



Analisa Perbandingan Algoritma Support Vector Machine dan K-Nearest Neighbors Terhadap Ulasan Aplikasi Vidio

Rizki Bintang Gumilar*, Yana Cahyana, Cici Emilia Sukmawati, Amril Mutoi Siregar

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

Jl. HS.Ronggo Waluyo, Puseurjaya, Telukjambe Timur, Karawang, Jawa Barat, Indonesia

Email: ¹*if20.rizkigumilar@mhs.ubpkarawang.ac.id, ²yana.cahyana@ubpkarawang.ac.id, ³cici.emilia@ubpkarawang.ac.id,

⁴amrilmutoi@ubpkarawang.ac.id

Email Penulis Korespondensi: if20.rizkigumilar@mhs.ubpkarawang.ac.id

Submitted: 20/07/2024; Accepted: 27/07/2024; Published: 27/07/2024

Abstrak—Penggunaan internet di Indonesia mencapai 77% dari total populasi pada Januari 2023, dengan layanan Over The Top (OTT) menunjukkan pertumbuhan pengguna sebesar 25% setiap tahun. Aplikasi Vidio, salah satu platform OTT populer dengan pengunduhan melebihi 50 juta, memiliki rating bintang 3,5 berdasarkan 649 ribu ulasan di Google Play Store. Meskipun populer, Vidio menghadapi keluhan terkait pilihan film yang terbatas, eror pembayaran, dan iklan yang berlebihan, sehingga mempengaruhi kepuasan pengguna. Riset ini bertujuan menganalisis opini komentar user aplikasi Vidio dengan menggunakan metode SVM (Support Vector Machine) dan metode KNN (K-Nearest Neighbors) guna menentukan model dengan akurasi terbaik. Data ulasan sebanyak 15.000 dikumpulkan melalui scraping, kemudian diproses dengan teknik text preprocessing dan vektorisasi TF-IDF. Evaluasi model menunjukkan bahwa SVM memiliki hasil dari akurasi 82%, nilai precision 82%, nilai recall 83%, dan nilai F1-score 82%, sedangkan KNN memiliki akurasi 69%, precision 74%, recall 73%, dan F1-score 69%. Hasil penelitian menunjukkan bahwa SVM lebih unggul dibandingkan KNN dalam klasifikasi sentimen ulasan aplikasi Vidio. Temuan ini diharapkan dapat digunakan oleh developer aplikasi dalam upaya untuk meningkatkan layanan dan kepuasan dari para pengguna aplikasi Vidio.

Kata Kunci: Sentimen; Vidio; Support Vector Machine (SVM); K-Nearest Neighbors (KNN); Pengguna Internet Indonesia; Kepuasan Pengguna

Abstract—Internet usage in Indonesia reached 77% of the total population in January 2023, with Over The Top (OTT) services showing user growth of 25% every year. The Vidio application, one of the popular OTT platforms with downloads exceeding 50 million, has a 3.5 star rating based on 649 thousand reviews on the Google Play Store. Despite its popularity, Vidio faces complaints regarding limited film selection, payment errors, and excessive advertising, which affects user satisfaction. This research aims to analyze the opinions of Vidio application user comments by applying the SVM (Support Vector Machine) method and the KNN (K-Nearest Neighbors) method to determine the model with the best accuracy. 15,000 review data were collected through scraping, then processed using text preprocessing and TF-IDF vectorization techniques. Model evaluation shows that SVM has an accuracy value of 82%, a precision value of 82%, a recall value of 83%, and an F1-score value of 82%, while KNN has an accuracy of 69%, precision 74%, recall 73%, and F1-score 69%. The research results show that SVM is superior to KNN in classifying the sentiment of Vidio application reviews. It is hoped that these findings can be used by application developers in an effort to improve service and satisfaction of Vidio application users.

Keywords: Sentiment; Vidio; Support Vector Machine (SVM); K-Nearest Neighbors (KNN); Indonesian Internet Users; User Satisfaction

1. PENDAHULUAN

Pengguna internet di Indonesia pada bulan Januari 2023 mencapai 77% dari total populasi 212,9 juta. Layanan media Over The Top (OTT), yang merupakan platform untuk berbagai konten televisi dan film melalui internet, pada saat ini mengalami peningkatan setiap tahunnya. Menurut laporan, masyarakat Indonesia cenderung lebih aktif dalam menonton (OTT), saat ini satu dari tiga masyarakat Indonesia menonton OTT streaming. Dampaknya bisa terlihat pada pertumbuhan pengguna OTT yang meningkat sebesar 25% di Indonesia setiap tahunnya[1].

