

Komparasi Algoritma Naïve Bayes Dan Support Vector Machine (SVM) Pada Analisis Sentimen Spotify

Ayu Sri Rahayu*, Ahmad Fauzi, Rahmat

Fakultas Ilmu Komputer, Program Studi Teknik Informatika, Universitas Buana Perjuangan Karawang, Karawang, Indonesia

Email: ^{1,*}if19.ayurahayu@mhs.ubpkarawang.ac.id, ²afauzi@ubpkarawang.ac.id, ³rahmat@ubpkarawang.ac.id

Email Penulis Korespondensi: : if19.ayurahayu@mhs.ubpkarawang.ac.id

Submitted: 23/12/2022; Accepted: 29/12/2022; Published: 31/12/2022

Abstrak-Aplikasi Spotify merupakan subjek yang menarik bagi komunitas jejaring sosial yang memiliki perbedaan pendapat atau sentimen yang signifikan. Analisis Sentimen adalah solusi untuk secara otomatis mengelompokkan opini atau peringkat menjadi opini negatif atau positif. Teknik yang digunakan dalam penelitian ini adalah Support Vector Machines (SVM) dan Naïve Bayes. Keunggulan Naïve Bayes adalah sederhana, cepat dan akurasi tinggi. SVM, di sisi lain, dapat mengidentifikasi hyperplane berbeda yang memaksimalkan margin antara dua kelas yang berbeda. Hasil klasifikasi dari penelitian ini mempunyai dua label kategori yaitu negatif dan positif. Nilai akurasi yang dihasilkan menunjukkan model uji terbaik untuk kasus klasifikasi sentimen. Keakuratan diukur dengan matriks konfusi dan hasilnya menunjukkan bahwa nilai akurasi algoritma SVM adalah 84% sedangkan nilai akurasi algoritma Naïve Bayes lebih tinggi dari SVM yaitu 86,4%.

Kata Kunci: SVM; Naïve Bayes; Analisis Sentimen

Abstract-The Spotify app is a subject of interest to social networking communities with significant disagreements or sentiments. Sentiment Analysis is a solution to automatically categorize opinions or ratings into negative or positive opinions. The techniques used in this research are Support Vector Machines (SVM) and Naïve Bayes. The advantages of Naïve Bayes are simple, fast and high accuracy. SVM, on the other hand, can identify different hyperplanes that maximize the margin between two different classes. The classification results of this study have two category labels, namely negative and positive. The resulting accuracy value indicates the best test model for sentiment classification cases. Accuracy is measured by the confusion matrix and the results show that the accuracy value of the SVM algorithm is 84% while the accuracy value of the Naïve Bayes algorithm is higher than SVM which is 86.4%.

Keywords: SVM; Naïve Bayes; Sentiment Analysis

1. PENDAHULUAN

Perkembangan teknologi semakin hari semakin maju, salah satunya pada bidang teknologi musik yang digemari seluruh kalangan. Musik sering digunakan sebagai media untuk mewakili emosi atau perasaan pendengarnya. Hal itu sering dijadikan sebagai peluang bisnis bagi industri musik. Saat ini, banyak orang telah berhenti menggunakan radio, TV, CD, DVD dan MP3 dikarenakan lebih mudah menggunakan smartphone. Spotify merupakan teknologi musik dengan layanan streaming yang memberikan layanan pengguna akses lebih mudah ke dalam informasi kapan saja dan di mana saja [1]. Dengan pesatnya perkembangan pada smartphone dan aplikasi dalam beberapa tahun terakhir, pengguna kini dapat berkomentar di berbagai platform melalui layanan internet seluler, media sosial, dll [2]. Play Store adalah layanan penyedia konten digital dari Google yang menawarkan berbagai toko online seperti aplikasi, game, film, musik. Play Store dapat diakses melalui website, aplikasi Android atau Google TV. Play Store memiliki beberapa fitur termasuk peringkat pengguna dan peringkat untuk aplikasi atau layanan. Ulasan adalah kalimat atau teks berisi komentar atau penilaian tentang karya seseorang. Rating ini sangat penting dijadikan sebagai tolak ukur untuk aplikasi apakah dapat di rekomendasikan atau tidak untuk pengguna baru. [3]

