

Penerapan Metode *Support Vector Machine* Dalam Menganalisis Sentimen Pengguna Aplikasi Sirekap 2024 Di Google Playstore

**Redho Aidil Iqrom¹, Muhammad Syahril², Pamuji Muhamad Jakak³,
Indra Irawan⁴, Yanita Febyani⁵**

^{1,5}Politeknik Prasetiya Mandiri, ²Universitas Muhammadiyah Palembang

^{3,4}Universitas Nurul Huda,

Penulis Korespondensi : redhoaidiliqrom@gmail.com¹,
msyahril@um-palembang.ac.id², jakak@unuha.ac.id³, indrairawan@unuha.ac.id⁴,
febyciby04@gmail.com⁵

Article Information:

Received: January 03, 2025, Accepted: January 09, 2025, Published: January 12, 2025

Abstract

Sirekap is a mobile application that was built to help the public monitor and oversee the development of the 2024 elections held in Indonesia. The research aims to apply the Support Vector Machine algorithm in analyzing sentiment about the use of the Sirekap application in 2024. The Support Vector Machine method is used to classify user sentiment into classes, namely positive, negative, and very negative. The amount of data used is 15,000 data sourced from Sirekap application reviews on Google PlayStore, with more detailed research stages including data collection, data preprocessing, data labeling, visualization, word weighting, and testing and analysis. The results show that the Support Vector Machine algorithm provides an accuracy of 88% for the Sirekap 2024 application. These results are expected to help developers to develop further the Sirekap 2024 application in improving the quality of the application and providing better user comfort Based on the results of the sentiment analysis of Sirekap 2024 application users on the Google Play store using the Support Vector Machine (SVM) method, an accuracy rate of 88% was obtained in classifying the sentiment of reviews into positive, negative, and very negative. This shows that the Support Vector Machine method is quite accurate for sentiment analysis of Indonesian text data. Overall, most reviews are very negative with a percentage reaching 76.9%, followed by negative reviews at 12.6%, and the least are positive reviews at 11%.

Keywords: Analysis Sentiment, Support Vector Machine, Sirekap

1. Pendahuluan

Pemilu merupakan salah satu momen yang sangat berarti bagi bangsa Indonesia, karena pemilu adalah sebuah implementasi dari sistem demokrasi sekaligus pengamalan sila keempat dalam

Published by:

pancasila [1]. Pada tahun 2024 ini Indonesia melaksanakan pemilu serentak di seluruh wilayah Indonesia mulai dari pemilihan presiden, pemilihan DPR RI, DPD RI, DPRD Provinsi, serta DPRD Kabupaten/Kota dalam kurun waktu yang bersamaan. Pelaksanaan pemilihan umum serentak akan memberikan keuntungan berupa penghematan waktu dan mengurangi konflik yang biasa terjadi pada masa pemilu. Namun di sisi lain, pemilu serentak dapat berimbas pada proses penghitungan dan pengawasan suara yang menjadi lebih sulit dan kompleks. Dengan kehadiran aplikasi Sirekap akan menjadi salah satu solusi bagi Masyarakat untuk mengawasi dan memantau perkembangan pemilu tahun 2024. Aplikasi Sirekap dilengkapi dengan fitur-fitur seperti informasi kandidat, berita *update* pemilu, hasil penghitungan suara dan survei serta forum diskusi [2].

Sirekap merupakan singkatan dari Sistem Informasi Rekapitulasi Elektronik. Aplikasi Sirekap merupakan perangkat aplikasi *mobile* yang diluncurkan oleh KPU (Komisi Pemilihan Umum) yang berfungsi untuk membantu Masyarakat untuk memantau dan mengawasi jalannya rekapitulasi suara pada pemilu tahun 2024. Meskipun bertujuan untuk memudahkan masyarakat, kenyataannya pada implementasi aplikasi Sirekap belum bisa memenuhi ekspektasi Masyarakat. Dengan banyaknya kekurangan aplikasi Sirekap justru menimbulkan ketidaknyamanan pada masyarakat, seperti kesalahan dalam proses rekapitulasi, perbedaan antara data yang di input dengan suara yang ada dalam TPS, selain itu fitur aplikasi yang terbatas dan fungsionalitas aplikasi yang belum optimal menurunkan kredibilitas pengguna terhadap aplikasi Sirekap.

