ANALISIS SENTIMEN ISU VIRAL 'FUFUFAFA' DI X DENGAN SVM DAN DECISION TREE

David Wahyu Setyo Aji [1], Dhea Maharani [2], Najwa Ratu Afi [3], My. Teguh Sulistyono [4]*
[1,2,3,4] Ilmu Komputer Sistem Informasi Universitas Dian Nuswantoro
teguh.sulistyono@dsn.dinus.ac.id

ABSTRAK

Perkembangan teknologi informasi yang pesat telah memudahkan masyarakat mengakses informasi melalui media sosial, terutama Twitter (kini X), yang berfungsi sebagai sumber berita terkini sekaligus wadah opini publik. Salah satu isu viral yang menjadi sorotan adalah kontroversi akun KasKus "Fufufafa" yang memicu beragam reaksi warganet di Indonesia. Penelitian ini bertujuan untuk menganalisis sentimen publik terhadap kata kunci "Fufufafa" pada platform X serta membandingkan kinerja dua algoritma klasifikasi, yaitu Support Vector Machine (SVM) dan Decision Tree. Data dikumpulkan melalui crawling sebanyak 1.007 tweet pada periode 1–30 September 2024, kemudian dilakukan preprocessing (case folding, normalisasi, stopword removal, stemming), labeling otomatis menggunakan VADER Lexicon, ekstraksi fitur TF-IDF, serta pembagian data menjadi 70% pelatihan dan 30% pengujian. Model SVM dan Decision Tree dilatih dan dievaluasi menggunakan metrik akurasi, presisi, recall, dan F1-score. Hasil penelitian menunjukkan bahwa kedua model mencapai akurasi 95,70%, namun Decision Tree memiliki F1-score tertimbang sedikit lebih tinggi (0,9515) dibandingkan SVM (0,9444), sehingga lebih efektif pada data dengan ketidakseimbangan kelas. Kontribusi penelitian ini adalah memberikan gambaran komposisi sentimen publik terhadap isu viral lokal serta mengevaluasi performa dua metode klasifikasi dalam konteks analisis sentimen media sosial di Indonesia.

Keyword : *Analisis Sentimen, X, Fufufafa, Support Vector Machine, Decision Tree*

1. PENDAHULUAN

Munculnya media sosial telah mengubah paradigma berkomunikasi,baik untuk penyebaran secara global maupun lokal. Platform seperti Facebook, Twitter, Instagram dan dalam hal ini platform X, telah mengaktifkan interaksi waktu nyata, memungkinkan informasi untuk menyebar dengan cepat dan mempengaruhi wacana publik hampir secara instan. Fenomena viral, khususnya isu-isu yang timbul dari masalah-masalah kontroversial, dapat dengan cepat memicu konflik diskusi dan memainkan peran penting dalam membentuk opini publik [1]. Salah satu isu viral di Indonesia yang menjadi perhatian masyarakat dimana akun KasKus "Fufufafa", yang telah menuai beragam tanggapan dari komunitas online. Dalam konteks ini [2], analisis sentimen menjadi kunci untuk memahami bagaimana reaksi publik terhadap isu-isu yang tengah hangat diperbincangkan.

Metode Support Vector Machine (SVM) telah dikenal luas dalam analisis sentimen karena kemampuannya untuk menangani data yang kompleks dan tidak terstruktur dengan baik. SVM bekerja dengan mencari hyperplane optimal [3], sehingga dapat digunakan untuk klasifikasi yang efektif, bahkan dalam penerapan dengan data yang tidak linear [1]. Sementara itu, Decision Tree dipilih karena kemudahannya dalam interpretasi dan penggunaannya, memungkinkan para peneliti dan pemangku kepentingan untuk memahami faktor-faktor yang berkontribusi dalam analisis data dengan lebih jelas [4].

Penelitian ini bertujuan untuk mengukur sentimen publik terhadap keyword "Fufufafa" pada platform Twitter (X) dalam periode tertentu dengan menggunakan pendekatan Support Vector Machine (SVM) dan Decision Tree [5]. Studi ini dapat diharapkan mampu memberikan gambaran umum mengenai opini masyarakat serta efektivitas model yang digunakan dalam konteks isu viral lokal[6].

