



Prediksi Analisis Sentimen Data Debat Pemilihan Presiden 2024 Menggunakan *Support Vector Machine* (SVM)

Prediction of Sentiment Analysis of 2024 Presidential Election Debate Data Using *Support Vector Machine* (SVM)

Vardina Nava Madya Kusman¹, Vanessa Metayani², Oscar Karnalim³

^{1,2,3}Universitas Kristen Maranatha

Email: ¹12279804@maranatha.ac.id, ²22279803@maranatha.ac.id, ³oscar.karnalim@it.maranatha.edu

INFO ARTIKEL

Sejarah Artikel:

Diterima 13 Februari 2024
Direvisi 28 Maret 2024
Disetujui 9 Juni 2024
Dipublikasi 10 Juni 2024

Katakunci:

Analisis Sentimen
Debat Pemilihan Presiden
Support Vector Machine

ABSTRAK

Penelitian ini bertujuan untuk mengembangkan model klasifikasi menggunakan *Support Vector Machine* untuk menganalisis sentimen pada data dialog debat Pemilihan Presiden tahun 2024. Sentimen dari ucapan tidak selalu dapat diketahui, sehingga model dalam penelitian ini diusulkan untuk menemukan sentimen dibalik ucapan. Untuk dapat memprediksi sentimen, model dilatih dengan data debat pilpres yang telah dikumpulkan. Model kemudian melakukan klasifikasi terhadap data tersebut, dan kemudian diuji tingkat akurasi. Setelah diuji menggunakan data tes, diperoleh nilai akurasi sebesar 52,5%. Hasil tersebut kurang memuaskan, maka dilakukan optimasi terhadap model dan data, Hasilnya, nilai akurasi meningkat menjadi sekitar 94%. Untuk kedepannya, mungkin data yang digunakan bisa semakin ditingkatkan dengan memperhatikan distribusi kelas dalam data.

Keyword:

Presidential Debate
Sentiment Analysis
Support Vector Machine

DOI Artikel:

<https://doi.org/10.35891/explorit.v16i1.4887>

ABSTRACT

This research aimed to develop a classification model using *Support Vector Machine* (SVM) to analyze sentiment in the dialogues of the 2024 Presidential Election debates. Sentiments expressed in sentences are not always straightforward, prompting the proposed model in this study to uncover sentiments behind the dialogues. To predict sentiments, the model was trained on collected data from presidential debate dialogues. The model then classified this data and was tested for its accuracy level. After being tested on the test dataset, an accuracy score of 52.5% was obtained. The results were unsatisfactory; therefore, optimization was performed on both the model and the data. As a result, the accuracy increased to around 94%. For future considerations, improving the data further may involve paying attention to the class distribution within the data.

Explore IT: Jurnal Keilmuan dan Aplikasi Teknik Informatika with CC-BY 4.0 license. Copyright © 2024, the author(s)

1. Pendahuluan

Debat pemilihan presiden adalah sarana bagi masyarakat untuk memperoleh informasi yang mempengaruhi pandangan mereka terhadap calon presiden negara. Di zaman modern ini, informasi dapat menyebar melalui berbagai sumber, seperti media sosial, televisi, dan berbagai platform daring di internet. Pemahaman masyarakat akan sentimen dan topik bahasan dalam debat adalah hal penting agar masyarakat dapat mengevaluasi dengan cermat calon-calon kandidat presiden. Analisis sentimen adalah proses menganalisis teks digital untuk menentukan apakah pesan memiliki makna atau nada emosional pesan (positif, negatif, atau netral) [1]. Analisis sentimen dapat dimanfaatkan untuk ekstraksi opini dari komentar, dokumen, sosial media, *review blog*, dan data lainnya. Penerapan analisis sentimen dapat menggunakan pendekatan pembelajaran mesin untuk proses klasifikasi, pemrosesan data hingga membangun model. Belakangan ini, teknik-teknik analisis sentimen menggunakan kecerdasan buatan menjadi semakin umum dipakai untuk mencari tahu sentimen dari teks. *Natural Language Processing* (NLP) merupakan cabang dari kecerdasan buatan yang memungkinkan komputer memahami, menghasilkan, dan memanipulasi bahasa manusia [2]. *Natural Language Processing* (NLP) memiliki kemampuan untuk mengintegrasikan data dengan teks atau suara. NLP berlaku baik untuk teks tertulis maupun ucapan dan dapat diterapkan pada semua bahasa manusia.

