

Peningkatan Metode Support Vector Machines (SVM) pada Data *Child-free* Menggunakan Oversampling

Rizki Astuti¹, Ratna Andini Husen², Agus Triono³, M Khairul Anam⁴

^{1,2,3,4}Program Studi Teknik Informatika STMIK Amik Riau

^{1,2,3,4}Jl. Purwodadi Indah KM. 10 Tampan, Pekanbaru, Riau, telp. (0761) 7047091

e-mail: ¹2010031802007@sar.ac.id, ²2010031802011@sar.ac.id, ³2010031802017@sar.ac.id,

⁴khairulanam@sar.ac.id

Abstrak

Keputusan tidak untuk memiliki anak, yang dikenal sebagai *child-free*, semakin relevan dalam masyarakat modern. Support Vector Machines (SVM) yaitu algoritma yang terawasi digunakan dalam menganalisis keputusan seperti ini. Namun, SVM dapat menghadapi tantangan ketika berurusan dengan data yang tidak seimbang, yang dapat mengakibatkan kinerja yang buruk dalam mengklasifikasikan kelas minoritas. Untuk mengatasi masalah ini, teknik oversampling dapat diterapkan. Studi ini bertujuan untuk membandingkan penerapan SVM dengan oversampling untuk meningkatkan akurasi analisis sentimen pada konsep *child-free*. Penelitian ini melibatkan langkah-langkah seperti pra-pemrosesan data, visualisasi data, pembobotan kata, pembagian data, oversampling menggunakan SMOTE, pemodelan klasifikasi menggunakan SVM, dan evaluasi model. Hasil penelitian menunjukkan bahwa SVM menggunakan pembobotan kata TF-IDF mencapai yang lebih akurat dibandingkan dengan BoW dan dapat disimpulkan bahwa SVM dengan pembobotan kata TF-IDF direkomendasikan untuk menganalisis proses pengambilan keputusan individu *child-free*.

Kata Kunci: Oversampling, Support Vector Machines (SVM), *Child-free*, TF-IDF, Bag of Words.

Abstract

The decision to have no children, known as *child-free*, is becoming increasingly relevant in modern society. SVM is a supervised learning algorithm used to analyze decisions like this. However, SVMs can face challenges when dealing with unbalanced data, which can result in poor performance in classifying minority classes. To solve this problem, oversampling techniques can be applied. This study aims to compare the application of SVM with oversampling to improve the accuracy of sentimental analysis on the *child-free* concept. The study involves measures such as data pre-processing, data visualization, word grinding, data division, oversampling using SMOTE, classification modeling using SVM, and model evaluation. The results showed that SVMs with TF-IDF word grading achieved higher accurations compared to BoW and it was concluded that the SVM's with the TF - IDF word Grading were recommended to analyze the individual *child-free* decision-making process.

Keywords: Oversampling, Support Vector Machines (SVM), *Child-free*, TF-IDF, Bag of Words

1. PENDAHULUAN

Keputusan individu atau pasangan untuk tidak memiliki anak, yang dikenal sebagai *child-free*. *Child-free* adalah konsep di mana seseorang memilih untuk tidak memiliki anak. Hal ini menjadi topik yang semakin relevan dalam masyarakat modern. Keputusan ini dipengaruhi oleh berbagai faktor, seperti perubahan gaya hidup, pertimbangan ekonomi, serta preferensi pribadi. Dalam konteks ini, analisis data memiliki peran penting dalam memahami faktor-faktor yang mempengaruhi keputusan *Child-free*[1].

Support Vector Machine (SVM), yang merupakan algoritma pembelajaran yang terawasi, adalah salah satu alat yang digunakan dalam analisis keputusan seperti ini. SVM dirancang untuk mempertimbangkan faktor-faktor umum dan mengurangi risiko struktural saat menemukan hyperplane terbaik untuk memisahkan data dari kelas yang

ditentukan. SVM adalah algoritma pembelajaran mesin yang kuat dan telah digunakan dalam berbagai konteks, termasuk analisis keputusan *Child-free*. Meskipun demikian, SVM dapat mengalami kendala ketika dihadapkan pada data yang tidak seimbang, yaitu ketika jumlah sampel dalam kelas-kelas yang berbeda tidak merata. Ketidakseimbangan ini dapat mengakibatkan kinerja yang buruk dalam mengklasifikasikan kelas minoritas, yang dalam konteks *Child-free*, mungkin mencakup individu atau pasangan yang memilih untuk tidak memiliki anak[2].

