

Penerapan Metode SMOTE Untuk Mengatasi *Imbalanced Data* Pada Klasifikasi Ujaran Kebencian

Ridwan¹, Eni Heni Hermaliani^{2*}, Muji Ernawati³

^{1,2,3}Universitas Nusa Mandiri
Jl. Kramat Raya No. 18 Jakarta Pusat, Indonesia

e-mail: ¹14210122@nusamandiri.ac.id, ^{2*}enie_h@nusamandiri.ac.id, ³14210225@nusamandiri.ac.id

(*) Corresponding Author

Artikel Info : Diterima : 01-12-2023 | Direvisi : 29-01-2024 | Disetujui : 31-01-2024

Abstrak - Ujaran kebencian adalah penyebaran kebencian terhadap individu atau kelompok atas dasar suku, agama, ras, dan karakteristik lain yang dapat menimbulkan diskriminasi, kekerasan, dan konflik sosial. Data yang tidak seimbang dapat menyebabkan hasil negatif pada hasil klasifikasi. *Metode Synthetic Minority Oversampling Technique* (SMOTE) digunakan untuk mengatasi data yang tidak seimbang. Ekstraksi fitur menggunakan *Bag of Words* dan TD-IDF, kemudian *data training* dilakukan *oversampling* menggunakan metode SMOTE, SVM-SMOTE, *Kmeans-SMOTE*, dan *Borderline-SMOTE*. Klasifikasi tersebut menggunakan algoritma *Random Forest*, *Support Vector Machine*, *Logistic Regression* dan *Naive Bayes* dengan menggunakan data *Twitter*. Hasil penelitian menunjukkan bahwa penerapan metode *Borderline-SMOTE* untuk menangani data *imbalanced* menghasilkan kinerja yang lebih baik daripada metode SMOTE lainnya berdasarkan nilai *accuracy*, *recall*, *precision* dan *F1-Score* dengan masing-masing nilai sebesar 84,09%, 85,25%, 84,55% dan 81,16%. Algoritma *Random Forest* menghasilkan nilai kinerja yang lebih tinggi dibandingkan algoritma lainnya.

Kata Kunci : *Data Imbalanced*, *Oversampling*, SMOTE, Ujaran Kebencian

Abstracts - *Hate speech is the spread of hatred towards individuals or groups on the basis of ethnicity, religion, race, and other characteristics that can lead to discrimination, violence, and social conflict. Unbalanced data can cause negative results in classification results. The Synthetic Minority Oversampling Technique (SMOTE) method is used to deal with unbalanced data. Feature extraction uses Bag of Words and TD-IDF, then the training data are oversampled using the SMOTE, SVM-SMOTE, Kmeans-SMOTE, and Borderline-SMOTE methods. This classification uses the Random Forest, Support Vector Machine, Logistic Regression, and Naive Bayes algorithms using Twitter data. The research results show that the application of the Borderline-SMOTE method to handle imbalanced data produces better performance than other SMOTE methods based on accuracy, recall, precision and F1-Score values with respective values of 84.09%, 85.25%, 84.55% and 81.16%. The Random Forest algorithm produces higher performance values than other algorithms.*

Keywords : *Imbalanced Data, Oversampling, SMOTE, Hate Speeches*

PENDAHULUAN

Ujaran kebencian adalah tindakan menyebarkan kebencian terhadap individu atau kelompok atas dasar suku, agama, ras, dan karakteristik lain yang dapat menimbulkan diskriminasi, kekerasan, dan konflik sosial (Marpaung et al., 2021). Tindakan tersebut dapat berupa hasutan terhadap individu atau kelompok yang dapat menimbulkan kekerasan, diskriminasi, konflik sosial, dan ancaman jiwa (Pratiwit et al., 2019). Saat ini, tindakan ujaran kebencian tidak hanya dilakukan secara langsung tetapi sudah banyak yang dilakukan secara online yaitu melalui media sosial. Media sosial merupakan salah satu platform yang paling sering digunakan oleh masyarakat untuk berkomunikasi, berbagi informasi, dan menyampaikan pendapat. Berdasarkan studi yang dilakukan oleh *We Are Social*, pada Januari 2022 tercatat bahwa di Indonesia pengguna aktif media sosial mengalami kenaikan sebesar 12,35% dari tahun sebelumnya yaitu ada 191 juta pengguna aktif media sosial (Murtopo et al., 2022). Menurut Bawaslu, masih belum ada pemahaman umum tentang ujaran kebencian dalam menangani konten media sosial terutama dengan teks ujaran kebencian yang bersifat sarkastik (Marpaung et al., 2021).

