *Industri pariwisata yang berkembang pesat menyebabkan kebutuhan akan hotel semakin meningkat. Hal ini memunculkan inovasi berupa virtual hotel operator yang salah satunya adalah OYO. OYO adalah aplikasi VHO dengan rating tertinggi yang terdapat di google playstore. OYO menjadi salah satu aplikasi dengan pengguna mencapai jutaan. Tentunya hal ini tidak lepas dari pemberian rating dan ulasan dari para pengguna. Ulasan yang terdapat di playstore sendiri bisa berisi opini positif, netral atau bahkan negatif. Mengingat pentingnya ulasan pengguna terhadap perkembangan aplikasi, pada penelitian ini dilakukan klasifikasi ulasan pada aplikasi OYO untuk menentukan sentimen pengguna. Dalam penelitian ini, data yang digunakan adalah sebanyak 2.000 data yang akan diklasifikasikan menjadi sentimen positif, netral, dan negatif. Algoritma Support vector machine dipilih dikarenakan mampu menghasilkan akurasi yang tinggi. Berdasarkan pengujian, kernel Radial basis function mampu menghasilkan akurasi tertinggi diantara kernel lainnya dan dengan menggunakan rasio pembagian dataset sebesar 80:20 hasil akurasi yang didapatkan adalah sebesar 78,98%. Sedangkan pengujian dengan menggunakan Confusion matrix menghasilkan akurasi sebesar 80,36%.*

Kata Kunci – OYO; Playstore; RBF; Support Vector Machine

I. PENDAHULUAN

Teknologi dalam bidang pariwisata, khususnya akomodasi penginapan, menjadi salah satu hal penting yang perlu diperhatikan oleh para pemilik hotel. Ini dikarenakan teknologi ini bisa membantu perkembangan hotel itu sendiri dan juga membantu konsumen yang menggunakan jasa ini [1]. Terdapat beberapa perkembangan mengenai bisnis perhotelan, dimana konsumen dapat melakukan pemesanan akomodasi penginapan dengan menggunakan sistem online. Hal tersebut disebut juga dengan VHO (*Virtual Hotel Operator*). Salah satu VHO di Indonesia adalah hotel OYO, yang pada awalnya merupakan singkatan dari *On Your Own*. OYO berasal dari India, yakni sebuah situs web yang menawarkan inventaris kamar hotel bermerk di seluruh India dengan fasilitas standar, seperti: wifi gratis, televisi layar datar, linen berkualitas, dan perlengkapan mandi bermerek. Pangsa pasar utama yang disasar adalah permintaan akomodasi, terutama di segmen hotel kecil, independen, dan *unbranded* [2].

Ulasan yang terdapat di *google playstore* dapat digunakan sebagai acuan dalam menilai suatu aplikasi. Aplikasi OYO sendiri memiliki beragam jenis ulasan, baik yang mengandung opini positif atau bahkan negatif. Mengingat pentingnya ulasan bagi suatu aplikasi, maka pada penelitian ini akan mengklasifikasikan ulasan menjadi ulasan dengan sentimen positif, netral, dan negatif. Pelabelan sentimen nantinya akan didasarkan atas rating yang diberikan oleh pengguna. Algoritma yang digunakan adalah *Support vector machine* yang akan diterapkan dalam proses analisis sentimen. Algoritma SVM merupakan sebuah algoritma *machine learning* dengan dasar teori struktural pembelajaran statistik untuk menemukan batas yang memisahkan antar kelas [3].

Penelitian ini bertujuan untuk mengukur tingkat akurasi yang dihasilkan oleh algoritma SVM dalam mengklasifikasikan ulasan ke dalam beberapa kelas sentimen.

A. Text mining

Text mining adalah sebuah proses pengetahuan intensif dimana pengguna berinteraksi dan bekerja dengan sekumpulan dokumen dengan menggunakan beberapa alat analisis. Tujuan utama dari *text mining* ini adalah untuk mengekstrak informasi dari data berupa bahasa tekstual dengan tujuan tertentu. Jika dibandingkan dengan data yang tersimpan pada suatu basis data, text (data berupa data tekstual) merupakan data yang tidak terstruktur dan sulit menemukan algoritma untuk menyelesaikannya [4].