2. TINJAUAN PUSTAKA

2.1. Analisis Sentimen

Analisis sentimen adalah metode untuk mengidentifikasi, mengekstraksi, dan mengklasifikasikan opini atau emosi yang terkandung dalam data teks, sehingga dapat menentukan kecenderungan sentimen suatu pernyataan menjadi positif, negatif, atau netral [7].Proses ini umum digunakan pada ulasan produk, tanggapan media sosial, hingga analisis opini publik terhadap suatu isu [8]. Analisis sentimen mengandalkan teknik pengolahan bahasa alami (Natural Language Processing/NLP) dan pembelajaran mesin (machine learning) untuk mengubah teks tidak terstruktur menjadi informasi yang terukur.

2.2. Text Mining

Text mining merupakan proses penemuan informasi dan pola tersembunyi dari kumpulan teks, baik yang bersifat semi-terstruktur maupun tidak terstruktu. Perbedaannya dengan data mining adalah jenis data yang diolah text mining berfokus pada teks, sedangkan data mining memproses data terstruktur seperti tabel atau basis data. Dalam analisis sentimen, text mining melibatkan

preprocessing teks, ekstraksi fitur, dan analisis statistik atau pembelajaran mesin [9].

2.3. Support Vector Machine

SVM salah satu metode klasifikasi yang dikembangkan oleh Vapnik pada tahun 1992 dan dikenal memiliki kinerja yang tinggi, khususnya dalam menangani permasalahan nonlinier [10]. SVM termasuk dalam kategori algoritma pembelajaran mesin modern yang muncul setelah metode sebelumnya, yakni Neural Network (NN). Keduanya telah terbukti efektif dalam penerapan pengenalan pola. Proses pembelajaran dalam SVM dilakukan dengan memanfaatkan data pelatihan yang terdiri atas pasangan input dan output yang merepresentasikan target yang diharapkan [11]. Secara umum, prinsip kerja SVM adalah mencari hyperplane optimal yang mampu memisahkan dua kelas data secara maksimal dalam ruang fitur [12], dibawah ini terdapat proses kerja SVM dalam analisis sentimen:

a. Representasi Data dimana Data teks diubah menjadi vektor numerik menggunakan metode seperti TF-IDF.
b. Pemilihan Kernel, Kernel linear digunakan jika data relatif linear; kernel non-linear seperti RBF digunakan untuk data yang kompleks.
c. Penentuan Hyperplane Optimal Algoritma menentukan garis atau bidang pemisah terbaik yang memaksimalkan jarak margin antar kelas.
d. Klasifikasi Data uji diklasifikasikan berdasarkan sisi hyperplane tempat data berada.

   Persamaan dasar linear:

   $$f(x) = w \cdot x + b \quad (1)$$

   Dengan w sebagai bobot, x sebagai fitur input, dan b sebagai bias.

2.4. Decesion Tree

Decision Tree merupakan algoritma induksi yang membentuk struktur pohon secara rekursif dari atas ke bawah, di mana pemilihan atribut yang akan diuji ditentukan berdasarkan ukuran seleksi atribut tertentu. Dalam proses pembentukannya, algoritma ini berupaya meningkatkan tingkat akurasi dengan mengeliminasi cabang-cabang pohon yang dianggap merepresentasikan noise atau gangguan dalam data [13], dibawah ini terdapat proses kerja Decision Tree dalam analisis sentimen:

a. Pemilihan Atribut Menentukan atribut terbaik untuk memisahkan data menggunakan ukuran seperti *information gain* atau *Gini index*.
b. Pemisahan Data Dimana Data dibagi ke dalam cabang berdasarkan nilai atribut yang dipilih.
c. Pembangunan Pohon Secara Rekursif Langkah 1 dan 2 diulang hingga semua data dalam node memiliki kelas yang sama atau memenuhi kriteria penghentian.
d. Pruning (Pemangkasan) Menghapus cabang yang tidak memberikan kontribusi signifikan untuk mengurangi overfitting.