Penelitian ini berfokus pada analisis sentimen tanggapan pengguna terhadap aplikasi Sirekap di Google Play Store. Analisis sentimen adalah proses untuk memahami, menelaah, dan mengekstraksi informasi dari pengguna yang disampaikan dalam bentuk teks tertulis dan penilaian (*rating*). Analisis sentimen bertujuan untuk memahami tanggapan publik mengenai suatu topik yang akan di klasifikasikan dalam sebuah penelitian [3].

Tujuan utama penelitian ini adalah menentukan akurasi metode *Support Vector Machine* yang digunakan dan melakukan analisis sentimen untuk memahami tanggapan pengguna terhadap aplikasi Sirekap 2024 melalui ulasan di *Google Play Store*. Bagi developer Penelitian ini penting untuk mengetahui sentimen pengguna mengenai kualitas aplikasi Sirekap 2024, sehingga dapat terus ditingkatkan kepuasan pengguna. Penelitian ini diharapkan dapat meningkatkan partisipasi masyarakat dalam mengawasi hasil penghitungan secara *realtime* pada pemilu 2024.

2. Tinjauan Literatur

2.1 Analisis Sentimen

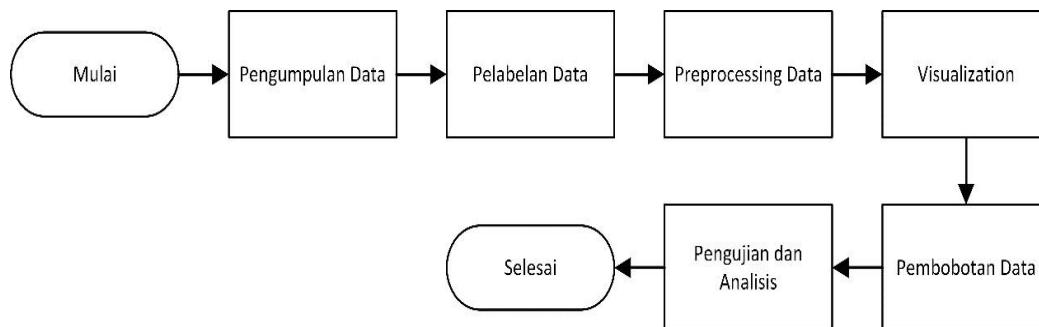
Analisis sentimen merupakan sebuah studi yang menggunakan komputasi untuk memahami pendapat, perasaan, dan emosi (positif dan negatif) orang-orang. Hal ini dilakukan dengan menganalisis atribut dan entitas yang mereka ungkapkan secara tekstual. Analisis sentimen menghasilkan kalimat opini yang terbentuk dari proses ekstraksi dan pengolahan data tekstual secara otomatis. Secara keseluruhan, analisis sentimen dapat membantu organisasi dalam berbagai aspek, seperti memahami publik, meningkatkan kepuasan pelanggan, meningkatkan penjualan dan pemasaran, memantau media sosial, meningkatkan penelitian dan pengembangan, serta meningkatkan keamanan dan intelijen[2].

2.2 Algoritma Support Vector Machine

Algoritma Support Vector Machines (SVM) merupakan metode klasifikasi dan regresi yang populer untuk menangani masalah linear dan nonlinear. Keunggulan SVM terletak pada kemampuannya untuk memisahkan data non-linear berdimensi tinggi secara linear dengan menggunakan fungsi kernel. Efektivitas SVM sangat bergantung pada jenis fungsi kernel yang dipilih dan disesuaikan dengan karakteristik data. Para peneliti yang mengatakan bahwa Algoritma Support Vector machine menjadi metode yang paling akurat dalam hal teks klasifikasi meskipun waktu yang dibutuhkan lebih lama daripada Algoritma lainnya[9].