B. Text preprocessing

Text preprocessing adalah tahap awal dari *text mining*. Tahap ini mencakup semua rutinitas, dan proses untuk mempersiapkan data yang akan digunakan pada operasi *knowledge discovery* sistem *text mining* [4]. Beberapa metode yang dilakukan untuk menyederhanakan dan mengurangi kata serta karakter special yang tidak dibutuhkan dalam data diantaranya yaitu, *case folding*, *cleaning*, *tokenizing*, *filtering stopword* dan *stemming*. Hal ini dilakukan agar data yang didapatkan menjadi lebih terstruktur dan sederhana untuk memudahkan data saat hendak diolah [5].

Case Folding

Case Folding merupakan proses untuk menyederhanakan huruf yang terdapat dalam suatu teks dengan cara merubah setiap huruf kapital menjadi huruf kecil serta menghilangkan tanda baca dan karakter special yang terdapat pada dokumen.

Tokenizing

Tokenizing merupakan salah satu proses untuk memisahkan tiap kata dengan potongan-potongan yang disebut token. Proses *tokenizing* yaitu dengan melakukan pemisahan pada karakter tertentu atau spasi untuk membuang beberapa karakter tertentu.

Filtering Stopword

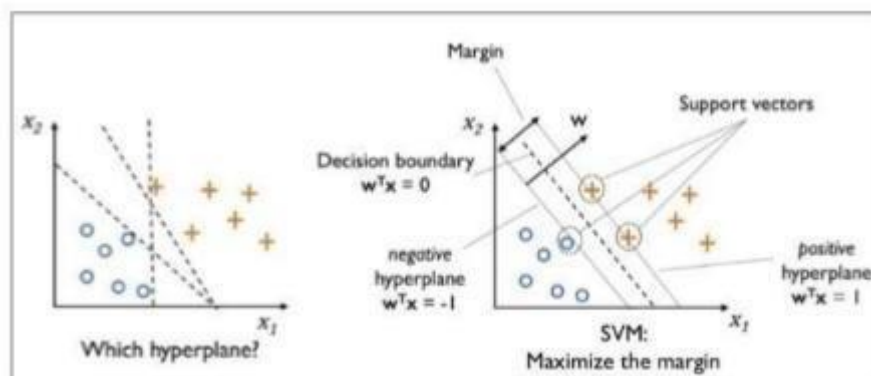
Stopword merupakan istilah untuk kata yang tidak memiliki makna yang krusial didalam *text mining*. *Filtering stopword* dilakukan dengan cara menyaring kata yang tidak memiliki makna krusial untuk selanjutnya dibuang dari teks. Hal ini bertujuan untuk meningkatkan kecepatan sistem dalam memproses suatu dokumen.

Stemming

Pada tahap stemming, pada data review, imbuhan pada kata-kata akan dihilangkan. Contohnya adalah kata “dihilangkan” diubah menjadi “hilang”.

C. Support vector machine (SVM)

SVM merupakan algoritma yang bekerja menggunakan pemetaan nonlinier untuk mengubah data pelatihan asli ke dimensi yang lebih tinggi. Dalam hal ini dimensi baru, akan mencari *hyperplane* untuk memisahkan secara linier dan dengan pemetaan nonlinier yang tepat ke dimensi lebih tinggi, data dari dua kelas selalu dapat dipisahkan dengan *hyperplane* tersebut. SVM menemukan ini menggunakan *support vector* dan margin. Dalam teknik ini, kita berusaha untuk menemukan fungsi pemisah (klasifier) yang optimal yang bisa memisahkan dua kelas yang berbeda. Teknik ini berusaha menemukan fungsi pemisah (*hyperplane*) terbaik diantara fungsi yang tidak terbatas jumlahnya untuk memisahkan dua macam obyek. *Hyperplane* terbaik adalah *hyperplane* yang terletak di tengah-tengah antara dua set obyek dari dua kelas. Gambar 1, memperlihatkan bagaimana SVM mencari *hyperplane* terbaik ekuivalen dengan memaksimalkan margin atau jarak dua set dari kelas yang berbeda.