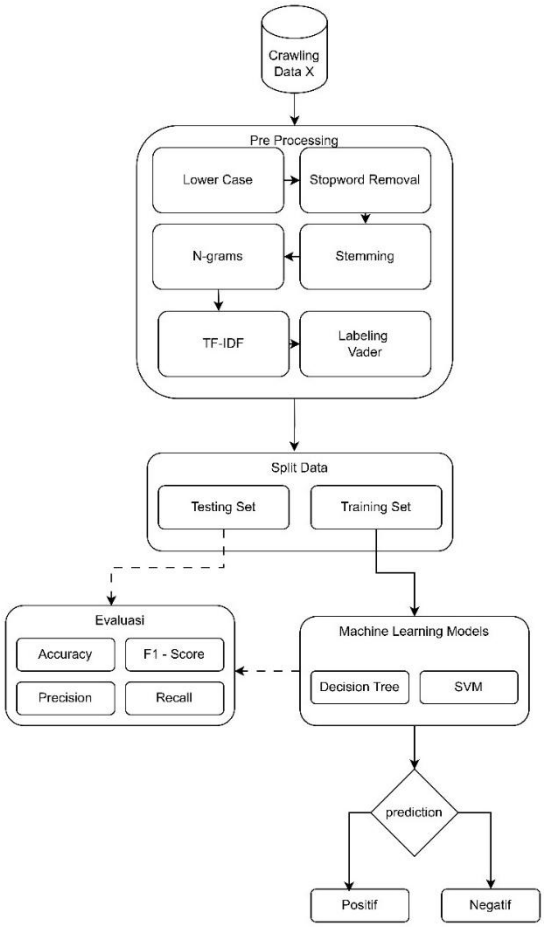
   Formula information gain:

   $$IG(S, A) = Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} \times Entropy(S_v) \quad (2)$$

   Dimana S himpunan data, A atribut yang di uji, dan S_v subset data dengan nilai atribut v.

3. METODE PENELITIAN

Penelitian ini terdiri dari beberapa tahap yang saling berkaitan, mulai dari pengumpulan data hingga evaluasi model. Tahapan metodologi dapat dilihat pada Gambar 1, yang menampilkan alur proses analisis sentimen isu viral "Fufufafa" di platform X menggunakan SVM dan Decision Tree.



Gambar 1. Metodologi Penelitian Analisis Sentimen

3.1. Crawling Data Twitter dan Labeling Data

Dalam penelitian ini, langkah awal yang diambil dimana mengumpulkan data dari platform media sosial Twitter. Pengumpulan data dilakukan dengan teknik crawling yang memanfaatkan Twitter API. Untuk membatasi jumlah data yang diperoleh, ditetapkan batasan sebanyak 1000 tweet. Namun, setelah proses crawling selesai, total tweet yang

berhasil dikumpulkan mencapai 1007. Labeling data memberi label pada tweet tersebut berdasarkan sentimen yang terkandung dalam masing-masing menggunakan Google Collab. Proses ini akan menggunakan model analisis sentimen VADER, yang akan mengevaluasi setiap tweet dan mengklasifikasikannya menjadi positif atau negatif. Dalam hal ini, jika skor compound untuk tweet tersebut lebih besar atau sama dengan 0.05, tweet tersebut akan dikategorikan sebagai positif [14]. Sebaliknya, jika skor compound lebih kecil atau sama dengan -0.05, tweet tersebut akan dikategorikan sebagai negatif, sesuai dengan metode yang diterapkan dalam penelitian ini. Dengan demikian, seluruh proses ini bertujuan untuk memperoleh tweet yang mengandung kata kunci "Fufufafa" dalam periode waktu dari 1 September 2024 hingga 30 September 2024 dimana juga memperoleh pemahaman yang lebih baik mengenai persepsi publik terhadap "Fufufafa" berdasarkan analisis sentimen dari tweet yang telah dikumpulka [15].