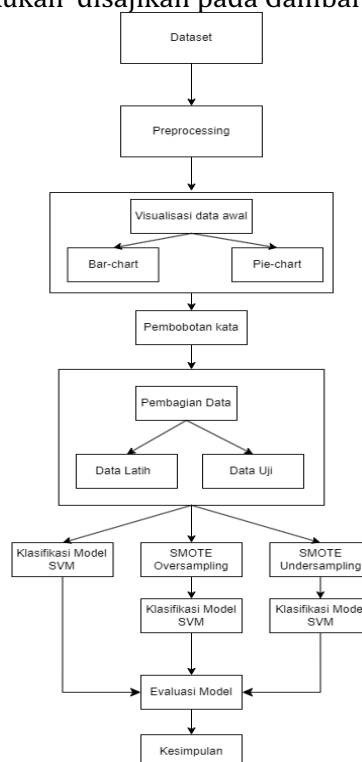
Untuk mengatasi masalah ketidakseimbangan data, teknik oversampling dapat diterapkan. *Oversampling* bertujuan untuk menyeimbangkan kelas-kelas dalam data dengan meningkatkan jumlah sampel dari kelas minoritas. Dengan menerapkan teknik *oversampling* secara bijaksana pada metode SVM dalam konteks *Child-free*, dapat diharapkan bahwa kinerja model akan meningkat secara signifikan[3].

Penelitian yang dilakukan oleh Wenhao et al [4] dengan judul “*An Improved Oversampling Algorithm Based on the Samples Selection Strategy for Classifying Imbalanced Data*” yang bertujuan untuk meningkatkan teknik pengambilan sampel terhadap pengklasifikasian data tidak seimbang. Hasil penelitian ini menunjukkan bahwa algoritma pengambilan sampel yang ditingkatkan dapat melakukan klasifikasi data tidak seimbang dengan lebih baik[3].

Tujuan penelitian ini adalah untuk membandingkan penggunaan metode SVM dengan oversampling dengan metode lain untuk mengatasi jumlah pengamatan yang tidak seimbang dalam mengklasifikasikan sentimen masyarakat terhadap gagasan tidak memiliki anak. Selain itu, tujuan adalah untuk mengetahui metode SVM dengan teknik oversampling yang dapat meningkatkan hasil akurasi yang lebih baik pembahasan konsep *child-free* oleh pengguna Twitter.

2. METODE PENELITIAN

Tahapan Penelitian yang dilakukan disajikan pada Gambar 1.



Gambar 1. Tahap penelitian

2.1. Dataset

Dataset yang digunakan pada penelitian ini adalah dataset *Child-free* yang diambil dari Platform Drone Emprit. Keyword yang dipakai untuk mengambil data *retweet* adalah "*Child-free*". Dataset ini berisi *retweet* mengenai *child-free* pada Maret 2023 yang mencakup 5000 data.

2.2. Preprocessing

Preprocessing Adalah tahap melengkapi kumpulan data yang digunakan sehingga ketika menggunakan algoritma klasifikasi dapat lebih mudah untuk dianalisis[4]. Pada tahap ini data yang dikumpulkan masih berupa data yang tidak terstruktur, sehingga preprocessing diperlukan agar data lebih terstruktur. Pada penelitian ini, preprocessing meliputi beberapa langkah, seperti Pembersihan data meliputi penghapusan kolom (atribut) yang tidak terpakai dan pengisian data *null* (*cleaning*), mengubah huruf kapital yang ada pada kolom menjadi huruf kecil (*Case folding*), memecah kalimat menjadi kata (*tokenizing*), Menghapus kata singkatan yang memiliki arti tidak penting (*Stopword*).