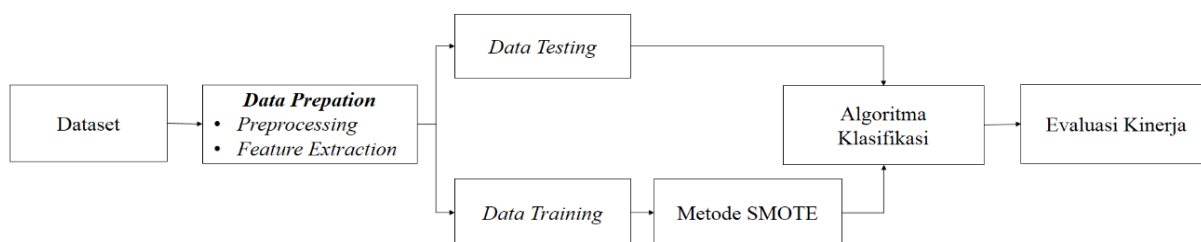


Machine Learning diperlukan untuk membantu media sosial dalam melacak ujaran kebencian, bahkan melakukan upaya preventif, yaitu menghindari ujaran kebencian tersebut tampil di media sosial (Antariksa et al., 2019). Algoritma *Machine Learning* telah banyak digunakan untuk klasifikasi ujaran kebencian diantaranya *Naive Bayes* dalam (Antariksa et al., 2019; Hairani et al., 2020; Khanday et al., 2022; Susanti et al., 2022), *Support Vector Machine* (SVM) dalam (Antariksa et al., 2019; Hairani et al., 2020; Khanday et al., 2022), *Logistic Regression* dalam (Antariksa et al., 2019; Khanday et al., 2022), *Random Forest* dalam (Khanday et al., 2022), *C4.5* dalam (Hairani et al., 2020), dan *K-Nearest neighbour* (KNN) dalam (Mullah & Zainon, 2021). Dalam *Machine Learning*, data *imbalanced* merupakan masalah penting untuk dipecahkan. Ketidakseimbangan kelas adalah suatu kondisi dimana jumlah kemunculan kelas mayoritas lebih besar dari jumlah kemunculan kelas minoritas (Hairani et al., 2020). Data yang tidak seimbang dapat menyebabkan hasil negatif pada hasil klasifikasi. Hal ini dikarenakan jumlah data yang tidak seimbang antara kelas mayoritas dan kelas minoritas, cenderung membuat kelas mayor tampil lebih baik daripada kelas minor.

Untuk mengatasi masalah ketidakseimbangan kelas, salah satu metode yang digunakan adalah *sampling*. *Oversampling* adalah pendekatan *sampling* umum lainnya yang digunakan untuk menangani masalah kelas yang tidak seimbang (Thabtah et al., 2019). *Synthetic Minority Oversampling Technique* (SMOTE) merupakan metode *oversampling* yang dapat melakukan duplikasi data secara sintetik sehingga masalah distribusi data yang berbeda dapat diatasi. Pengembangan metode SMOTE telah dilakukan diantaranya *Borderline-SMOTE* (T. Lee et al., 2020), *Kmeans-SMOTE* (Last et al., 2017) dan *SVM-SMOTE* (Almajid & Arifudin, 2022). Penelitian ini bertujuan untuk melakukan perbandingan hasil kinerja dari masing-masing metode SMOTE dalam mengatasi masalah data yang tidak seimbang. Algoritma *Machine Learning* yang akan digunakan untuk klasifikasi ujaran kebencian dalam penelitian ini diantaranya SVM, *Random Forest*, *Logistic Regression* dan *Naive Bayes*. Kemudian hasil kinerja yang diperoleh dengan menggunakan metode SMOTE akan dibandingkan berdasarkan nilai akurasi, *precision*, *recall* dan *F1-Score*. Pembaharuan pada penelitian ini terletak pada penggunaan teknik SMOTE dan teknik lainnya seperti *Borderline-SMOTE*, *Kmeans-SMOTE* dan *SVM-SMOTE*.

METODE PENELITIAN

Penelitian ini dimulai dengan pengumpulan *dataset*, yang kemudian dilakukan *preprocessing* untuk membersihkan data yang tidak dibutuhkan dan menerapkan teknik ekstraksi fitur. Setelah itu dilakukan *split data* dengan ketentuan 80% data *training* dan 20% data *testing*. Kemudian, metode SMOTE diterapkan pada data *training* untuk menangani data *imbalance*. Algoritma *Machine Learning* diterapkan untuk melakukan klasifikasi ujaran kebencian. Evaluasi dilakukan dengan melihat kinerja model berdasarkan nilai *Accuracy*, *Recall*, *Precision* dan *F1-Score*. Adapun alur penelitian yang diuraikan diatas, dapat dilihat pada gambar 1 berikut ini :



Sumber : Hasil Penelitian (2023)

Gambar 1. Alur Penelitian

1. Dataset

Dataset yang digunakan adalah data Twitter yang dikumpulkan dari beberapa penelitian sebelumnya yang terdiri dari (Ibrohim & Budi, 2019), (Alfina et al., 2017), (Putri et al., 2020) dan (Ibrohim & Budi, 2018). Dataset ini dikumpulkan selama kurang lebih tujuh bulan, dari 20 Maret 2018 hingga 10 September 2018. Untuk menganotasi kumpulan data, para peneliti sebelumnya menggunakan 30 anotator dari latar belakang demografis yang berbeda. Para anotator terdiri dari 14 pria dan 16 wanita dari berbagai usia dan latar belakang pendidikan. Selain itu, para anotator juga berasal dari berbagai profesi, suku, dan agama. Penentuan kelas pada dataset ini dilakukan anotasi dua langkah. Proses anotasi dua langkah ini menghasilkan 13.169 tweet yang dapat digunakan untuk eksperimen penelitian, yang terdiri dari 7.608 tweet *non-hate speech* (6.187 tweet dari fase anotasi pertama dan 1.421 tweet dari (Ibrohim & Budi, 2018)) dan 5.561 tweet ujaran kebencian.