3.2. Prepocessing Data Set

Untuk memastikan kualitas data dan efektivitas pemodelan, dilakukan serangkaian proses pra-pemrosesan teks [16].

a. Lowercase
   Setelah melakukan tahapan Nominal to Text selanjutnya yaitu memakai fungsi lowercases untuk mengonversi seluruh karakter dalam tweet menjadi huruf kecil agar konsisten.
b. Stopwords Removal
   Setelah melakukan tahapan Nominal to Text selanjutnya yaitu memakai fungsi lowercases untuk mengonversi seluruh karakter dalam tweet menjadi huruf kecil agar konsisten.
c. Stemming
   Fungsi Stemming digunakan setelah tahap Stopwords Removal. Fungsi ini bertujuan untuk mengubah kata ke bentuk dasarnya misalnya "memarahi" menjadi "marah".
d. N-Gram Generation
   Selanjutnya Fungsi n-Gram Generation yaitu membentuk kombinasi kata berdasarkan urutan kemunculannya misalnya unigram, bigram.
e. TF-IDF Weighting
   Dalam tahapan praproses, metode TF-IDF dimanfaatkan untuk merepresentasikan teks ke dalam vektor numerik berdasarkan tingkat kepentingan kata terhadap keseluruhan dokumen.

3.3. Split Data

Selanjutnya data dibagi menjadi dua bagian yaitu training dan testing dengan proporsi 70:30. Bertujuan agar model dapat belajar dari sebagian besar data (70%) dan kemudian diuji dengan data yang tidak pernah dilihat sebelumnya (30%) untuk mengevaluasi performa dan generalisasi model secara objektif.

3.4. Apply Model

Tahap selanjutnya menerapkan model klasifikasi pada data. Dua model pembelajaran mesin yang digunakan diantaranya Support Vector Machine (SVM) dan Decision Tree.

a. Support Vector Machine (SVM)
   Algoritma klasifikasi yang dikembangkan oleh Vapnik pada tahun 1992, dan dikenal memiliki performa yang baik dalam mengatasi permasalahan klasifikasi, terutama pada data yang tidak terdistribusi secara linear.
   Rumus dasar Support Vector Machine (SVM) [17]:

$$f(x) = w \cdot x + b \tag{3}$$

   Dimana w bobot model, x fitur input dan b sebagai bias.

$$\min_{w,b,\xi} \frac{1}{2}|w|^2 + C\sum_{i=1}^{n} \xi_i \tag{4}$$

   Dengan syarat :

$$y_i(w \cdot x_i + b) \geq 1 - \xi_i, \quad \xi_i \geq 0 \tag{5}$$

b. Decsion Tree
   Decision Tree merupakan metode klasifikasi yang membentuk struktur hierarkis berupa cabang dan daun untuk merepresentasikan alur pengambilan keputusan [18]:

$$Entropy(S) = -\sum_{i=1}^{c} p_i \log_2 p_i \tag{6}$$

   S : *Himpunan data yang dianalisis*
   c : *Jumlah kelas atau kategori dalam data*
   p_i : *Probabilitas dari elemen data yang termasuk ke dalam kelas ke-iii*
   log_2 : *Logaritma basis 2 digunakan untuk mengukur informasi dalam bit*

   Berikutnya, information Gain dari setiap fitur dihitung :

$$\text{Gain}(S, A) = Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} \cdot Entropy(S_v) \tag{7}$$

   Dimana, atribut dengan nilai information Gain tertinggi dipilih sebagai node pada Decision Tree.

3.5. Evaluasi

Setelah proses pelatihan dan pengujian model dilakukan, tahap berikutnya adalah mengevaluasi

performa model guna mengukur tingkat akurasi serta efektivitas prediksi yang dihasilkan. Evaluasi ini menggunakan empat metrik utama, yaitu accuracy, precision, recall, dan f1-score [19]. Dalam penelitian ini, dua algoritma klasifikasi diterapkan, yaitu Support Vector Machine (SVM) dan Decision Tree. Selain itu, confusion matrix dimanfaatkan sebagai alat bantu untuk melihat detail hasil klasifikasi, dengan membedakan antara true positive, true negative, false positive, dan false negative. Keempat komponen ini kemudian dijadikan dasar dalam penghitungan dan analisis metrik evaluasi yang digunakan.

a. Accuracy
Accuracy mengukur seberapa besar proporsi prediksi yang benar dibandingkan dengan seluruh data [20].