2.3 Visualisasi data

Gambaran dari data yang dikumpulkan adalah apa yang disebut visualisasi data. Objek-objek ini unik dan menarik, sehingga menarik perhatian dan skeptisisme. [5]. Tahap visualisasi data berfungsi untuk melihat data *retweet* positif, *neutral*, dan negatif. Untuk mengetahui apakah data dalam penelitian ini seimbang atau tidak, kami melakukan dua visualisasi yaitu *bar-chart* dan *pie-chart*.

2.4 Pembobotan kata

Setelah visualisasi data selanjutnya dilakukan Pembobotan kata adalah teknik untuk menilai jumlah kata yang muncul dalam dokumen teks[6]. Pada titik ini, terdapat dua pembobotan kata: Frekwensi Term-Inverse Dokumen (TF-IDF) dan Bag of Word (BoW).

2.4.1 Term Frequency-Inverse Document Frequency (TF-IDF)

Pembobotan kata TF-IDF adalah algoritma untuk menggambarkan data yang dikumpulkan atau kumpulan dokumen untuk mengetahui seberapa penting kata-kata tersebut dalam konteks dokumen tersebut.[7] Selanjutnya, data teks diubah menjadi bentuk vektor sehingga dapat mempertimbangkan urutan kata yang tepat. Angka TF-IDF berkorelasi dengan setiap kata dalam korpus, menunjukkan seberapa signifikan setiap kata[8]. Untuk perhitungan TF-IDF pada penelitian ini menggunakan *library Sklearn Python*, kita dapat menggunakan *TfidfVectorizer*.

2.4.2 Bag of Word (BoW)

Teknik Bag of Word adalah model yang mengatur teks berdasarkan urutan di mana ia dimasukkan ke dalam dokumen. Teknik ini sangat baik untuk klasifikasi teks dan pemodelan bahasa. pada penelitian ini *Library "CountVectorizer"* *CountVectorizer* digunakan untuk menjalankan BOW. Ini menghitung jumlah kata kata yang muncul dan membuat matriks database cadangan kata kata. [9].

2.5 Pembagian data

Setelah melalui tahap *preprocessing* kemudian dibagi menjadi dua kumpulan data yaitu data latih dan data uji. Pada tahap ini Pemisahan data dilakukan dengan membagi dataset menjadi 3 bagian data latih dan data uji yaitu 90:10, 80:20, 70:30. Data latih adalah data yang berperan sebagai model pelatihan, yaitu membangun pengetahuan untuk proses klasifikasi. Data uji digunakan untuk menguji keakuratan algoritma klasifikasi menggunakan model yang dilatih. Data latih digunakan dalam penelitian ini untuk melatih model klasifikasi yang digunakan. Model klasifikasi yang digunakan dalam penelitian ini terdiri dari beberapa classifier yaitu *Support Vector Machine* (SVM). Setelah melatih data selanjutnya melakukan pengujian data dengan tujuan untuk mengukur kinerja model atau algoritma yang telah dibangun. Ini membantu untuk memahami sejauh mana model mampu membuat prediksi yang akurat dan relevan.

2.6 SMOTE Oversampling

Synthetic Minority Over-sampling Technique (SMOTE) mensintesis dataset minoritas hingga jumlah dataset pada kelas mayoritas seimbang. Untuk mengatasi ketidakseimbangan dalam penambahan data, penelitian ini menggunakan *SMOTE oversampling*. ketidakseimbangan yang dimaksud disini adalah ada beberapa kelas memiliki jumlah data yang relatif kecil, sementara beberapa lainnya memiliki sejumlah besar data. Tujuan dari metode oversampling adalah untuk membuat sampel kelas minoritas sebanding atau setara dengan sampel kelas mayoritas lainnya yang menduplikasi secara acak sampel kelas minoritas. Metode ini juga memungkinkan untuk membuat dataset yang lebih seimbang dan meningkatkan kinerja model untuk mengidentifikasi kelas minoritas[10]. yang disajikan dalam bentuk tabel nantinya.