Dalam pemrosesan teks dan analisis dokumen terdapat metode yang umum digunakan dalam ekstraksi dan pembobotan kata, yaitu TF-IDF (Term Frequency-Inverse Document Frequency). Dalam NLP, TF-IDF digunakan untuk memberikan bobot pada kata-kata berdasarkan seberapa sering kata tersebut muncul dalam dokumen tertentu (TF) dan seberapa umum kata tersebut di seluruh koleksi dokumen (IDF). Hal ini membantu mengidentifikasi kata-kata yang unik dan relevan untuk masing-masing dokumen. Menggabungkan TF-IDF dan konsep NLP dapat menghasilkan sistem yang lebih cerdas dan mampu memahami makna dalam teks, membantu dalam pemahaman, klasifikasi, dan ekstraksi informasi dari dokumen-dokumen teks. Penelitian ini bertujuan memprediksi sentimen dari data debat pemilihan presiden 2024. Model akan menggunakan SVM (*Support Vector Machine*) untuk memprediksi atau mengklasifikasikan sentimen (*positive, netral, negative*) berdasarkan teks dari dialog-debat sebelumnya dan sentimen sebelumnya.

2. Kajian Teori

2.1 Analisis Sentimen

Analisis sentimen adalah salah satu proses komputasi dengan menganalisis teks digital yang digunakan untuk menentukan apakah kata-kata atau kalimat yang disampaikan memiliki makna atau emosional pesan. Analisis sentimen merupakan salah satu aplikasi yang menerapkan teknologi NLP yang melatih perangkat lunak komputer untuk memahami teks dengan cara yang mirip dengan manusia. Saat ini, analisis sentimen banyak digunakan untuk mengukur kepuasan pelanggan, *brand monitoring*, sentiment politik, analisis pasar, pemantauan media sosial, dan lain-lain [1].

2.2 TF-IDF

TF-IDF merupakan metode algoritma yang berguna untuk menghitung bobot setiap kata yang umum digunakan. Metode ini juga terkenal efisien, mudah dan memiliki hasil yang akurat. Metode ini akan menghitung nilai TF dan IDF pada setiap token (kata) di setiap dokumen dalam korpus. Secara sederhana, metode TF-IDF digunakan untuk mengetahui berapa sering suatu kata muncul di dalam dokumen [3].

2.3 SVM

SVM merupakan algoritma supervised learning yang dapat digunakan untuk klasifikasi dan regresi. *Structural Risk Minimization* yang dirancang untuk mengolah data menjadi Hyperplane yang mengklasifikasikan ruang input menjadi dua kelas. Teori SVM diawali dengan pengelompokan kasus-kasus linier yang dapat dipisahkan dengan hyperplane dan dibagi menurut kelasnya. Konsep SVM diawali dengan masalah klasifikasi dua kelas sehingga membutuhkan set pelatihan positif dan negatif. SVM akan berusaha mendapatkan hyperplane (pemisah) sebaik mungkin untuk memisahkan kedua kelas dan memaksimalkan margin kedua kelas tersebut [4].

2.4 Studi Literatur

Terdapat beberapa studi terkait analisa sentiment, namun tidak pada konteks debat calon presiden 2024. Widi Widayat [5] membahas tentang analisis sentimen terhadap data teks menggunakan metode *deep learning*, khususnya LSTM dan Word2Vec. Studi ini berfokus pada penggunaan metode *deep learning* untuk mengatasi tantangan dalam menganalisis sentimen terhadap data teks yang panjang dan kompleks. Rahman Isnain *et. al.* [6] melakukan penelitian tentang analisis sentimen terhadap kebijakan pembatasan wilayah yang dilakukan oleh pemerintah Jakarta dalam menghadapi COVID-19. Penelitian menggunakan metode *Support Vector Machine* (SVM) dengan ekstraksi fitur TF-IDF untuk menganalisis opini pengguna Twitter terhadap kebijakan tersebut. Marcelo Jenny *et. al.* [7] menggunakan prosedur analisa sentimen untuk mengukur tingkat ketidaksantunan dalam debat parlemen dengan fokus pada identifikasi dan analisis perilaku atau ucapan yang dianggap tidak pantas atau tidak santun. Peneliti menggunakan pendekatan untuk memvalidasi hasil analisis sentimen dengan membandingkan dengan panggilan tertib (Calls to Order) yang tercatat dalam parlemen Austria.