Gambar 1. Hyperplane Pemisah Kelas Positif (+1) Dan Negatif (-1)

D. Python

Python merupakan bahasapemrograman yang fleksibel dan sederhana yang didefinisikan dalam dokumen-dokumennya sebagai berikut:

1. *Python* merupakan bahasa pemrograman tujuan umum yang sangat tingkat tinggi, dinamis, berorientasi terhadap objek yang sering digunakan.
2. Dapat diimplementasikan ke dalam program yang luas.
3. *Python* merupakan bahasa pemrograman yang mensupport banyak model pemrograman diantaranya struktural dan berorientasi objek.
4. *Python* sangat fleksibel, karena kemampuannya untuk menggunakan komponen modular yang dirancang dalam bahasa pemrograman lainnya. Kita dapat membuat program di C ++ dan mengimpornyake *python* sebagai modul.

Beberapa karakteristik yang menjadi pembeda *python* adalah:

1. *Python* merupakan bahasa pemrograman yang cepat dan kuat serta menyediakan fasilitas pemrograman dari operasi dasar sampai fungsi lanjutan.
2. *Python* kompatibel dengan teknologi lainnya seperti mendukung COM, .Net, dan objek lainnya. Selain itu, beberapa alternatif dan pelengkap dibuat untuk *python* yang memudahkan pekerjaan dengan objek-objek dalam mode yang saling terhubung.
3. *Python* dapat berjalan di sitem operasi yang berbeda-beda diantaranya Windows, Linux, Mac Oportabe
4. *Python* bersifat *open source*

Beberapa modul yang dipakai pada machine learning adalah sebagai berikut:

1. Scikit-learn, yaitu sebuah modul *Python* yang digunakan sebagai sarana integrasi antar algoritma-algoritma mesin state-of-the-art. Paket ini berfokus pada membawa pembelajaran mesin ke non-spesialis menggunakanbbahasa tingkat tinggi tujuan umum[6]
2. Pandas, yaitu sebuah package yang tersedia dalam *Python* untuk mengolah struktur data yang cepat, fleksibel, dan ekspresif yang dirancang untuk mempermudah pekerjaan yang melibatkan data "relasional" atau "berlabel" [7].

E. Klasifikasi

Klasifikasi merupakan salah satu metode untuk mengkategorikan sebuah objek ke dalam suatu kelompok maupun kelas tertentu [8]. Algoritma yang umum digunakan adalah *Decision* atau *Classification Trees*, Analisis Statik, *Naive Bayes Classifier*, Algoritma Genetika, *Rough Sets*, Metode *Rule Based*, *Memory Based Reasoning*, K-Nearest Neighbor dan SVM [9]

F. Analisis Sentimen

Analisis sentimen atau opinion mining merupakan proses memahami, mengekstrak dan mengolah data tekstual secara otomatis untuk mendapatkan informasi sentimen yang terkandung dalam suatu kalimat opini. Analisis sentimen dilakukan untuk melihat pendapat atau kecenderungan opini terhadap sebuah masalah atau objek oleh seseorang, apakah cenderung berpandangan atau beropini negatif atau positif. Salah satu contoh penggunaan analisis sentimen dalam dunia nyata adalah identifikasi kecenderungan pasar dan opini pasar terhadap suatu objek barang. Besarnya pengaruh dan manfaat dari analisis sentimen menyebabkan penelitian dan aplikasi berbasis analisis sentimen berkembang pesat. Bahkan di Amerika terdapat sekitar 20-30 perusahaan yang memfokuskan pada layanan analisis sentimen [10]

G. Confusion matrix

Confusion matrix menyajikan sebuah tabel yang menyatakan jumlah data uji yang benar dan jumlah data uji yang salah diklasifikasikan [11]. Tabel *confusion matrix* digunakan untuk mengukur tingkat akurasi suatu sistem orientasi sentimen. Contoh confussion matrix dalam klasifikasi biner ditunjukkan pada Tabel 1 berikut:

Tabel 1. *Confusion Matrix*

Kelas Sebenarnya	Kelas hasil prediksi	
	Positif = 1	Negatif = 0
Positif = 1	TP	FN
Negatif = 0	FP	TN