2. Preprocessing

Dataset diproses terlebih dahulu agar proses klasifikasi lebih efisien dan memberikan hasil yang lebih baik. Penelitian ini melakukan lima proses dalam data preprocessing yang terdiri dari *case folding*, *data cleaning*, *text normalization*, *stemming*, dan *stopword removal*.

a. Data Cleaning

Data Cleaning atau pembersihan data dilakukan untuk menghapus karakter yang tidak perlu seperti simbol *retweet* ("RT", "rt"), *mention*, *username* ("USER", "user"), URL, dan tanda baca. Selain itu, karena pada ekstraksi fitur tidak menggunakan emotikon maka emotikon juga dihilangkan dalam proses pembersihan data.

b. Case Folding

Case Folding adalah tahapan untuk mengubah text menjadi suatu bentuk yang standar. Pada tahap ini dipilih *lowercase* untuk membuat huruf kapital menjadi huruf kecil (Ibrohim & Budi, 2019).

c. Text Normalization

Text Normalization yaitu tahap yang dilakukan untuk mengubah kata-kata nonformal menjadi kata-kata formal. Dalam penelitian ini, teks dinormalisasi menggunakan kamus yang berasal dari gabungan kamus dari beberapa karya sebelumnya (Alfina et al., 2017; Ibrohim & Budi, 2018; Salsabila et al., 2018).

d. Stopword Removal

Penghapusan *stopword* adalah bagian dari langkah *preprocessing* teks yang digunakan untuk menghapus kata-kata yang tidak relevan dari kalimat berdasarkan daftar *stopword*. Menghapus *stopword* tidak hanya akan membantu sistem meminimalkan interupsi, tetapi juga dapat mengurangi waktu pemrosesan sistem. Untuk menghilangkan *stopword*, penelitian ini menggunakan daftar *stopword* dari penelitian sebelumnya (Ibrohim & Budi, 2019).

e. Stemming

Stemming digunakan untuk menghapus artikel, preposisi, dan konjungsi dalam tweet yang dikumpulkan untuk mengambil akar kata. Lemmatisasi dilakukan untuk mengambil akar kata dari setiap tweet dengan menghapus semua lampiran menggunakan *library*. *Library* yang digunakan adalah *library* Sastrawi.

3. Feature Extraction

Ekstraksi fitur digunakan untuk membantu komputer dalam mengenali data yang akan digunakan dalam pembelajaran mesin. *Bag Of Words* (BoW) dan *Term Frequency - Revised Document Frequency* (TF-IDF) adalah dua dari beberapa teknik ekstraksi fitur yang digunakan untuk melatih model pembelajaran mesin.

a. Term Frequency - Revised Document Frequency (TF-IDF)

TF-IDF adalah metrik peringkat yang digunakan untuk meringkas dan mengambil informasi dengan tujuan mengukur pentingnya istilah dalam teks yang diberikan (E. Lee et al., 2022). Fungsi ekstraksi TF-IDF mengambil dua input : IDF dan TF. TF-IDF menyediakan token yang tampaknya tidak umum di dataset. Jika kata-kata yang tidak biasa muncul di banyak dokumen, relevansinya meningkat.

Kisaran parameter n-gram dibagi dengan TF-IDF. TF-IDF digunakan untuk menghitung bobot kata, yang mengembalikan bobot korpus untuk kata tertentu. Hasilnya adalah matriks kata tertimbang. Pendekatan TF banyak digunakan untuk ekstraksi fitur dan banyak digunakan untuk kategorisasi teks. Istilah frekuensi kejadian digunakan sebagai parameter dalam pelatihan klasifikasi. Fungsi TF tidak mempertimbangkan arti dari kata-kata yang jarang, tidak seperti TF-IDF, yang memberi bobot lebih kecil pada istilah yang lebih umum.

b. Bag Of Words (BoW)

BoW adalah salah satu fungsi ekstraksi yang digunakan dalam tugas NLP. Pendekatan ini adalah yang paling nyaman dan mudah beradaptasi untuk mendapatkan fitur dokumen (E. Lee et al., 2022). Histogram dari kata diperiksa di BoW dalam teks. Frekuensi kata digunakan sebagai fungsi untuk pelatihan kalimat. Dalam penelitian ini, pendekatan BoW diimplementasikan menggunakan *Count Vectorizer* dari *library scikit-learn Python*. Vektorisasi disebut teknik memperoleh vektor numerik dengan mengubah sekumpulan data tekstual. Frekuensi kata yang dihitung menunjukkan bahwa token telah dihitung dan membuat vektor token. BoW memberikan nilai untuk setiap atribut berdasarkan frekuensi fungsi ini.

4. Synthetic Minority Oversampling Technique (SMOTE)

SMOTE adalah teknik di mana *oversampling* kelas minoritas dilakukan dengan menghasilkan sampel sintetik baru (Thabtah et al., 2019). SMOTE didasarkan pada *k-nearest neighbor* untuk menghasilkan sampling sintetik baru dalam ruang fitur berdasarkan persentase tertentu untuk kelas minoritas (Swana et al., 2022). SMOTE dapat menghasilkan data sintetik baru berdasarkan data kelas minoritas yang ada tanpa mereplikasinya untuk mengatasi tantangan *overfitting*. SMOTE pertama-tama memilih setiap sampel dari sampel minoritas secara berurutan sebagai sampel akar untuk sintesis sampel tambahan untuk mensintesis sampel baru dan diulang sebanyak n kali. Terakhir, interpolasi linier dilakukan antara sampel dan setiap sampel tambahan untuk menghasilkan n sampel yang disintesis.