Yogev Matalon *et. al.* [8] dalam penelitiannya meneliti penggunaan analisis sentimen untuk memprediksi "opinion inversion" (pembalikan opini) dalam *tweet* komunikasi publik. Opinion Inversion dapat merujuk pada perubahan sikap atau pendapat yang signifikan dalam tweet. Teknik analisa sentimen digunakan untuk memahami dan mengklasifikasi sentimen dari *tweet* (positif, negatif, atau netral). Farha Nausheen dan Sayyada Hajer Begum [9], mempelajari atau mengidentifikasi korelasi antara sentimen yang terekspresi di media sosial atau platform online dengan hasil pemilihan politik. Teknik analisis sentimen digunakan untuk mengekstrak opini atau perasaan dari teks yang terkait dengan pemilihan politik. Metodenya dapat melibatkan teknik pemrosesan bahasa alami dan pembelajaran mesin untuk mengklasifikasikan sentimen sebagai positif, negatif, atau netral. Gavin Abercrombie dan Riza Batista Navarro [10] melakukan studi literatur sistematis dari 61 studi yang isinya membahas analisis sentimen dan pendapat yang diungkapkan, serta posisi yang diambil oleh pembicara dalam debat parlemen di Amerika. Metode yang dilakukan adalah dengan menganalisis transkrip debat parlemen dan legislatif lainnya untuk mengakses informasi tentang opini, posisi, dan preferensi kebijakan politisi terpilih.

Pansy Nandwani dan Rupali Verma [11] melakukan penelitian tentang tingkat analisis sentimen, berbagai model emosi, dan proses analisis sentimen dan deteksi emosi dari teks. Para peneliti memanfaatkan *machine learning* untuk mencari tahu sentimen dari teks. Metode *machine learning* yang digunakan adalah NLP untuk memahami, memproses bahasa manusia, menganalisis sentimen, dan pengenalan emosi. Parasian DP Silitonga *et. al.*, [12] melakukan penelitian yang bertujuan untuk mengukur tingkat popularitas program kampus merdeka yang telah diluncurkan oleh Kementerian Pendidikan, Kebudayaan, Riset dan Teknologi Informasi Republik Indonesia berdasarkan data komentar atau opini masyarakat di media sosial. Metode yang dilakukan adalah dengan menggunakan model machine learning SVM. Ulfah *et. al.* [13] membahas penggunaan metode SVM untuk menganalisis hate speech dalam portal berita *online* di Indonesia. Tujuan penelitiannya adalah untuk mengukur akurasi metode SVM dalam mendeteksi komentar hate speech, dengan tingkat akurasi sebesar 53.88%. Penelitian juga melibatkan pengolahan data dan klasifikasi komentar untuk menentukan apakah mereka mengandung hate speech. Netizen di Indonesia. Azhaer Ahmed Efat *et. al.* [14] melakukan analisis sentiment untuk mendeteksi emosi dalam pidato politik PBB. Hasil yang dibuat adalah *Empoliticon*, sebuah model berbasis NLP untuk melakukan tugas tersebut.

Federico Neri *et. al.* [15], menjelaskan cara memantau aktivitas media sosial untuk mengukur loyalitas pelanggan dengan menganalisis sentimen mereka terhadap merek atau produk. Penelitian yang dilakukan menggambarkan sebuah studi Analisis Sentimen yang dilakukan terhadap lebih dari 1000 postingan Facebook tentang siaran berita yang berisi sentimen untuk Rai (layanan broadcast Italia) yang kemudian dibandingkan dengan sentimen terhadap perusahaan swasta yang masih baru dan lebih dinamis, La7. Lin Jiang dan Yoshimi Suzuki [16], membahas isu bahwa tidak semua orang di internet bersifat positif dan menuliskan kata-kata yang tidak menyinggung apalagi menyerang seorang individu atau grup. Hasilnya, metode dengan akurasi deteksi terbaik adalah model yang dibuat dengan menggunakan Bidirectional Recurrent Neural Network (BiRNN). Primandani Arsi dan Retno Waluyo [17], mengusulkan metode SVM untuk diterapkan pada tweets dengan topik pemindahan ibu kota Indonesia yang bertujuan untuk mengklasifikasi macam sentimen pada media sosial twitter. Terdapat 2 kelas untuk klasifikasi yaitu positif dan negatif.

