

SENTIMENT ANALYSIS OF OYO APP REVIEWS USING THE SUPPORT VECTOR MACHINE ALGORITHM (Case Study: Play Store Comments/Reviews)

ANALISIS SENTIMEN TERHADAP ULASAN APLIKASI OYO MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE (Studi Kasus: Komentar/Ulasan Play Store)

Zaenal¹⁾, Ika Ratna Indra Astutik²⁾

¹⁾Program Studi Teknik Informatika, Universitas Muhammadiyah Sidoarjo, Indonesia

²⁾ Program Studi Teknik Informatika, Universitas Muhammadiyah Sidoarjo, Indonesia

Email Penulis Korespondensi: ikaratna@umsida.ac.id

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 90:10 the accuracy obtained is 90.96%. While testing using the Confusion Matrix produces an accuracy of 89.83%.

Keywords - OYO,, Playstore, Support Vector Machine, RBF

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 90:10 hasil akurasi yang didapatkan adalah sebesar 90,96%. Sedangkan pengujian dengan menggunakan Confusion Matrix menghasilkan akurasi sebesar 89,83%.

Kata Kunci - OYO, Playstore, Support Vector Machine, RBF

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]. Kusumawati, F. (2021). Tren Virtual Hotel Operator (VHO) di Yogyakarta. Media Wisata, 18(1), 90–100. <https://doi.org/10.36276/mws.v18i1.80>

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 teoristruktural pembelajaran statistik untuk menemukanbatas yang memisahkan antar kelas [3].

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

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].

Text Preprocessing

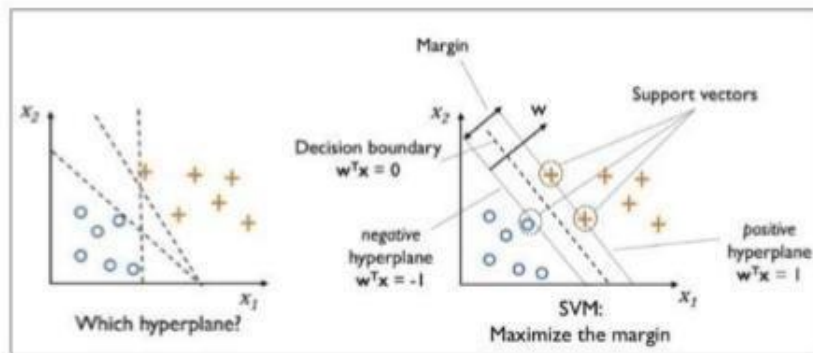
Text preprocessing adalah tahap awal dari text mining. Tahap ini mencakup semua rutinitas, dan proses untuk mempersiapkan data yang akan digunakan pada operasi knowledge discovery sistem text mining [4]. Beberapa metode yang dilakukan untuk menyederhanakan dan mengurangi kata serta karakter special yang 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].

1. 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.
2. 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.
3. 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.
4. Stemming
Pada tahap stemming, pada data review, imbuhan pada kata-kata akan dihilangkan. Contohnya adalah kata “dihilangkan” diubah menjadi “hilang”.

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 kedimensilebih 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 ekivalen dengan memaksimalkan margin atau jarak dua set dari kelas yang berbeda.



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

Python

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

1. Python merupakan bahasa pemrograman tujuan umum yang sangatbtingkat 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 strukturalbdan berorientasi objek.
4. Python sangat fleksibel, karena kemampuannya untuk menggunakan komponenbmodular yang dirancang dalam bahasabpemrograman 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 pelengkapdbiduat 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 mesinbstate-of-the-art. Paket ini berfokus pada membawa pembelajaranm esin ke non-spesialis menggunakanbbahasa tingkat tinggi tujuan umum[6]
2. Pandas, yaitu sebuah package yang tersedia dalam Python untuk mengolah struktur data yang cepat,bfleksibel, dan ekspresif yang dirancang untuk mempermudah pekerjaan yang melibatkan data "relasional" atau "berlabel" [7].

