

## Comparative Sentiment Analysis of Election News Articles with Smote using Classification Algorithm

Fathir Fathir<sup>1\*</sup>, Afsa Rizki<sup>1</sup>, Yuliyanti Yuliyanti<sup>1</sup>, Siti Mutmainah<sup>1</sup>

<sup>1</sup>Universitas Muhammadiyah Bima, Bima, Indonesia

[fathirpuncak@gmail.com](mailto:fathirpuncak@gmail.com)\*

| Received: 23/07/2024 | Revised: 29/07/2024 | Accepted: 30/07/2024 |

Copyright©2024 by authors, all rights reserved. Authors agree that this article remains permanently open access under the terms of the Creative Commons Attribution License 4.0 International License

### Abstrak

Penelitian ini berfokus pada analisis sentimen terhadap artikel pemberitaan mengenai pemilihan umum terutama presiden dan wakil presiden dengan membandingkan performa algoritma klasifikasi, khususnya *Decision Tree* dan *K-Nearest Neighbors* (KNN), serta mengevaluasi efektivitas teknik SMOTE (*Synthetic Minority Over-sampling Technique*) dalam mengatasi masalah ketidakseimbangan data atau dataset menunjukkan bahwa jumlah data yang memiliki sentimen positif lebih banyak dari pada sentimen berupa negatif. Tujuan utama dari penelitian ini adalah untuk menentukan algoritma yang lebih unggul dalam klasifikasi sentimen dan melihat bagaimana SMOTE dapat meningkatkan kinerja model tersebut. Dataset di scraping dan dilakukan untuk normalisasi teks, penghapusan *stop words*, dan ekstraksi fitur. SMOTE diterapkan untuk menyeimbangkan kelas dalam dataset, sehingga mengatasi ketidakseimbangan yang sering terjadi pada data sentimen. Algoritma digunakan *Decision Tree* dan KNN. Hasil penelitian menunjukkan bahwa *Decision Tree* secara konsisten memberikan performa yang lebih baik dibandingkan KNN dalam hal akurasi sebesar 85%, presisi 44%, recall 47%, dan skor F1 45%. Penerapan SMOTE terbukti meningkatkan kinerja kedua algoritma, namun pengaruhnya lebih signifikan pada *Decision Tree*. Dengan demikian, penelitian ini menyimpulkan bahwa *Decision Tree*, dikombinasikan dengan SMOTE, merupakan pendekatan yang lebih efektif dan andal untuk analisis sentimen artikel pemilihan umum dibandingkan dengan KNN. Hasil ini memberikan kontribusi penting dalam pengembangan metode analisis sentimen yang dapat diterapkan untuk memahami dinamika opini publik dalam konteks politik.

Kata kunci: Analisis sentiment, SMOTE, KNN, *decision tree*.

### Abstract

*This research focuses on sentiment analysis of news articles about general elections, especially the president and vice president by comparing the performance of classification algorithms, especially Decision Tree and K-Nearest Neighbors (KNN), and evaluating the effectiveness of the SMOTE (Synthetic Minority Over-sampling Technique) technique in overcoming the problem of data imbalance or the dataset shows that the amount of data that has positive sentiment is more than negative sentiment. The main objective of this research is to determine which algorithm is*

*superior in sentiment classification and see how SMOTE can improve the performance of the model. The dataset was scraped and subjected to text normalization, stop words removal, and feature extraction. SMOTE was applied to balance the classes in the dataset, thus overcoming the imbalance that often occurs in sentiment data. Decision Tree and KNN algorithms were used. The results showed that Decision Tree consistently performed better than KNN in terms of 85% accuracy, 44% precision, 47% recall, and 45% F1 score. The application of SMOTE is proven to improve the performance of both algorithms, but the effect is more significant on Decision Tree. Thus, this study concludes that Decision Tree, combined with SMOTE, is a more effective and reliable approach for sentiment analysis of election articles than KNN. These results make an important contribution to the development of sentiment analysis methods that can be applied to understand the dynamics of public opinion in a political context.*

*Keywords: sentiment analysis, SMOTE, KNN, decision tree.*

## **1. Pendahuluan**

Media pemberitaan online menjadi konsumsi utama masyarakat dalam pergerakan informasi dan komunikasi yang ada di Indonesia, sehingga informasi mengalami pergerakan yang cepat di kalangan masyarakat, terutama, pemberitaan seputar pelaksanaan pemilu serentak presiden dan dewan perwakilan rakyat (Mudjiyanto & Dunan, 2020). Konsumsi berita online yang meningkat pesat dan tidak terkendali menyebabkan asumsi masyarakat menjadi sulit terkendali (Eddyono, 2021) (Keputusan Dirjen Penguatan Riset dan Pengembangan Ristek Dikti et al., 2017; Mudjiyanto & Dunan, 2020), baik mencakup berita-berita bersifat positif maupun negatif atau pemberitaan yang mencakup fakta dan berita palsu (Habibah, 2021). Dampaknya terlihat pada pemahaman masyarakat terhadap politik, khususnya terkait pelaksanaan pemilu serentak (Al-Azani & El-Alfy, 2017).