Dirumuskan sebagai berikut:

$$Akurasi = \frac{TP+TN}{TP+FP+FN+TN}$$

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

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

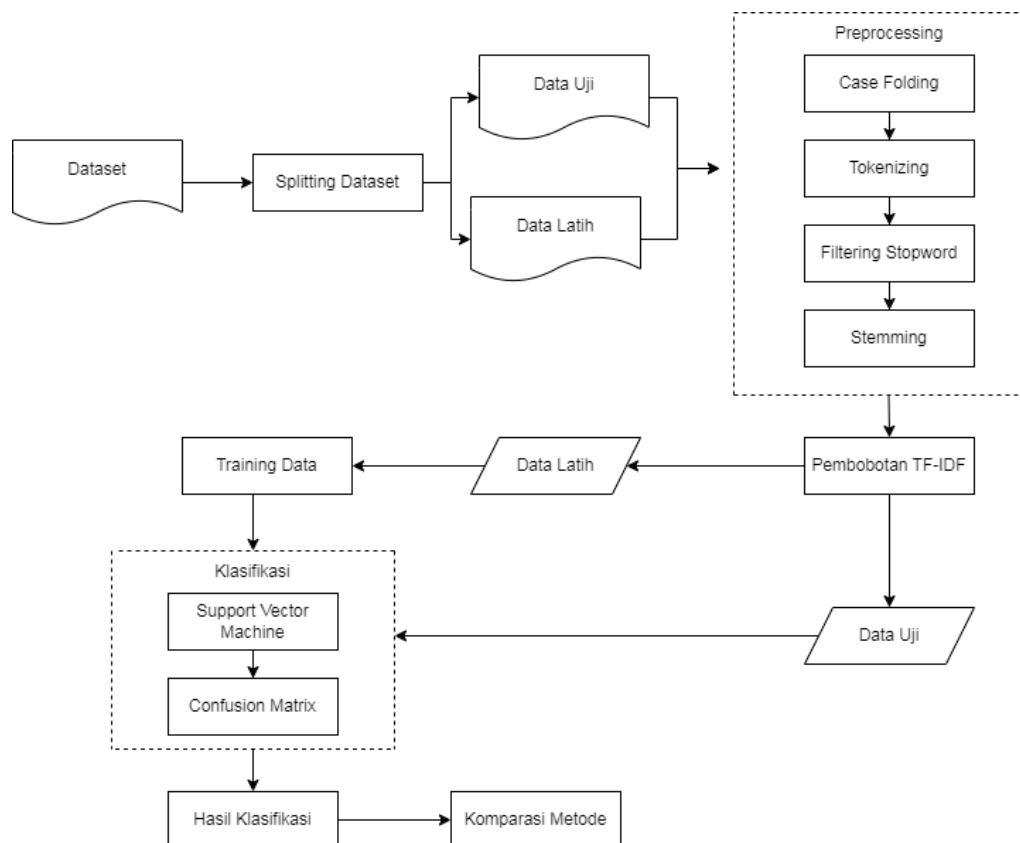
$$F1\ Score = 2 \times \frac{precision \times recall}{precision + recall}$$

Keterangan :

- A. True Positif (TP), yaitu jumlah dokumen dari kelas 1 yang benar dan diklasifikasikan sebagai kelas 1.
- B. True Negatif (TN), yaitu jumlah dokumen dari kelas 0 yang benar diklasifikasikan sebagai kelas 0.
- C. False Positif (FP), yaitu jumlah dokumen dari kelas 0 yang salah diklasifikasikan sebagai kelas 1.
- D. False Negatif (FN) yaitu jumlah dokumen dari kelas 1 yang salah diklasifikasikan sebagai kelas 0.

II. METODE

Flowchart merupakan gambaran jalannya sebuah program dari satu proses ke proses lainnya. Sehingga, alur dari system menjadi simpel dan mudah dipahami oleh semua orang. Selain itu, fungsi lain *flowchart* adalah menyederhanakan rangkaian prosedur pemahaman terhadap informasi tersebut. *Flowchart* metode *honeypot* dapat dilihat pada Gambar 2 dibawah ini.