Klasifikasi

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

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 sentiment [10]

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 confusion matrix dalam klasifikasi biner ditunjukkan pada tabel berikut:

Table 1. Confusion Matrix

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

Dirumuskan sebagai berikut :

$$\text{Akurasi} = \frac{TP+TN}{TP+FP+FN+TN} \times 100\%$$

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

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

$$\text{F1 Score} = \frac{2 \times (\text{Recall} \times \text{Presisi})}{(\text{Recall} + \text{Presisi})}$$

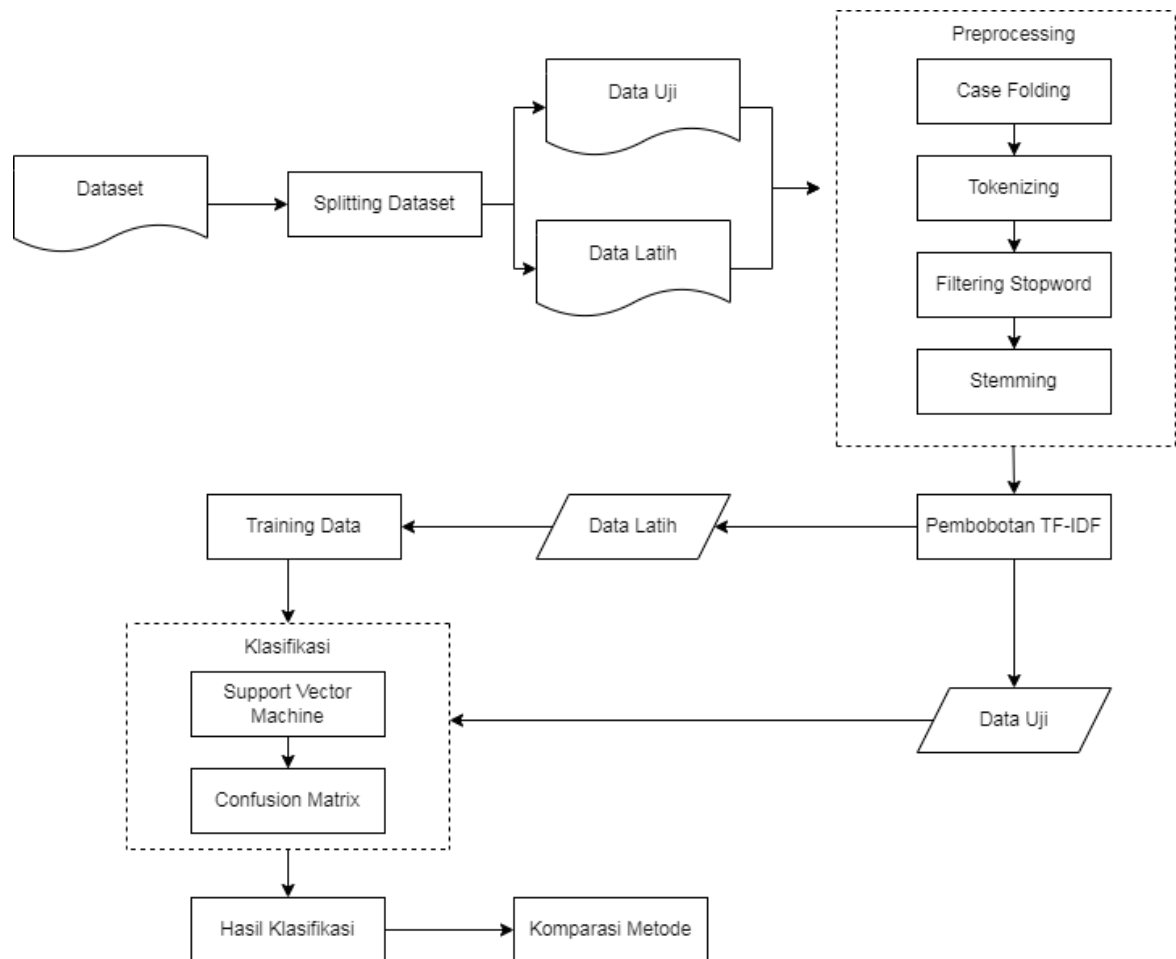
Keterangan :

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

II. METODE

Flowchart

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



Gambar 2. Flowchart Arsitektur Umum Penelitian

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 kedalam 3 kelas sentimen yang dapat dilihat pada tabel 2. Kemudian penulis melabeli data tersebut menggunakan bantuan dari Pakar (Dosen Bahasa Indonesia Murni Universitas Mayjend Sungkono, Bapak Irham Miftakhul Jannah S.Pd., M.Pd.). Jumlah data yang dilabeli pakar berjumlah 200 data yang akan di gunakan sebagai data latih untuk melabeli data-data yang belum ter-label. Dataset disimpan dalam format .xlsx yang nantinya akan diinputkan kedalam program. Berikut total data yang telah dilabeli :

Tabel 2. Ulasan Berdasarkan Sentimen

No.	Sentimen	Total
1.	Positif	134
2.	Netral	10
3.	Negatif	83
Total		200

Berikut contoh dataset yang digunakan pada penelitian ini :

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	userName	score	at	content									
2	cup-cep. c	5	#####	bagusssss									
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 referensi 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 saaaat 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 sani	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 bntuan teman									
18	wildan kal	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 Dataset

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 semakn baik [12].