Saat ini, tidak sedikit aplikasi hiburan yang menyediakan layanan streaming yang mudah diakses. Keanekaragaman aplikasi hiburan ini membuat pengguna lebih berhati-hati dalam menentukan pilihan mereka, mencari aplikasi mana yang lebih cocok ataupun lebih unggul untuk digunakan [2]. Keterlibatan ulasan yang tersedia di halaman unduhan aplikasi bisa mempengaruhi keputusan pengguna. Ulasan tersebut sering kali menjadi faktor utama yang dipertimbangkan dalam menilai mutu aplikasi hiburan [3]. Jika ulasan-ulasan ini dikumpulkan dan dianalisis secara cermat, maka hasilnya bisa membantu dan membentuk kesimpulan mengenai aplikasi hiburan mana yang mempunyai reputasi ataupun tanggapan pengguna terbaik [2].

Aplikasi Vidio ataupun Vidio.com, merupakan salah satu aplikasi hiburan untuk streaming film Indonesia dan streaming pertandingan sepak bola, namun untuk menikmati streaming sepenuhnya, pengguna dapat berlangganan terlebih dahulu. Secara keseluruhan, aplikasi ini mendapatkan rating bintang sebesar 3,5 berdasarkan pendapat pengguna pada google play store pada tahun 2023. Lebih dari 50 juta unduhan dan telah menerima 649 ribu ulasan, Vidio menjadi pilihan populer di kalangan pengguna. Dengan beberapa keunggulan yang ditawarkan oleh aplikasi Vidio, namun masih terdapat beberapa permasalahan, seperti pilihan film yang terbatas, terjadi eror saat pembayaran member, iklan yang terlalu banyak masalah-masalah ini akan mempengaruhi kepuasan terhadap pengguna. [4].

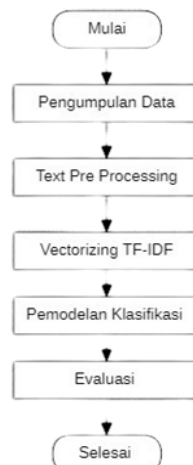
Dari usulan di atas maka dilaksanakan proses analisis sentimen, analisis sentimen adalah komponen text mining yang dimaksudkan untuk menemukan pendapat dan kecenderungan pendapat seseorang atau kelompok tentang suatu objek atau masalah. [5]. Melalui analisis sentimen kita dapat memprediksi potensi hasil peluang terkait dengan data tersebut[6]. Selain itu, penelitian digunakan sebagai referensi untuk perusahaan yang membuat keputusan tentang perbaikan aplikasi dengan mengutamakan kebutuhan dan harapan pengguna. Sehingga penelitian ini memiliki dampak konsisten pada peningkatan kualitas layanan aplikasi, kepuasan pelanggan, dan persaingan perusahaan di pasar yang sangat kompetitif. [7].

Penelitian sebelumnya Irma Surya Kumala Idris pada tahun (2023) Penelitian ini fokus pada Analisis Sentimen Berbasis Aspek (ABSA), yang mengklasifikasikan dokumen teks berdasarkan berbagai aspek seperti pendapat, emosi, dan evaluasi. Solusi yang diambil dengan menggunakan algoritma Support Vector Machine (SVM) untuk menganalisis sentimen pada aplikasi Shopee. Algoritma ini dipilih untuk secara efektif mengklasifikasikan komentar pengguna. hasil dari pengambilan data sebanyak 3000 sudah terbukti bisa mendapatkan kinerja yang baik melalui hasil akurasi 98% [8]. Kemudian studi Antonius Mbay Ndapamuri pada tahun (2023) masalah dari penelitian ini mencari tahu apakah umpan balik pengguna cenderung positif atau negative solusi yang diambil penelitian ini menggunakan algoritma pembelajaran mesin (SVM, K-NN, Naive Bayes) dan alat Orange untuk pemodelan klasifikasi untuk mencapai hasil analisis sentimen yang akurat hasil studi ini berupa dataset yang terkumpul sebanyak 1000 hasil SVM akurasi yaitu 89,8% serta KNN 89,02% [9]. pada penelitian selanjutnya dari Piqih Aditya pada tahun (2022) Analisis ulasan aplikasi MyIM3 diperlukan untuk mengetahui pendapat masyarakat tentang aplikasi tersebut karena semakin banyak pengguna yang menggunakannya untuk memberikan komentar atau ulasan. Sehingga perlu dilakukan analisis sentimen menggunakan SVM dengan jumlah data 3484. Hasil dari algoritma SVM mendapatkan 87% akurasi [10].