Gambar 2. Flowchart Sistem

Data pada penelitian ini didapatkan dengan metode *web scrapping* pada situs *google playstore*, data berupa kumpulan ulasan beserta *rating* terhadap aplikasi OYO dengan total data sebanyak 2.000 data. Ulasan akan dikelompokkan ke dalam 3 kelas sentimen yang dapat dilihat pada Tabel 1. Dimana pelabelan didasarkan terhadap *rating* yakni *rating* 4-5 sebagai ulasan positif, *rating* 3 untuk ulasan netral dan *rating* 1-2 untuk ulasan negatif.

Tabel 1. Jumlah Data

No.	Sentimen	Total
1.	Positif	1.179
2.	Netral	61
3.	Negatif	760
Total		2.000

Berikut contoh data yang digunakan pada penelitian ini ditunjukkan Gambar 3.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	userNameScore	at	content										
2	cup-cep.c	5	#####	bagussss									
3	Wiwied W	5	#####	Mantap kamar luas, bersih, pelayanan baik									
4	Muhamm	1	#####	Apk mempersulit tah kayak mana daftarnya segala cara udah di buat jijin kadang minta kode refernsi uda									
5	Andari Wu	5	#####	terbaik									
6	Reynaldi A	5	#####	keren									
7	Hilya Hilya	5	#####	Jos									
8	abdul roki	5	#####	Mantap									
9	Herry Mar	4	#####	sangat baik									
10	Jayadi 78	5	#####	ga perlu repot pada saat keluar kota cari hotel.tingal klik									
11	Dena Char	5	#####	Bagus									
12	Aluna Gali	1	#####	Aplikasi jelek, saya kecewa. Promo tidak ada kejelasan. Saya sudah seneng booking dapet diskon dan me									
13	Sekitarkit	1	#####	PAYAH OYO BERTELE TELE BUAT. PENGEMBALIAN UANG NYA, NGGA IKLAS DUNIA AKHIRAT									
14	Berith Sav	5	#####	terbaik									
15	bagas san	5	#####	hemat									
16	Dori Ansy	5	#####	apk nya bgus cpat pemesanan nya harganya bersahabat dan sangat membantu									
17	Yuda Ferd	5	#####	Saya baru tau ada hotel berkualitas dengan harga murah. Berkt bantuan teman									
18	wildan ka	1	#####	Pelayan nya tidak lah sangat adil									
19	Fitri Fitri	5	#####	sip dan mantap									
20	Faisal Adr	5	#####	saya puas									
21	DJ 20D 16	5	#####	ok									

Gambar 3. Contoh Data

Dari total data yang mencapai jumlah 2.000 data, penulis nantinya akan membagi dataset menjadi data training dan juga data testing dengan beberapa perbandingan, Hal tersebut disebabkan karena jumlah data training akan mempengaruhi tingkat akurasi. Semakin banyak data training maka model akan semakin banyak belajar sehingga ketelitian akan semakin baik [12].

III. HASIL DAN PEMBAHASAN

A. Preprocessing data

Proses diawali dengan tahap *preprocessing* karena data tidak sepenuhnya menggunakan kata baku. Tahap *preprocessing* dilakukan dengan menggunakan bantuan library pada bahasa pemrograman *Python 3*. Setelah data dilakukan *preprocessing*, selanjutnya hasil *preprocessing* disimpan menjadi file baru yang nantinya akan digunakan sebagai dataset dalam proses klasifikasi. Adapun hasil *preprocessing* dapat dilihat pada Gambar 4.