2.7 SMOTE Undersampling

Undersampling mengambil data dari kelas mayoritas secara acak untuk dihilangkan dari kumpulan data latih. Dengan melakukan undersampling secara acak, jumlah data latih dari kelas mayoritas akan berkurang. Untuk mendapatkan distribusi kelas yang diinginkan pada data latih, prosedur *undersampling* dapat diulang. Metode ini dapat diterapkan pada kumpulan data dengan kelas yang tidak seimbang di mana kelas minoritas cukup untuk membuat model. Dalam *undersampling*, jumlah data dari kelas mayoritas dan minoritas mungkin sama atau data mayoritas mungkin hanya dikurangi hingga batas tertentu.

2.8 Klasifikasi Model

Penelitian ini menggunakan *Support Vector Machine (SVM)* sebagai algoritma klasifikasi untuk mengevaluasi dan membandingkan bagaimana metode-metode mengklasifikasikan sentimen *tweet*. Data latih dan data ujinya diperoleh dari masing-masing hasil ekstraksi fitur (BoW, dan TF-IDF). Selanjutnya dilakukan prediksi klasifikasi dengan algoritma SVM. Dari setiap metode klasifikasi yang digunakan didapat nilai *accuracy*.

2.9 Evaluasi Model

Pengujian model diperlukan untuk mengevaluasi kinerja metode Support Vector Machine (SVM). Dengan menggunakan library *sklearn.metrics.confusion_matrix*, matrik kekacauan dapat digunakan untuk menampilkan informasi tentang perbandingan hasil klasifikasi dengan data sebenarnya. Confusion Matrix dapat menampilkan nilai akurasi, ketepatan, recall, dan skor f1 untuk masing-masing skenario model klasifikasi. Nilai akurasi merupakan matriks evaluasi yang digunakan untuk mengevaluasi seberapa baik suatu model klasifikasi atau prediksi dapat memberikan hasil yang akurat atau sesuai. prediksi mereka. Untuk *accuracy*, *precision*, *recall*, dan *f1-score*, perhitungan yang disediakan oleh library *sklearn.metrics.accuracy_score* dapat digunakan untuk menghitung nilai *accuracy*, *precision*, dan *recall*.

$$Akurasi = \frac{TP+TN}{TP+FN+FP+TN} \quad (1)$$

$$Precesion = \frac{TP}{TP+FP} \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

$$F1 - Score = \frac{Precesion * Recall}{Precesion+Recall} \quad (4)$$

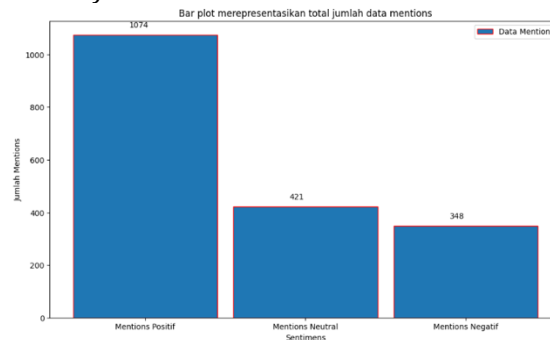
3. HASIL DAN PEMBAHASAN

Ada 5000 data yang dikategorikan menjadi positif, negatif, dan *neutral*. Untuk membuat data yang digunakan lebih akurat pada akhirnya, teks akan diproses terlebih dahulu sebelum data digunakan. *Preprocessing* data mencakup penghapusan kolom (atribut) yang tidak terpakai dan pengisian data *null* (*cleaning*), mengubah huruf kapital yang ada pada kolom menjadi huruf kecil (*Case folding*), memecah kalimat menjadi kata (*tokenizing*), Menghapus kata singkatan yang memiliki arti tidak penting (*Stopword*).