3. Metode Penelitian

Penelitian ini menggunakan perbandingan algoritma Random Forest, Naïve Bayes, dan Support Tahapan penelitian merupakan rangkaian langkah-langkah yang dilakukan peneliti secara sistematis dalam melaksanakan proses penelitian [4]. Secara lebih rinci tahapan penelitian meliputi pengumpulan data, pelabelan data, preprocessing data, visualization, pembobotan kata serta pengujian dan analisis [5]. Tahapan penelitian dapat dilihat pada gambar 1.



Gambar 1. Tahapan Penelitian

3.1 Pengumpulan Data

Pada penelitian ini pengumpulan data dilakukan dengan metode *web scraping*. *Web Scraping* merupakan suatu proses yang melibatkan pengambilan data semi terstruktur dari sebuah Alamat website atau aplikasi, dalam penelitian ini menggunakan bantuan *google collab* dengan Bahasa pemrograman *python* [6]. Data yang terkumpul dan digunakan sejumlah 15.000 data yang disimpan dalam format csv.

3.2 Pelabelan Data

Data yang diperoleh dari tahap *scraping* tidak bisa langsung teranalisis, oleh karena itu data diperlukan sebuah label/kelas. Pada proses pelabelan akan dilakukan dengan dasar nilai rating dan juga ulasan yang diberikan oleh pengguna aplikasi Sirekap 2024. Pelabelan terdiri dari 3 buah kelas yaitu positif, negatif, dan sangat negatif.

3.3 Preprocessing Data

Tahapan *preprocessing* data akan menjadi proses yang paling penting karena akan menentukan kualitas data yang digunakan. Pada dasarnya *preprocessing* data diterapkan pada data mentah sebelum data digunakan untuk klasifikasi dan identifikasi [7]. *Preprocessing* meliputi tahap *cleaning*, *normalization*, *tokenizing*, *Stopword removal*, dan *steaming*.

3.4 Visualization

Visualization merupakan kegiatan untuk mengungkapkan makna tersembunyi dan menjabarkan kepada orang lain [8] *Visualiziation* berfungsi untuk merepresentasikan hasil temuan yang ditemukan pada setiap penelitian.

Visualization yang di pakai di penelitian ini yaitu *wordcloud* dan diagram sentimen.

3.5 Pembobotan Data

Pada Titik ini, metode TF-IDF (*Term Frequency-Inverse Document Frequency*) digunakan. TF-IDF mengevaluasi hubungan antara kata dalam data dengan menghitung seberapa sering kata tertentu muncul dan seberapa umumnya kata tersebut dalam seluruh dokumen [9].

3.6 Pengujian dan Analisis

Dalam proses ini, data dipisahkan menjadi dua bagian, dimana 90% digunakan sebagai data pelatihan dan 10% sebagai data pengujian. Data pelatihan digunakan untuk melatih algoritma *Support Vector Machine* (SVM) dan membuat model klasifikasi. Data pengujian digunakan untuk mengevaluasi kinerja model SVM [10]. Cara kerja SVM yaitu dengan mencari *hyperplane* terbaik yang memisahkan dua kelas data dengan margin maksimum. SVM bertujuan mencari posisi *hyperplane* optimal sehingga margin antar kelas data menjadi semaksimal mungkin. Dengan margin maksimum, model SVM diharapkan memiliki akurasi klasifikasi yang tinggi [11].

4. Hasil dan Pembahasan

4.1 Pengumpulan Data

Data ulasan SIREKAP 2024 di *Google Play Store* dikumpulkan dengan teknik *scraping* bahasa pemrograman *Python* dan *library google-play-scrapper*. Parameter yang digunakan adalah *lang* = 'id', *country* = 'id', *sort* = *Sort.MOST_RELEVANT*, *count* = 15000 dan *filter_score_with*=None. Data yang diperoleh adalah 15000 data ulasan dengan beberapa atribut yakni *reviewId*, *userName*, *userImage* *content*, *score*, dan lain-lain. Selanjutnya dilakukan seleksi, atribut yang digunakan yaitu hanya *score* dan *content*. *Dataset* setelah dilakukan seleksi data setelah di dapat dilihat pada tabel 1.