Ilham *et. al.* [18] melakukan penelitian yang bertujuan untuk mengevaluasi F1-Score dalam mengklasifikasikan ujaran kebencian di Twitter menggunakan ekstraksi fitur GloVe dan SVM dengan kernel RBF dan *sigmoid*. Studi menggunakan dataset dari penelitian sebelumnya dan melakukan langkah-langkah preprocessing seperti menambahkan kolom ID, mengonversi teks menjadi huruf kecil, dan menghapus karakter *non-alphanumeric*. Darwis *et. al.* [19] membahas penggunaan algoritma *Support Vector Machine* dan ekstraksi fitur TF-IDF untuk mengklasifikasikan sentimen publik di Twitter terkait kinerja Komisi Pemberantasan Korupsi (KPK) Republik Indonesia. Penelitian juga melibatkan proses *preprocessing*, *labeling* data, *clustering*, dan pengujian untuk mencapai akurasi tinggi dalam klasifikasi sentimen. Styawati *et. al.* [20] membahas analisis sentimen masyarakat Indonesia terhadap program Kartu Prakerja, yang merupakan upaya pemerintah dalam mengatasi pengangguran dan korban PHK tenaga kerja. Metode yang digunakan adalah SVM untuk menganalisis opini masyarakat berdasarkan data yang diperoleh dari media sosial Twitter.

3. Metodologi Penelitian

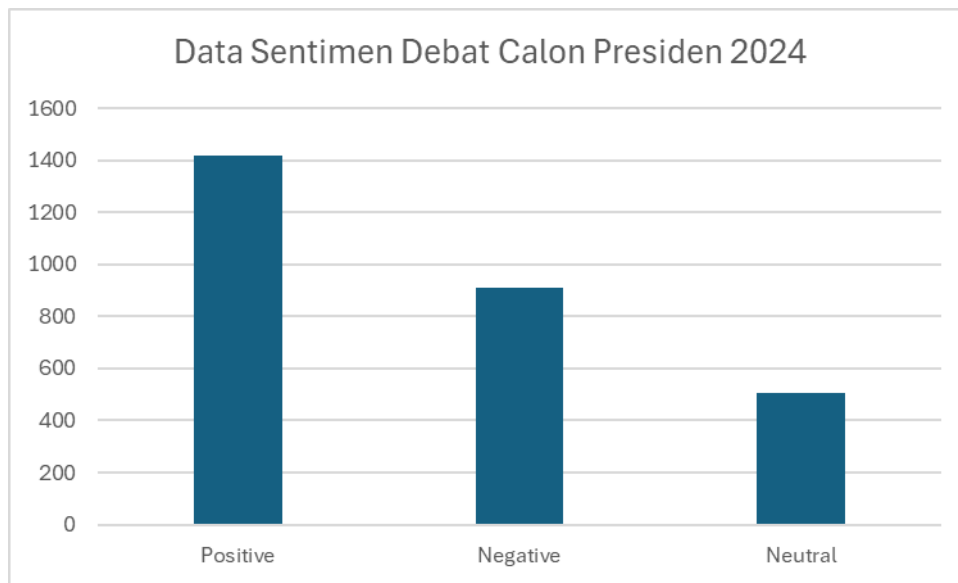
Penelitian terdiri dari beberapa tahap seperti yang diperlihatkan pada Gambar 1. Penelitian diawali dengan pengumpulan data debat pemilihan presiden 2024. Data kemudian dirapikan melalui tahap *preprocessing*. Data kemudian dilengkapi dengan kolom-kolom tambahan data pembanding. Selanjutnya, model analisis sentimen akan dijalankan terhadap data. Hasil eksekusi akan digunakan untuk optimasi model juga optimasi data. Jika dibandingkan dengan studi-studi sejenis, studi ini merupakan studi pertama yang menganalisa sentiment data debat pemilihan presiden Indonesia pada tahun 2024.



Gambar 1. Tahapan Penelitian

3.1 Pengumpulan Data

Data yang digunakan dalam penelitian berupa kalimat-kalimat dari video debat pemilihan presiden yang diambil melalui kanal YouTube dengan bantuan *automatic transcript generator*. Data tersebut kemudian dipecah menjadi kumpulan kalimat berdasarkan tanda baca akhir kalimat. Selanjutnya, data disimpan dalam sebuah dokumen excel berformat *comma separated values* (CSV). Setiap baris akan berisi satu kalimat. Untuk setiap kalimat, akan ditambahkan kolom kedua yaitu nilai sentimen kalimat (*positive*, *neutral*, dan *negative*). Sentimen diberikan dengan cara melakukan observasi manual terhadap kalimat yang diberikan dan konteks percakapan yang ada. Terdapat total 2836 baris data dengan proporsi sentimen dapat dilihat pada Gambar 2. Kalimat dengan sentimen *positive* cukup mendominasi dengan jumlah 1418 kalimat. Sentimen kedua terbanyak ialah sentimen *neutral* yang ditemukan pada 912 kalimat. Sisanya merupakan kalimat dengan sentimen *negative*.