Oleh karena itu, media penyampai informasi perlu memberikan berita yang independen dan disampaikan dengan baik tanpa unsur-unsur yang dapat merugikan masyarakat. Perlu ada pendidikan politik di tengah masyarakat untuk mencegah penyebaran berita online yang bersifat hoaks yang dapat menyebabkan perpecahan. Hal ini penting agar beberapa topik pembahasan dapat diklasifikasikan oleh algoritma machine learning berdasarkan sentimen positif dan negatif (Lazuardi & Juarna, 2023; Surya Gemilang et al., 2024; Toruan et al., 2023).

Sentimen analisis sering digunakan untuk melihat pandangan positif dan negatif terhadap sebuah topik yang dibahas, khususnya pemodelan topik pemberitaan untuk kebutuhan sentimen analisis topik-topik hangat, pemberitaan menjadi sangat diminati, bukan hanya sebagai sarana untuk menyampaikan informasi terkait kebijakan politik dan ekonomi suatu pemerintahan, tetapi juga sebagai bagian dari ekspresi masyarakat luas (Supriatna & Rohman, 2024). Pemberitaan di negara demokrasi dianggap sebagai wujud dari kebebasan berpendapat yang dijamin oleh undang-undang kebebasan pers, khususnya Pasal 28F. Undang-undang ini melibatkan hak masyarakat dan media pers untuk mencari, memperoleh, memiliki, menyimpan, mengolah, dan menyampaikan informasi (Setiawan & Isnain, 2024). Meskipun demikian, perlu dicatat bahwa beberapa media mungkin melakukan tindakan sebaliknya demi kepentingan tertentu (Satrio et al., 2024).

Algoritma KNN juga sering digunakan dalam model klasifikasi(A'yuniyah & Reza, 2023; Loka & Marsal, 2023), hasilnya sangat baik di beberapa dataset yakni 96% dengan jumlah dataset 3500, sehingga model ini patut untuk di coba pada beberapa dataset pemberitaan mengenai pemilu(Yani et al., 2020). Dan Model *lexicon based* sering digunakan pada penelitian machine learning dalam mengoptimasi hasil-hasil opini dalam analisis sosial media baik di tweet sosial media lainnya(Azhar, 2017, 2018; Ibrahim et al., 2022; Kholifah et al., 2024), dipadukan dengan model *Latent Dirichlet Allocation* (LDA) memadukan topik-topik hangat yang di bahas dalam struktur opini yang diteliti sehingga beberapa penelitian menunjukkan bahwa pendekatan identifikasi dan pelabelan topik yang digunakan cukup dinamis dan efektif dalam menghasilkan akurasi(Winarso et al., 2021).

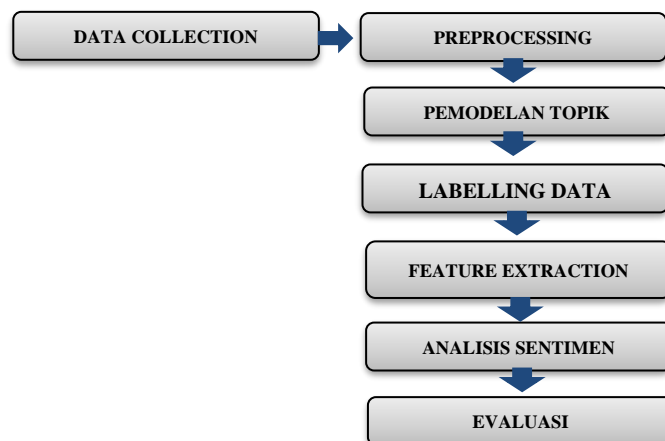
Algoritma K-Nearest Neighbor juga digunakan dalam beberapa sentimen analisis seperti pada penelitian Yuniyah dalam klasifikasi, dimana hasil yang di dapatkan akurasi maksimal sejumlah nilai k=3 dengan akurasi optimal 93% dan rata-rata presisi 88% jumlah itu menjadi nilai paling tinggi yang dapatkan(A'yuniyah & Reza, 2023; Loka & Marsal, 2023; Putra et al., 2024; Yani et al., 2020).