Berdasarkan latar belakang di atas, maka masalah penelitian dapat dirumuskan sebagai berikut.: perbandingan antara dua algoritma dalam melaksanakan analisis sentimen terhadap ulasan pengguna pada suatu aplikasi, serta memprediksi algoritma mana yang lebih unggul tingkat akurasi.[11] Pada proses klasifikasi menggunakan metode perbandingan algoritma, yaitu metode Support Vector Machine (SVM) dan metode K-Nearest Neighbors (KNN) untuk memberikan pemahaman lebih mendalam terkait performa dan akurasi klasifikasi sentimen pada ulasan aplikasi Vidio[12]. Alasan penggunaan SVM dipilih karena algoritma ini telah terbukti memberikan hasil klasifikasi yang akurat dan handal dalam berbagai kasus, menjadikannya pilihan yang kuat untuk menganalisis dan prediksi[13]. Dipilihnya K-Nearest Neighbor (KNN) sebagai metode untuk menganalisis opini komentar pengguna aplikasi Vidio karena KNN cocok, mempunyai kapasitas untuk mengolah dataset besar dengan tingkat kinerja komputasi tinggi, serta dikenal mempunyai akurasi dan performa yang baik, sambil tetap mempertahankan sifat yang sederhana dan mudah dipelajari [14]. Diharapkan bahwa temuan penelitian ini akan memberikan informasi mengenai evaluasi sentimen pengguna aplikasi Vidio, yang nantinya dapat dipergunakan sebagai referensi oleh penelitian selanjutnya maupun pengembang aplikasi Vidio dalam upaya meningkatkan kualitas pelayanan mereka [15].

2. METODOLOGI PENELITIAN

Pada proses penelitian ini dilakukan melalui beberapa prosedur dari pengumpulan data melalui proses scraping. Langkah berikutnya melibatkan pre-processing teks, yang mencakup case folding, tokenizing, filtering, dan stemming. Data ini kemudian dilakukan proses pembobotan menggunakan vectorizing TF-IDF. Kemudian klasifikasi dengan menerapkan algoritma SVM dan KNN, dan kemudian diikuti oleh tahap evaluasi. Detail langkah-langkah studi bisa ditemukan di Gambar 2 seperti berikut.



Gambar 1. Flowcart penelitian



2.1 Pengumpulan Data

Dataset diambil di bulan Januari 2024 merupakan komentar dari bulan maret 2023 sampai bulan januari 2024 berasal dari penilaian pemakai aplikasi Vidio yang diunduh melalui Google Play Store. Proses pengumpulan data melibatkan analisis ulasan yang terbaru serta relevan dengan tujuan penelitian. Total data yang berhasil diperoleh dari Google Play Store mencapai 15.000, yang kemudian digunakan sebagai basis analisis guna menggali pandangan serta pengalaman pengguna terkait aplikasi Vidio. Metode yang digunakan untuk pengumpulan data yaitu scraping, merupakan metode yang menggunakan program komputer untuk mengumpulkan data dari website secara otomatis, pada penelitian ini teknik scraping di gunakan untuk pengumpulan data komentar pada web Google paly store.

2.2 Text Pre Processing

Text preprocessing yaitu serangkaian langkah untuk membersihkan dan mengorganisir teks sebelum dianalisis, tujuannya yaitu membuat teks lebih mudah diolah, meningkatkan kualitas analisis, dan kinerja model. Berikut tahapan yang di pakai pada penelitian ini :

- a. Case Folding
merupakan langkah yang melibatkan transformasi huruf ataupun kata dalam suatu ulasan menjadi bentuk huruf kecil ataupun huruf besar.
- b. Tokenizing
merupakan langkah pemrosesan yang melibatkan pemecahan setiap kata dari suatu kalimat, sehingga kata-kata tersebut dipisahkan menjadi urutan yang terpisah. Contohnya: “Hari ini cuacanya cerah sekali” menjadi (“Hari”, “ini”, “cuacanya”, “cerah”, “sekali”]
- c. Filtering
Yaitu tindakan yang dipakai untuk menghilangkan kata-kata yang dianggap tidak sesuai dengan analisa sentimen. Pada konteks studi ini, metode stopwords dipergunakan guna mengeliminasi kata-kata yang diasumsikan tidak sesuai berdasarkan data yang sedang diolah.
- d. Stemming
yaitu tahap di mana kata-kata diubah menjadi bentuk dasar ataupun dihapus imbuhan. Pada konteks ini, library sastra dipergunakan sebagai panduan aturan bahasa Indonesia untuk memastikan bahwasannya proses stemming mengikuti ketentuan yang terdapat dalam library tersebut.

