

Studi Perbandingan *Naïve Bayes* dan *Support Vector Machine* (SVM) dalam Analisis Sentimen Pengguna Metaverse

Sang Dara Parameswari¹, Muharman Lubis², Sinung Suakanto³, Yumna Zahran Ramadhan⁴, Raisyah Nurul Amanah⁵, Revyolla Ananta Dila⁶

^{1,2,3,4,5,6} Jurusan Sistem Informasi, Fakultas Rekayasa Industri, Universitas Telkom

Jl. Telekomunikasi No. 1, Terusan Buahbatu, Bojongsong, Bandung, Jawa Barat, 40257

Email: sangdara@student.telkomuniversity.ac.id, muharmanlubis@telkomuniversity.ac.id, sinung@telkomuniversity.ac.id, yumnazahrn@telkomuniversity.ac.id, raisyahnr1@student.telkomuniversity.ac.id, rvyolla@student.telkomuniversity.ac.id

ABSTRAK

Penelitian ini bertujuan mengevaluasi persepsi publik di Indonesia terhadap isu metaverse melalui analisis sentimen berbasis *text mining*. Metaverse, yang memadukan media sosial, permainan daring, *augmented reality* (AR), *virtual reality* (VR), serta aset digital seperti *cryptocurrency*, semakin mendapat perhatian sejak pengumuman perubahan nama Facebook menjadi Meta pada tahun 2021 dan memunculkan beragam opini publik. Data diperoleh dari Twitter (X) dan dianalisis menggunakan dua algoritma klasifikasi teks, yaitu *Naïve Bayes* dan *Support Vector Machine* (SVM). Dalam penerapannya, *Naïve Bayes* menggunakan fungsi MultinomialNB, sedangkan SVM dijalankan dengan LinearSVC yang lebih sesuai untuk data teks berdimensi tinggi. Hasil penelitian menunjukkan bahwa SVM memberikan kinerja lebih baik dengan akurasi 78,3% dan Macro-F1 78,3%, dibandingkan *Naïve Bayes* yang memperoleh akurasi 72,4% dan Macro-F1 sebesar 60,2%. Selain itu, SVM lebih seimbang dalam mengenali seluruh kelas sentimen, khususnya kategori negatif, sementara *Naïve Bayes* tetap relevan sebagai *baseline* karena kesederhanaan dan efisiensinya. Penelitian ini berkontribusi dalam menyajikan perbandingan komparatif kedua algoritma pada analisis sentimen metaverse di Indonesia, sekaligus membuka ruang bagi pengembangan metode yang lebih mutakhir pada studi berikutnya.

Kata kunci: Analisis Sentimen, Metaverse, *Naïve Bayes*, *Support Vector Machine* (SVM)

ABSTRACT

This study aims to evaluate public perception in Indonesia regarding the metaverse issue through text mining-based sentiment analysis. The metaverse, which combines social media, online games, augmented reality (AR), virtual reality (VR), and digital assets such as cryptocurrency, has gained increasing attention since Facebook announced its name change to Meta in 2021, sparking various public opinions. Data was obtained from Twitter (X) and analysed using two text classification algorithms, Naïve Bayes and Support Vector Machine (SVM). Naïve Bayes uses the MultinomialNB function in its application, while SVM runs with LinearSVC, which is more suitable for high-dimensional text data. The results show that SVM performed better with an accuracy of 78.3% and a Macro-F1 score of 78.3%, compared to Naïve Bayes, which achieved an accuracy of 72.4% and a Macro-F1 score of 60.2%. Furthermore, SVM is more balanced in recognising all sentiment classes, particularly the negative category, while Naïve Bayes remains relevant as a baseline due to its simplicity and efficiency. This research contributes to presenting a comparative analysis of the two algorithms in metaverse sentiment analysis in Indonesia, while also opening up opportunities for developing more advanced methods in future studies.