Tabel 1. *Dataset* Awal

No	Score	Content
1	1	Aplikasi Buatan GORONG GORONG,,, YANG COCOK UNTUK MEMANIPULASI HASIL DAN DATA PASLON TERTENTU 😂😂😂😂😂
2	1	Tolonglah dicoba dulu sebelum dipakai, ini udah banyak bug malah nyalahin usernya hadeh.
...	...	...
15000	5	Semoga dengan adanya aplikasi ini pemilu akan lebih terbuka ke pada masyarakat. Terimakasih 🙏🙏🙏

4.2 Labeling

Proses pemberian label dilakukan dengan menggunakan rating ulasan sebagai referensi sentimen, dimana rating 1 dikelompokkan ke dalam sentimen sangat negatif (-1), sedangkan rating 2 dan 3 dikelompokkan ke dalam sentimen negatif (1) serta rating 4 dan 5 dikelompokkan ke dalam sentimen positif (0). Pelabelan data tersebut menghasilkan 11.056 ulasan sangat negatif, 2.113 ulasan negatif dan 1.831 ulasan positif. Hasil pada tahap *labeling* dapat dilihat pada tabel 2.

Tabel 2. Hasil Labeling

No	Score	Content	Label
1	1	Aplikasi Buatan GORONG GORONG ,, YANG COCOK UNTUK MEMANIPULASI HASIL DAN DATA PASLON TERTENTU 😞😞😞😞😞	-1
2	1	Tolonglah dicoba dulu sebelum dipakai, ini udah banyak bug malah nyalahin usernya hadeh.	-1
...	...	...	...
15000	5	Semoga dengan adanya aplikasi ini pemilu akan lebih terbuka ke pada masyarakat. Terimakasih 🙏🙏🙏	0

4.3 Preprocessing Data

Cleaning

Dalam proses pembersihan data, dilakukan penghapusan karakter-karakter selain huruf, seperti menghilangkan simbol pagar (#), tautan, dan tanda baca seperti garis miring (/), titik koma (;), titik dua (:), dan titik (.). ("). Seluruh teks juga dikonversi menjadi huruf kecil, menghapus *mention* akun (@...), menstandarisasi kata dengan menghilangkan apostrof (') dan karakter alfanumerik yang menempel, serta menghapus spasi ganda agar diperoleh data teks yang lebih bersih untuk analisis selanjutnya. Hasil pada tahap *cleaning* dapat dilihat pada tabel 3.

Tabel 3. Hasil *Cleaning*

No	Score	Content	Label
1	1	aplikasi buatan gorong gorong yang cocok untuk memanipulasi hasil dan data paslon tertentu	-1
2	1	tolonglah dicoba dulu sebelum dipakai ini udah banyak bug malah nyalahin usernya hadeh	-1
...	...	...	...
15000	5	semoga dengan adanya aplikasi ini pemilu akan lebih terbuka ke pada masyarakat terimakasih	0

Normalization

Dalam proses normalisasi data, dilakukan standarisasi ejaan kata-kata yang sering disingkat atau ditulis dengan ejaan tidak baku menjadi bentuk baku. Sebagai contoh, kata "gue" dinormalisasi menjadi "saya", "dgn" menjadi "dengan", "bgmn" menjadi "bagaimana", dan seterusnya. Tujuan

normalisasi ini adalah untuk menyeragamkan penulisan kata agar lebih baku sehingga dapat dianalisis dengan lebih akurat. Hasil pada tahap *cleaning* dapat dilihat pada tabel 4.