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \quad (8)$$

b. Preceion
Precision menunjukkan ketepatan model dalam memprediksi kelas positif dari seluruh prediksi positif yang dihasilkan[11].

$$Precision = \frac{TP}{TP + FP} \quad (9)$$

c. Recall
Recall menggambarkan kemampuan model dalam menemukan semua data yang benar-benar termasuk dalam kelas positif[11].

$$Recall = \frac{TP}{TP + FN} \quad (10)$$

d. F1 Score
F1 Score merupakan rata-rata harmonis antara nilai precision dan recall, yang berfungsi sebagai indikator umum dalam mengevaluasi performa model, khususnya ketika menangani data dengan distribusi kelas yang tidak seimbang.

$$F1 - Score = 2 \times \frac{precision \times recall}{precision + recall} \quad (11)$$

Hasil evaluasi terhadap kedua model tersebut akan dianalisis pada bab berikutnya guna mengetahui model mana yang memberikan performa terbaik dalam mendeteksi sentimen.

4. HASIL DAN PEMBAHASAN

4.1. Crawling Data Twitter dan Labeling data

Pengumpulan data dilakukan melalui teknik crawling menggunakan Twitter API, dengan kata kunci "Fufufafa". Total 1.007 tweet berhasil diperoleh dan Labeling otomatis dilakukan menggunakan VADER Sentiment Analyzer di Google Colab. Denagn aturan pelabelan Positif skor compund ≥ 0,05 dan Negatif skor compound ≤ -0,05.

Tabel 1. Dataset

No	Name	Type
1	Conversation_id_str	Real
2	Created_at	Nominal
3	Favorite_count	Integer
4	Full_text	Nominal
…	……	…….
16	Sentiment	Nominal

4.2. Preprocessing Data Set

Prepocesing menampilkan 4 komentar pertama dari dataset, baik dalam format aslinya maupun setelah melalui proses pembersihan menggunakan Lowercase, Stopword Removal, N-Grams, Stemming, dan TF-IDF. Ini untuk membandingkan teks sebelum dan sesudah preprocessing, seperti penghapusan emoji, URL, mentions, hashtags, karakter spesial, case folding, normalisasi karakter berulang, tokenisasi, stopword removal, dan stemming. Setelah menampilkan contoh-contoh tersebut, kode mencetak pesan konfirmasi bahwa proses pembersihan data dan preprocessing telah selesai.

Tabel 2. Prepoccesing dataset

No	Original	Cleaned
1	@fajarnugros Bisa usul pak besok pake kaos adiknya fufufafa atau negara keluarga atau mulyono jujur tidak pernah bo'ong.	usul besok pake kao adiknya fufufafa negara keluarga mulyono jujur boong
2	@nabiylarisfa Bener juga kenapa mereka bilang adik/adik ipar fufufafa	bener bilang adikadik ipar fufufafa
3	@tekarok007 Enakan pendukung fufufafa kalo ngitung hasil kinerja tinggal X0 maka hasilnya 0	enakan pendukung fufufafa kalo ngitung hasil kinerja tinggal x0 hasilnya 0
4	@ARSIPAJA Belum semukatembok itu klo blm sanggup ngomong kami keluarganya fufufafa	semukatembok blm sanggup ngomong keluarganya fufufafa

4.3. Splitt Data

Dataset yang digunakan terdiri dari 1.006 sampel, yang kemudian dibagi menjadi dua bagian: 70% untuk data pelatihan (704 sampel) dan 30% untuk data pengujian (302 sampel). Setelah melalui proses *feature extraction*, diperoleh sebanyak 1.815 fitur yang merepresentasikan setiap sampel. Dataset ini mengandung dua label klasifikasi, yaitu 'negatif' dan 'positif'. Distribusi label menunjukkan dominasi kelas 'negatif' secara signifikan, baik pada data pelatihan (675 sampel negatif dan 29 positif) maupun pada data pengujian (290 sampel negatif dan 12 positif). Ketimpangan distribusi ini mengindikasikan adanya masalah ketidakseimbangan kelas (*class imbalance*) dalam

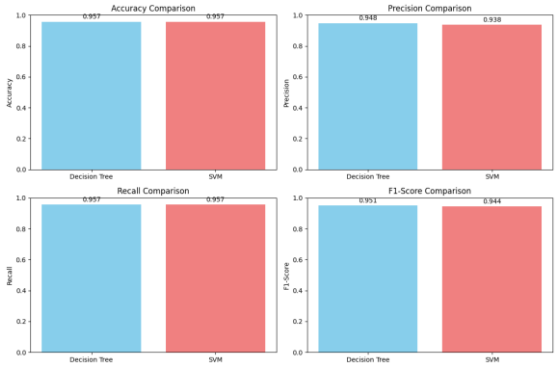
dataset, yang berpotensi mempengaruhi kinerja model klasifikasi.