Algoritma SMOTE secara efektif memecahkan masalah *overfitting* yang disebabkan oleh replikasi buta teknik *oversampling* acak (Gao et al., 2020). Pengembangan SMOTE telah banyak dilakukan oleh para peneliti sebelumnya, berikut ini merupakan beberapa pengembangan teknik SMOTE yang akan dievaluasi kinerjanya pada penelitian ini :

a. *Support Vector Machine* SMOTE (SVM-SMOTE)

SVMSMOTE merupakan teknik *oversampling* yang digunakan untuk mengatasi masalah ketidakseimbangan dataset dengan cara memodifikasi dataset training untuk menghasilkan dataset training yang seimbang untuk setiap kelas (Almajid & Arifudin, 2022). Teknik SVM-SMOTE menggunakan teknik interpolasi, ekstrapolasi, dan SVM untuk membuat *instance* sintetik baru.

b. Kmeans-SMOTE

Metode Kmeans-SMOTE. K-means-SMOTE terdiri dari tiga langkah: *clustering*, *filtering*, dan *oversampling* (Last et al., 2017). Pada langkah *clustering*, ruang input dikelompokkan menjadi k grup menggunakan k-means *clustering*. Langkah *filtering* memilih *cluster* untuk *oversampling* dan menentukan berapa banyak sampel yang akan dihasilkan di setiap cluster. Langkah filter mengalokasikan lebih banyak sampel yang dihasilkan ke cluster minoritas yang jarang daripada yang padat. Untuk setiap kluster yang difilter, hitung bobot sampling berdasarkan *density* minoritasnya. Terakhir, pada langkah *oversampling*, melakukan *oversample* setiap cluster yang difilter menggunakan SMOTE. Jumlah sampel yang akan dihasilkan dihitung dengan menggunakan bobot sampel.

c. *Borderline*-SMOTE

Borderline-SMOTE merupakan pengembangan algoritma SMOTE yang mengatasi sampel kelas yang tumpang tindih karena lokasi data kelas mayoritas yang berdekatan saat mensintesis data kelas minoritas (T. Lee et al., 2020). Algoritma *Borderline*-SMOTE melakukan *oversampling* pada *instance* kelas minoritas di dekat garis batas. Setelah mengekstrak m tetangga terdekat dari *instance* kelas minoritas, lebih dari setengah m membangun himpunan dengan memilih *instance* kelas minoritas yang sesuai dengan kelas mayoritas.

5. Algoritma Klasifikasi

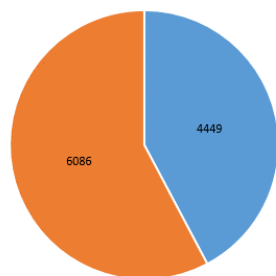
Penelitian ini menggunakan algoritma *Machine Learning* untuk melakukan klasifikasi ujaran kebencian. Algoritma *Machine Learning* yang digunakan antara *Random Forest*, *Support Vector Machine*, *Logistic Regression* dan *Naive Bayes*. *Random Forest* memiliki potensi untuk menganalisis karakteristik klasifikasi dari interaksi yang kompleks dan sangat kuat terhadap data derau dan data dengan nilai yang hilang (Xu et al., 2020). Dalam klasifikasi teks, *Naive Bayes* banyak digunakan karena kejelasan dan keefektifannya pada kinerja yang dihasilkan. Penerapan *Support Vector Machine* pada klasifikasi teks juga memberikan hasil presisi yang sangat baik. Berdasarkan penelitian sebelumnya menyatakan bahwa regresi logistik dapat menghasilkan hasil yang lebih baik dalam mengklasifikasikan teks daripada metode lain (Aborisade & Anwar, 2018).

6. Evaluasi Kinerja

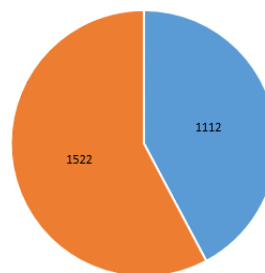
Metrics Performance adalah konstruksi logis-matematis yang diperoleh dengan selisih antara nilai aktual dan nilai prediksi (Mullah & Zainon, 2021). *Metric performance* yang digunakan untuk mengukur kinerja dalam penelitian ini meliputi *accuracy* yang merupakan rasio sampel yang diprediksi dengan benar terhadap jumlah total sampel (Challa et al., 2021). *Precision* merupakan proporsi sampel positif yang diprediksi dengan benar dari jumlah total sampel yang diprediksi sebagai positif (Challa et al., 2021). *Recall* merupakan proporsi sampel positif yang diprediksi dengan benar dari total sampel positif (Challa et al., 2021). *F1-Score* merupakan estimasi menyeluruh dari akurasi model dan dapat dihitung sebagai rata-rata dari presisi dan *recall* (Challa et al., 2021).

HASIL DAN PEMBAHASAN

Penelitian ini melakukan percobaan menggunakan *Google Colabs* dengan lima pendekatan yaitu dengan *Data Imbalanced*, SMOTE, SVM-SMOTE, Kmeans-SMOTE, *Borderline*-SMOTE. *Dataset* dibagi menjadi 80% data *training* dan 20% data *testing*. Lima Pendekatan tersebut diterapkan pada data *training* untuk mengatasi data yang tidak seimbang. Pembagian dataset dapat dilihat pada gambar 2 dan gambar 3 dimana daerah berwarna kuning mewakili data *non-hate speech* dan daerah warna biru mewakili data *hate speech*.