Tabel 4. Hasil *Normalization*

No	Score	Content	Label
1	1	aplikasi buatan gorong gorong yang cocok untuk memanipulasi hasil dan data paslon tertentu	-1
2	1	tolonglah dicoba dulu sebelum dipakai ini udah banyak bug malah nyalahin usernya hadeh	-1
...	...	...	...
15000	5	semoga dengan adanya aplikasi ini pemilu akan lebih terbuka ke pada masyarakat terimakasih	0

Tokenizing

Dalam proses *tokenizing*, teks dipotong menjadi token berupa kata-kata individual dengan menggunakan *RegexTokenizer*. Teks dipotong pada setiap kata yang terdiri dari karakter alfanumerik (\w+). Dengan *tokenizing* ini, teks yang awalnya berupa kalimat utuh dapat dipecah menjadi kata per kata sehingga lebih mudah dianalisis secara leksikal. Hasil *tokenizing* ditambahkan ke dalam kolom baru 'text_token' pada data frame untuk digunakan pada proses analisis selanjutnya. Hasil pada tahap *cleaning* dapat dilihat pada tabel 5.

Tabel 5. Hasil *Tokenizing*

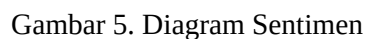
No	Score	Content	Label
1	1	['aplikasi', 'buatan', 'gorong', 'gorong', 'yang', 'cocok', 'untuk', 'manipulasi', 'hasil', 'dan', 'data', 'paslon', 'tertentu']	-1
2	1	['tolonglah', 'dicoba', 'dulu', 'sebelum', 'dipakai', 'ini', 'udah', 'banyak', 'bug', 'malah', 'nyalahin', 'usernya', 'hadeh']	-1
...	...	...	...
15000	5	['semoga', 'dengan', 'adanya', 'aplikasi', 'ini', 'pemilu', 'akan', 'lebih', 'terbuka', 'ke', 'pada', 'masyarakat', 'terimakasih']	0

Stopword Removal

Dalam proses *Stopword removal*, kata-kata yang termasuk *Stopword* dalam bahasa Indonesia dihapus dari hasil tokenisasi agar tidak terbawa ke proses analisis selanjutnya. *Stopword* adalah kata-kata yang sangat umum seperti 'yang', 'dan', 'di', 'dll' yang kurang memiliki makna penting secara leksikal. Dengan menghapus *Stopword*, diharapkan kata-kata yang tersisa lebih merepresentasikan substansi dan topik teks sebenarnya. *Stopword* bahasa Indonesia diunduh dari corpus NLTK kemudian daftar *Stopword* ini digunakan untuk menyaring kata hasil tokenisasi dengan menghapus kata yang termasuk *Stopword*. Hasilnya adalah kumpulan kata tanpa *Stopword* yang lebih bermakna untuk analisis. Hasil pada tahap *Stopword removal* dapat dilihat pada tabel 6.



Dalam visualisasi diagram pie sentimen, data label sentimen dihitung jumlah kemunculannya untuk masing-masing kategori positif, negatif, dan sangat negatif menggunakan fungsi *value_counts()*. Dari data ulasan yang ada, terdapat 1.831 ulasan dengan sentimen positif, 2.113 ulasan dengan sentimen negatif, dan 11.056 ulasan dengan sentimen netral. Persentase ulasan yang telah diberi label sebagai sentimen positif, negatif, dan netral dapat dilihat pada Gambar 5.



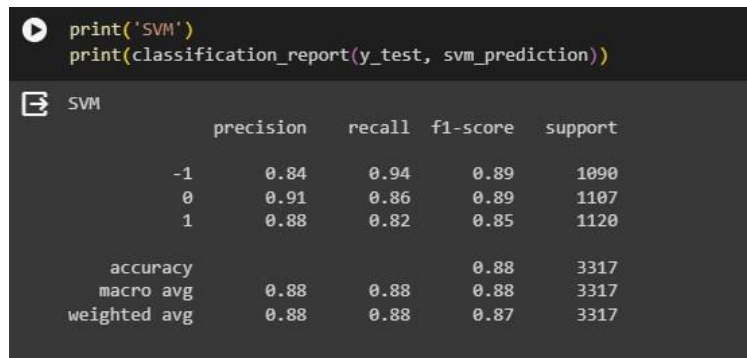
Berdasarkan analisis pada Gambar 4, dapat disimpulkan bahwa persentase data tertinggi berada pada sentimen sangat negatif, yaitu mencapai 76,9%. Untuk sentimen dengan tingkat menengah, persentase terbanyak ada pada sentimen negatif sebesar 12,6%, sedangkan sentimen positif memiliki persentase paling rendah yaitu 10,5%