Penerapan smote pada beberapa dataset yang tidak seimbang sering kali menjadi alternatif untuk memperbaiki model algoritma yang dihasilkan berupa akurasi dan presisi algoritma(Hasan, 2024; Lase et al., 2021; Nooryuda Prasetya et al., 2021; Winarso et al., 2021), sehingga pada penelitian ini penting untuk menerapkan algoritma smote tanpa merubah jumlah dataset mayoritas(Shahriar et al., 2023), ketidakseimbangan yang terjadi di akibatkan kelas positif dan negatif dengan menghasilkan sampel sintesis untuk kelas minoritas sehingga distribusi kelas menjadi lebih seimbang(Mohasseb et al.,2018).

Sehingga tujuan penelitian ini adalah untuk memahami pola-pola sentimen dalam beberapa artikel tentang pemilu, dengan membandingkan algoritma klasifikasi yang digunakan dengan beberapa teknik smote atau teknik *Synthetic Minority Over-sampling Technique* dalam menyelesaikan data-data yang tidak seimbang bisa mengukur pengaruh distribusi data dan performas sebuah model yang dihasilkan.

## 2. Metodologi Penelitian

Pada bagian ini ada beberapa tahapan-tahapan penelitian yang akan di dijelaskan pada diagram alir pada gambar 2.1 berikut ini.



Gambar 1 Tahapan Penelitian

## 2.1. Data Collection

Pada penelitian ini pengumpulan data dilakukan dengan cara web scraping yakni mengambil data artikel berita, data artikel berita yang di scraping berjumlah 5900 data, dengan periode pengambilan mulai dari agustus 2023 sampai dengan desember 2023, selanjutnya melakukan preprocessing dan menganalisis dengan algoritma klasifikasi.

Metode scraping yang dilakukan menggunakan python dengan google colab, proses pengambilan data sebagai berikut:

<i>No</i>	<i>content</i>
1	<i>Ketua Umum Partai NasDem Surya Paloh menanggapi...</i>
2	<i>Masa jabatan Anies Baswedan menjadi Gubernur D...</i>
3	<i>Gubernur Jawa Tengah (Jateng) Ganjar Pranowo m...</i>
4	<i>Ketua Umum Partai NasDem Surya Paloh mengungkap...</i>
5	<i>Wakil Ketua Umum Partai Gelora Fahri Hamzah be...</i>
6	.....

## 2.2. Preprocessing

Pada tahap preprocessing dilakukan untuk memahami dan meningkatkan kualitas data yang digunakan seperti membersihkan data dan mengembalikan kata dalam data ke bentuk standar sehingga dapat mempermudah komputer dalam memproses data.

- 1) Pembersihan Teks: Melakukan penghapusan tanda baca, emotikon, angka, tautan/URL, dan pengonversian huruf kecil untuk menyamakan huruf pada umpan balik.
- 2) Tokenisasi Teks: Memisahkan kalimat pada umpan balik menggunakan karakter spasi menjadi kata-kata.
- 3) Normalisasi: Mengubah kata dalam umpan balik dari bentuk singkatan, slang, atau bahasa alay menjadi kata standar menggunakan kamus slang berbahasa Indonesia.
- 4) Penghapusan Stopword: Menghapus kata-kata yang umumnya besar dan sering muncul tetapi tidak memiliki makna.
- 5) Stemming: Mereduksi awalan dan akhiran kata pada umpan balik.

## 2.3. Labelling

Setelah melakukan pengumpulan data dan preprocessing proses selanjutnya dalam penelitian ini yaitu melakukan pelabelan data. Lexicon based merupakan metode feature klasifikasi data yang dapat digunakan pada data yang tidak memiliki label (Utomo et al., 2024). Proses labelisasi menggunakan algoritma lexicon based. Pada proses pelabelan didapatkan hasil bawah konten yang termasuk dalam class positif sebanyak 5000 content class negatif sebanyak 900 konten berita.

## 2.4. Feature Extraction

Penelitian ini memanfaatkan Word2Vec untuk melakukan embedding kata. Word2Vec mengonversi teks menjadi representasi dalam bentuk vektor (Mustasaruddin et al., 2023). Hasil

dari proses ini adalah vektor representasi kata-kata dalam feedback yang akan digunakan untuk melatih model.

## 2.5. Analisis Sentimen

Pembagian dataset untuk training dan testing pada penerapan algoritma Decision Tree (Es-Sabery et al., 2021; Syamala & Nalini, 2020) dan algoritma Algoritma K-Nearest Neighbor (Harun & Putri Ananda, 2021; Putu et al., 2021) yakni sejumlah 80/20 sampel dataset, menggunakan data yang sudah diproses untuk membangun model machine learning yang dapat mengklasifikasikan sentimen teks.