	Mentions	Sentiment	Mentions_remove_user	Mentions_cleaning	Mentions_case_folding	Mentions_tokenization	Mentions_stopwords	Mentions_clean
0	Pengen bangel gu gulai rahim penganut sette ch...	Negative	Pengen bangel gu gulai rahim penganut sette ch...	Pengen bangel gu gulai rahim penganut sette ch...	pengen bangel gu gulai rahim penganut sette ch...	ipengen, bangel, gu, guai, rahim, penganut, s...	pengen bangel gu gulai rahim penganut sette ch...	pengen bangel gu gulai rahim penganut sette ch...
1	RT @glik yak bagus, video ini semakin meneg...	Positive	RT yak bagus, video ini semakin menegaskan ba...	yak bagus video ini semakin menegaskan bahwa c...	yak bagus video ini semakin menegaskan bahwa c...	[yak, bagus, video, ini, semakin, menegaskan...	yak bagus video childfree pilihan yg menegkah...	yak bagus video childfree pilihan yg menegkah...
2	RT @linal_20 @laryakani Nani alesannya "Rez...	Negative	RT Nani alesannya "Rezeki mah ada aja seka...	Nani alesannya Rezeki mah ada aja sekarang mu...	nani alesannya rezeke mah ada aja sekarang mu...	[nani, alesannya, rezeke, mah, ada, aja, seka...	alesannya rezeke mah aja mumpung finansial bag...	alesannya rezeke mah aja mumpung finansial bag...
3	RT @SoundOfopi Video anti childfree tapi isi...	Negative	RT Video anti childfree tapi isinya contoh o...	Video anti childfree tapi isinya contoh yg ora...	video anti childfree tapi isinya contoh yg ora...	[video, anti, childfree, tapi, isinya, contoh...	video anti childfree isinya contoh yg oranglusa...	video anti childfree isinya contoh yg oranglusa...
4	RT @SoundOfopi Video anti childfree tapi isi...	Negative	RT Video anti childfree tapi isinya contoh o...	Video anti childfree tapi isinya contoh yg ora...	video anti childfree tapi isinya contoh yg ora...	[video, anti, childfree, tapi, isinya, contoh...	video anti childfree isinya contoh yg oranglusa...	video anti childfree isinya contoh yg oranglusa...
4995	RT @AREAJULIO Orang yang notabennya hidup tercukupi meni...	Positive	RT Orang yang notabennya hidup tercukupi meni...	Orang yang notabennya hidup tercukupi memilih...	orang yang notabennya hidup tercukupi memilih...	[orang, yang, notabennya, hidup, tercukupi, me...	orang notabennya hidup tercukupi memilih child...	orang notabennya hidup tercukupi memilih child...
4996	RT @Khodamharadi Seharusnya bapak ini childfree https://...	Negative	RT Seharusnya bapak ini childfree https://...	Seharusnya bapak ini childfree RE...	seharusnya bapak ini childfree re khodamharadi	[seharusnya, bapak, ini, childfree, re, khodam...	childfree re khodamharadi	childfree re khodamharadi

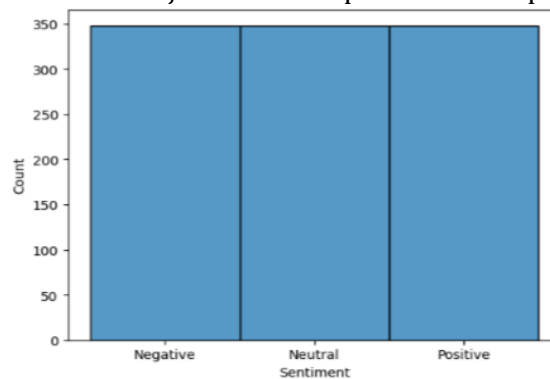
Gambar 2. Hasil Preprocessing text

Setelah *preprocessing* teks selesai, Selanjutnya visualisasi *bar-chart* label sentimen dari data yang sudah dibersihkan akan dihitung untuk mengetahui seberapa besar reaksi publik terhadap *retweet child-free*.



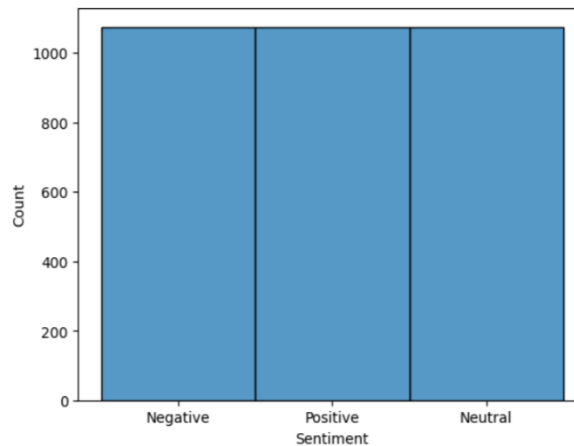
Gambar 3. Visualisasi *bar-chart* Jumlah data reetwet