4.5 Pembobotan Kata

Dalam pembobotan kata dengan TF-IDF, setiap kata dalam dokumen teks diberi bobot yang mencerminkan seberapa penting kata tersebut dalam dokumen/korpus. Bobot dihitung berdasarkan frekuensi kemunculan kata dalam dokumen (TF) dikalikan dengan logaritma dari invers frekuensi dokumen yang mengandung kata tersebut (IDF). Semakin tinggi TF-IDF suatu kata, semakin relevan kata itu dengan topik dokumen. Pertama, nilai NaN diganti dengan *string* kosong agar tidak mengganggu perhitungan. Kemudian *TfidfVectorizer* digunakan untuk mengkonversi data teks menjadi vektor fitur TF-IDF. Hasilnya adalah representasi angka dari data teks asli, di mana setiap kata direpresentasikan oleh bobot TF-IDFnya. Representasi vektor ini yang akan digunakan oleh model analisis teks selanjutnya.

4.6 Pengujian dan Analisis

Dalam tahap pengujian dan evaluasi model, data dipisahkan menjadi data latih dan data uji menggunakan *train_test_split* dengan rasio 90:10. Karena data tidak seimbang, *oversampling* dilakukan dengan SMOTE agar jumlah data proporsional untuk setiap kelas. Model SVM kemudian dilatih dengan data latih dan dievaluasi performanya menggunakan data uji. Performa diukur dengan *metric accuracy, precision, recall*, dan *f1-score* yang dihasilkan *classification report*. Hasil evaluasi pada gambar 6 menunjukkan performa model SVM dalam mengklasifikasikan sentimen teks. Model terbaik selanjutnya dapat digunakan untuk prediksi data baru.



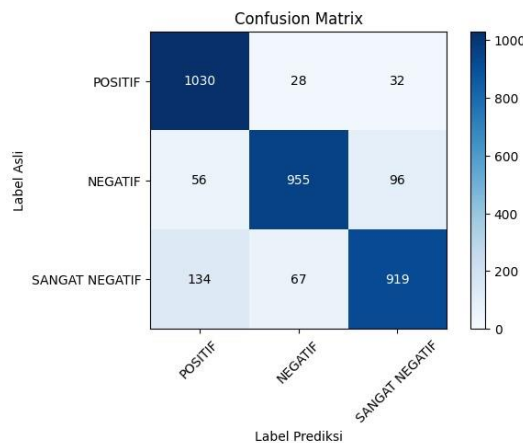
```
print('SVM')
print(classification_report(y_test, svm_prediction))
```

SVM	precision	recall	f1-score	support
-1	0.84	0.94	0.89	1090
0	0.91	0.86	0.89	1107
1	0.88	0.82	0.85	1120
accuracy			0.88	3317
macro avg	0.88	0.88	0.88	3317
weighted avg	0.88	0.88	0.87	3317

Gambar 6. Hasil Pengujian

Berdasarkan hasil pengujian model SVM pada data uji yang terlihat pada gambar 6, didapatkan nilai akurasi sebesar 0.88 atau 88%. Hal ini menunjukkan bahwa secara keseluruhan, model SVM mampu mengklasifikasikan sentimen teks dengan akurat sebesar 88% pada data uji. Secara lebih rinci, model SVM menghasilkan nilai presisi atau *precision* sebesar 0.84 untuk kelas sangat negatif (-1), 0.91 untuk kelas positif (0), dan 0.88 untuk kelas negatif (1). Artinya, model cukup akurat dalam memprediksi masing-masing kelas sentimen. Untuk nilai *recall* atau sensitivitas, model SVM menghasilkan 0.94 untuk kelas sangat negatif, 0.86 untuk kelas positif, dan 0.82 untuk kelas negatif. Hal ini berarti kemampuan model untuk mendeteksi sampel positif