	userName	score	at	content	Case_folded	Tokenized	Stemmed	No_Stop
0	Sri Maulinda	5	2022-02-13 12:11:03	Sangat membantu	sangat membantu	[sangat, membantu]	[sangat, bantu]	[bantu]
1	dydi irwandi	5	2022-02-13 06:40:49	singkat, padat, cepat..	singkat padat cepat	[singkat, padat, cepat]	[singkat, padat, cepat]	[singkat, padat, cepat]
2	Sabrina chanel AJANI	1	2022-02-13 02:03:06	Wah parah sih aplikasi ga bisa refund gopay	wah parah sih aplikasi ga bisa refund gopay ap...	[wah, parah, sih, aplikasi, ga, bisa, refund, ...]	[wah, parah, sih, aplikasi, ga, bisa, refund, ...]	[parah, sih, aplikasi, ga, refund, gopay, apli...
3	dj. Weed	1	2022-02-13 00:38:34	Ga perlu bintang harusnya, Pelayanan ga Profes...	ga perlu bintang harusnya pelayanan ga profesi...	[ga, perlu, bintang, harusnya, pelayanan, ga, ...]	[ga, perlu, bintang, harus, layan, ga, profesi...	[ga, bintang, layan, ga, profesional, sesuai, ...]
4	Khoiri N.A	1	2022-02-13 00:00:15	Baru pertama kali nyoba. Ehhh malah error. Uda...	baru pertama kali nyoba ehheh malah error udah ...	[baru, pertama, kali, nyoba, ehheh, malah, erro...]	[baru, pertama, kali, nyoba, ehheh, malah, erro...]	[kali, nyoba, ehheh, error, udah, transfer, gim...]

Gambar 4. Hasil *Preprocessing*

B. Pembagian data training dan data testing

Data training digunakan untuk membentuk model klasifikasi. Model ini merupakan representasi *knowledge* yang akan digunakan untuk prediksi kelas data baru. Sedangkan data testing digunakan untuk mengukur performa dari model yang telah didapatkan.

Perbandingan data training sebesar 70% dan data testing 30%

Tabel 2. Perbandingan Data Training dan Data Testing 1

Jenis Data	Presentase	Jumlah Data
Data Training	70%	1.357,30 \approx 1.357
Data Testing	30%	581,7 \approx 582

Perbandingan data training sebesar 80% dan data testing 20%

Tabel 3. Perbandingan Data Training dan Data Testing 2

Jenis Data	Presentase	Jumlah Data
Data Training	80%	1551,2 \approx 1.151
Data Testing	20%	387,8 \approx 388

Perbandingan data training sebesar 90% dan data testing 10%

Tabel 4. Perbandingan Data Training dan Data Testing 3

Jenis Data	Presentase	Jumlah Data
Data Training	90%	1.754,1 \approx 1.754
Data Testing	10%	193,9 \approx 194

C. Penerapan Algoritma Support Vector Machine

Pada metode SVM, terdapat beberapa kernel yang digunakan, seperti kernel *Linear*, *Polynomial*, *Radial basis function* (RBF), dan *Sigmoid*. Dalam hal ini pengujian akan menggunakan perbandingan dataset 80% : 20% yang masing-masing akan dilakukan pengujian dengan dataset yang sama. Dari setiap kernel akan dicari akurasi terbaik yang nantinya akan digunakan.

Berikut merupakan hasil perbandingan dari ke empat kernel yang telah dilakukan pengujian:

Tabel 5. Perbandingan Kernel

Kernel	Akurasi
Linear	72,76%
Polynomial	70,16%
<i>Radial basis function</i> (RBF)	78,98%
Sigmoid	68,13%

Pada Tabel 2. dapat diketahui bahwa kernel *Radial basis function* (RBF) memiliki akurasi tertinggi dibandingkan kernel yang lainnya. Oleh karena itu, untuk proses klasifikasi selanjutnya akan digunakan kernel RBF dalam proses klasifikasi metode SVM. Proses klasifikasi dilakukan dengan membuat *confusion matrix* untuk mengetahui tingkat akurasi, recall, dan presisi. Matriks ini digunakan untuk mengevaluasi performansi model yang dibentuk oleh setiap algoritma klasifikasi. Dalam melakukan evaluasi model, dilakukan lima kali percobaan dataset untuk mendapatkan nilai akurasi terbaik.