Sumber : Hasil Penelitian (2023)
Gambar 2. Jumlah Data *Training*



Sumber : Hasil Penelitian (2023)
Gambar 3. Jumlah Data *Testing*

Data *training* yang digunakan sebanyak 10.535 data *text* yang terdiri dari 4449 data *hate speech* dan 6086 data *non-hate speech*. Data *testing* yang digunakan untuk evaluasi sebanyak 2634 data *text* terdiri dari 1522 *non-hate speech* dan 1112 data *hate speech*. Untuk memverifikasi kinerja model, penelitian ini menggunakan empat pengklasifikasi yang umum yaitu *Random Forest* (RF), *Logistic Regression* (LR), *Support Vector Machine* (SVM), dan *Naive Bayes* (NB). Keempat algoritma tersebut digunakan karena memiliki keunggulan seperti *Random Forest* memiliki potensi untuk menganalisis karakteristik klasifikasi dari interaksi yang kompleks dan sangat kuat terhadap data derau dan data dengan nilai yang hilang (Xu et al., 2020). *Naive Bayes* banyak digunakan karena kejelasan dan keefektifannya pada kinerja yang dihasilkan. *Support Vector Machine* mampu memberikan hasil presisi yang sangat baik. Dan berdasarkan penelitian sebelumnya regresi logistik menghasilkan hasil yang lebih baik dalam mengklasifikasikan teks daripada metode lain (Aborisade & Anwar, 2018). Konfigurasi algoritma *machine learning* yang digunakan pada penelitian ini dapat dilihat pada tabel 1.

Tabel 1. Konfigurasi Parameter Algoritma *Machine Learning*

Algoritma	Konfigurasi
<i>Random Forest</i>	<code>RandomForestClassifier()</code>
<i>Logistic Regression</i>	<code>LogisticRegression()</code>
<i>Support Vector Machine</i>	<code>svm.LinearSVC()</code>
<i>Naive Bayes</i>	<code>MultinomialNB()</code>

Sumber : Hasil Penelitian (2023)

Konfigurasi parameter yang digunakan pada setiap metode SMOTE dalam penelitian ini dapat dilihat pada tabel 2. Penerapan metode SMOTE diterapkan pada algoritma di tabel 1 secara bergantian. Oleh karena itu, terdapat 4 kali eksperimen yang dilakukan dalam penelitian ini.

Tabel 2. Konfigurasi Parameter Metode SMOTE

Metode	Konfigurasi
SMOTE	<code>SMOTE(sampling_strategy='auto', random_state=None, k_neighbors=10, n_jobs=None)</code>
SVM-SMOTE	<code>SVMSMOTE(sampling_strategy='auto', random_state=None, k_neighbors=10, n_jobs=None, m_neighbors=10, svm_estimator=None, out_step=0.5)</code>
Kmeans-SMOTE	<code>KMeansSMOTE(sampling_strategy='auto', random_state=0, k_neighbors=10, n_jobs=None, kmeans_estimator=50, cluster_balance_threshold='auto', density_exponent='auto')</code>
Borderline-SMOTE	<code>BorderlineSMOTE(sampling_strategy='auto', random_state=None, k_neighbors=10, n_jobs=None, m_neighbors=10, kind='borderline-1')</code>

Sumber : Hasil Penelitian (2023)

1. Hasil Kinerja *Imbalanced Data*

Berdasarkan penerapan pada data *imbalanced*, hasil kinerja dari klasifikasi yang telah dilakukan dapat dilihat pada tabel 3. Dari hasil yang diperoleh algoritma *Random Forest* menjadi algoritma klasifikasi yang terbaik dengan nilai *accuracy* 84,58%, *recall* 77,96%, *precision* 84,33% dan *F1-Score* 81,02%.

Tabel 3. Hasil Kinerja Penerapan Imbalanced Data

Algoritma	Accuracy	Recall	Precision	F1-Score
RF	84,58%	77,96%	84,33%	81,02%
SVM	82,11%	77,51%	79,59%	78,54%
LR	82,38%	73,47%	82,86%	77,88%
NB	80,97%	73,29%	79,98%	76,48%

Sumber : Hasil Penelitian (2023)

2. Hasil Kinerja Metode SMOTE

Setelah dilakukan eksperimen dengan parameter yang digunakan berdasarkan konfigurasi pada tabel 2, hasil kinerja dari penerapan metode SMOTE bahwa hasil kinerja yang diperoleh algoritma *Random Forest* menjadi algoritma klasifikasi yang terbaik dengan nilai *accuracy* 83,86%, *recall* 81,02%, *precision* 80,80% dan *F1-Score* 80,91% seperti yang terlihat pada tabel 4.

Tabel 4. Kinerja Penerapan SMOTE

Algoritma	Accuracy	Recall	Precision	F1-Score
RF	83,86%	81,02%	80,80%	80,91%
SVM	80,78%	80,03%	75,80%	77,86%
LR	82,68%	80,93%	78,67%	79,78%
NB	79,95%	85,07%	72,32%	78,18%

Sumber : Hasil Penelitian (2023)

Sedangkan hasil kinerja yang diperoleh algoritma *Random Forest* menjadi algoritma klasifikasi yang terbaik dengan nilai *accuracy* 84,01%, *recall* 79,94%, *precision* 81,78% dan *F1-Score* 80,85% seperti yang terlihat pada tabel 5.

