

ANALISIS SENTIMEN TERHADAP ULASAN APLIKASI OYO MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE

Oleh:

Zaenal,

Ika Ratna Indra Astutik

Program Studi Informatika

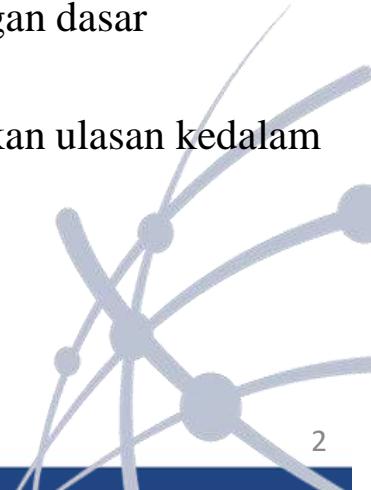
Fakultas Sains & Teknologi

Universitas Muhammadiyah Sidoarjo

Februari, 2023

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 teorivstruktural pembelajaran statistik untuk menemukanbbatas yang memisahkan antar kelas [3].
- Penelitian ini bertujuan untuk mengukur tingkat akurasi yang dihasilkan oleh algoritma SVM dalam mengklasifikasikan ulasan kedalam beberapa kelas sentimen.



Pertanyaan Penelitian (Rumusan Masalah)

Berdasarkan latar belakang, bahwa masalah adalah sebagai berikut :

1. Bagaimana penerapan metode *Support Vector Machine* dalam mengklasifikasikan suatu ulasan sebagai ulasan positif, negatif dan netral?
2. Berapa tingkat akurasi metode *Support Vector Machine* untuk mengklasifikasikan ulasan pengguna aplikasi OYO?



Metode

01

02

03

04

Pengumpulan Data

Data diperoleh melalui metode web scrapping pada ulasan aplikasi OYO terdapat di google playstore. Data dilabeli berdasarkan rating yang diberikan oleh pengguna

Preprocesssing

Case Folding
Tokenizing
Filtering Stopword
Stemming

Pembobotan Term

Pembobotan term menggunakan metode TFIDF (Term Frequency-Inverse Document Frequency)

Algoritma SVM

Implementasi algoritma SVM MultiClass

Hasil

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.

	userNmae	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...

Hasil

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 :

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%

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%

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%

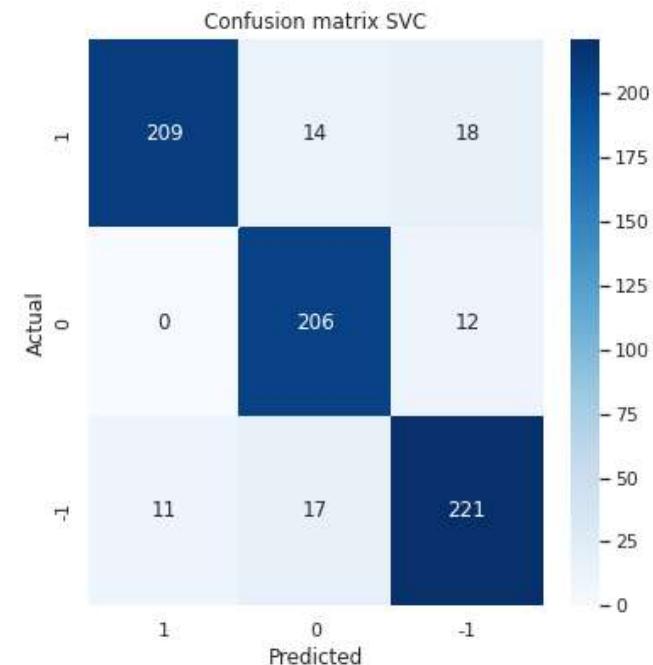
Pada Tabel diatas dapat diketahui bahwa kernel Radial Basis Function (RBF) memiliki akurasi tertinggi dibandingkan kernel yang lainnya. Oleh karena itu, untuk proses klasifikasi selanjutnya akan digunakan kernel RBF dalam proses klasifikasi metode SVM. Proses klasifikasi dilakukan dengan membuat confusion 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 %.



Hasil

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.



```
[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

Pembahasan

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:

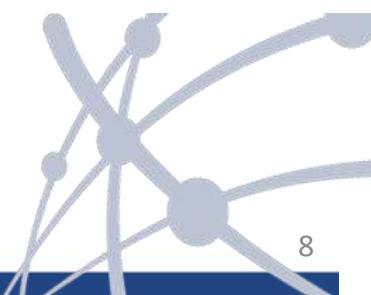
$$\begin{aligned} \text{Recall Sentimen Positif (1)} &= \frac{TP}{TP+FN} = \frac{209}{123+118} = 0.8672 \\ \text{Recall Sentimen Netral (0)} &= \frac{TP}{TP+FN} = \frac{206}{206+12} = 0.9449 \\ \text{Recall Sentimen Negatif (-1)} &= \frac{TP}{TP+FN} = \frac{221}{221+28} = 0.8875 \end{aligned}$$

Kalkulasi F1-Score Sentimen:

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

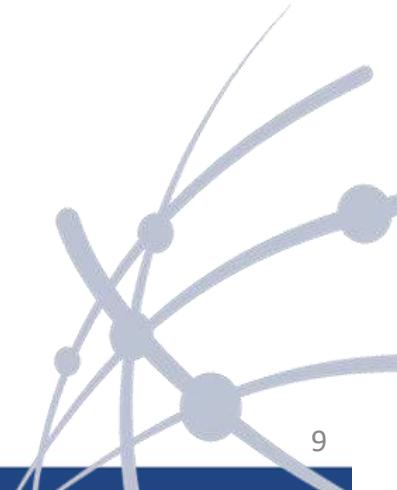
Kalkulasi akurasi keseluruhan sentimen :

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



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%.



Manfaat Penelitian

Manfaat penelitian adalah sebagai berikut:

a. Bagi Penulis

1. Untuk memenuhi salah satu syarat kelulusan stara satu (S1) Program Studi Teknik Informatika Fakultas Sains dan Teknologi.
2. Menerapkan ilmu-ilmu yang diperoleh selama perkuliahan

b. Bagi Universitas

1. Mengetahui kemampuan mahasiswa dalam menguasai materi yang telah diperoleh selama masa kuliah.
2. Mengetahui kemampuan mahasiswa dalam menerapkan ilmunya dan sebagai bahan evaluasi.

c. Bagi Pembaca

1. Menambah wawasan pembaca mengenai algoritma *Support Vector Machine* dalam melakukan analisis sentimen terhadap ulasan suatu aplikasi.

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 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.