Keywords: Sentiment Analysis, Metaverse, *Naïve Bayes*, *Support Vector Machine* (SVM)

Pendahuluan

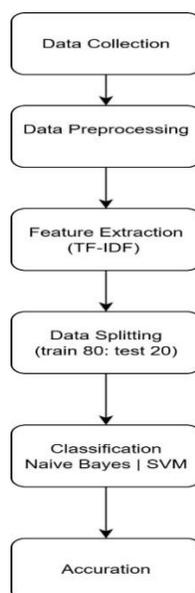
Perkembangan teknologi komunikasi dan informasi yang berlangsung sangat cepat telah membawa dampak signifikan terhadap berbagai bidang kehidupan. Perubahan ini tidak hanya memengaruhi pola interaksi sosial, melainkan juga sistem ekonomi, pendidikan, hingga gaya hidup masyarakat modern. Transformasi tersebut menjadi ciri revolusi industri 4.0, sebuah fase yang menekankan digitalisasi dan otomatisasi dalam hampir seluruh aspek kehidupan manusia. Saat ini, masyarakat juga mulai bergerak menuju era *Society* 5.0, di mana manusia ditempatkan sebagai pusat inovasi untuk menciptakan keseimbangan antara pertumbuhan ekonomi dan penyelesaian persoalan sosial melalui keterhubungan antara dunia nyata dan dunia maya [1]. Salah satu inovasi digital yang mendapat sorotan besar adalah metaverse. Konsep ini menggambarkan ruang virtual yang mengintegrasikan berbagai elemen, termasuk media sosial, gim daring, *augmented reality* (AR), *virtual reality* (VR), dan mata uang kripto. Metaverse memungkinkan penggunaannya untuk berinteraksi, berkolaborasi, sekaligus memperoleh pengalaman yang menyerupai realitas [2] Popularitas konsep ini semakin meningkat setelah Mark Zuckerberg mengumumkan perubahan

nama Facebook menjadi Meta pada tahun 2021, yang menandai komitmen perusahaan dalam mengembangkan ekosistem metaverse [3][4]. Di Indonesia, pemanfaatan metaverse mulai terlihat pada sejumlah bidang, misalnya pendidikan dengan penggunaan *virtual reality* dalam pembelajaran, serta pariwisata dengan simulasi destinasi melalui teknologi *virtual 3D* [5].

Diskursus mengenai metaverse berkembang pesat di ruang digital, khususnya di media sosial Twitter (X). Platform ini menjadi salah satu cara utama bagi masyarakat untuk menyampaikan pendapat, baik berupa dukungan terhadap kesempatan yang ditawarkan maupun kritik terhadap risiko yang mungkin muncul [6][7]. Oleh karena itu, analisis sentimen terhadap percakapan di Twitter (X) relevan dilakukan guna memahami persepsi publik secara lebih komprehensif. Selain itu, studi semacam ini juga penting dalam mengkaji faktor-faktor yang dapat memengaruhi penerimaan masyarakat terhadap inovasi teknologi digital [8]. Mengingat volume data yang sangat besar, analisis manual terhadap opini di media sosial tidaklah efisien. Diperlukan pendekatan berbasis *machine learning* dan *text mining* untuk mengolah data serta mengklasifikasikan sentimen publik [9]. Analisis sentimen, sebagai bagian dari *Natural Language Processing* (NLP), berfungsi untuk menentukan arah opini (positif, negatif, atau netral) yang terkandung dalam teks yang bersumber dari media sosial maupun berbagai dokumen digital. Dalam konteks ini, penelitian memanfaatkan dua algoritma klasifikasi, yakni *Naïve Bayes* dan *Support Vector Machine* (SVM). *Naïve Bayes* dikenal efektif dalam penerapan klasifikasi probabilistik karena modelnya sederhana serta efisien [10]. Di sisi lain, *Support Vector Machine* (SVM) terbukti memiliki kinerja tinggi untuk data berdimensi besar, sebab algoritma ini mampu membentuk *hyperplane* yang optimal dalam memisahkan kelas data [11]. Penelitian Aulia Rahman (2021), misalnya, menggunakan *Support Vector Machine* (SVM) dengan pembobotan TF-IDF untuk menganalisis opini publik terkait kebijakan lockdown di Jakarta melalui Twitter. Hasilnya menunjukkan tingkat akurasi 74%, presisi 75%, *recall* 92%, dan *F1-score* 83%. Kendati demikian, masing-masing algoritma memiliki keterbatasan. Kelebihan *Naïve Bayes* terletak pada kesederhanaannya, tetapi kelemahannya adalah asumsi independensi antar fitur yang sering kali tidak sesuai dengan karakteristik data teks [12], sementara *Support Vector Machine* (SVM) unggul dari sisi akurasi, namun membutuhkan komputasi yang lebih kompleks ketika jumlah dimensi fitur sangat besar [13], [14]. Penelitian ini difokuskan untuk membandingkan kinerja *Naïve Bayes* dan *Support Vector Machine* (SVM) dalam analisis sentimen terkait metaverse di Twitter (X). Perbandingan ini diharapkan dapat memberikan kontribusi terhadap literatur akademik sekaligus menghasilkan pemahaman praktis mengenai bagaimana masyarakat Indonesia merespons fenomena metaverse.