Tabel 5. Kinerja Penerapan SVM-SMOTE

Algoritma	Accuracy	Recall	Precision	F1-Score
RF	84,01%	79,94%	81,78%	80,85%
SVM	81,43%	79,58%	77,15%	78,35%
LR	82,87%	80,48%	79,27%	79,87%
NB	62,98%	80,93%	54,11%	64,86%

Sumber : Hasil Penelitian (2023)

Kemudian berdasarkan tabel 6, hasil kinerja yang diperoleh algoritma *Random Forest* menjadi algoritma klasifikasi yang terbaik dengan nilai *accuracy* 83,94%, *recall* 75,80%, *precision* 84,55% dan *F1-Score* 79,94% adalah :

Tabel 6. Kinerja Penerapan Kmeans-SMOTE

Algoritma	Accuracy	Recall	Precision	F1-Score
RF	83,94%	75,80%	84,55%	79,94%
SVM	82,19%	77,60%	79,68%	78,63%
LR	82,72%	74,10%	83,14%	78,36%
NB	81,05%	78,14%	77,24%	77,69%

Sumber : Hasil Penelitian (2023)

Berdasarkan tabel 7, hasil kinerja yang diperoleh algoritma *Random Forest* menjadi algoritma klasifikasi yang terbaik dengan nilai *accuracy* 84,09%, *recall* 81,20%, *precision* 81,13% dan *F1-Score* 81,16% yaitu :

Tabel 7. Kinerja Penerapan Borderline-SMOTE

Algoritma	Accuracy	Recall	Precision	F1-Score
RF	84,09%	81,20%	81,13%	81,16%
SVM	80,48%	79,58%	75,51%	77,49%
LR	82,87%	81,56%	78,66%	80,08%

Algoritma	Accuracy	Recall	Precision	F1-Score
NB	80,03%	85,25%	72,36%	78,28%

Sumber : Hasil Penelitian (2023)

3. Hasil Analisis

Berdasarkan eksperimen pada penerapan metode SMOTE, berikut ini hasil analisis dari perbandingan *metric performance* untuk kinerja algoritma pada setiap metode SMOTE yang diterapkan. Berdasarkan tabel 8, perbandingan dari nilai *accuracy* dalam penerapan metode SMOTE yang digunakan untuk klasifikasi ujaran kebencian. Berdasarkan perbandingan hasil akurasi, algoritma *Random Forest* menjadi algoritma terbaik dengan nilai akurasi rata-rata diatas 83%. Penerapan metode *Borderline-SMOTE* memiliki nilai *accuracy* tertinggi dari metode SMOTE lainnya yaitu 84,09% seperti yang terlihat di tabel 8.

Tabel 8. Perbandingan Berdasarkan Nilai Accuracy

Algoritma	Borderline-SMOTE	Kmeans-SMOTE	SVM-SMOTE	SMOTE	Data Imbalanced
RF	84,09%	83,04%	84,01%	83,86%	84,58%
SVM	80,48%	82,19%	81,43%	80,78%	82,11%
LR	82,87%	82,72%	82,87%	82,68%	82,38%
NB	80,03%	81,05%	62,98%	79,95%	80,97%

Sumber : Hasil Penelitian (2023)

Adapun perbandingan dari nilai *recall* dalam penerapan metode SMOTE yang digunakan untuk klasifikasi ujaran kebencian. Berdasarkan perbandingan nilai *recall*, penerapan metode *Borderline-SMOTE* pada algoritma *Naive Bayes* menjadi model dengan nilai *recall* tertinggi yaitu 85,25%, terlihat pada tabel 9.

Tabel 9. Perbandingan Berdasarkan Nilai Recall

Algoritma	Borderline-SMOTE	Kmeans-SMOTE	SVM-SMOTE	SMOTE	Data Imbalanced
RF	81,20%	75,80%	79,94%	81,02%	77,96%
SVM	79,58%	77,60%	79,58%	80,03%	77,51%
LR	81,56%	74,10%	80,48%	80,93%	73,40%
NB	85,25%	78,14%	80,93%	85,00%	73,29%

Sumber : Hasil Penelitian (2023)

Sedangkan perbandingan dari nilai *F1-Score* dalam penerapan metode SMOTE yang digunakan untuk klasifikasi ujaran kebencian. Berdasarkan perbandingan nilai *F1-Score*, penerapan metode *Borderline-SMOTE* pada algoritma *Random Forest* menjadi model dengan nilai *F1-Score* tertinggi yaitu 81,16% seperti pada tabel 10.

Tabel 10. Perbandingan Berdasarkan Nilai F1-Score

Algoritma	Borderline-SMOTE	Kmeans-SMOTE	SVM-SMOTE	SMOTE	Data Imbalanced
RF	81,16%	79,84%	80,85%	80,91%	81,02%
SVM	77,49%	78,63%	78,35%	77,86%	78,54%
LR	80,08%	78,36%	79,87%	79,78%	77,88%
NB	78,28%	77,69%	64,86%	78,18%	76,48%

Sumber : Hasil Penelitian (2023)

Dari perbandingan nilai *precision* dalam penerapan metode SMOTE yang digunakan untuk klasifikasi ujaran kebencian. Berdasarkan perbandingan nilai *precision*, penerapan metode *Kmeans-SMOTE* pada algoritma *Random Forest* menjadi model dengan nilai *precision* tertinggi yaitu 84,55% seperti pada tabel 11.