Gambar 2. Proporsi data berdasarkan sentimen

3.2 Preprocessing Data

Untuk setiap baris data, kalimat akan dipreproses menjadi kumpulan fitur melalui beberapa tahapan. Pertama, kalimat akan dikonversi menjadi kumpulan kata dengan bantuan *library NLTK*. Pemisahan kata berdasarkan spasi atau tanda baca. Sebagai contoh, kalimat “argument anda baik sekali” akan dikonversi menjadi empat kata: “argument”, “anda”, “baik”, dan “sekali”. Kedua, setiap kata akan dicari bentuk dasarnya menggunakan Sastrawi *Stemmer*, sebuah algoritma *stemming* Bahasa Indonesia yang merupakan turunan dari *porter stemmer*. Sebagai contoh, kata-kata “berdiskusi” dan “didiskusikan” akan memiliki bentuk dasar yang serupa dengan “diskusi”. Ketiga, kata-kata yang ada akan divalidasi apakah termasuk kategori *stopwords*: kata-kata umum yang sering muncul namun memiliki sedikit kontribusi dalam pemahaman konteks kalimat. Jika termasuk kategori tersebut, kata-kata ini akan dibuang guna efektifitas dan efisiensi. Daftar kata *stopwords* berbahasa Indonesia diperoleh dari *library NLTK*. Contoh kata-kata *stopwords* antara lain “dan”, “kami”, juga “adalah”. Keempat, kata-kata dari semua kalimat digabungkan dan dibentuk menjadi kumpulan kata unik. Kelima, vektor fitur pada setiap kalimat akan dibentuk dengan menghitung kemunculan kata-kata tersebut.

3.3 Pembentukan Data Pembeding

Pada tahap ini, data set yang telah diproses kemudian dibentuk kembali menjadi sebuah file CSV dan tiga kolom ditambahkan untuk menyimpan hasil perbandingan algoritma-algoritma model analisis sentimen. Terdapat tiga kolom yang ditambahkan untuk hasil model TF-IDF, model RBF, dan model RBF dengan optimasi.

3.4 Pembentukan Model

Langkah berikutnya adalah membagi data yang sudah diproses menjadi data *train* dan data *test*. Kemudian dilakukan pembobotan TF-IDF untuk setiap kalimat. Pembobotan ini bertujuan untuk memberikan nilai prioritas lebih tinggi pada kata-kata yang jarang muncul namun representatif. Klasifikasi sentimen menggunakan SVM dengan kernel *linear*. Berdasarkan data yang ada, akurasi yang diperoleh ialah 0,525 atau sebesar 52,5%.

3.5 Optimasi Model

Model yang telah dibuat belum menghasilkan nilai akurasi yang baik, sehingga dilakukan optimasi lebih lanjut. Setelah mencoba beberapa kernel untuk model SVM, diperoleh hasil bahwa menggunakan kernel *radial basis function* (RBF) untuk model SVM memberikan peningkatan nilai akurasi senilai sekitar 3%, atau menjadi sekitar 55,5%. Peningkatan ini tidak tergolong substansial namun tetap dipandang positif.

3.6 Optimasi Data

Setelah mencoba mengoptimasi model, maka selanjutnya dilakukan optimasi data guna memperoleh akurasi yang lebih baik. Data yang digunakan mungkin masih kurang seragam, oleh karena itu dilakukan proses *scaling*. Ketika diuji kembali akurasinya, hasil akurasi berkurang meski menggunakan kernel RBF. SVM adalah algoritma yang mengukur jarak antar titik, sehingga kemungkinan besar data setelah di-*scaling* memiliki pengaruh yang signifikan pada kinerja SVM, terutama pada kernel RBF yang sangat sensitif terhadap perbedaan skala. Akan tetapi, efek *scaling* dapat bervariasi tergantung pada karakteristik spesifik dari dataset dan algoritma yang digunakan.