Metode Penelitian

Metode penelitian yang diterapkan pada penelitian ini yaitu menggunakan algoritma *Naïve Bayes* dan *Support Vector Machine* (SVM). Algoritma *Naïve Bayes* diusulkan sebagai metode yang dapat memperbaiki masalah ketidakseimbangan akurasi klasifikasi positif dan akurasi klasifikasi negatif. Dengan algoritma ini kesenjangan antara akurasi klasifikasi positif dan akurasi klasifikasi negatif dapat diminimalkan sekaligus meningkatkan akurasi [15]. Dalam teknik *Naïve Bayes*, ide dasarnya adalah mencari probabilitas kategori yang diberikan pada dokumen teks dengan menggunakan probabilitas gabungan kata dan kategori. Hal ini didasarkan pada asumsi independensi kata [16], sementara algoritma *Support Vector Machine* (SVM) merupakan model baru pada teori pelatihan statistik yang diterapkan untuk memprediksi berbeda parameter. Model pada algoritma *Support Vector Machine* (SVM) merupakan sebuah representasi dari objek data input dalam ruang grafis dengan celah yang jelas antara kelompok titik yang mewakili kategori yang berbeda [17]. Berikut tahapan yang dilakukan:



Gambar 1. Diagram alur metode

Data Collection

Data Collection merupakan tahapan penting dalam mendapatkan informasi mengenai pendapat atau perasaan seseorang terhadap suatu topik, produk, atau fenomena tertentu. Umumnya, data yang dikumpulkan berupa teks yang menyatakan sentimen positif, negatif, atau netral, seperti tweet, ulasan produk, atau komentar di platform media sosial. Tujuan dari proses pengumpulan data ini adalah untuk memperoleh data yang cukup serta mewakili secara baik agar dapat memahami perasaan masyarakat secara lebih mendalam [18] [19]. Dataset diperoleh dari Twitter (X). Jumlah tweet mentah sebanyak 10.779, setelah dibersihkan jumlah data yang diperoleh sejumlah 10.575. Proses pelabelan dilakukan oleh lima anotator independen dan menghasilkan tingkat kesepakatan sangat tinggi dengan rata-rata *Cohen's k* sebesar 1,00. Distribusi kelas setelah pembersihan terdiri dari 4.487 tweet positif, 4.957 tweet netral, dan 1.133 tweet negatif.

Data Preprocessing

Data preprocessing merupakan tahapan krusial dalam analisis sentimen yang bertujuan untuk mempersiapkan data agar siap digunakan oleh algoritma *machine learning*. Tahapan ini dimulai dengan pembersihan teks, yang mencakup penghapusan elemen yang tidak relevan seperti URL, *mention*, hashtag, dan simbol lainnya [20]. Selanjutnya, dilakukan tokenisasi, yaitu pemecahan teks menjadi unit terkecil seperti kata atau token. Penghapusan kata berhenti (*stop words*) juga penting dilakukan untuk menghilangkan kata-kata yang tidak memberikan kontribusi signifikan pada analisis sentimen. Stemming dan lematisasi digunakan untuk mengubah kata-kata menjadi bentuk dasarnya, sehingga mengurangi variasi yang tidak diperlukan [21].

Feature Extraction

Feature Extraction merupakan sebuah proses perubahan data teks menjadi sebuah numerik yang dapat digunakan oleh *machine learning* untuk mengklasifikasi sentimen. Teknik yang digunakan dalam studi ini yaitu dengan melakukan *Term Frequency-Inverse Document Frequency* (TF-IDF), yaitu mengukur pentingnya kata dalam suatu dokumen dengan mempertimbangkan frekuensinya dalam dokumen serta seberapa jarang kata tersebut dalam korpus. Pemilihan teknik *feature extraction* yang tepat sangat bergantung pada sifat data dan tujuan analisis yang diinginkan, dan sering kali beberapa metode dikombinasikan untuk meningkatkan akurasi klasifikasi sentimen [22].