2.3 Vectorizing TF-IDF

Mengkonversi data teks ke data numerik guna memberi nilai ke setiap teks ataupun fitur. Tingkat kepentingan kata dalam dokumen ditentukan dengan ukuran statistik TF-IDF. Nilai TF yaitu kebalikan dari nilai DF, sedangkan komponen DF pada pembobotan TF-IDF mengindikasikan seberapa sering suatu kata muncul dalam dokumen yang diberikan, sementara nilai IDF yaitu kebalikan dari nilai DF. Ketika kata mempunyai frekuensi kemunculan yang tinggi dalam suatu dokumen, nilai TF-IDF-nya akan naik karena perkalian nilai TF dengan IDF. Ini berarti bahwasannya kata yang sering terlihat pada dokumen tersebut mempunyai bobot besar dalam analisis. Kebalikannya, apabila kata tersebut terlihat pada banyak dokumen, nilai IDF-nya akan rendah, mengindikasikan bahwasannya kata tersebut kurang signifikan dalam membedakan dokumen satu dengan yang lain. Dengan demikian, pembobotan TF-IDF membantu mengidentifikasi kata-kata kunci yang paling relevan dan mewakili dokumen secara lebih akurat. [16]

$$Wt, d = tft, d \times idft = tft, d \times \log N df \frac{N}{dft} \quad (1)$$

Pada Penjelsan rumus Skor Tf-Idf (Wt, d), Jumlah Kata yang muncul (tft, d), hasil yang berlawanan dengan frekuensi dokumen untuk setiap kata ($idft$), Seberapa sering setiap kata muncul dalam dokumen (dft), Hasil keseluruhan dokumen (N). Pada proses ini membaca data teks dari file Excel, kemudian menggunakan TF-IDF untuk menghasilkan representasi numerik dari teks tersebut. Hasilnya adalah sebuah DataFrame yang berisi nilai-nilai TF-IDF guna masing-masing kata dalam teks beserta labelnya.[17]

2.4 Pemodelan Klasifikasi

a. Support Vector machine (SVM)

Algoritma Support Vector Machine (SVM) yaitu Algoritma pembelajaran yang cukup populer karena keunggulannya yang beragam. Salah satu keunggulan utamanya yaitu dasar teoritis yang kuat yang menjadi landasan bagi keandalan serta konsistensinya dalam memberikan hasil klasifikasi yang akurat. Selain itu, SVM juga populer dalam menangani berbagai jenis data serta masalah klasifikasi yang kompleks, membuatnya menjadi salah satu algoritma yang paling populer dalam berbagai aplikasi di berbagai bidang, mulai dari pengenalan pola hingga analisis data besar.. Pada proses ini menghitung sejauh mana kualitas algoritma SVM.

b. K-Nearest Neighbor (KNN)

Metode ini mengklasifikasi objek menurut training data dengan mempertimbangkan tingkat kesamaan ataupun jarak terdekat antara objek-objek dalam ruang fitur. Algoritma ini secara khusus hanya mempertahankan



klasifikasi dari data latihan selama proses pengajaran serta vektor-vektor fitur, tanpa mempertimbangkan informasi lain yang mungkin tersedia terkait dengan objek ataupun variabel lain yang tidak relevan dengan proses klasifikasi. Untuk data uji yang belum diklasifikasikan, fitur yang sama dihitung pada fase klasifikasi. Dihitunglah jarak antara vektor data latihan dan vektor baru ini, dan sejumlah K yang paling dekat dipilih. Sebagian besar klasifikasi titik tersebut menentukan klasifikasi baru ini [18]. Dalam proses pemodelan klasifikasi SVM dan KNN, kinerja model diukur selama pembacaan data untuk mengevaluasi seberapa baik model dapat mengklasifikasikan data.

2.5 Evaluasi

Tujuan terakhir dari bagian ini dari penelitian yaitu untuk mengevaluasi sebgas dan efektif apa model klasifikasi yang telah digunakan pada proses sebelumnya. Oleh karena itu, evaluasi ini bertujuan untuk meningkatkan pemahaman tentang seberapa baik dan efektif model tersebut digunakan dalam praktik klasifikasi. Evaluasi ini penting untuk menentukan apakah model yang dikembangkan dapat memberikan hasil yang cukup relevan[9]. Confusion Matrix yaitu alat bantu yang bisa mengindikasikan serta dibandingkan nilai nyata ataupun nilai sebenarnya dengan nilai hasil perkiraan model. Ini berguna guna menciptakan matriks penilaian yang memberikan gambaran lebih rinci mengenai performa model dalam klasifikasi objek ke dalam berbagai kategori[[19].