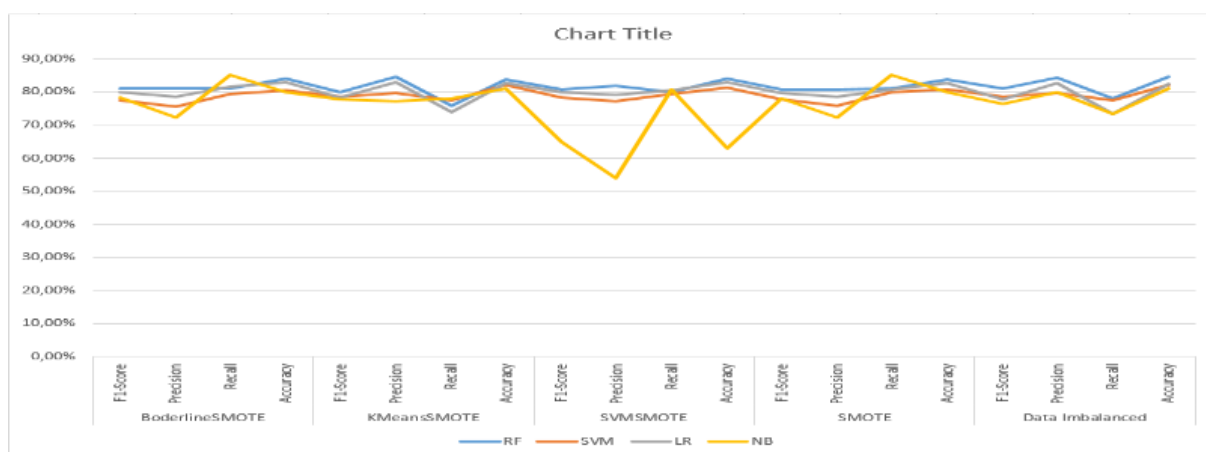
Tabel 11. Perbandingan Berdasarkan Nilai Precision

Algoritma	Borderline-SMOTE	Kmeans-SMOTE	SVM-SMOTE	SMOTE	Data Imbalanced
RF	81,13%	84,55%	81,78%	80,80%	84,33%
SVM	75,51%	79,68%	77,15%	75,80%	79,59%

Algoritma	<i>Borderline-SMOTE</i>	<i>Kmeans-SMOTE</i>	<i>SVM-SMOTE</i>	<i>SMOTE</i>	<i>Data Imbalanced</i>
LR	78,66%	83,14%	79,27%	78,67%	82,86%
NB	72,36%	77,24%	54,11%	72,32%	79,98%

Sumber : Hasil Penelitian (2023)

Berdasarkan analisis dari tiap tabel maka algoritma klasifikasi yang memiliki kinerja lebih baik untuk melakukan klasifikasi ujaran kebencian adalah algoritma *Random Forest*. Grafik menunjukkan hasil kinerja yang diperoleh algoritma *Random Forest* secara keseluruhan lebih tinggi dari algoritma lainnya. Berdasarkan penelitian ini, penerapan *Borderline-SMOTE* menjadi metode *oversampling* terbaik karena hasil *accuracy*, *recall* dan *F1-Score* mendapatkan nilai tertinggi dari metode *SMOTE* lainnya. Oleh karena itu, dengan melakukan *oversampling* pada data *training* menggunakan metode *Borderline-SMOTE* dapat meningkatkan hasil kinerja klasifikasi dengan hasil yang lebih baik dapat terlihat pada Gambar 4.



Sumber : Hasil Penelitian (2023)

Gambar 4. Perbandingan Hasil Kinerja Klasifikasi Ujaran Kebencian

KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa penerapan metode *Borderline-SMOTE* dalam menangani data *imbalanced* menghasilkan kinerja yang lebih baik dari metode *SMOTE* lainnya berdasarkan nilai *accuracy*, *recall*, *precision* dan *F1-Score* dengan masing-masing nilai sebesar 84,09%, 85,25%, 84,55% dan 81,16%. Kemudian, algoritma *Random Forest* merupakan algoritma *machine learning* yang memiliki kinerja yang bagus dari algoritma *Logistic Regression*, *Support Vector Machine* dan *Naive Bayes*. Untuk penelitian lebih lanjut, peneliti dapat menambahkan *Glove Embedding* dalam ekstraksi fitur untuk meningkatkan akurasi model. Selain itu, augmentasi data dapat diterapkan pada kumpulan data untuk menambah jumlah data.

REFERENSI

- Aborisade, O. M., & Anwar, M. (2018). Classification for authorship of tweets by comparing logistic regression and naive bayes classifiers. *Proceedings - 2018 IEEE 19th International Conference on Information Reuse and Integration for Data Science, IRI 2018*, 269–276. <https://doi.org/10.1109/IRI.2018.00049>
- Alfina, I., Mulia, R., Fanany, M. I., & Ekanata, Y. (2017). Hate Speech Detection in the Indonesian Language: A Dataset and Preliminary Study. *2017 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*, 233–238. <https://doi.org/10.1109/ICACSIS.2017.8355039>
- Almajid, A. S., & Arifudin, R. (2022). Multilayer Perceptron Optimization on Imbalanced Data Using SVM-SMOTE and One-Hot Encoding for Credit Card Default Prediction. *Journal of Advances in Information Systems and Technology*, 3(2), 67–74. <https://doi.org/10.15294/jaist.v3i2.57061>
- Antariksa, K., Wp, Y. S. P., & Ernawati, D. (2019). Klasifikasi Ujaran Kebencian pada Cuitan dalam Bahasa Indonesia. *Jurnal Buana Informatika*, 10(2), 164–171. <https://doi.org/10.24002/jbi.v10i2.2451>
- Challa, S. K., Kumar, A., & Semwal, V. B. (2021). A multibranch CNN-BiLSTM model for human activity recognition using wearable sensor data. *Visual Computer*, 0123456789. <https://doi.org/10.1007/s00371-021-02283-3>
- Gao, Q., Jin, X., Xia, E., Wu, X., Gu, L., Yan, H., Xia, Y., & Li, S. (2020). Identification of Orphan Genes in