III. HASIL DAN PEMBAHASAN

Preprocessing Data

Proses diawali dengan tahap preprocessing karena data tidak sepenuhnya menggunakan kata baku. Tahap preprocessing dilakukan dengan menggunakan bantuan library pada bahasa pemrograman Python3. 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 chenel 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 ehhh malah error udah ...	[baru, pertama, kali, nyoba, ehhh, malah, erro...]	[baru, pertama, kali, nyoba, ehhh, malah, erro...]	[kali, nyoba, ehhh, error, udah, transfer, gim...]

Gambar 4. Hasil Preprocessing

Penerapan Algoritma Support Vector Machine

Pada metode SVM, terdapat beberapa kernel yang digunakan, seperti kernel Linear, Polynomial, Radial Basis Function (RBF), dan Sigmoid. Dari setiap kernel akan dicari akurasi terbaik yang nantinya akan digunakan. Pada penelitian kali ini akan dilakukan pembagian dataset sebanyak 3 kali yang selanjutnya akan diterapkan metode SVM pada tiap kernelnya. Perbandingan dataset dapat dilihat pada tabel berikut :

- a. Perbandingan data *training* dan data *testing* sebesar 70% : 30% :

Tabel 3. Perbandingan 70% : 30% pada Kernel SVM

Kernel	Akurasi
Linear	79,19%
Polynomial	62,61%
Radial Basis Function (RBF)	90,11%
Sigmoid	77,11%

- b. Perbandingan data *training* dan data *testing* sebesar 80% : 20% :

Tabel 4. Perbandingan 80% : 20% pada Kernel SVM

Kernel	Akurasi
Linear	80,08%
Polynomial	63,13%
Radial Basis Function (RBF)	90,25%
Sigmoid	77,11%

- c. Perbandingan data *training* dan data *testing* sebesar 90% : 10% :

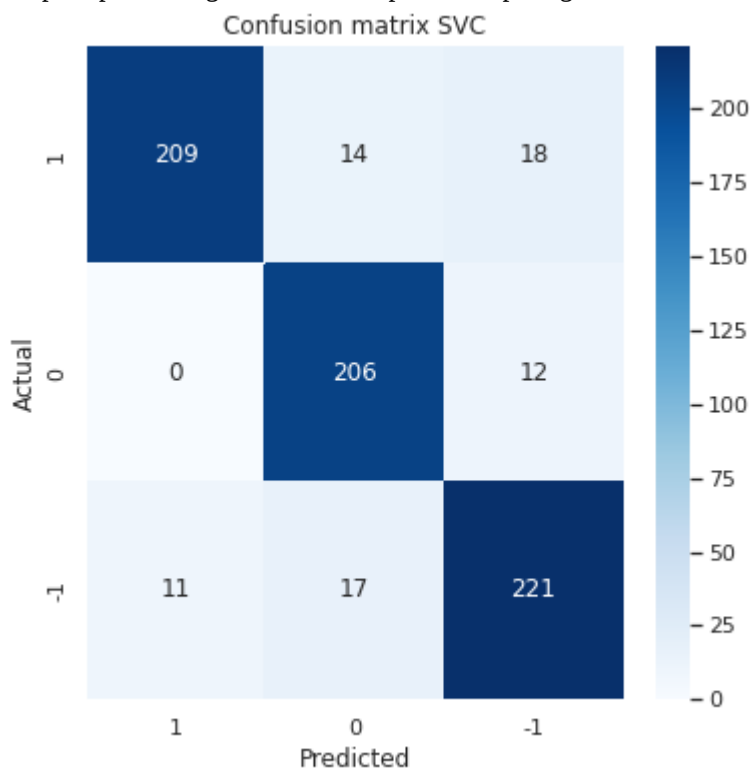
Tabel 5. Perbandingan 90% : 10% pada Kernel SVM

Kernel	Akurasi
Linear	81,92%
Polynomial	65,81%
Radial Basis Function (RBF)	90,96%
Sigmoid	79,37%

Pada Tabel diatas 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. Berdasarkan tabel 5. pengujian metode SVM dalam beberapa percobaan menunjukkan akurasi yang paling tinggi di perbandingan 90% : 10% yakni dengan akurasi sebesar 90,96 %.