pada masing-masing kelas sudah cukup baik, terutama pada kelas sangat negatif. Sedangkan nilai *F1-score* yang merupakan rata-rata harmonik dari presisi dan *recall* adalah 0.89 untuk kelas sangat negatif, 0.89 untuk kelas positif, dan 0.85 untuk kelas negatif. *F1-score* yang mendekati 1 mengindikasikan performa klasifikasi yang baik secara keseluruhan. Hasil penelitian menunjukkan bahwa metode *Support Vector Machine* memberikan tingkat akurasi yang tinggi dalam menganalisis sentimen pengguna SIREKAP, selanjutnya dari hasil tersebut *confusion matrix* juga memberikan informasi sebagai berikut



Gambar 7. Hasil Confusion Matrix

Pada gambar 7 bisa dilihat terdapat 1030 *True Positive* (TP), yang merupakan ulasan positif yang diklasifikasikan dengan benar. Sebanyak 955 *True Negative* (TN) menunjukkan ulasan negatif yang diklasifikasikan dengan benar. Sebanyak 919 *True Very Negative* (TVN) menunjukkan ulasan sangat negatif yang diklasifikasikan dengan benar. Disisi lain, terdapat 28 dan 32 *False Positive* (FP) yang menunjukkan ulasan negatif dan sangat negatif yang keliru diklasifikasikan sebagai positif. Sebanyak 56 dan 96 *False Negative* (FN) yang menunjukkan ulasan positif dan sangat negatif yang keliru diklasifikasikan sebagai ulasan negatif dan sebanyak 134 dan 67 *False Very Negatif* (FVN) yang menunjukkan ulasan positif dan negatif yang keliru diklasifikasikan sebagai sangat negatif.

Berikut cara perhitungan manual mencari nilai akurasi dengan menggunakan rumus berikut:

$$\begin{aligned}
 \text{Accuracy} &= \frac{TP+TN+TVN}{TP+TN+TVN+FP+FN+FVN} \times 100\% \\
 &= \frac{1030+955+919}{1030+955+919+60+152+201} \times 100\% \\
 &= \frac{2904}{3317} \times 100\% \\
 &= 88\%
 \end{aligned}$$

Gambar 8. Hasil Perhitungan Akurasi

Secara umum performa model SVM dalam klasifikasi sentimen teks sudah cukup baik ditinjau dari nilai akurasi, presisi, *recall* dan *F1-score* untuk semua kelas. Model ini layak digunakan pada kasus nyata untuk prediksi sentimen teks yang baru. Namun demikian, performa masih dapat ditingkatkan lebih lanjut dengan penyesuaian *hiperparameter* dan algoritma lainnya.

5. Kesimpulan

Berdasarkan hasil analisis sentimen pengguna aplikasi Sirekap 2024 di *Google Play Store* dengan menggunakan metode *Support Vector Machine* (SVM), diperoleh tingkat akurasi sebesar 88% dalam mengklasifikasikan sentimen ulasan menjadi positif, negatif, dan sangat negatif. Hal ini menunjukkan metode SVM cukup akurat untuk analisis sentimen data teks berbahasa Indonesia. Secara keseluruhan, mayoritas ulasan bersifat sangat negatif dengan persentase mencapai 76,9%, diikuti ulasan negatif sebesar 12,6%, dan paling sedikit adalah ulasan positif yaitu 10,5%. Ulasan negatif menunjukkan banyaknya keluhan pengguna terhadap *bug*, *error*, gagal *login*, dan lainnya yang membuat penggunaan aplikasi Sirekap 2024 menjadi tidak nyaman. Oleh karena itu, pengembang Sirekap 2024 perlu memperhatikan masukan pengguna untuk meningkatkan kualitas aplikasi. Dengan analisis sentimen, developer dapat memperbaiki bagian yang dirasa kurang agar kepuasan pengguna meningkat. Hasil penelitian diharapkan meningkatkan kepercayaan masyarakat dalam mengawasi Pemilu 2024 melalui Sirekap. Penelitian selanjutnya dapat membandingkan akurasi beberapa metode analisis sentimen lainnya.