## 2.6. Evaluasi

Pada tahap evaluasi atau tahap terakhir akan diuji model yang dibangun. Algoritma tersebut kemudian dievaluasi menggunakan metrik akurasi, presisi, recall, dan f1-score untuk perbandingan yang lebih komprehensif. Pada parameter yang prediksi yaitu TP = Jumlah kelas Positif yang diprediksi benar, TN = Jumlah kelas yang diprediksi Negative benar, Tn = Jumlah kelas Neutral yang diprediksi benar, FN = Jumlah kelas Positif yang diprediksi salah, FP = Jumlah Kelas Negative yang diprediksi salah.

Adapun rumus confusion matrix yang digunakan adalah sebagai berikut:

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (2.1)$$

$$Precision = \frac{TP}{TP+FP} \quad (2.2)$$

$$Recall = \frac{TP}{TP+FN} \quad (2.3)$$

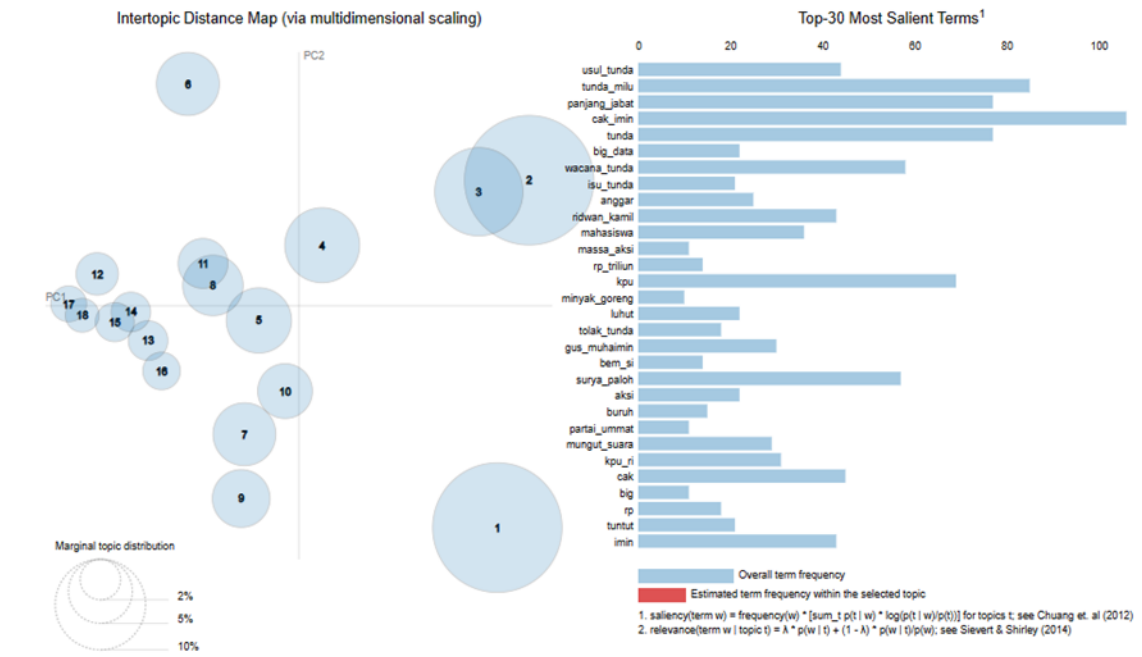
$$Specificity = \frac{TN}{TN+FP} \quad (2.4)$$

$$F1 - Score = \frac{2TP}{2TP+FP+FN} \quad (2.5)$$

## 3. Hasil dan Pembahasan

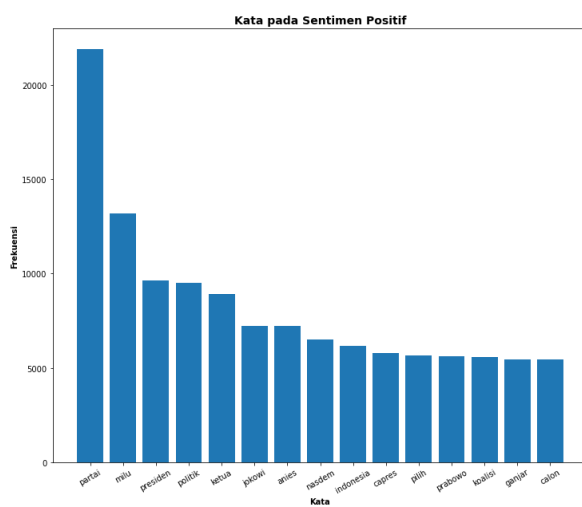
### 3.1 Hasil Pengumpulan Data

Fokus penelitian ini adalah pada uji coba algoritma dengan membandingkan dengan penggunaan algoritma smote sebagai pembanding, melihat dan menilai hasil evaluasi dengan confusion matriks berupa akurasi, presisi dari algoritma klasifikasi tersebut.