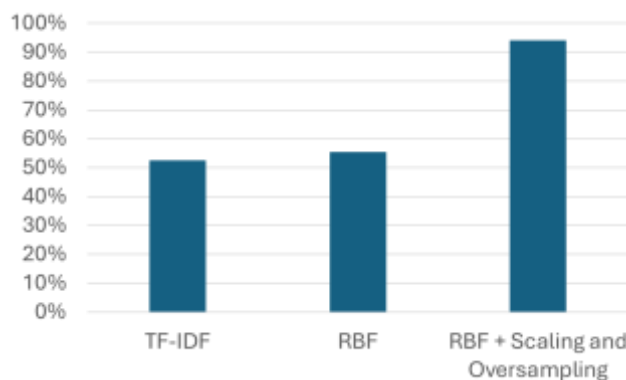
Perlu dilakukan analisis lebih lanjut untuk mencari penyebab *scaling* tidak meningkatkan kinerja model. Dalam beberapa kasus, faktor-faktor seperti distribusi data, varian, dan proporsi kelas mungkin juga berperan dalam hasil yang diamati. Faktor yang diuji adalah faktor masalah distribusi data, sehingga untuk selanjutnya dilakukan proses *oversampling*.

Oversampling bertujuan untuk meningkatkan jumlah sampel pada kelas minoritas, yang mungkin memiliki dampak signifikan terhadap pembelajaran model pada data yang awalnya tidak seimbang. Dengan menghasilkan sampel-sampel buatan atau menambahkan salinan dari sampel-sampel kelas minoritas, *oversampling* menciptakan distribusi data yang lebih seimbang, mengurangi potensi bias terhadap kelas mayoritas, dalam kasus ini data berlabel positif dan netral yang jumlahnya jauh lebih banyak dibanding yang berlabel negatif. Hasilnya, setelah menerapkan *oversampling* menggunakan data yang sudah di-*scaling*, nilai akurasi meningkat menjadi sekitar 94%. Peningkatan ini tergolong sangat substansial dan dirasa lebih dari cukup untuk efektifitas sebuah model analisis sentimen.

4. Hasil Uji Coba Dan Pembahasan

4.1 Hasil Uji Coba

Rangkuman hasil uji coba dapat dilihat pada Gambar 3. Cara kerja model adalah dengan menggunakan data yang telah di-*preprocess* kemudian data dibagi menjadi data train dan data test, lalu membuat vektor TF-IDF dari data tersebut. Dari model yang telah dibuat, diperoleh persentase akurasi sebesar 52,5%. Nilai tersebut tidak begitu baik dan menunjukkan bahwa model belum bekerja optimal. Untuk mencari penyebabnya, dilakukan uji coba parameter dalam pembuatan model untuk melihat apakah nilai akurasi tetap sama. Setelah dicoba, nilai akurasi mengalami peningkatan yang paling tinggi mencapai 55,5% dengan menggunakan kernel RBF sebagai parameter kernel dari model. Sebelumnya kernel yang digunakan adalah *linear* atau garis lurus, sehingga mungkin tidak cocok dengan data yang digunakan. Akan tetapi, peningkatan sebesar 3% tidak begitu signifikan, sehingga penyebab nilai akurasi rendah mungkin disebabkan oleh faktor lain.



Gambar 3. Hasil Uji Coba

Faktor lain yang kemungkinan besar menjadi penyebab dari rendahnya akurasi adalah kualitas dari data yang digunakan, sehingga kemudian beberapa uji coba optimasi data dilakukan. Oleh karena itu, untuk meningkatkan hasil akurasi maka perlu dilakukan pemrosesan data lanjutan. Setelah dilakukan pemrosesan data lanjutan seperti *scaling* dan *oversampling*, hasil akurasi meningkat drastis hingga sebesar 94%. Hasil ini tergolong baik dan model dapat dikategorikan sebagai efektif.

4.2 Pembahasan

Pemilihan kernel linear pada model awal, yang merepresentasikan pendekatan *linear* atau garis lurus, mungkin tidak sepenuhnya cocok dengan kompleksitas pola data yang ada. Kernel RBF, dengan kemampuan untuk menangkap hubungan non-linear dalam data, tampaknya sedikit lebih sesuai dalam konteks ini. Hal ini dikonfirmasi dengan hasil peningkatan akurasi sebesar 3%.