Data Splitting

Penerapan *machine learning*, pembagian data (*data splitting*) merupakan langkah penting yang memisahkan dataset menjadi data latih (*training set*) dan data uji (*testing set*). Data latih berfungsi untuk membangun model, sementara data uji digunakan untuk menilai kemampuan model dalam melakukan generalisasi pada data baru. Tahap ini sangat krusial untuk mencegah terjadinya *overfitting*, yakni ketika model bekerja sangat baik pada data latih tetapi gagal pada data yang belum pernah ditemui. Rasio pembagian yang umum digunakan adalah 80% untuk pelatihan dan 20% untuk pengujian, meskipun dalam beberapa kasus dapat digunakan variasi lain, seperti 70:30 atau pendekatan *k-fold cross-validation*. Penelitian terkini menunjukkan bahwa pemilihan rasio memiliki dampak yang cukup signifikan terhadap kinerja model, dan rasio 80:20 dianggap sebagai standar yang banyak diadopsi dalam berbagai studi [23]. Di sisi lain, metode alternatif seperti *k-fold cross-validation*, Kennard–Stone, maupun SPXY dapat digunakan untuk meningkatkan validitas evaluasi, khususnya pada dataset berukuran kecil atau tidak seimbang [24].

Klasifikasi Menggunakan Algoritma Naïve Bayes dan Support Vector Machine (SVM)

Pada klasifikasi sentimen, teks dikategorikan menjadi tiga kelas utama, yakni positif, negatif, dan netral, untuk menggambarkan opini pengguna. *Naïve Bayes* sering digunakan karena sederhana dan efisien, bekerja dengan menghitung probabilitas kata dalam setiap kategori dan memilih kategori dengan probabilitas tertinggi [25]. Sementara itu, *Support Vector Machine* (SVM) mampu menangani data berdimensi tinggi dan biasanya memberikan akurasi lebih tinggi dengan cara menemukan hyperplane optimal untuk memisahkan kelas-kelas data. Meski demikian, *Support Vector Machine* (SVM) memiliki kelemahan berupa kebutuhan komputasi yang lebih besar saat jumlah fitur tinggi. Penggunaan teknik *preprocessing* dan *feature extraction* yang tepat sangat penting untuk meningkatkan akurasi kedua algoritma dalam analisis sentimen [26][27].

Hasil Dan Pembahasan

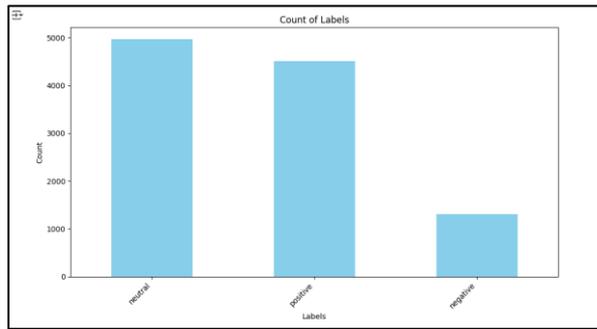
Interpretasi Perilaku Pengguna berdasarkan Penggunaan dan Lima Kata Teratas dari Ulasan Pengguna Metaverse

Interpretasi terhadap perilaku pengguna menunjukkan bahwa sebagian besar responden cenderung bersikap netral terhadap Metaverse. Hal ini diduga disebabkan oleh kurangnya pengetahuan dan pemahaman masyarakat mengenai konsep dan penerapan Metaverse, sehingga sebagian besar data sentimen yang diperoleh menunjukkan respons netral dibandingkan respons positif maupun negatif. Analisis lebih lanjut terhadap kata-kata yang paling sering muncul dalam cuitan Twitter mengidentifikasi lima kata teratas, yang dapat dilihat pada Gambar 2, antara lain:

id	date	lang	tweet	Hilda	Rani	Arul	Akbar	Dita	labeling;;;;;;;;;;	labels
0	1.47E+18	12/23/2021 11:18	en Snoop Dogg Joins The Bored Ape Yacht Club And ...	negative	negative	negative	neutral	neutral	negative;;;;;;;;;	0
1	1.47E+18	12/23/2021 12:31	en @BitSpectra It's time to be part of something ...	positive	neutral	positive	neutral	neutral	neutral;;;;;;;;;	1
2	1.47E+18	12/23/2021 12:37	en So this happened today. I discovered that @Bor...	positive	positive	positive	positive	positive	positive;;;;;;;;;	2
3	1.47E+18	12/23/2021 12:37	en Check out my new Quint Village on OpenSea! h...	positive	positive	positive	positive	positive	positive;;;;;;;;;	2
5	1.47E+18	12/23/2021 12:40	en RUG & SCAM TEST FOR META DECENTRALAND !! C...	negative	negative	negative	negative	negative	negative;;;;;;;;;	0