Table 3. Data Summary

DATA SUMMARY	
Total Samples	1006
Training Samples	704 (70.0%)
Testing Samples	302 (30.0%)
Number Of Features	1815
Numbuer Of Unique Labels	2
Train Label Distribution	'negatif': 675, 'positif': 29
Test Label Distribution	'negatif': 290, 'positif': 12

4.4. Evaluasi

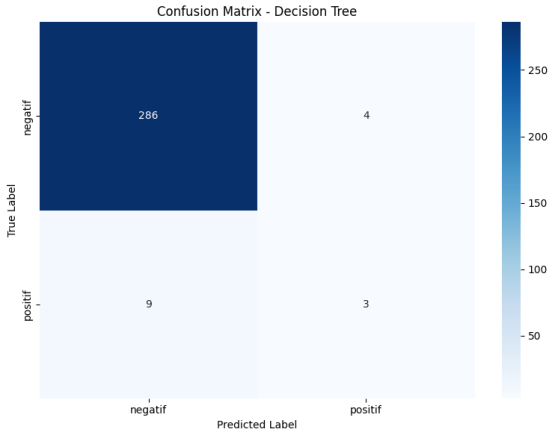
Evaluasi dilakukan untuk membandingkan performa dua algoritma klasifikasi, yaitu Decision Tree dan Support Vector Machine (SVM), dalam konteks analisis sentimen. Kedua model menunjukkan tingkat akurasi yang sama, yaitu sebesar 0,9570. Namun, dengan mempertimbangkan metrik evaluasi lain seperti Precision, Recall, dan F1-Score—yang lebih representatif untuk data yang mengalami ketidakseimbangan kelas—model Decision Tree menunjukkan performa yang sedikit lebih baik. Hal ini ditunjukkan oleh nilai *weighted F1-Score* sebesar 0,9515, lebih tinggi dibandingkan SVM yang memperoleh skor 0,9444. Oleh karena itu, berdasarkan F1-Score tertimbang, Decision Tree dapat diidentifikasi sebagai model dengan kinerja terbaik di antara keduanya.



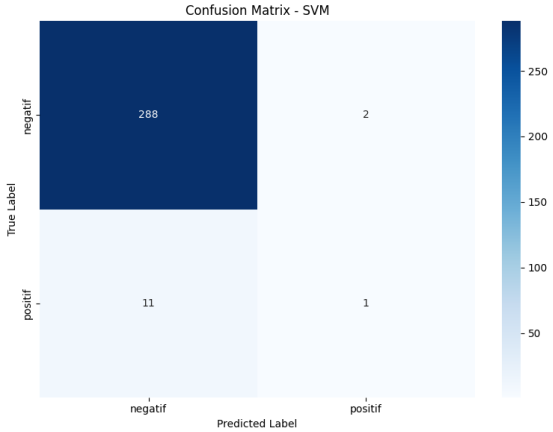
Gambar 2. Hasil evaluasi

4.5. Prediction

Prediction memberikan rincian tentang bagaimana model Decision Tree dan SVM melakukan klasifikasi pada set pengujian yang terdiri dari 302 sampel. Kedua model sama-sama berhasil melakukan 289 prediksi yang benar dan 13 prediksi yang salah secara total. Untuk model Decision Tree, dari 290 sampel 'negatif' yang sebenarnya, 286 berhasil diprediksi dengan benar (True Positives), sementara 4 sampel 'negatif' salah diprediksi sebagai 'positif' (False Negatives). Dari 12 sampel 'positif' yang sebenarnya, hanya 3 yang berhasil diprediksi dengan benar (True Positives), sementara 9 sampel 'positif' salah diprediksi sebagai 'negatif' (False Negatives). Terdapat juga 9 sampel yang sebenarnya 'negatif' namun salah diprediksi sebagai 'positif' (False Positives). Untuk model SVM, dari 290 sampel 'negatif', 288 berhasil diprediksi dengan benar (True Positives), sementara 2 sampel 'negatif' salah diprediksi sebagai 'positif' (False Negatives). Dari 12 sampel 'positif', hanya 1 yang berhasil diprediksi dengan benar (True Positives), sementara 11 sampel 'positif' salah diprediksi sebagai 'negatif' (False Negatives). Terdapat juga 11 sampel yang sebenarnya 'negatif' namun salah diprediksi sebagai 'positif' (False Positives).