Dari penelitian ini, terdapat beberapa kemungkinan penyebab nilai akurasi yang kurang memuaskan. Faktor-faktor seperti ketidakseimbangan kelas, kekurangan fitur, atau ketidakakuratan label dapat menjadi faktor penyebab akurasi yang belum mencapai tingkat optimal. Beberapa analisis dan uji coba dilakukan agar hasil akurasi dapat meningkat. Dari faktor-faktor tersebut, kemungkinan besar penyebab akurasi rendah disebabkan oleh kualitas data. Setelah dilakukan proses *scaling* dan *oversampling*, hasilnya adalah nilai akurasi yang meningkat menjadi 94%. Hal tersebut mungkin terjadi karena algoritma model SVM, terutama dengan kernel RBF, sangat responsif terhadap distribusi dan perbandingan skala dalam data.

Scaling membantu mengoreksi ketidakseimbangan skala antar fitur, sementara *oversampling* memberikan model lebih banyak data untuk dipelajari pada kelas minoritas, mengurangi kemungkinan ketidakseimbangan kelas. Kombinasi kedua teknik ini memberikan perbaikan signifikan dalam kinerja model, menunjukkan bahwa penanganan ketidakseimbangan kelas dan perbedaan skala sangat penting dalam meningkatkan akurasi SVM pada penelitian ini.

5. Kesimpulan

Penelitian ini mencoba untuk membuat sebuah model menggunakan SVM untuk memprediksi sentimen dari dialog debat pemilihan presiden 2024. Penelitian bertujuan untuk memprediksi sentimen dari dialog. Setelah dilakukan uji akurasi, diperoleh nilai akurasi sebesar 52,5% yang kemudian meningkat menjadi 55,5% dengan mengganti parameter kernel SVM menjadi RBF. Optimasi model menyebabkan kenaikan akurasi. Selain optimasi model, optimasi data lanjutan juga dilakukan agar akurasi dapat meningkat. Setelah melakukan proses *scaling* dan *oversampling* terhadap data, diperoleh peningkatan akurasi menjadi sebesar 94%.

Dari hasil penelitian dan upaya yang dilakukan untuk meningkatkan akurasi model, dapat disimpulkan bahwa mencari model yang sesuai dan pemrosesan data adalah hal-hal penting untuk membuat model yang akurat. Dalam penelitian ini, data berlabel negatif jauh lebih sedikit jika dibandingkan dengan data berlabel positif dan netral.

Saran untuk penelitian ke depannya adalah untuk menggunakan data yang lebih seimbang, karena masalah data yang timbul dalam penelitian ini diakibatkan oleh data yang kurang terstandarisasi dan adanya ketidakseimbangan sampel. Selain itu, nilai akurasi model dalam penelitian ini pun juga mungkin masih dapat ditingkatkan kembali dengan memperhitungkan faktor lainnya.

6. Daftar Pustaka

[1] Y. Kurniawati, "How Does Sentiment Analysis Work?," *School of Information Systems*, Nov. 24, 2023. <https://sis.binus.ac.id/2023/11/24/how->

does-sentiment-analysis-work/ (accessed Jan. 12, 2024).