Gambar 2. Lima kata cuitan teratas dari ulasan pengguna

Gambar 2, memperlihatkan lima data teratas dari kumpulan cuitan Twitter yang disajikan dalam bentuk tabel. Dari data tersebut menunjukkan adanya variasi sentimen, mulai dari positif (misalnya cuitan terkait keterlibatan dalam komunitas) hingga negatif (seperti cuitan yang mengandung istilah “SCAM TEST”) serta sentimen netral.



Gambar 3. Plot data sentimen

Gambar 3, memperlihatkan plot data sentimen yang menggambarkan jumlah distribusi tiga kategori sentimen, yaitu netral, positif, dan negatif. Berdasarkan grafik, jumlah cuitan dengan sentimen netral dan positif terlihat mendominasi dibandingkan dengan sentimen negatif yang relatif lebih sedikit. Adapun proses label encoding dilakukan dengan penetapan kode numerik, yaitu negatif diberi label 0, netral diberi label 1, dan positif diberi label 2.

Analisis dan Evaluasi Performa Algoritma Support Vector Machine (SVM) dan Algoritma Naïve Bayes dan menggunakan Confusion Matrix

Berdasarkan hasil penelitian, peneliti melakukan perbandingan antara algoritma *Support Vector Machine* (SVM) dan algoritma *Naïve Bayes* (NB) dengan memperoleh hasil sebagai berikut:

Algoritma Support Vector Machine (SVM)

```

Classification Report:
              precision    recall  f1-score   support

 negative     0.739     0.466     0.571     262
  neutral     0.759     0.843     0.799     993
  positive     0.821     0.809     0.815     901

 accuracy                   0.783     2156
 macro avg     0.773     0.706     0.728     2156
 weighted avg     0.785     0.783     0.783     2156

Confusion Matrix (rows=true, cols=pred) [neg, neu, pos]:
[[122 106  34]
 [ 31 837 125]
 [ 12 168 729]]
    
```

Gambar 4. Akurasi support vector machine (SVM)

Model yang digunakan untuk *Support Vector Machine* (SVM) yaitu menggunakan fungsi *LinearSVC* yang digunakan dengan representasi fitur *TF-IDF* menghasilkan akurasi sebesar 78,3% pada data uji. Nilai ini lebih tinggi dibandingkan *MultinomialNB* yang hanya mencapai 72,4%. Berdasarkan *classification report*, performa SVM terlihat lebih seimbang di setiap kelas. Untuk kelas negatif, *LinearSVC* memiliki *precision* sebesar 0,739 dan *recall* 0,466, dengan *F1-score* 0,571. Meskipun *recall* pada kelas negatif belum optimal, nilai ini jauh lebih baik dibandingkan *MultinomialNB* yang hanya memperoleh *recall* 0,172. Pada kelas netral, performa *LinearSVC* cukup tinggi dengan *precision* 0,759, *recall* 0,843, dan *F1-score* 0,799. Sementara itu, pada kelas positif, model menunjukkan hasil terbaik dengan *precision* 0,821, *recall* 0,809, dan *F1-score* 0,815. Secara keseluruhan, *macro-F1* mencapai 0,728, menunjukkan bahwa model mampu menjaga keseimbangan performa antar kelas meskipun terdapat perbedaan jumlah data (*imbalance*). *Weighted-F1* sebesar 0,783 juga konsisten dengan hasil akurasi, menandakan prediksi yang stabil sesuai distribusi kelas.

Hasil confusion matrix Gambar 4, memperlihatkan bahwa dari 262 data kelas negatif, sebanyak 122 berhasil diprediksi dengan benar, 106 salah diprediksi sebagai netral, dan 34 salah sebagai positif. Untuk kelas netral (993 data), sebanyak 837 terprediksi benar, sedangkan sisanya 156 salah masuk ke kelas lain. Pada kelas positif (901 data), sebanyak 729 benar diklasifikasikan, 160 salah ke netral, dan 12 ke negatif. Temuan ini menunjukkan bahwa LinearSVC lebih unggul dibanding *Naïve Bayes* dalam mendeteksi sentimen netral dan positif, serta jauh lebih baik dalam mengenali sentimen negatif meskipun masih terdapat tantangan pada *recall*. Kelebihan ini sejalan dengan literatur yang menyebutkan bahwa SVM unggul untuk data teks berdimensi tinggi karena kemampuannya menemukan *hyperplane* pemisah yang optimal.