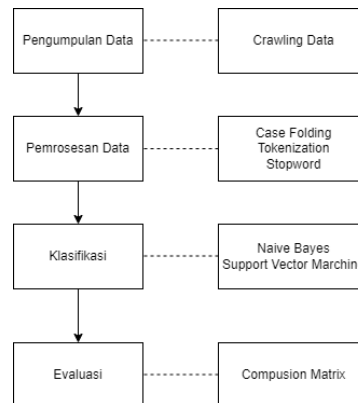
Musik menjadi subjek yang menarik bagi banyak orang di komunitas jejaring sosial, musik membuat perbedaan besar dalam opini dan sentimen bagi pendengarnya. Opini ulasan musik diukur lebih menantang dari pada opini dari kategori ulasan lainnya, seperti ulasan produk [4]. Analisis sentimen adalah studi komputer tentang pendapat, perilaku dan perasaan manusia tentang keseluruhan. Entitas ini dapat perasaan manusia tentang keseluruhan [5]. Penelitian terdahulu terkait sentimen dilakukan oleh [3] tentang analisis sentimen ulasan aplikasi Spotify untuk meningkatkan layanan menggunakan algoritma Naïve Bayes dengan menggunakan data sebanyak 3600 yang mendapatkan hasil nilai sentimen negatifnya sebesar 46,4% dan positifnya sebesar 53,6%. Penelitian Selanjutnya oleh [6] mengenai analisis prediksi mood genre musik pop menggunakan algoritma C4.5 dan K-means. mendapatkan nilai mood terbaik sebesar 91,9%. Penelitian selanjutnya dilakukan oleh [7] tentang komparasi algoritma klasifikasi genre musik pada Spotify menggunakan CRISP-DM dengan mengumpulkan total sebesar 42.305. hasil tertinggi nilai akurasi didapatkan oleh algoritma klasifikasi Naïve Bayes yaitu sebesar 58,91%. Kemudian penelitian analisis klasifikasi artist music menggunakan model regresi logistik biner dan analisis diskriminan oleh [8] Model regresi logistik biner yang dibangun, menggunakan nilai 1-APER dengan proporsi data 90:10 memiliki ketepatan klasifikasi untuk data testing sebesar 90,00% dan training adalah sebesar 88,00%. Penelitian selanjutnya, yaitu perbandingan metode regresi logistik dan random forest untuk klasifikasi fitur mode audio Spotify oleh [9] dengan menghasilkan nilai akurasi random forest sebesar 71%.

Berdasarkan uraian di atas dan beberapa penelitian terdahulu dapat disimpulkan maka metode yang menghasilkan akurasi terbaik adalah metode Naive Bayes. Namun pada penelitian ini, penulis menggunakan dua

algoritma yaitu SVM dan Naïve Bayes dalam menyelesaikan permasalahan pengelompokan opini menjadi positif dan negatif secara otomatis pada ulasan aplikasi Spotify. Penelitian ini dilakukan untuk membandingkan tingkat akurasi terbaik pada algoritma SVM dan Naive Bayes untuk meningkatkan layanan pada aplikasi Spotify.

2. METODOLOGI PENELITIAN

Dalam penelitian ini, penulis menggunakan SVM dan Naïve Bayes untuk mencapai akurasi tertinggi dalam analisis label. Opini tentang ulasan pengguna aplikasi Spotify di Play Store. Berikut adalah proses yang dilakukan untuk mencapai nilai akurasi terbaik.