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 6.

Gambar 5. Hasil *Confusion Matrix* SVM

```
[374] print(metrics.classification_report(y_test, y_pred_svc))
```

	precision	recall	f1-score	support
-1	0.87	0.44	0.59	362
0	1.00	1.00	1.00	356
1	0.61	0.93	0.74	344
accuracy			0.79	1062
macro avg	0.83	0.79	0.77	1062
weighted avg	0.83	0.79	0.77	1062

Gambar 6. Hasil *Confusion Matrix*

Berdasarkan hasil analisis sentimen terhadap data meme pada tabel 4.6, 4.7, dan 4.8, evaluasi dilakukan dengan menghitung *classification report* berikut :

Kalkulasi Precision Sentimen:

$$\begin{aligned}
 \text{Precision Sentimen Positif (1)} &= \frac{TP}{TP+FP} = \frac{209}{209+32} = 0.95 \\
 \text{Precision Sentimen Netral (0)} &= \frac{TP}{TP+FP} = \frac{206}{206+31} = 0.8691 \\
 \text{Precision Sentimen Negatif (-1)} &= \frac{TP}{TP+FP} = \frac{221}{221+30} = 0.8804
 \end{aligned}$$

Kalkulasi Recall Sentimen:

$$\begin{aligned}
 \text{Recall Sentimen Positif (1)} &= \frac{TP}{TP+FN} = \frac{209}{123+118} = 0.8672 \\
 \text{Recall Sentimen Netral (0)} &= \frac{TP}{TP+FN} = \frac{206}{206+12} = 0.9449 \\
 \text{Recall Sentimen Negatif (-1)} &= \frac{TP}{TP+FN} = \frac{221}{221+28} = 0.8875
 \end{aligned}$$

Kalkulasi F1-Score Sentimen:

$$\begin{aligned}
 \text{F1-Score Sentimen Positif (1)} &= 2 \times \frac{\text{Precision 1} \times \text{Recall 1}}{\text{Precision 1} + \text{Recall 1}} = 2 \times \frac{0.95 \times 0.8672}{0.95 + 0.8672} = 0.9068 \\
 \text{F1-Score Sentimen Netral (0)} &= 2 \times \frac{\text{Precision 0} \times \text{Recall 0}}{\text{Precision 0} + \text{Recall 0}} = 2 \times \frac{0.8691 \times 0.9449}{0.8691 + 0.9449} = 0.9054 \\
 \text{F1-Score Sentimen Negatif (-1)} &= 2 \times \frac{\text{Precision -1} \times \text{Recall -1}}{\text{Precision -1} + \text{Recall -1}} = 2 \times \frac{0.8804 \times 0.8875}{0.8804 + 0.8875} = 0.8838
 \end{aligned}$$

Kalkulasi akurasi keseluruhan sentimen :

$$\frac{\text{Jumlah data yang diprediksi benar}}{\text{total data}} \times 100\% = \frac{206 + 209 + 221}{708} \times 100\% = \frac{636}{708} = 0.8983$$

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 Machin dengan total data sebanyak 2.000 data mampu memberikan hasil akurasi yang cukup tinggi yaitu mencapai 90,96%. Selain itu dilakukan juga pengujian menggunakan Confusion Matrix yang menghasilkan akurasi sebesar 89,83%.

UCAPAN TERIMA KASIH

Penulis mengucapkan terimakasih yang sebesar besarnya kepada pihak-pihak yang berkontribusi dalam penelitian ini. Terimakasih kami sampaikan kepada :

1. Laboratorium Informatika Fakultas Sains dan Teknologi Universitas Muhammadiyah Sidoarjo.
2. Dosen Bahasa Indonesia Universitas Mayjen Sungkono
3. Rekan mahasiswa Universitas Muhammadiyah Sidoarjo

REFERENSI

- [1] E. Gunawan, G. O. Sebastian, and A. Harianto, "ANALISA PENGARUH KUALITAS LAYANAN TERHADAP KEPUASANPELANGGAN 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 BEMOTOR," 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.