Algoritma *Naïve Bayes*

```

Classification Report:
              precision    recall  f1-score   support

negative     0.918      0.172   0.289     262
neutral     0.667      0.856   0.750     993
positive     0.799      0.738   0.767     901

 accuracy          0.724     2156
macro avg     0.795      0.589   0.602     2156
weighted avg  0.753      0.724   0.701     2156

Confusion Matrix (rows=true, cols=pred) [neg, neu, pos]:
[[ 45 190  27]
 [  3 850 140]
 [  1 235 665]]
    
```

Gambar 5. Akurasi *naïve bayes* (NB)

Gambar 5 menunjukkan hasil implementasi algoritma *Naïve Bayes* pada data uji. Model menggunakan fungsi MultinomialNB dengan representasi TF-IDF menghasilkan akurasi sebesar 72,4% pada data uji, lebih rendah dibandingkan LinearSVC yang mencapai 78,3%. Berdasarkan *classification report*, kinerja MultinomialNB cenderung kurang seimbang antar kelas. Untuk kelas negatif, meskipun *precision* sangat tinggi (0,918), nilai *recall* hanya 0,172 sehingga *F1-score* hanya 0,289. Hal ini menandakan model cenderung jarang mengenali tweet negatif, meskipun prediksi yang dibuat relatif tepat. Untuk kelas netral, performa relatif lebih baik dengan *precision* 0,667, *recall* 0,856, dan *F1-score* 0,750. Pada kelas positif, hasilnya juga cukup tinggi dengan *precision* 0,799, *recall* 0,738, dan *F1-score* 0,767. Secara keseluruhan, macro-F1 sebesar 0,602, lebih rendah dibandingkan LinearSVC (0,728). Nilai *weighted-F1* sebesar 0,701 juga menunjukkan bahwa model masih cenderung bias terhadap kelas dengan jumlah data yang lebih besar (netral dan positif), sementara kelas negatif sering diabaikan. Confusion matrix memperlihatkan bahwa dari 262 data kelas negatif, hanya 45 yang diprediksi dengan benar, sedangkan 190 salah diprediksi sebagai netral dan 27 sebagai positif. Untuk kelas netral (993 data), sebanyak 850 berhasil diprediksi dengan benar, sisanya salah ke positif (140) atau negatif (3). Pada kelas positif (901 data), sebanyak 665 benar diprediksi, tetapi 235 salah diklasifikasikan sebagai netral dan 1 sebagai negatif.

Analisis Perbandingan Hasil

Tabel 1. Analisa perbandingan hasil

Model	Akurasi	Macro-F1	Negatif	Netral	Positif
LinearSVC (C=1,0)	78,3%	72,8%	57,1%	79,9%	81,5%
MultinomialNB ($\alpha=0,5$)	72,4%	60,2%	28,9%	75,0%	76,7%

Hasil evaluasi menunjukkan bahwa SVM (LinearSVC) lebih unggul dibandingkan *Naïve Bayes* (MultinomialNB) dalam analisis sentimen metaverse. LinearSVC mencapai akurasi 78,3% dan Macro-F1 72,8%, lebih tinggi dibandingkan MultinomialNB dengan akurasi 72,4% dan Macro-F1 60,2%. Keunggulan utama LinearSVC terlihat pada kelas negatif, di mana *recall* meningkat signifikan (46,6% vs 17,2%), sehingga *F1-score* lebih baik. Untuk kelas netral dan positif, kedua algoritma cukup baik, namun LinearSVC tetap lebih seimbang. Secara keseluruhan, LinearSVC lebih efektif pada teks berdimensi tinggi, sedangkan MultinomialNB tetap berguna sebagai baseline karena kesederhanaannya.