Gambar 1. Alur Penelitian

2.1 Pengumpulan Data

Pengumpulan data yang diperoleh dalam penelitian ini dari ulasan pengguna aplikasi Spotify di Play Store. Pengumpulan data dilakukan dengan melakukan scrapping menggunakan google collab. Terdapat empat informasi yang diperoleh dari hasil scrapping dari playstore yaitu username, tanggal dan waktu, content, dan skor yang diberikan dari penulis ulasan.

2.2 Pemrosesan Data

Langkah selanjutnya adalah menyiapkan data yang diperoleh untuk diproses ke dalam klasifikasi. Tahap preprocessing atau pengolahan data, meliputi kegiatan yang mengubah teks dan membersihkan data agar siap untuk dikelola ke klasifikasi. Langkah-langkah pemrosesan data tercantum di bawah ini.

- 1) Case Folding : mengubah semua data dari huruf besar menjadi kecil atau sebaliknya.
- 2) Tokenization : Proses tokenisasi pada data teks terdiri dari memecah rangkaian kalimat menjadi bagian kata-kata, yang sering disebut suku token, sesuai kebutuhan untuk menjadi kata-kata dengan makna tertentu
- 3) Filter Stopwords : menghilangkan kata-kata yang tidak masuk akal berupa konjungsi.

2.3 Klasifikasi

Setelah data diproses maka tahap selanjutnya data akan di klasifikasikan menjadi label negatif dan positif untuk ulasan pengguna Spotify, menentukan klasifikasi pendapat atau pandangan dari hasil pesan singkat yang telah di kumpulkan. Pada tahap ini melakukan analisis data berdasarkan algoritma Naïve Bayes dan SVM. Total data yang di ambil sekitar 1500 lalu data akan dibagi dua bagian di antaranya data testing dan data training menggunakan algoritma Naïve Bayes dan SVM untuk mendapatkan nilai akurasi terbaik. Klasifikasi Naïve Bayes adalah klasifikasi yang bersifat supervised learning karena memiliki supervisor (orang yang secara manual mengklasifikasikan data yang digunakan dalam pelatihan) sebagai guru dalam proses pembelajaran. Kinerja Naïve Bayes mempunyai waktu klasifikasi yang singkat sehingga percepat dalam memproses sistem analisis sentimen. Berikut adalah rumus untuk menghitung Naive Bayes [10] :

$$P(H|X) = \frac{P(H|X) P(H)}{P(X)} \quad (1)$$

Keterangan :

X = Data kelas tidak diketahui

H = Data hipotesis X adalah kelas yang terpisah

P(H|X) = Probabilitas hipotesis H berdasarkan kondisi (posteriori probability)

P(H) = Probabilitas hipotesis H

P(X|H) = Probabilitas X berdasarkan kondisi pada hipotesis

P(X) = Probabilitas X

Meskipun SVM memiliki keunggulan kemampuan generalisasi yang baik dan dapat menghasilkan klasifikasi yang baik dan diuji dengan dataset yang relatif kecil pada pengaturan parameter yang sederhana, SVM mempunyai konsep dan formulasi yang relatif jelas dengan beberapa parameter yang dapat disesuaikan dengan mudah,relatif mudah diimplementasikan karena penentuan SVM bisa dirumuskan ke dalam QP (quadratic programming).Berikut adalah rumus perhitungan SVM. Konsep pengoperasian SVM adalah mencari hyperplane atau garis pemisah optimal untuk memisahkan dua kelas. Dalam memperoleh hyperline pada SVM, rumus beberapa kernel dapat dilihat pada Tabel berikut ini:

Tabel 1. Rumus Kernel

Kernel	Persamaan
Linear	$K(x_i, x) = x^T x$
Polynomial	$K(x_i, x) = (y \cdot x^T x + r)^p, y > 0$
RBF	$K(x_i, x) = \exp(-\gamma x_i - x ^2), \gamma > 0$
Sigmoid	$K(x_i, x) = \tanh(\gamma x^T x + r)$

2.5 Evaluasi

Setelah tahap klasifikasi selesai, selanjutnya evaluasi hasil pemodelan dengan membandingkan dua hasil algoritma SVM dan Naïve Bayess untuk nilai akurasi. Tahap evaluasi ini adalah tentang menemukan hasil terbaik dari hasil uji. Pengujian dilakukan untuk mengukur akurasi hasil Pengukuran akurasi model menggunakan confusion matrix berikut pengukuran evaluasi menggunakan confusion matrix disajikan pada Tabel 2.