Dapat diketahui pada gambar diatas terdapat 1071 *reetwet* dengan label positif, 421 *reetwet* label neutral dan 348 *reetwet* label negatif. Dengan total setelah dilakukan *preprocessing* dan dilabeli terdapat 1840 data. Gambar diatas juga menunjukkan ternyata mayoritas pengguna *Twitter* menunjukkan reaksi positif terhadap *child-free*.



Gambar 4. Hasil grafik setelah Undersampling

Gambar 4 menunjukkan distribusi kelas setelah dilakukan undersampling pada data. Dalam gambar tersebut, dapat dilihat bahwa jumlah data dari kelas mayoritas telah dikurangi secara acak untuk mencapai distribusi kelas yang diinginkan.



Gambar 5. Hasil grafik setelah Oversampling

Distribusi kelas setelah oversampling ditunjukkan pada gambar 5 Teknik oversampling meningkatkan jumlah sampel kelas minoritas secara signifikan. Ini menghasilkan distribusi kelas yang lebih seimbang antara kelas minoritas dan mayoritas, yang akan membantu model memprediksi kelas minoritas dengan lebih baik. Perbandingan antara kedua grafik ini menunjukkan bagaimana oversampling dapat memperbaiki ketidakseimbangan kelas, sehingga meningkatkan kinerja model dalam tugas klasifikasi pada data Child-free. Dengan menerapkan teknik undersampling dan oversampling, distribusi kelas dalam data dapat diatur sehingga memungkinkan model untuk mengklasifikasikan sentimen dengan lebih baik dan akurat.

TABEL 1 Hasil komparasi Spliting data dengan Tf-IDF

TI-IDF			
Spliting data	Tanpa SMOTE	SMOTE Oversampling	SMOTE Undersampling
90:10	61%	87%	58%
80:20	60%	88%	54%
70:30	60%	84%	50%

Berdasarkan pada tabel 1 dengan beberapa *spliting* data yang digunakan, SMOTE *Oversampling* dengan metode pembobotan kata TF-IDF dengan *spliting* data 80:20 mencapai akurasi yang paling tinggi yaitu 88%.

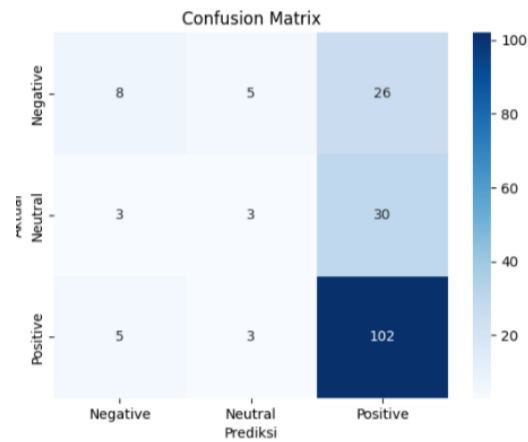
Table 2. Hasil Komparasi *spliting* data dengan BoW

BOW			
Spliting data	Tanpa SMOTE	SMOTE Oversampling	SMOTE Undersampling
90:10	57%	67%	44%
80:20	54%	63%	45%
70:30	54%	60%	46%

Tabel 2 menunjukkan bahwa algoritma SVM dengan pembobotan kata BoW memiliki tingkat akurasi yang lebih rendah ketika *SMOTE oversampling* ditambahkan ke dalamnya. Pembobotan kata TF-IDF dengan *spliting* data 80:20 menunjukkan tingkat akurasi terbaik sebesar 88% dibandingkan dengan pembobotan kata BoW.