Gambar 3. Hasil prediction decision tree



Gambar 4. Hasil prediction svm

5. KESIMPULAN DAN SARAN

Penelitian ini berhasil menganalisis sentimen publik terhadap isu viral “Fufufafa” di platform X serta membandingkan performa algoritma Support Vector Machine (SVM) dan Decision Tree. Berdasarkan data sebanyak 1.007 tweet yang diperoleh melalui crawling dan diproses menggunakan teknik preprocessing, labeling VADER, serta pembobotan TF-IDF, ditemukan bahwa mayoritas opini publik bersentimen negatif. Kedua model mencapai akurasi yang sama yaitu 95,70%, namun Decision Tree menunjukkan F1-score tertimbang sedikit lebih tinggi (0,9515) dibandingkan SVM (0,9444), yang

mengindikasikan keunggulannya pada data dengan distribusi kelas tidak seimbang. Keterbatasan penelitian ini terletak pada ukuran dataset yang relatif kecil dan belum diterapkannya teknik penyeimbangan kelas seperti SMOTE. Untuk penelitian selanjutnya disarankan memperluas periode pengambilan data, menerapkan metode penyeimbangan data, mengeksplorasi algoritma ensemble seperti Random Forest atau Gradient Boosting, serta menguji model pada isu viral lain untuk mengukur kemampuan generalisasi.

DAFTAR PUSTAKA

[1] S. Suryanto and W. Andriyani, "Sentiment Analysis of X Platform on Viral 'Fufufafa' Account Issue in Indonesia Using SVM," *IJCCS (Indonesian J. Comput. Cybern. Syst.*, vol. 19, no. 1, p. 95, 2025, doi: 10.22146/ijccs.104158.

[2] S. W. Ritonga, . Y., M. Fikry, and E. P. Cynthia, "Klasifikasi Sentimen Masyarakat di Twitter terhadap Ganjar Pranowo dengan Metode Naïve Bayes Classifier," *Build. Informatics, Technol. Sci.*, vol. 5, no. 1, 2023, doi: 10.47065/bits.v5i1.3535.

[3] K. Zerrouki, R. M. Hamou, and A. Rahmoun, "Sentiment Analysis of Tweets Using Naïve Bayes, KNN, and Decision Tree," *Int. J. Organ. Collect. Intell.*, vol. 10, no. 4, pp. 35–49, 2020, doi: 10.4018/ijoci.2020100103.

[4] C. Cholifah, Hanny Hikmayanti Handayani, and Ayu Ratna Juwita, "Analisis Sentimen Twitter Terhadap Uu Omnibus Law Menggunakan Metode Support Vector Machine (Svm) Dan Naïve Bayes Classifier (Nbc)," *J. Inform. Teknol. dan Sains*, vol. 4, no. 4, pp. 483–488, 2023, doi: 10.51401/jinteks.v4i4.2191.

[5] D. Duei Putri, G. F. Nama, and W. E. Sulistiono, "Analisis Sentimen Kinerja Dewan Perwakilan Rakyat (DPR) Pada Twitter Menggunakan Metode Naive Bayes Classifier," *J. Inform. dan Tek. Elektro Terap.*, vol. 10, no. 1, pp. 34–40, 2022, doi: 10.23960/jitet.v10i1.2262.

[6] N. Hadi and D. Sugiarto, "Analisis Sentimen Pembangunan IKN pada Media Sosial X Menggunakan Algoritma SVM, Logistic Regression dan Naïve Bayes," *J. Inform. J. Pengemb. IT*, vol. 10, no. 1, pp. 37–49, 2025, doi: 10.30591/jpit.v10i1.7106.

[7] T. Walasary, "Survey Paper tentang Analisis Sentimen," *KONSTELASI Konvergensi Teknol. dan Sist. Inf.*, vol. 2, no. 1, pp. 201–206, 2022, doi: 10.24002/konstelasi.v2i1.5378.