Tabel 2. Rumus Confusion Matrix

Data Aktual	TRUE	Data Prediksi		Total
		TRUE	FALSE	
TRUE	TP	FN	P	
FALSE	FP	TN	N	
Total	P'	N'	P+N	

Keterangan:

TP (True Positive) = Data positif diklasifikasikan secara benar.

TN (True Negative) = Data negatif diklasifikasikan secara benar.

FP (False Positive) = Data negatif diklasifikasikan menjadi positif.

FN (False Negative) = Data positif diklasifikasikan menjadi negatif.

3. HASIL DAN PEMBAHASAN

3.1 Pengambilan Data

Pengumpulan data untuk penelitian ini dilakukan dari ulasan pengguna aplikasi Spotify di Play store menggunakan teknik web scraping Google Colab, dan terkumpul 1.500 ulasan.

	userName	score	at	content
0	No Vi	5	2022-11-18 15:23:07	Bgussss
1	rahmat nasution	5	2022-11-18 15:21:16	Mudah digunakan.
2	Raihan Syuhada Firansyah	1	2022-11-18 15:13:09	Updatean terbaru gk jelas to! 🙄
3	ahmad sururi	5	2022-11-18 14:58:54	ok
4	intan permata sari	5	2022-11-18 14:52:24	Bagus bgt
...
1495	Giritami Store	5	2022-11-08 07:53:51	Ok
1496	Pricillia Humaira	5	2022-11-08 07:41:11	Saya suka sekali karena setiap malam saya itu ...
1497	Hapid Awalul	3	2022-11-08 07:24:30	eeere see aeerre eh hh rwrcc VII g64442##\$\$_
1498	Ramdani	5	2022-11-08 06:30:49	Bagus
1499	Boy Boy	4	2022-11-08 06:00:52	Bgus

1500 rows x 4 columns

Gambar 2. Isi Dataset

3.2 Pemrosesan Data

Setelah mengumpulkan data dari Playstore,langkah berikutnya yaitu persiapan data yang diperoleh dan diolah untuk diklasifikasikan. Tahap pemrosesan data ini mencakup kegiatan data dan juga membersihkan data sehingga siap untuk langkah selanjutnya yaitu klasifikasi. Dalam tahap pemrosesan data dibagi menjadi beberapa langkah: Case Folding, Tokenizing, Stopword.

3.2.1 Case Folding

Pada Tahap case folding ini data akan di proses untuk mengubah seluruh huruf menjadi huruf kecil.

Tabel 3. Hasil Case Folding

Before	After
Mudah Digunakan	mudah digunakan

3.2.2 Tokenizing

Pada tahap Tokenizing data ini berguna untuk memisahkan kata dalam suatu kalimat dengan tujuan untuk di analisis lebih lanjut. Proses awal dalam melakukan tokenizing yaitu menghapus URL, menghilangkan simbol emoji dan menghapus double spasi dalam teks.

Tabel 4. Hasil Tokenizing

Before	After
Updatenya bagus banget ☺	updatenya bagus banget

3.2.3 Stopword

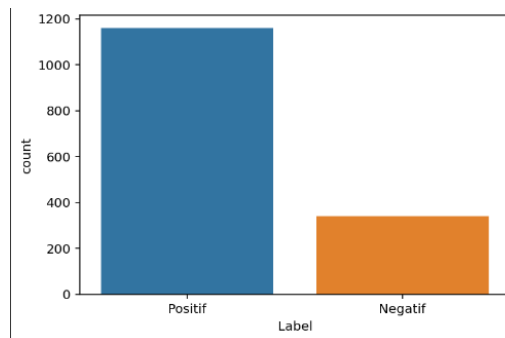
Stop list ini berisi daftar kata-kata umum yang memiliki fungsi tetapi tidak memiliki arti. Sehingga stopwords ini digunakan untuk memisahkan kata yang tidak penting di dalam data.