Referensi

- [1] Yusrin, Y. and Salpina, S. (2023) '*Partisipasi Generasi Millennial dalam Mengawasi Tahapan Pemilu 2024*', *Journal on Education*, 5(3), pp. 9646– 9653. Available at: <https://doi.org/10.31004/joe.v5i3.1842>.
- [2] Marzellina, H. et al. (2022) '*Urgensi Sistem E-Voting dan Sirekap Dalam Penyelenggaraan Pemilu 2024*', 7(2).
- [3] VF. Anindya, A.W. (2023) '*Analisis Sentimen Ulasan Aplikasi Capcut Menggunakan Metode Naive Bayes*', *Statistika*, 23(2), pp. 164–175. Available at: <https://doi.org/10.29313/statistika.v23i2.2353>.
- [4] Styawati, S., Hendrastuty, N. and Isnain, A.R. (2021) '*Analisis Sentimen Masyarakat Terhadap Program Kartu Prakerja Pada Twitter Dengan Metode Support Vector Machine*', *Jurnal Informatika: Jurnal Pengembangan IT*, 6(3), pp. 150–155. Available at: <https://doi.org/10.30591/jpit.v6i3.2870>.
- [5] Hasibuan, S.S. and Saputra, E. (2024) '*Sentimen Analisis Terhadap Fitur Tiktok Shop Menggunakan Naive Bayes dan K-Nearest Neighbor*', 8, pp. 303–311. Available at: <https://doi.org/10.30865/mib.v8i1.7238>.
- [6] Mufidah, U. and Siahaan, M. (2021) '*Perancangan Aplikasi Perbandingan Harga Produk (Historical Data) Menggunakan Teknik Web Scraping*', *Pusdansi.org*, 1(1), pp. 1–14.

- [7] Rahman, A. (2019) '*Statistics-based data preprocessing methods and machine learning algorithms for big data analysis*', International Journal of Artificial Intelligence, 17(2), pp. 44–65.
- [8] Wang, Q. et al. (2022) '*A Survey on ML4VIS: Applying Machine Learning Advances to Data Visualization*', IEEE Transactions on Visualization and Computer Graphics, 28(12), pp. 5134–5153. Available at:
- [9] Sintia, Defit, S. and Nurcahyo, G.W. (2021) '*Product Codefication Accuracy With Cosine Similarity and Weighted Term Frequency and Inverse Document Frequency (Tf-Idf)*', Journal of Applied Engineering and Technological Science, 2(2), pp. 14–21. Available at: <https://doi.org/10.37385/jaets.v2i2.210>.
- [10] Kosasih, R. (2021) '*Penggunaan Metode Linear Discriminant Analysis Untuk Pengenalan Wajah Dengan Membandingkan Banyaknya Data Latih*', Jurnal Ilmiah Teknologi dan Rekayasa, 26(1), pp. 25–34. Available at: <https://doi.org/10.35760/tr.2021.v26i1.3520>.
- [11] Tinages, R., Triayudi, A. and Sholihati, I.D. (2020) '*Analisis Sentimen Terhadap Layanan Indihome Berdasarkan Twitter Dengan Metode Klasifikasi Support Vector Machine (SVM)*', Jurnal Media Informatika Budidarma, 4(3), p. 650. Available at: <https://doi.org/10.30865/mib.v4i3.2181>.
- [12] C. G. Indrayanto, D. E. Ratnawati, dan B. Rahayudi, "Analisis Sentimen Data Ulasan Pengguna Aplikasi MyPertamina di Indonesia pada Google Play Store menggunakan Metode Random Forest," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 3, hal. 1131–1139, 2023.
- [13] E. Indrayuni, "Komparasi Algoritma Naive Bayes Dan Support Vector Machine Untuk Analisa Sentimen Review Film," *J. Pilar Nusa Mandiri*, vol. 14, no. 2, hal. 175, 2018, doi: 10.33480/pilar.v14i2.918.

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>)