Tabel 1. Confusion Matrix

	Actual Positif	Actual Negatif
Predicted Positif	TP	FP
Predicted Negatif	FN	TN

.Dari tabel di atas, confusion matrix yang berisi True Positif (TP), True Negatif (TN), False Positif (FP) dan False Negatif (FN). TP yaitu persentase dari kelas positif yang sudah diklasifikasikan menjadi kelas positif. Ini berarti model berhasil mengidentifikasi dengan benar kasus-kasus yang benar-benar positif. Di sisi lain, TN yaitu persentase dari kelas negatif yang sudah diklasifikasikan menjadi kelas negatif, yang mengindikasikan bahwasannya model juga dapat membedakan dengan baik kasus-kasus yang sebenarnya negatif. Dengan demikian, TP dan TN memberikan gambaran mengenai kemampuan model dalam mengenali dan membedakan kelas positif serta negatif dengan akurasi yang baik. Dari capaian kalkulasi “Confusion matrix”, misalnya akurasi “accuracy”, presisi “precision”, “recall”, serta “f1-score” , dapat dihitung memakai persamaan yang disajikan.[20]

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} * 100\% \tag{1}$$

$$Precision = \frac{TP}{TP+FP} * 100\% \tag{2}$$

$$Recall = \frac{TP}{TP+FN} * 100\% \tag{3}$$

$$F1-Score = \frac{2*Precision*Recall}{Precision+Recall} * 100\% \tag{4}$$

3. HASIL DAN PEMBAHASAN

Pada Penelitian ini menghasilkan analisis sentimen yang mengklasifikasikan komentar pengguna pada Google PlayStore tentang aplikasi Vidio dengan menggunakan algoritma “Support Vector Machine” (SVM) dan “K-Nearest Neighbor” (KNN). Penelitian ini bertujuan untuk mendapatkan hasil klasifikasi data yang paling akurat dengan metode SVM dan KNN berdasarkan opini pengguna aplikasi Vidio di Google PlayStore.

3.1 Pengumpulan Data

Jumlah keseluruhan data ulasan berhasil didapati melalui proses scraping yaitu sebanyak 15,000, mencakup ulasan pengguna terbaru. Contoh konkret dari data yang berhasil diambil mempergunakan metode scraping juga disertakan dalam kumpulan informasi berikut, yang memberikan gambaran mendalam tentang pendapat dan pengalaman pengguna

Tabel 2. Hasil Pengambilan Dataset

	Content	Score	At
0	Aplikasi apaan ini, udh beli paket tpi percuma...	1	2023-03-09 05:00:31
1	Pidio	5	2023-03-09 05:32:07
2	Andi tangsa	5	2023-03-09 06:07:14
3	Tiga bintang dulu	3	2023-03-09 06:13:43
4	parah sih video ny macet terus jd ganggu pas n...	1	2023-03-09 06:54:08
...
14995	Kenapa sekarang lemot bgt muter2 mluk .	3	2024-01-28 18:46:54

	Content	Score	At
14996	Tolonglah, masa nonton film ngelag ngelag, kek...	1	2024-01-28 20:00:56
14997	Kenapa serial kartun spongebob dan Avatar hila...	1	2024-01-28 23:50:22
14998	Aplikasi tidak berguna dan merugikan	1	2024-01-29 00:50:59
14999	Good	5	2024-01-29 01:20:50

3.2 Text Pre Processing

Pada tahapan proses ini ada empat langkah dalam text-preprocessing yang dijalankan, yakni “case folding”, “tokenizing”, “filtering”, serta “stemming”. Case folding yaitu proses untuk menyederhanakan seluruh teks data ulasan, mengubahnya menjadi huruf kecil, serta tidak menampilkan karakter selain huruf. Hal ini perlu dilaksanakan karena data tidak konsisten dalam penggunaan huruf besar ataupun kecil, serta terdapat angka, tanda baca, dan emotikon, sebagaimana tampak pada Tabel 3.

Tabel 3 . Case Folding

Case Folding	Tokenizing
paketan masih ada tp gabisa dibuka streaming	['paketan', 'masih', 'ada', 'tp', 'gabisa', 'dibuka', 'streaming']

Pada tahap tokenizing, dilakukan proses penguraian kalimat menjadi perkata, sebagaimana bisa diperlihatkan contoh dalam Tabel 4.