Tabel 5. Hasil Stopword

guna gess akankah amat joss gg

3.3 Klasifikasi

Pada tahap proses klasifikasi data yang di ambil sebanyak 1500 ditentukan ke dalam 2 kelas yaitu negatif dan positif jumlah positif sebanyak 1159 dan negatif 341. Proses klasifikasi menggunakan Platform Google Colab dengan bahasa pemrograman Python. Kemudian Data dibagi menjadi data testing dan data training, jumlah data training sebanyak 1050 dan jumlah data testing 450 menggunakan ratio 0.3 dengan 70% data dipakai untuk training dan 30% data untuk testing.



Gambar 3. Hasil Klasifikasi

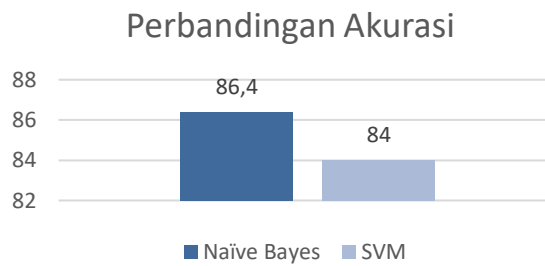
3.4 Evaluasi

Hasil akurasi klasifikasi teks opini menggunakan algoritma Naive Bayes dan SVM disajikan pada tabel berikut:

Tabel 6. Hasil Perbandingan Tingkat Performa

No	Algoritma	Akurasi (%)	Precision (%)	Recall (%)
1	Naïve Bayes	86.4	0.69	0.75
2	SVM	84.0	0.71	0.51
	Selisih	0.4	0.02	4.3

Berdasarkan tabel 5 diatas dapat disimpulkan bahwa nilai performansi akurasi model algoritma Naive Bayes dengan nilai akurasi 86% memiliki nilai akurasi yang lebih baik dengan selisih 2,4% dibandingkan dengan algoritma SVM dengan nilai lebih rendah 84% dalam konfigurasi dataset dengan 1500 dataset. Grafik yang membandingkan nilai akurasi ditunjukkan pada Gambar 4.



Gambar 4. Hasil Perbandingan Akurasi

Berdasarkan percobaan, hasil pengujian 100 data dengan algoritma Naive Bayes pada tabel Confusion Matrix ditunjukkan pada tabel berikut.

Tabel 7. Confusion Matrix Untuk Algoritma Naive Bayes

Accuracy 86.4%		
	TP	TN
Prediksi Positif	312	35
Prediksi Negatif	26	77

Berdasarkan tabel diatas terlihat bahwa jumlah TP sebanyak 312 opini dan FN sebanyak 26 opini. Berikut 77 pernyataan TN dan 35 FP Hasilnya disajikan pada Tabel 8.

Tabel 8. Hasil Perhitungan Naive Bayes

Dalam Persen (%)	
Accuracy	86.00
Sensitivity	92.00
Specificity	68.00
Ppv	87.00
Npv	74.00

Berikut adalah tabel Confusion Matrix hasil pengujian data testing menggunakan SVM.

Tabel 9. Confusion Matrix Untuk Algoritma SVM

Accuracy 84%		
	TP	TN
Prediksi Positif	325	22
Prediksi Negatif	50	53

Berdasarkan tabel di atas terlihat bahwa jumlah TP sebanyak 325 ulasan dan FN sebanyak 50 opini. 53 opini berikut adalah TN dan 22 FP. Hasilnya disajikan pada Tabel 10.