- Unbalanced Datasets Based on Ensemble Learning. *Frontiers in Genetics*, 11(October). <https://doi.org/10.3389/fgene.2020.00820>
- Hairani, H., Saputro, K. E., & Fadli, S. (2020). K-means-SMOTE for handling class imbalance in the classification of diabetes with C4.5, SVM, and naive Bayes. *Jurnal Teknologi Dan Sistem Komputer*, 8(2), 89–93. <https://doi.org/10.14710/jtsiskom.8.2.2020.89-93>
- Ibrohim, M. O., & Budi, I. (2018). A Dataset and Preliminaries Study for Abusive Language Detection in Indonesian Social Media. *3rd International Conference on Computer Science and Computational Intelligence 2018*, 135, 222–229. <https://doi.org/10.1016/j.procs.2018.08.169>
- Ibrohim, M. O., & Budi, I. (2019). Multi-label Hate Speech and Abusive Language Detection in Indonesian Twitter. *Proceedings Ofthe Third Workshop on Abusive Language Online*, 46–57. <https://doi.org/10.18653/v1/w19-3506>
- Khanday, A. M. U. D., Rabani, S. T., Khan, Q. R., & Malik, S. H. (2022). Detecting twitter hate speech in COVID-19 era using machine learning and ensemble learning techniques. *International Journal of Information Management Data Insights*, 2(2), 100120. <https://doi.org/10.1016/j.jjime.2022.100120>
- Last, F., Douzas, G., & Bacao, F. (2017). *Oversampling for Imbalanced Learning Based on K-Means and SMOTE*. 1–19. <https://doi.org/https://doi.org/10.48550/arXiv.1711.00837>
- Lee, E., Rustam, F., Washington, P. B., Barakaz, F. El, Aljedaani, W., & Ashraf, I. (2022). Racism Detection by Analyzing Differential Opinions Through Sentiment Analysis of Tweets Using Stacked Ensemble GCR-NN Model. *IEEE Access*, 10, 9717–9728. <https://doi.org/10.1109/ACCESS.2022.3144266>
- Lee, T., Kim, M., & Kim, S. P. (2020). Data augmentation effects using borderline-SMOTE on classification of a P300-based BCI. *8th International Winter Conference on Brain-Computer Interface, BCI 2020*, 9–12. <https://doi.org/10.1109/BCI48061.2020.9061656>
- Marpaung, A., Rismala, R., & Nurrahmi, H. (2021). Hate Speech Detection in Indonesian Twitter Texts using Bidirectional Gated Recurrent Unit. *International Conference Knowledge and Smart Technology (KST)*, 186–190.
- Mullah, N. S., & Zainon, W. M. N. W. (2021). Advances in Machine Learning Algorithms for Hate Speech Detection in Social Media: A Review. *IEEE Access*, 9, 88364–88376. <https://doi.org/10.1109/ACCESS.2021.3089515>
- Murtopo, A. A., H, R. C. S., & Yustira. (2022). Systematic Literature Review : Klasifikasi Ujaran Kebencian Sosisal Media Dengan Algoritma Naïve Bayes. *Smart Comp*, 12(2), 19–28.
- Pratiwit, N. I., Budi, I., & Jiwanggi, M. A. (2019). Hate speech identification using the hate codes for Indonesian tweets. *ACM International Conference Proceeding Series*, 128–133. <https://doi.org/10.1145/3352411.3352432>
- Putri, T. T. A., Sriadhi, S., Sari, R. D., Rahmadani, R., & Hutahaean, H. D. (2020). A comparison of classification algorithms for hate speech detection. *IOP Conference Series: Materials Science and Engineering*, 830(3), 1–6. <https://doi.org/10.1088/1757-899X/830/3/032006>
- Salsabila, N. A., Winatmoko, Y. A., Septiandri, A. A., & Jamal, A. (2018). Colloquial Indonesian Lexicon. *2018 International Conference on Asian Language Processing (IALP)*, 236–239. <https://doi.org/10.1109/IALP.2018.8629151>
- Susanti, N. A., Walid, M., & Hoiriyah. (2022). KLASIFIKASI DATA TWEET UJARAN KEBENCIAN DI MEDIA SOSIAL. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 6(2), 538–543. <https://doi.org/10.36040/jati.v6i2.5174>
- Swana, E. F., Doorsamy, W., & Bokoro, P. (2022). Tomek Link and SMOTE Approaches for Machine Fault Classification with an Imbalanced Dataset. *Sensors*, 22(9). <https://doi.org/10.3390/s22093246>
- Thabtah, F., Hammoud, S., & Kamalov, F. (2019). Data Imbalance in Classification : Experimental Evaluation. *Information Sciences*. <https://doi.org/10.1016/j.ins.2019.11.004>
- Xu, Z., Shen, D., Nie, T., & Kou, Y. (2020). A hybrid sampling algorithm combining M-SMOTE and ENN based on Random forest for medical imbalanced data. *Journal of Biomedical Informatics*, 107(June), 103465. <https://doi.org/10.1016/j.jbi.2020.103465>