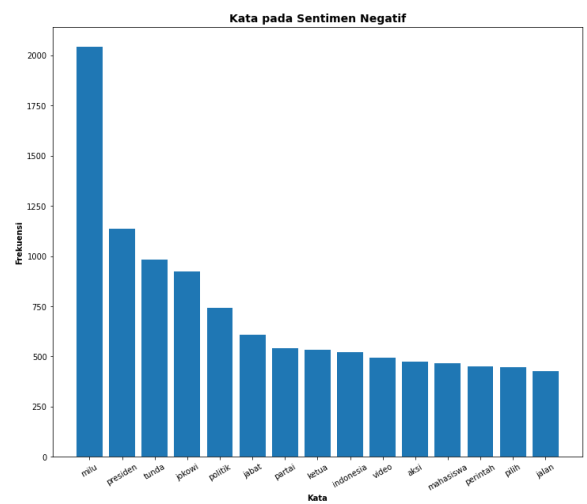


Gambar 2 mapping intertopik media

Berdasarkan gambar 2 diatas menggambarkan atau menghitung ukuran dan visualisasi jarak antara kata atau kalimat untuk memahami titik dan objek dalam suatu domain tertentu, dari gambar tersebut kita pahami bahwa sentimen yang paling banyak di bahas adalah persoalan tunda pemilu dan beberapa nama tokoh politisi yang berpartisipasi dalam ajang pemilihan presiden 2024, term kata-kata tertinggi yang dibahas dan di filter menjadi 30 kata paling populer.



Gambar 3 presentasi kata-kata sentimen



Gambar 4 presentasi kata-kata sentimen

Pada gambar 3 dan gambar 4 di atas menggambarkan jumlah kata yang di tampilkan menggunakan visualisasi berupa partai politik menjadi keyword yang banyak di bahas dalam artikel pemberitaan. Persebaran kata sentimen dapat dilihat dari visualisasi wordcloud pada Gambar 3.2 dan Gambar 3.3 berdasarkan kalimat-kalimat yang mengandung tunda pemilu misalnya, dan pada feedback sentimen positive menggambarkan wordcloud sentimen positif berdasarkan data preprocessing (Yang, 2018; Zelina & Afiyati, 2024).



Gambar 5 presentasi cloud sentimen

Sentimen dalam teks, seperti ulasan, komentar media sosial, artikel berita, dan berbagai bentuk teks lainnya. Analisis sentimen adalah proses mengidentifikasi dan mengkategorikan opini yang diekspresikan dalam sepotong teks.

### 3.2 Evaluasi Performa

Pengukuran performa peningkatan kinerja berdasarkan memberikan umpan balik konstruktif untuk meningkatkan kinerja algoritma, pertama menggunakan algoritma smote dan yang kedua yakni tanpa menggunakan algoritma smote, seperti di bawah ini:

#### a. Smote

Algoritma *decicision tree* dalam mengklasifikasi membagi data menjadi beberapa subset dengan membuat keputusan berdasarkan fitur-fitur data pada setiap node sampai pada proses selanjutnya, sedangkan KNN dalam mengklasifikasi data baru harus menentukan nilai “K” dari data terdekat dan menggunakan suara mayoritas dari kelas tetangga terdekat.

Tabel 1 algoritma decision tree

		<i>True class</i>	
		<i>Positif</i>	<i>Negatif</i>
<i>Predict</i>	<i>Positif</i>	44	60
	<i>Negatif</i>	92	965

<i>Decision Tree Accuracy</i>	85%
<i>Decision Tree Precision</i>	44%
<i>Decision Tree Recall</i>	47%
<i>Decision Tree f1_score</i>	45%

Dari tabel tersebut maka nilai akurasi dari algoritma Decision Tree adalah 85%, kemudian hasil precision adalah 44%, dan nilai recall 47%, nilai f1-score adalah 45%. Pada penelitian ini juga digambarkan dari matrik di atas, dimana nilai true positive (TP) adalah 35 data, dan nilai sebaliknya adalah false positive (FP) 74 data, dan nilai true negative (TN) 1054 dan nilai false negative (FN) yakni 10 data.