Tabel 4. Tokenezing

Tokenizing	Filtering
['paketan', 'masih', 'ada', 'tp', 'gabisa', 'dibuka', 'streaming']	['paketan', 'masih', 'ada', 'gabisa', 'dibuka', 'streaming']

Dalam Tabel 5 terlihat contoh-contoh penggunaan kata konjungsi seperti "dan", "dengan", "yang", serta kata-kata serupa lainnya yang digunakan untuk menghubungkan ide atau informasi dalam teks tersebut.

Tabel 5. Filtering

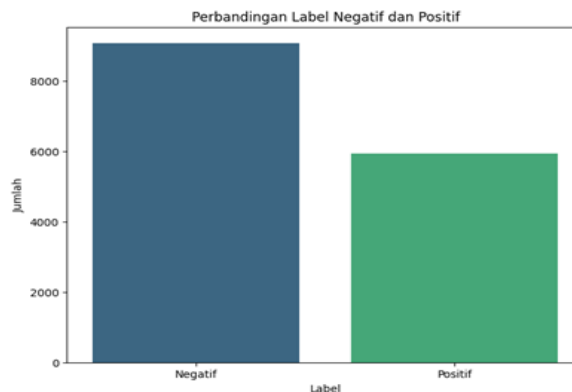
Filtering	Stemming
['paketan', 'masih', 'ada', 'gabisa', 'dibuka', 'streaming']	['paket', 'ada', 'gabisa', 'buka', 'streaming']

Langkah terakhir yaitu melaksanakan stemming, di mana kata-kata diubah menjadi bentuk dasarnya, seperti yang diindikasikan pada contoh dalam Tabel 6

Tabel 6. Stemming

Sebelum	Case Folding
Paketan masih ada tp gabisa dibuka streaming ☺	paketan masih ada tp gabisa dibuka streaming

Pada Gambar 2 mengindikasikan prediksi jumlah data antara kelas positif dan negatif. Dari segi jumlah, terlihat bahwasannya kelas sentimen positif dan negatif tidak seimbang (imbalanced). Karena ketidak seimbangan ini, metode oversampling dipergunakan untuk menyeimbangkan data.. Tujuannya yaitu untuk meningkatkan akurasi saat membangun model.



Gambar 2. Perbandingan Label

3.3 Vectorizing TF-IDF

Vektorisasi data dipergunakan untuk mengubah token kata menjadi representasi numerik, Kegunaannya yaitu untuk memungkinkan algoritma machine learning ataupun analisis data untuk bekerja dengan data teks, yang pada

dasarnya hanya terdiri dari kata-kata, dan mengambil informasi penting dari teks tersebut.berikut hasil dari Vectorizing TF-IDF

Tabel 7 . Vectorizing TF-IDF

Kata	numerik
bagus	966.139589
iklan	728.198171
aplikasi	543.001327
langganan	319.439675
Vidio	283.654405

3.4 Pemodelan Klasifikasi

Setelah proses pembobotan kata selesai, langkah selanjutnya yaitu menerapkan algoritma klasifikasi SVM dan KNN untuk melaksanakan analisis sentimen. Tujuannya yaitu untuk menentukan model mana yang paling optimal untuk tugas tersebut. Di bawah ini pada Gambar 3 dan Gambar 4 yaitu hasil evaluasi dari penggunaan algoritme SVM dan KNN. Evaluasi dilaksanakan dengan mempergunakan indikator “precision”, “recall”, “f1-score”, serta “accuracy”

```

Classification report:
              precision    recall  f1-score   support

   Negatif    0.89         0.80         0.85         2706
   Positif    0.74         0.86         0.79         1794

 accuracy          0.82         0.82         0.82         4500
 macro avg         0.82         0.83         0.82         4500
 weighted avg     0.83         0.82         0.83         4500
    
```