Tabel 3. Perbandingan Akurasi SVM

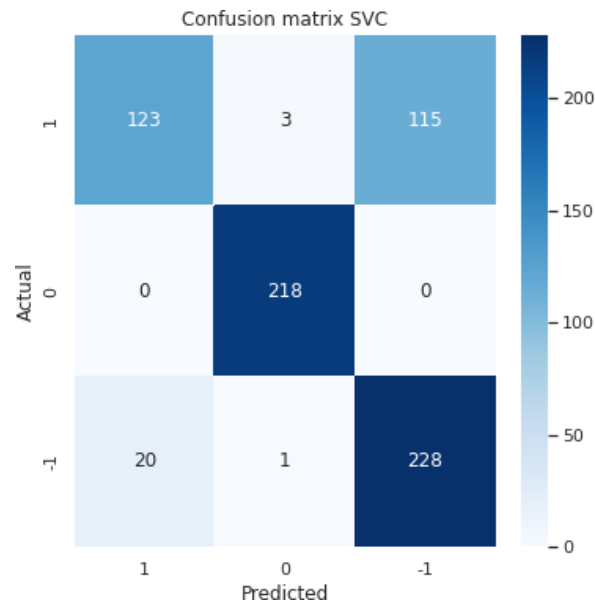
Percobaan	Akurasi
Perbandingan 70% : 30%	
Percobaan 1	78,62%
Percobaan 2	80,32%
Percobaan 3	81,16%
Percobaan 4	79,00%
Percobaan 5	80,41%
Perbandingan 80% : 20%	
Percobaan 1	80,79%
Percobaan 2	80,22%
Percobaan 3	80,08%
Percobaan 4	79,37%
Percobaan 5	79,80%
Perbandingan 90% : 10%	
Percobaan 1	80,79%
Percobaan 2	81,63%
Percobaan 3	79,09%
Percobaan 4	83,61%
Percobaan 5	81,92%
Rata-rata total	82,20%

Berdasarkan Tabel 3. pengujian metode SVM dalam beberapa percobaan menunjukkan akurasi yang cukup tinggi yakni dengan rata-rata total sebesar 82,20%.

D. Tahapan Pengujian

Evaluasi dilakukan untuk melihat tolak ukur terhadap kinerja algoritma *Support vector machine* (SVM) dalam melakukan analisis sentimen. Metode evaluasi dilakukan dengan *Confusion matrix* yang akan menarik

kesimpulan mengenai nilai *precision*, *f-score*, *recall* terhadap tiap kelas sentiment serta nilai keseluruhan *accuracy*. Hasil *confusion matrix* pada perbandingan 80%:20% dapat dilihat pada Gambar 5.



Gambar 5. Hasil *Confusion Matrix*

Evaluasi dilakukan dengan menghitung *classification report* berikut :

Kalkulasi *Precision Sentimen*:

$$\text{Precision Sentimen Positif (1)} = \frac{TP}{TP+FP} = \frac{123}{123+20} = 0,8601$$

$$\text{Precision Sentimen Netral (0)} = \frac{TP}{TP+FP} = \frac{218}{218+4} = 0,9819$$

$$\text{Precision Sentimen Negatif (-1)} = \frac{TP}{TP+FP} = \frac{228}{228+115} = 0,6647$$

Kalkulasi *Recall Sentimen*:

$$\text{Recall Sentimen Positif (1)} = \frac{TP}{TP+FN} = \frac{123}{123+118} = 0,5103$$

$$\text{Recall Sentimen Netral (0)} = \frac{TP}{TP+FN} = \frac{218}{218+0} = 1,00$$

$$\text{Recall Sentimen Negatif (-1)} = \frac{TP}{TP+FN} = \frac{228}{228+21} = 0,9156$$

Kalkulasi *F1-Score Sentimen*:

$$\text{F1 - Score Sentimen Positif (1)} = 2 \times \frac{\text{Precision (1)} \times \text{Recall (1)}}{\text{Precision (1)} + \text{Recall (1)}} = 2 \times \frac{0,8601 \times 0,5103}{0,8601 + 0,5103} = 0,6405$$

$$\text{F1 - Score Sentimen Netral (0)} = 2 \times \frac{\text{Precision (0)} \times \text{Recall (0)}}{\text{Precision (0)} + \text{Recall (0)}} = 2 \times \frac{0,9819 \times 1,00}{0,9819 + 1,00} = 0,9908$$

$$\text{F1 - Score Sentimen Negatif (-1)} = 2 \times \frac{\text{Precision (-1)} \times \text{Recall (-1)}}{\text{Precision (-1)} + \text{Recall (-1)}} = 2 \times \frac{0,6647 \times 0,9156}{0,6647 + 0,9156} = 0,7702$$