Tabel 2 algoritma knn

		<b><i>True class</i></b>	
		<i>Positif</i>	<i>Negatif</i>
<b><i>Predict</i></b>	<i>Positif</i>	55	21
	<i>Negatif</i>	82	501
<hr/>			
<i>KNN Accuracy</i>		47%	
<i>KNN Precision</i>		89%	
<i>KNN Recall</i>		47%	
<i>KNN f1_score</i>		60%	

Dari tabel tersebut maka nilai akurasi dari algoritma *K-Nearest Neighbor* adalah 47%, kemudian hasil precision adalah 89 %, dan nilai recall 47%, nilai f1-score adalah 60%. Pada penelitian ini juga digambarkan dari matrik di atas, dimana nilai true positive (TP) adalah 55 data, dan nilai sebaliknya adalah false positive (FP) 21 data, dan nilai true negative (TN) 82 dan nilai false negative (FN) yakni 501 data.

b. Without smote

Tabel 3 algoritma decision tree

		<b><i>True class</i></b>	
		<i>Positif</i>	<i>Negatif</i>
<b><i>Predict</i></b>	<i>Positif</i>	47	61
	<i>Negatif</i>	55	1004
<hr/>			
<i>Decision Tree Accuracy</i>		88%	
<i>Decision Tree Precision</i>		88%	
<i>Decision Tree Recall</i>		89%	
<i>Decision Tree f1_score</i>		89%	



Dari tabel tersebut maka nilai akurasi dari algoritma Decision Tree adalah 88%, kemudian hasil precision adalah 88%, dan nilai recall 89%, nilai f1-score adalah 89%. Pada penelitian ini juga digambarkan dari matrik di atas, dimana nilai true positive (TP) adalah 47 data, dan nilai sebaliknya adalah false positive (FP) 61 data, dan nilai true negative (TN) 55 dan nilai false negative (FN) yakni 1004 data.

Tabel 4 algoritma knn

		<i>True class</i>	
		<i>Positif</i>	<i>Negatif</i>
<i>Predict</i>	<i>Positif</i>	26	36
	<i>Negatif</i>	25	564

<i>KNN Accuracy</i>	50%
<i>KNN Precision</i>	89%
<i>KNN Recall</i>	50%
<i>KNN f1_score</i>	64%

Dari tabel tersebut maka nilai akurasi dari algoritma K-Nearest Neighbor adalah 50 %, kemudian hasil precision adalah 89%, dan nilai recall 50%, nilai f1-score adalah 64%. Pada penelitian ini juga digambarkan dari matrik di atas, dimana nilai true positive (TP) adalah 26 data, dan nilai sebaliknya adalah false positive (FP) 36 data, dan nilai true negative (TN) 25 dan nilai false negative (FN) yakni 564 data.

#### 4. Kesimpulan

Komparasi yang dilakukan pada penelitian ini yakni dengan menggunakan algoritma decision tree dan algoritma K-Nearest Neighbor, dimana dataset mengalami ketidakseimbangan sehingga perlu menerapkan smote (Synthetic Minority Over-sampling Technique), pembagian datanya 80% training dan data testing 20% sehingga menghasilkan eksperimen pertama tanpa menggunakan algoritma smote adalah 88% akurasi pada algoritma decision tree, sedangkan algoritma knn menghasilkan akurasi sebesar 50% dan eksperimen kedua menggunakan algoritma smote 47% dengan algoritma knn, sedangkan algoritma decision tree menghasilkan akurasi 85%, Hasil penelitian ini menunjukkan bahwa kombinasi antara analisis sentimen dan teknik klasifikasi memiliki potensi besar dalam menghasilkan pemahaman yang lebih akurat mengenai sikap publik terhadap pemilihan umum. Sehingga perlu menerapkan parameter lain seperti fitur ekstraksi dan mencoba dengan algoritma machine learning atau deep learning lainnya.

#### Daftar Pustaka

A'yuniyah, Q. A., & Reza, M. (2023). Penerapan Algoritma K-Nearest Neighbor Untuk Klasifikasi Jurusan Siswa Di Sma Negeri 15 Pekanbaru. *Indonesian Journal of Informatic Research and Software Engineering (IJIRSE)*, 3(1), 39–45.