Gambar 3. Klasifikasi SVM

Hasil penilaian algoritma pada Gambar 3 menunjukkan kinerja SVM (Support Vector Machine) yang cukup baik. SVM mencapai akurasi sebesar 82%, yang menampilkan nilai keakuratan model pada proses klasifikasi seluruh data. Selain itu, nilai precision sebesar 82% menunjukkan seberapa banyak hasil positif yang benar-benar positif dari keseluruhan hasil yang diprediksi positif oleh model, sementara recall sebesar 83% menunjukkan seberapa baik model dapat mengidentifikasi data positif secara keseluruhan. F1-score sebesar 82% menunjukkan hasil yang cukup baik antara precision dan recall, menegaskan bahwa SVM dapat memberikan prediksi yang akurat dan konsisten dalam klasifikasi data.

```

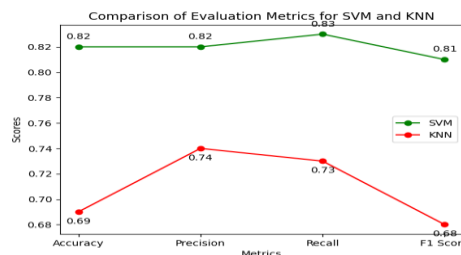
Classification report:
              precision    recall  f1-score   support

   Negatif    0.91         0.53         0.67         2706
   Positif    0.57         0.93         0.70         1794

 accuracy          0.69         0.73         0.69         4500
 macro avg         0.74         0.73         0.69         4500
 weighted avg     0.78         0.69         0.68         4500
    
```

Gambar 4. Klasifikasi KNN

Hasil evaluasi algoritma KNN pada Gambar 4 mengindikasikan bahwasannya algoritma tersebut mempunyai tingkat akurasi sebesar 69%, yang mengindikasikan seberapa baik model dapat mengklasifikasikan data secara benar. Rata-rata nilai precision sebesar 74% mengindikasikan persentase hasil positif yang benar-benar positif dari keseluruhan hasil yang diprediksi positif oleh model, sedangkan recall sebesar 73% mengindikasikan kemampuan model dalam mengidentifikasi data positif secara keseluruhan. F1-score sebesar 69% mencerminkan keseimbangan antara precision dan recall, menegaskan bahwasannya algoritma KNN mempunyai kemampuan yang cukup baik dalam memberikan prediksi yang akurat dan konsisten dalam klasifikasi data. Secara keseluruhan,