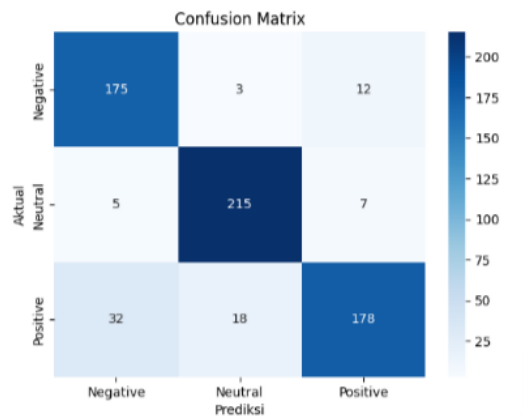
Selain itu, penelitian ini menampilkan hasil uji data dalam bentuk Matrik Konflik.

Matrik ini berfungsi sebagai dasar untuk perhitungan skor Precision, Recall, dan F1, dengan perhitungan sebelumnya yang menghasilkan hasil 3 tertinggi.



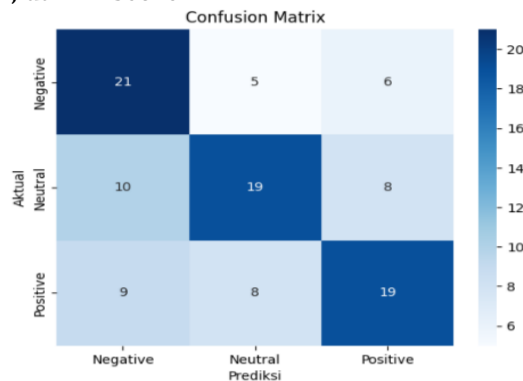
Gambar 6. Confusion Matriks Tanpa SMOTE

Gambar 6 menunjukkan confusion matrix tanpa SMOTE. Matriks ini berisi informasi tentang perbandingan hasil klasifikasi dengan data sebenarnya. Untuk skenario model klasifikasi tanpa SMOTE, nilai akurasi, ketepatan, recall, dan skor f1 ditunjukkan pada gambar.



Gambar 7. Confusion Matriks SMOTE Oversampling

Confusion matrix dengan SMOTE oversampling ditunjukkan pada Gambar 6. SMOTE adalah metode oversampling yang digunakan untuk mengatasi ketidakseimbangan dalam jumlah pengamatan yang dikategorikan dalam kategori sentimen mengenai gagasan tidak memiliki anak. Setelah menerapkan SMOTE oversampling, gambar menunjukkan nilai akurasi, ketepatan, recall, dan f1-score.



Gambar 8. Confusion Matriks SMOTE Undersampling

Confusion matrix dengan SMOTE undersampling ditunjukkan pada Gambar 7. Metode undersampling SMOTE digunakan untuk mengatasi ketidakseimbangan jumlah pengamatan dalam kategori sentimen terhadap gagasan tidak memiliki anak. Setelah menerapkan SMOTE undersampling, gambar menunjukkan nilai accuracy, precision, recall, dan f1-score.

Selain itu, hasil pengujian model terbaik akan ditampilkan. Hasilnya akan menunjukkan model mana yang memiliki skor F1, akurasi, ketepatan, dan recall yang paling tinggi.

TABEL 3. Hasil Pengujian Model Tertinggi

Model	Accuracy	Precesion	recall	F1-Score
Tanpa SMOTE	58%	47%	48%	58%
SMOTE Oversampling	88%	88%	88%	88%
SMOTE Undersampling	56%	56%	57%	56%

Hasil pengujian model tertinggi dapat dilihat pada Tabel 3. Model dengan SMOTE oversampling memiliki nilai accuracy, precision, recall, dan f1-score tertinggi, yaitu 88%.

4. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan dapat disimpulkan dapat disimpulkan bahwa metode *Support Vector Machine* (SVM) dengan pembobotan kata TF-IDF memiliki tingkat akurasi yang lebih baik dalam analisis keputusan *child-free*. Penggunaan metode *oversampling* SMOTE pada SVM dengan pembobotan kata *Bag of Words* (BoW) menghasilkan tingkat akurasi yang lebih rendah. Oleh karena itu, pembobotan kata TF-IDF lebih disarankan dalam analisis keputusan *child-free*. Metode SVM dengan pembobotan kata TF-IDF memiliki tingkat akurasi yang lebih baik dalam analisis keputusan *child-free*. Tahap *preprocessing* data, visualisasi data, dan pembobotan kata juga penting dalam analisis data.