[8] J. Sarwar, S. A. Khan, M. Azmat, and F. Khan, *An Application of Hybrid Bagging-Boosting Decision Trees Ensemble Model for Riverine Flood Susceptibility Mapping and Regional Risk Delineation*, vol. 39, no. 2. 2025. doi: 10.1007/s11269-024-03995-6.

[9] A. Madasu and S. Elango, "Efficient feature selection techniques for sentiment analysis," *Multimed. Tools Appl.*, vol. 79, no. 9–10, pp. 6313–6335, 2020, doi: 10.1007/s11042-019-08409-z.

[10] J. J. A. Limbong, I. Sembiring, and K. D. Hartomo, "Analisis Klasifikasi Sentimen Ulasan pada E-Commerce Shopee Berbasis Word Cloud dengan Metode Naive Bayes dan K-Nearest Neighbor," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 9, no. 2, p. 347, 2022, doi: 10.25126/jtiik.2022924960.

[11] F. A. Dolf, N. Safriadi, and T. Tursina, "Implementasi Sentiment Analysis Berdasarkan Tweets Masyarakat Terhadap Kinerja Presiden dalam Aspek Penanganan Covid-19," *J. Sist. dan Teknol. Inf.*, vol. 10, no. 3, p. 303, 2022, doi: 10.26418/justin.v10i3.54503.

[12] J. M. Moguerza and A. Muñoz, "Support vector machines with applications," *Stat. Sci.*, vol. 21, no. 3, pp. 322–336, 2006, doi: 10.1214/088342306000000493.

[13] Y. Dhebar, S. Gupta, and K. Deb, "Evaluating Nonlinear Decision Trees for Binary Classification Tasks with Other Existing Methods," *2020 IEEE Symp. Ser. Comput. Intell. SSCI 2020*, pp. 2806–2813, 2020, doi: 10.1109/SSCI47803.2020.9308505.

[14] M. Farid, N. #1, S. Ferdiana Kusuma, J. Ngagel, and J. Selatan, "JEPIN (Jurnal Edukasi dan Penelitian Informatika) Analisis Sentimen pada Media Sosial Twitter Terhadap Kebijakan Pemberlakuan Pembatasan Kegiatan Masyarakat Berbasis Deep Learning," *JEPIN (Jurnal Edukasi dan Penelit. Inform.*, vol. 8, no. 1, pp. 44–49, 2022.

[15] H. Najjichah, A. Syukur, and H. Subagyo, "Pengaruh Text Preprocessing dan Kombinasinya Pada Peringkas Dokumen Otomatis Teks Berbahasa Indonesia," *J. Teknol. Inf.*, vol. 15, no. 1, pp. 1–11, 2019, [Online]. Available: http://research.

[16] Jurafsky, "Ch. 2: Regular Expressions, Text Normalization, Edit Distance," 2021.

[17] V. K. Chauhan, K. Dahiya, and A. Sharma, "Problem formulations and solvers in linear SVM: a review," *Artif. Intell. Rev.*, vol. 52, no. 2, pp. 803–855, 2019, doi: 10.1007/s10462-018-9614-6.

[18] S. Suraji, A. C. Fauzan, and H. Harliana, "Penerapan Algoritma Decision Tree C5.0 Untuk Memprediksi Tingkat Kematian Pasien Penyakit Gagal Jantung," *J. Ilm. Intech Inf. Technol. J. UMUS*, vol. 4, no. 02, pp. 216–222, 2022, doi: 10.46772/intech.v4i02.682.

[19] A. Khazaie, N. B. Seghouani, and F. Bugiotti, *Smart Crawling: A New Approach toward*

Focus Crawling from Twitter, vol. 1, no. 1. Association for Computing Machinery, 2021. [Online]. Available: http://arxiv.org/abs/2110.06022

[20] V. W. D. Thomas and F. Rumaisa, "Analisis Sentimen Ulasan Hotel Bahasa Indonesia Menggunakan Support Vector Machine dan TF-IDF," *J. Media Inform. Budidarma*, vol. 6, no. 3, p. 1767, 2022, doi: 10.30865/mib.v6i3.4218.