Simpulan

Penelitian ini menunjukkan bahwa *Support Vector Machine* (SVM, LinearSVC) memiliki performa lebih baik dibandingkan *Naïve Bayes* (MultinomialNB) dalam analisis sentimen pengguna Twitter terkait metaverse. *Support Vector Machine* (SVM) mencapai akurasi dan Macro-F1 yang lebih tinggi, serta lebih seimbang dalam mengenali semua kelas sentimen, khususnya kategori negatif, sementara *Naïve Bayes* tetap relevan sebagai baseline karena kesederhanaan dan

kestabilannya. Kontribusi utama penelitian ini adalah memberikan evaluasi komparatif yang transparan antara dua algoritma klasik machine learning pada isu metaverse, yang masih jarang dibahas dalam konteks Indonesia. Untuk penelitian selanjutnya, disarankan memperluas ukuran dan periode dataset, mengoptimalkan tahapan pra-pemrosesan teks, serta mengeksplorasi algoritma lain seperti Random Forest, Logistic Regression, atau metode berbasis deep learning. Penerapan hyperparameter tuning dan analisis tambahan melalui *word cloud* atau topik modeling juga direkomendasikan untuk memperoleh hasil yang lebih komprehensif dan representatif.

Daftar Pustaka

- [1] F. Mahiri, A. Najoua, S. Souda Ben, And N. Amini, "From Industry 4.0 To Industry 5.0: The Transition to Human Centricity And Collaborative Hybrid Intelligence," *Journal Of Hunan University Natural Sciences*, Vol. 50, No. 4, Apr. 2023, Doi: 10.55463/Issn. 1674-2974.50.4.8.
- [2] A. Ertanto, V. C. Mawardi, And N. J. Perdana, "Analisa Topik Terhadap Komentar Mengenai Metaverse Menggunakan Metode Clustering K-Means," *Jurnal Ilmu Komputer Dan Sistem Informasi*, Vol. 11, No. 1, Pp. 1–6, 2023, Doi: 10.24912/Jiksi.V11i1.24073.
- [3] A. Nur And Y. D. Putra, "Manajemen Risiko Ekonomi Pada Penerapan Metaverse di Indonesia," 2023.
- [4] S. Yang, H. Joo, And J. Kim, "Metaverse Search System: Architecture, Challenges, And Potential Applications," *ICT Express*, Dec. 2023.
- [5] Casey Newton, "Facebook's CEO On Why the Social Network Is Becoming 'A Metaverse Company'," *The Verge*. Accessed: Jan. 03, 2024. [Online]. Available: <https://www.theverge.com/22588022/Mark-Zuckerberg-Facebook-Ceo-Metaverse-Interview>
- [6] Adhitya Karel Maulaya and Junadhi, "Analisis Sentimen Menggunakan Support Vector Machine Masyarakat Indonesia di Twitter Terkait Bjorka," *Jurnal Coscitech (Computer Science and Information Technology)*, Vol. 3, No. 3, Pp. 495–500, 2022, Doi: 10.37859/Coscitech.V3i3.4358.
- [7] M. Fuad Nasvian and R. Afif, "Public Opinion on Facebook Rebrand to Meta: A Twitter Big Data Analysis On The First 24 Hours After Meta Launched Opini Publik Terhadap Rebrand Facebook Ke Meta: Analisis Data Besar Twitter Pada 24 Jam Pertama Setelah Meta Diluncurkan," 2022.
- [8] P. M. Nirmala Dharmapatni and N. L. P. Merawati, "Penerapan Algoritma Support Vector Machine Dalam Sentimen Analisis Terkait Kenaikan Tarif BPJS Kesehatan," *Jurnal Bumigora Information Technology (Bite)*, Vol. 2, No. 2, Pp. 105–112, Sep. 2020, Doi: 10.30812/Bite.V2i2.904.
- [9] D. Darwis, N. Siskawati, And Z. Abidin, "Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Review Data Twitter BMKG Nasional," Vol. 15, No. 1, 2021.
- [10] V. B. Kobayashi, S. T. Mol, J. Vrolijk, And G. Kismihók, "Text Mining In Career Studies: Generating Insights From Unstructured Textual Data1," *Handbook Of Research Methods In Careers*, Pp. 139–163, 2021, Doi: 10.4337/9781788976725.00015.
- [11] A. R. Isnain, A. I. Sakti, D. Alita, And N. S. Marga, "Sentimen Analisis Publik Terhadap Kebijakan Lockdown Pemerintah Jakarta Menggunakan Algoritma SVM," *Jurnal Data Mining Dan Sistem Informasi*, 2021, [Online]. Available: <https://api.semanticscholar.org/Corpusid:233922818>
- [12] P. J. B. Pajila, B. G. Sheena, A. Gayathri, J. Aswini, M. Nalini, And R. Siva Subramanian, "A Comprehensive Survey On Naive Bayes Algorithm: Advantages, Limitations And Applications," In *Proceedings Of The 4th International Conference On Smart Electronics And Communication, ICOSEC 2023*, Institute Of Electrical And Electronics Engineers Inc., 2023, Pp. 1228–1234. Doi: 10.1109/ICOSEC58147.2023.10276274.
- [13] N. Ardeshir, C. Sanford, And D. Hsu, "Support Vector Machines and Linear Regression Coincide with Very High-Dimensional Features," Oct. 2021, [Online]. Available: <http://arxiv.org/abs/2105.14084>
- [14] R. Guido, S. Ferrisi, D. Lofaro, And D. Conforti, "An Overview on The Advancements of Support Vector Machine Models in Healthcare Applications: A Review," *Information (Switzerland)*, Vol. 15, No. 4, Apr. 2024, Doi: 10.3390/Info15040235.
- [15] M. M. Hamad, M. A. Salih, And R. A. Jaleel, "Sentiment Analysis of Restaurant Reviews In Social Media Using Naïve Bayes," 2021. [Online]. Available: <http://arqiiipubl.com/ams>
- [16] R. Kamali, Y. S. Sari, I. Aldmour, And R. Budiarto, "Verification of Covid-19 Social Assistance Recipients Using Naïve Bayes Classifier," *International Journal of Emerging Multidisciplinaries: Computer Science & Artificial Intelligence*, Vol. 1, No. 2, Pp. 1–12, Sep. 2022, Doi: 10.54938/Ijemdcsai.2022.01.2.100.
- [17] M. H. Santoso, D. A. Larasati, F. Teknik, And U. M. Area, "Wayang Image Classification Using SVM Method and GLCM Feature Extraction," Vol. 4, No. January, 2021.
- [18] P. C. Shilpa, R. Shereen, S. Jacob, And P. Vinod, "Sentiment Analysis Using Deep Learning," In *Proceedings of the 3rd International Conference on Intelligent Communication Technologies and Virtual*