UCAPAN TERIMAKASIH

Kami mengucapkan terima kasih kepada teman-teman penelitian kami yang telah mendukung, bekerja sama, dan memberikan kontribusi yang sangat penting untuk penelitian ini. Tanpa kerja sama dan kolaborasi dari teman-teman, penelitian ini tidak akan berhasil. Kami juga berterima kasih kepada dosen pembimbing kami yang telah memberikan saran, arahan, dan bimbingan yang sangat penting selama proses penelitian ini. Kami sangat menghargai pengetahuan dosen kami dan waktu mereka. Kami juga ingin mengucapkan terima kasih kepada STMIK Amik Riau karena telah memberi kami bantuan dan fasilitas yang diperlukan untuk melakukan penelitian ini. STMIK Amik Riau telah menyediakan lingkungan akademik yang ramah dan sumber daya yang cukup untuk mendukung penelitian kami. Selain itu, kami berterima kasih atas kesempatan untuk berkembang dan berkontribusi pada dunia peneletian.

DAFTAR PUSTAKA

- [1] D. Siregar, F. Ladayya, N. Z. Albaqi, and B. M. Wardana, "Penerapan Metode Support Vector Machines (SVM) dan Metode Naïve Bayes Classifier (NBC) dalam Analisis Sentimen Publik terhadap Konsep Child-free di Media Sosial Twitter," *Jurnal Statistika dan Aplikasinya*, vol. 7, no. 1, 2023.
 - [2] J. M. Teknologi Informatika dan Komputer Thamrin, F. Syofiani, S. Alam, M. S. Imam Sulisty, P. Studi Teknik Informatika, and S. Tinggi Teknologi Wastukencana,
-

- “Analisis Sentimen Penilaian Masyarakat Terhadap Child-free Berdasarkan Komentar di Youtube Menggunakan Algoritma Naïve Bayes”, doi: 10.37012/jtik.v9i2.1661.
- [3] “InfoTekJar: Jurnal Nasional Informatika dan Teknologi Jaringan”, doi: 10.30743/infotekjar.v6i2.4920.
- [4] Kurnianto, D. D., & Waluyo, S. (2023, October). Analisis Sentimen Masyarakat Terhadap Eks Pejabat Pajak Diperiksa Kpk Pada Youtube Menggunakan Metode K-Nearest Neighbor. In *Prosiding Seminar Nasional Mahasiswa Fakultas Teknologi Informasi (SENAFTI)* (Vol. 2, No. 2, pp. 632-641).
- [5] Fahri, A., & Ramdhani, Y. (2023). Visualisasi Data dan Penerapan Machine Learning Menggunakan Decision Tree Untuk Keputusan Layanan Kesehatan COVID-19. *Jurnal Tekno Kompak*, 17(2), 50-60.
- [6] Hadna, N. M. S., Santosa, P. I., & Winarno, W. W. (2016). Studi literatur tentang perbandingan metode untuk proses analisis sentimen di Twitter. *Semin. Nas. Teknol. Inf. dan Komun*, 2016, 57-64.
- [7] Fadhillah, P. N., & Indriyanti, A. D. (2023). Analisis Sentimen terhadap Opini Publik Mengenai Child-free dalam Pernikahan pada Twitter Menggunakan K-Nearest Neighbor (K-NN). *Journal of Informatics and Computer Science (JINACS)*, 5(01), 58-62.
- [8] Rupapara, V., Rustam, F., Shahzad, H. F., Mehmood, A., Ashraf, I., & Choi, G. S. (2021). Impact of SMOTE on imbalanced text features for toxic comments classification using RVVC model. *IEEE Access*, 9, 78621-78634.
- [9] Eshan, S. C., & Hasan, M. S. (2017, December). An application of machine learning to detect abusive bengali text. In 2017 20th International Conference of Computer and Information Technology (ICCIT)(pp. 1-6). IEEE.
- [10] Heranova, O. (2019). Synthetic Minority Oversampling Technique pada Averaged One Dependence Estimators untuk Klasifikasi Credit Scoring. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 3(3), 443-450.