Tabel 10. Hasil Perhitungan SVM

Dalam Persen (%)	
Accuracy	72.00
Sensitivity	86.00
Specificity	7.00
Ppv	39.00
Npv	51.00

Setelah rating Spotify review dibagi menjadi kategori opini negatif dan positif sehingga ditampilkan nilai akurasinya, maka dapat diuji tingkat akurasinya untuk melihat performa dari hasil pengujian diatas. membuktikan bahwa nilai akurasi algoritma Naive Bayes lebih tinggi dibandingkan dengan algoritma SVM.

4. KESIMPULAN

Komparasi algoritma Naive bayes dan SVM mendapatkan hasil terbaik dengan algoritma Naive Bayes dikarenakan pada algoritma Naive Bayes melakukan proses pencarian bobot dengan melakukan perhitungan probabilitas pada data dan mendapatkan hasil nilai akurasi sebesar 86.4% sedangkan pada algoritma Support Vector Machine (SVM) memiliki nilai akurasi sebesar 84% jumlah data yang di gunakan sebanyak 1500 data Spotify. Data training yang digunakan pada penelitian ini sebesar 1050 sedangkan data testing sebesar 450 data.

Dapat disimpulkan bahwa Naïve Bayes memberikan unjuk kerja lebih baik daripada Support Vector Machine (SVM) untuk klasifikasi sentimen review Spotify dengan menghasilkan nilai akurasi yang tinggi. Algoritma Naïve Bayes merupakan model pengujian algoritma yang terbaik dan akurat untuk permasalahan klasifikasi sentimen review Spotify.

REFERENCES

- [1] R. Kusumah, M. Ariyanti dan S. , “ANALIS PERBANDINGAN POSITIONING APLIKASI MUSIK DIGITAL,” e-Proceeding of Management, p. 2511, 2017.
- [2] L. Zhang, K. Hua, . Wang dan G. Qian, “Sentiment Analysis on Reviews of Mobile Users,” *Procedia Computer Science*, p. 458 – 465, 2014.
- [3] M. D. Rhajendra dan N. Trianasar, “Analisis Sentimen Ulasan Aplikasi Spotify Untuk Peningkatan Layanan Menggunakan,” e-Proceeding of Management, vol. 8, p. 4367, 2021.
- [4] A. S. H. Basari, G. Pramudya dan B. Hussin, “Opinion Mining of Movie Review using Hybrid Method of Support,” *Procedia Engineering*, p. 453 – 462, 2013.
- [5] W. Medhat, A. Hassan dan H. Korashy, “Sentiment analysis algorithms and applications:,” *ELECTRICAL ENGINEERING*, 2014.
- [6] L. Nurhalimah, T. I. Hermanto dan I. Kaniawulan, “Analisis Prediksi Mood Genre Musik Pop Menggunakan Algoritma,” *JURIKOM (Jurnal Riset Komputer)*, vol. 9, p. 1006–1013, 2022.
- [7] S. Navisa, L. Hakim dan A. Nabilah, “Komparasi Algoritma Klasifikasi Genre Musik,” *Jurnal Sistem Cerdas*, vol. 04, pp. 114 - 125, 2021.
- [8] A. T. Rian Dani, V. Ratnasari, L. Ni’matuzzahroh, I. . C. Aviantholib, R. Novidianto dan N. Y. Adrianingsih, “ANALISIS KLASIFIKASI ARTIST MUSIC MENGGUNAKAN MODEL,” *JAMBURA JOURNAL OF PROBABILITY AND STATISTICS*, vol. 3, 2022.
- [9] L. B. Christina Tanujayaa, B. Susanto dan A. Saragiha, “Perbandingan Metode Regresi Logistik dan Random Forest,” *Indonesian Journal of Data and Science (IJODAS)*, vol. 3, pp. 68-78, 2020.
- [10] S. Diantika, W. Gata dan H. Nalatissifa, “Komparasi Algoritma SVM Dan Naive Bayes Untuk Klasifikasi,” *JURNAL ILMIAH ELEKTRONIKA DAN KOMPUTER*, vol. 14, p. 10, 2021.