Kalkulasi akurasi keseluruhan sentimen :

$$\frac{\text{jumlah data yang diprediksi benar}}{\text{total data}} \times 100\% = \frac{123+218+228}{2.075} \times 100\% = \frac{569}{708} \times 100\% = 80,36$$

IV. KESIMPULAN

Berdasarkan hasil evaluasi, penggunaan kernel *Radial Basic Function* (RBF) pada penelitian kali ini memberikan akurasi tertinggi dibandingkan dengan kernel lain. Sehingga penerapan algoritma *Support Vector Machine* dengan total data sebanyak 2.000 data mampu memberikan hasil akurasi yang cukup tinggi yaitu mencapai rata-rata total

82,20% dengan pengujian sebanyak lima kali percobaan. Data training dan data testing juga dibuat bervariasi untuk mendapatkan hasil akurasi tertinggi. Selain itu dilakukan juga pengujian menggunakan *confusion matrix* yang menghasilkan akurasi sebesar 80,36%.

UCAPAN TERIMA KASIH

Penulis mengucapkan terimakasih yang sebesar besarnya kepada pihak-pihak yang berkontribusi dalam penelitian ini. Terimakasih kami sampaikan kepada Laboratorium Informatika Fakultas Sains dan Teknologi Universitas Muhammadiyah Sidoarjo dan rekan mahasiswa Universitas Muhammadiyah Sidoarjo

REFERENSI

- [1] E. Gunawan, G. O. Sebastian, and A. Harianto, "Analisa Pengaruh Kualitas Layanan Terhadap Kepuasan pelanggan Menginap di Empat Virtual Hotel Operator di Surabaya," *Journal of Indonesian Tourism, Hospitality and Recreation*, vol. 2, no. 2, pp. 145–153, 2019.
- [2] F. Kusumawati, "Tren Virtual Hotel Operator (Vho) Di Yogyakarta (Studi Kasus Hotel OYO)," *Media Wisata*, vol. 18, no. 8, pp. 90–100, 2020, doi: 10.36275/mws.
- [3] 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 of Informatics Journal*, vol. 4, no. 2, p. 41, Nov. 2019, doi: 10.21111/fij.v4i2.3573.
- [4] R. Feldman and J. Sanger, *The text mining handbook : advanced approaches in analyzing unstructured data*. Cambridge University Press, 2007.
- [5] S. Sanjaya and E. A. Absar, "Pengelompokan Dokumen Menggunakan Winnowing Fingerprint dengan Metode K-Nearest Neighbour," 2015.
- [6] D. K. Barupal and O. Fiehn, "Generating the blood exposome database using a comprehensive *text mining* and database fusion approach," *Environ Health Perspect*, vol. 127, no. 9, Sep. 2019, doi: 10.1289/EHP4713.
- [7] Jerryl Jeovano, "2D Data Visualization Tools Menggunakan Flask dan AngularJS," 2020.
- [8] H. Leidiyana, "Penerapan Algoritma K-Nearest Neighbor Untuk Penentuan Resiko Kredit Kepemilikan Kendaraan Bermotor," 2013.
- [9] B. Sartono and U. D. Syafitri, "Metode Pohon Gabungan: Solusi Pilihan untuk Mengatasi Kelemahan Pohon Regresi dan Klasifikasi Tunggal," *Forum Statistika dan Komputasi*, vol. 15, no. 1, 2010.
- [10] G. A. Buntoro, "Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter," 2016. [Online]. Available: <https://t.co/jrvaMsgBdH>
- [11] M. Windarti, "Perbandingan Kinerja Algoritma Naive Bayes dan Bayesian Network dalam Klasifikasi Masa Studi Mahasiswa," 2018.
- [12] E. N. Arrofiqoh and H. Harintaka, "Implementasi Metode Convolutional Neural Network untuk Klasifikasi Tanaman pada Citra Resolusi Tinggi," *GEOMATIKA*, vol. 24, no. 2, p. 61, Nov. 2018, doi: 10.24895/jig.2018.24-2.810.