- Mobile Networks, ICICV 2021*, Institute Of Electrical And Electronics Engineers Inc., Feb. 2021, Pp. 930–937. Doi: 10.1109/ICICV50876.2021.9388382.
- [19] R. Varghese, “A Survey on Sentiment Analysis and Opinion Mining.” [Online]. Available: [Http://Www.Ijret.Org](http://www.ijret.org)
- [20] H. T. Duong and T. A. Nguyen-Thi, “A Review: Preprocessing Techniques and Data Augmentation for Sentiment Analysis,” *Comput Soc Netw*, Vol. 8, No. 1, Dec. 2021, Doi: 10.1186/S40649-020-00080-X.
- [21] M. A. Palomino and F. Aider, “Evaluating the Effectiveness of Text Pre-Processing in Sentiment Analysis,” *Applied Sciences (Switzerland)*, Vol. 12, No. 17, Sep. 2022, Doi: 10.3390/App12178765.
- [22] H. J. Aleqabie, M. S. Sfoq, R. A. Albeer, And E. H. Abd, “A Review of Text Mining Techniques: Trends, And Applications in Various Domains,” 2024, *College Of Education, Al-Iraqia University*. Doi: 10.52866/Ijcs. 2024.05.01.009.
- [23] V. R. Joseph, “Optimal Ratio for Data Splitting,” Feb. 2022, Doi: 10.1002/Sam. 11583.
- [24] D. Eliane Birba, “A Comparative Study of Data Splitting Algorithms for Machine Learning Model Selection.”
- [25] S. M. Mohammad, “Sentiment Analysis,” In *Emotion Measurement*, Elsevier, 2021, Pp. 323–379. Doi: 10.1016/B978-0-12-821124-3.00011-9.
- [26] N. Malik and S. Jain, “Comparative Study of Machine Learning Algorithms for Social Media Text Analysis,” In *Communications In Computer And Information Science*, Springer, 2020, Pp. 223–235. Doi: 10.1007/978-981-15-5830-6_19.
- [27] A. Rahman Isnain, N. Hendrastuty, And L. Andraini, “Comparison of Support Vector Machine And Naïve Bayes On Twitter Data Sentiment Analysis,” Vol. 6, No. 1, 2021.