- [2] “Apa itu Pemrosesan Bahasa Alami (NLP)? - Penjelasan tentang NLP - AWS,” *Amazon Web Services, Inc.*, 2023. <https://aws.amazon.com/id/what-is/nlp/> (accessed Jan. 12, 2024).
- [3] D. Sierra, “Algoritma TF—IDF - Delta Sierra,” *Medium*, Feb. 13, 2019. Accessed: Jan. 12, 2024. [Online]. Available: <https://dltsierra.medium.com/algoritma-tf-idf-633e17d10a80>
- [4] K. Huang and E. Purnomo Putra, “Support Vector Machine Algorithm,” *School of Information Systems*, Feb. 14, 2022. <https://sis.binus.ac.id/2022/02/14/support-vector-machine-algorithm/> (accessed Jan. 12, 2024).
- [5] W. Widayat, “Analisis Sentimen Movie Review menggunakan Word2Vec dan metode LSTM Deep Learning,” *Jurnal Media Informatika Budidarma*, vol. 5, no. 3, p. 1018, Jul. 2021, doi: 10.30865/mib.v5i3.3111.
- [6] 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*, vol. 2, no. 1, p. 31, Feb. 2021, doi: 10.33365/jdmsi.v2i1.1021.
- [7] M. Jenny, M. Haselmayer, and D. Kapla, “Measuring incivility in parliamentary debates: validating a sentiment analysis procedure with calls to order in the Austrian Parliament,” in *Political Incivility in the Parliamentary, Electoral and Media Arena*, Routledge, 2021, pp. 56–66. Accessed: Jan. 12, 2024. [Online]. Available: <http://dx.doi.org/10.4324/9781003029205-3>
- [8] Y. Matalon, O. Magdaci, A. Almozlino, and D. Yamin, “Using sentiment analysis to predict opinion inversion in Tweets of political communication,” *Scientific Reports*, vol. 11, no. 1, Mar. 2021, doi: 10.1038/s41598-021-86510-w.
- [9] F. Nausheen and S. H. Begum, “Sentiment analysis to predict election results using Python,” in *2018 2nd International Conference on Inventive Systems and Control (ICISC)*, Jan. 2018. Accessed: Jan. 12, 2024. [Online]. Available: <http://dx.doi.org/10.1109/icisc.2018.8399007>
- [10] G. Abercrombie and R. Batista-Navarro, “Sentiment and position-taking analysis of parliamentary debates: a systematic literature review,” *Journal of Computational Social Science*, vol. 3, no. 1, pp. 245–270, Jan. 2020, doi: 10.1007/s42001-019-00060-w.
- [11] P. Nandwani and R. Verma, “A review on sentiment analysis and emotion detection from text,” *Social Network Analysis and Mining*, vol. 11, no. 1, Aug. 2021, doi: 10.1007/s13278-021-00776-6.
- [12] P. D. Silitonga, I. S. Morina, M. Hasibuan, and U. Lestari, “Analisis Sentimen Kampus Merdeka Menggunakan Machine Learning,” *Jurnal ICT: Information Communication & Technology*, vol. 21, no. 1, Jul. 2022.
- [13] A. N. Hidayat, “Analisis Sentimen Terhadap Wacana Politik Pada Media Massa Online Menggunakan Algoritma Support Vector Machine dan Naive Bayes,” *Jurnal Elektronik Sistem Informasi dan Komputer (JESIK)*, vol. 1, no. 1, 2015.
- [14] A. A. Efat, A. Atiq, A. S. Abeed, A. Momin, and Md. G. R. Alam, “Empoliticon: NLP and ML Based Approach for Context and Emotion Classification of Political Speeches From Transcripts,” *IEEE Access*, vol. 11, pp. 54808–54821, 2023, doi: 10.1109/access.2023.3282162.
- [15] F. Neri, C. Aliprandi, F. Capeci, M. Cuadros, and T. By, “Sentiment Analysis on Social Media,” in *2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, Aug. 2012. Accessed: Jan. 12, 2024. [Online]. Available: <http://dx.doi.org/10.1109/asonam.2012.164>
- [16] L. Jiang and Y. Suzuki, “Detecting hate speech from tweets for sentiment analysis,” in *2019 6th International Conference on Systems and Informatics (ICSAI)*, Nov. 2019. Accessed: Jan. 12, 2024. [Online]. Available: <http://dx.doi.org/10.1109/icsai48974.2019.9010578>
- [17] P. Arsi and R. Waluyo, “Analisis Sentimen Wacana Pemindahan Ibu Kota Indonesia Menggunakan Algoritma Support Vector Machine (SVM),” *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 8, no. 1, p. 147, Feb. 2021, doi: 10.25126/jtiik.0813944.
- [18] A. Ilham, N. Azmi Verdikha, and A. Johar Latipah, “Klasifikasi Ujaran Kebencian di Twitter Menggunakan Fitur Ekstraksi Glove dengan Support Vector Machine (SVM),” *Jurnal Keilmuan dan Aplikasi Teknik Informatika*, vol. 15, no. 2, pp. 64–72, Dec. 2023, doi: 10.35891/explorit.v15i2.4108.
- [19] D. Darwis, E. S. Pratiwi, and A. F. O. Pasaribu, “Penerapan Algoritma SVM Untuk Analisis Sentimen Pada Data Twitter Komisi Pemberantasan Korupsi Republik Indonesia,” *EduTic - Scientific Journal of Informatics Education*, vol. 7, no. 1, Nov. 2020, doi: 10.21107/edutic.v7i1.8779.
- [20] Styawati, N. Hendrastuty, and A. R. Isnain, “Analisis Sentimen Masyarakat Terhadap Program Kartu Prakerja Pada Twitter Dengan Metode Support Vector Machine,” *Jurnal Informatika: Jurnal Pengembangan IT*, vol. 6, no. 3, pp. 150–155, Oct. 2021, doi: 10.30591/jpit.v6i3.2870.