<https://doi.org/10.57152/ijirse.v3i1.484>

- Al-Azani, S., & El-Alfy, E. S. M. (2017). Using Word Embedding and Ensemble Learning for Highly Imbalanced Data Sentiment Analysis in Short Arabic Text. *Procedia Computer Science*, 109, 359–366. <https://doi.org/10.1016/j.procs.2017.05.365>
- Azhar, Y. (2017). METODE LEXICON-LEARNING BASED UNTUK IDENTIFIKASI TWEET OPINI BERBAHASA INDONESIA. In *Jurnal Nasional Pendidikan Teknik Informatika* / (Vol. 6, Issue 3).
- Azhar, Y. (2018). Metode Lexicon-Learning Based Untuk Identifikasi Tweet Opini Berbahasa Indonesia. *Jurnal Nasional Pendidikan Teknik Informatika (JANAPATI)*, 6(3), 237. <https://doi.org/10.23887/janapati.v6i3.11739>
- Eddyono, A. S. (2021). Pers Alternatif pada Era Orde Baru: Dijinakkan hingga Dibungkam. *Komunika*, 8(1), 53–60.
- Es-Sabery, F., Es-Sabery, K., Qadir, J., Sainz-De-Abajo, B., Hair, A., García-Zapirain, B., & De La Torre-Díez, I. (2021). A MapReduce Opinion Mining for COVID-19-Related Tweets Classification Using Enhanced ID3 Decision Tree Classifier. *IEEE Access*, 9, 58706–58739. <https://doi.org/10.1109/ACCESS.2021.3073215>
- Habibah, A. F. (2021). Era masyarakat informasi sebagai dampak media baru. *Jurnal Teknologi Dan Sistem Informasi Bisnis*, 3(2), 350–363.
- Harun, A., & Putri Ananda, D. (2021). Analisa Sentimen Opini Publik Tentang Vaksinasi Covid-19 di Indonesia Menggunakan Naïve bayes dan Decision Tree. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 1(1), 58–64. <https://doi.org/10.57152/malcom.v1i1.63>
- Hasan, F. N. (2024). Analisis Sentimen Masyarakat Terhadap Fenomena Childfree ( Kehidupan Tanpa Anak ) Pada Twitter Menggunakan Algoritma Naïve Bayes. 5(3), 853–861. <https://doi.org/10.47065/josh.v5i3.5064>
- Ibrahim, N. M., Yafooz, W. M. S., Emara, A. H. M., & Abdel-Wahab, A. (2022). Utilizing Deep Learning in Arabic Text Classification Sentiment Analysis of Twitter. *International Journal of Advanced Computer Science and Applications*, 13(12), 830–838. <https://doi.org/10.14569/IJACSA.2022.0131297>
- Keputusan Dirjen Penguatan Riset dan Pengembangan Ristek Dikti, S., Ari Kristanto, A., Harjoseputro, Y., Eric Samodra, J., & Jaya Yogyakarta yuliusharjoseputro, A. (2017). Terakreditasi SINTA Peringkat 2 Implementasi Golang dan New Simple Queue pada Sistem Sandbox Pihak Ketiga Berbasis REST API. *Masa Berlaku Mulai*, 1(3), 745–750.
- Kholifah, B., Thoib, I., Sururi, N., & Kurnia, N. D. (2024). Analisis Sentimen Warganet Terhadap Isu Layanan Transportasi Online Berbasis InSet Lexicon Menggunakan Logistic Regression. 11(1), 14–25.
- Lase, S. M. N., Adinda, A., Yuliantika, R. D., & Al, E. (2021). Kerangka Hukum Teknologi Blockchain Berdasarkan Hukum Siber di Indonesia. *Padjajaran Law Review*, 9(1), 1–20. <https://hbr.org/2017/02/a-brief-history-of->
- Lazuardi, J. U. S., & Juarna, A. (2023). Analisis Sentimen Ulasan Pengguna Aplikasi Joox Pada