Gambar 5. Hasil kedua matrik



- [2] E. Fitri, “Analisis Sentimen Terhadap Aplikasi Ruangguru Menggunakan Algoritma Naive Bayes, Random Forest Dan Support Vector Machine,” *J. Transform.*, vol. 18, no. 1, p. 71, 2020, doi: 10.26623/transformatika.v18i1.2317.
- [3] M. Gharzouli, A. K. Hamama, and Z. Khattabi, “Topic-based sentiment analysis of hotel reviews,” *Curr. Issues Tour.*, vol. 25, no. 9, pp. 1368–1375, 2022, doi: 10.1080/13683500.2021.1940107.
- [4] M. Fudhail Ferio Supeli and S. Setiaji, “Klasifikasi Sentimen Positif Dan Negatif Pada Aplikasi Vidio Dengan Algoritma K-Nearest Neighbor,” *Indones. J. Comput. Sci.*, vol. 2, no. 1, pp. 7–15, 2023, doi: 10.31294/ijcs.v2i1.1874.
- [5] B. Supriyanto, Sumijan, and Yuhandri, “Audit Keamanan Website Menggunakan Acunetix Web Vulnerability (Studi Kasus Di SMK Muhammadiyah 3 Terpadu Pekanbaru),” *J. CoSciTech (Computer Sci. Inf. Technol.)*, vol. 5, no. 1, pp. 134–143, 2024, doi: 10.37859/coscitech.v5i1.6705.
- [6] M. T. Waluyan and K. D. Hartomo, “Analisis Sentimen Kebutuhan Fast Track Pada Originals Vidio Menggunakan Support Vector Machine,” *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 9, no. 3, pp. 2153–2162, 2022, doi: 10.35957/jatisi.v9i3.2348.
- [7] N. Lestari, E. Haerani, and R. M. Candra, “Analisa Sentimen Ulasan Aplikasi Wetv Untuk Peningkatan Layanan Menggunakan Metode Naive Bayes,” *J. Inf. Syst. Res.*, vol. 4, no. 3, pp. 874–882, 2023, doi: 10.47065/josh.v4i3.3355.
- [8] I. S. K. Idris, Y. A. Mustofa, and I. A. Salihi, “Analisis Sentimen Terhadap Penggunaan Aplikasi Shopee Menggunakan Algoritma Support Vector Machine (SVM),” *Jambura J. Electr. Electron. Eng.*, vol. 5, no. 1, pp. 32–35, 2023, doi: 10.37905/jjee.v5i1.16830.
- [9] A. M. Ndapamuri, D. Manongga, and A. Iriani, “Analisis Sentimen Ulasan Aplikasi Tripadvisor Dengan Metode Support Vector Machine, K-Nearest Neighbor, Dan Naive Bayes,” *INOVTEK Polbeng - Seri Inform.*, vol. 8, no. 1, p. 127, 2023, doi: 10.35314/isi.v8i1.3260.
- [10] P. Aditya, U. Enri, and I. Maulana, “Analisis Sentimen Ulasan Pengguna Aplikasi Myim3 Pada Situs Google Play Menggunakan Support Vector Machine,” *JURIKOM (Jurnal Ris. Komputer)*, vol. 9, no. 4, p. 1020, 2022, doi: 10.30865/jurikom.v9i4.4673.
- [11] F. A. Larasati, D. E. Ratnawati, and B. T. Hanggara, “Analisis Sentimen Ulasan Aplikasi Dana dengan Metode Random Forest,” ... *Tekno. Inf. dan ...*, vol. 6, no. 9, pp. 4305–4313, 2022, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [12] T. Tinaliah and T. Elizabeth, “Analisis Sentimen Ulasan Aplikasi PrimaKu Menggunakan Metode Support Vector Machine,” *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 9, no. 4, pp. 3436–3442, 2022, doi: 10.35957/jatisi.v9i4.3586.
- [13] S. Fide, S. Suparti, and S. Sudarno, “Analisis Sentimen Ulasan Aplikasi Tiktok Di Google Play Menggunakan Metode Support Vector Machine (Svm) Dan Asosiasi,” *J. Gaussian*, vol. 10, no. 3, pp. 346–358, 2021, doi: 10.14710/j.gauss.v10i3.32786.
- [14] D. Asfi Warrailan, I. Permama, R. Novita, and A. Marsal, “Analisis Sentimen Pengguna Transportasi Online Maxim Pada Instagram Menggunakan Naive Bayes Classifier dan K-Nearest Neighbor,” *J. Media Inform. Budidarma*, vol. 7, no. 3, pp. 1134–1143, 2023, doi: 10.30865/mib.v7i3.6336.
- [15] A. A. Arifiyanti, N. R. Shantika, and A. O. Syafira, “Analisis Sentimen Ulasan Pengguna Bsi Mobile Pada Google Play Dengan Pendekatan Supervised Learning,” *J. Inform. Polinema*, vol. 9, no. 3, pp. 283–288, 2023, doi: 10.33795/jip.v9i3.1003.
- [16] J. A. Septian, T. M. Fachrudin, and A. Nugroho, “Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepkabolaan Indonesia Menggunakan Pembobotan TF-IDF dan K-Nearest Neighbor,” *J. Intell. Syst. Comput.*, vol. 1, no. 1, pp. 43–49, 2019, doi: 10.52985/insyst.v1i1.36.
- [17] Wahyu Sejati, Ankur Singh Bist, and Amirsyah Tambunan, “Pengembangan Analisis Sentimen dalam Rekayasa Software Engineering menggunakan tinjauan literatur sistematis,” *J. MENTARI Manajemen, Pendidik. dan Tekno. Inf.*, vol. 2, no. 1, pp. 95–103, 2023, doi: 10.33050/mentari.v2i1.377.
- [18] R. R. A. Siregar, Z. U. Siregar, and R. Arianto, “Klasifikasi Sentiment Analysis Pada Komentar Peserta Diklat Menggunakan Metode K-Nearest Neighbor,” *Kilat*, vol. 8, no. 1, pp. 81–92, 2019, doi: 10.33322/kilat.v8i1.421.
- [19] A. M. Siregar, S. Faisal, and B. Widiharto, “Analisis Sentimen Masyarakat Terhadap Universitas Buana Perjuangan Karawang Dengan Algoritme SVM dan Naive Bayes,” *Pros. Konf. Nas. Penelit. Dan Pengabd. Univ. Buana Perjuangan Karawang*, vol. 3, no. 1, pp. 25–36, 2023, [Online]. Available: <https://journal.ubpkarawang.ac.id/index.php/ProsidingKNPP/article/view/4894>
- [20] A. Maulana, Inayah Khasnaputri Afifah, Asghafi Mubarrak, Kiagus Rachmat Fauzan, Ardhan Dwintara, and B. P. Zen, “Comparison of Logistic Regression, Multinomialnb, Svm, and K-Nn Methods on Sentiment Analysis of Gojek App Reviews on the Google Play Store,” *J. Tek. Inform.*, vol. 4, no. 6, pp. 1487–1494, 2023, doi: 10.52436/1.jutif.2023.4.6.863.