- Android Menggunakan Metode Bidirectional Encoder Representation From Transformer (Bert). *Jurnal Ilmiah Informatika Komputer*, 28(3), 251–260. <https://doi.org/10.35760/ik.2023.v28i3.10090>
- Loka, S. K. P., & Marsal, A. (2023). Perbandingan Algoritma K-Nearest Neighbor dan Naïve Bayes Classifier untuk Klasifikasi Status Gizi Pada Balita. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 3(1), 8–14. <https://doi.org/10.57152/malcom.v3i1.474>
- Mohasseb, A., Bader-El-Den, M., Cocea, M., & Liu, H. (n.d.). *IMPROVING IMBALANCED QUESTION CLASSIFICATION USING STRUCTURED SMOTE BASED APPROACH*. [http://trec.nist.gov/data/qa/t2007\\_qadata.html](http://trec.nist.gov/data/qa/t2007_qadata.html)
- Mudjiyanto, B., & Dunan, A. (2020). Media mainstream jadi rujukan media sosial. *Majalah Semi Ilmiah Populer Komunikasi Massa*, 1(01).
- Mustasaruddin, M., Budianita, E., Fikry, M., & Yanto, F. (2023). Klasifikasi Sentiment Review Aplikasi MyPertamina Menggunakan Word Embedding FastText dan SVM (Support Vector Machine). *Jurnal Sistem Komputer Dan Informatika (JSON)*, 4(3), 526. <https://doi.org/10.30865/json.v4i3.5695>
- Nooryuda Prasetya, Y., Winarso, D., & Syahril. (2021). Penerapan Lexicon Based Untuk Analisis Sentimen Pada Twiter Terhadap Isu Covid-19. *Jurnal Fasilkom*, 11(2), 97–103.
- Putra, F., Tahiyat, H. F., Ihsan, R. M., Rahmaddeni, R., & Efrizoni, L. (2024). Penerapan Algoritma K-Nearest Neighbor Menggunakan Wrapper Sebagai Preprocessing untuk Penentuan Keterangan Berat Badan Manusia. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 4(1), 273–281. <https://doi.org/10.57152/malcom.v4i1.1085>
- Putu, N. L. P. M., Ahmad Zuli Amrullah, & Ismarmiaty. (2021). Analisis Sentimen dan Pemodelan Topik Pariwisata Lombok Menggunakan Algoritma Naive Bayes dan Latent Dirichlet Allocation. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 5(1), 123–131. <https://doi.org/10.29207/resti.v5i1.2587>
- Satrio, B., Dahlan, B. F., Fathan, F., Muwafa, F. Z., & Reyhan, M. (2024). *Klasifikasi Sentimen Emosi Pada Dataset Goemotion Menggunakan LSTM*. 7(1), 21–25.
- Setiawan, S. B., & Isnain, A. R. (2024). *Sentimen Analisis Masyarakat Terhadap Pembangunan IKN Menggunakan Algoritma Lexicon Based Approach dan Naïve Bayes*. 8(April 2019), 1019–1030. <https://doi.org/10.30865/mib.v8i2.7605>
- Shahriar, K. T., Islam, M. N., Moni, M. A., & Sarker, I. H. (2023). A dynamic topic identification and labeling approach for COVID-19 tweets. *Applied Intelligence for Industry 4.0, December 2019*, 227–239. <https://doi.org/10.1201/9781003256083-18>
- Supriatna, R., & Rohman, D. (2024). *PENERAPAN NATURAL LANGUAGE PROCESSING DALAM ANALISIS SENTIMEN CAWAPRES 2024 MENGGUNAKAN ALGORITMA NAIVE BAYES*. 8(1), 1109–1115.
- Surya Gemilang, W., Purwanto, P., & Carudin, C. (2024). Analisis Sentimen Pengguna Instagram Pada Calon Presiden 2024 Menggunakan Algoritma Support Vector Machine.

*JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(4), 2849–2855.  
<https://doi.org/10.36040/jati.v7i4.7256>

- Syamala, M., & Nalini, N. J. (2020). A filter based improved decision tree sentiment classification model for real-time amazon product review data. *International Journal of Intelligent Engineering and Systems*, 13(1), 191–202. <https://doi.org/10.22266/ijies2020.0229.18>
- Toruan, C. R. A., Yudistra, N., & Perdana, R. S. (2023). Analisis Sentimen Tokocrypto pada Twitter menggunakan Metode Long Short-Term Memory. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 7(2), 719–726. <http://j-ptiik.ub.ac.id>
- Utomo, P. B., Wahyudi, D., & Nalendra, A. K. (2024). Implementasi Convolution-Augmented Transformer Berbasis Kecerdasan Buatan dalam Analisis Sentimen Teks Hasil Konversi Suara ke Teks. 8(1), 63–71.
- Winarso, D., Yanda Noor Yudha, & Syahril. (2021). Analisis Sentimen Masyarakat Pada Twiter Terhadap Isu Covid-19 Menggunakan Metode Lexicon Based. *Jurnal Fasilkom*, 11(2), 97–103. <https://doi.org/10.37859/jf.v11i2.2772>
- Yang, S. (2018). Text Mining of Twitter Data Using a Latent Dirichlet Allocation Topic Model and Sentiment Analysis. *International Journal of Computer and Information Engineering*, 12(7), 525–529.
- Yani, S., Jumeilah, F. S., & Kadafi, M. (2020). Algoritma K-Nearest Neighbor Untuk Menentukan Kelayakan Keluarga Penerima Bantuan Pangan Non Tunai (Studi Kasus : Kelurahan Karya Jaya). *Journal of Information Technology Ampera*, 1(2), 75–87. <https://doi.org/10.51519/journalita.volume1.issue2.year2020.page75-87>
- Zelina, N., & Afiyati, A. (2024). Analisis Sentimen Ulasan Pengguna Aplikasi M- Banking Menggunakan Algoritma Support Vector Machine dan Decision Tree. 7(1), 31–37.