

ANALISIS PENANGANAN DATA TIDAK SEIMBANG TERHADAP KINERJA KLASIFIKASI SENTIMEN MULTIKELAS PADA ULASAN MARKETPLACE TOKOPEDIA

Nauval Alfarizi¹, Satria Sinurat², Adi Putra³, Muhammad Amin⁴, Prima Lydia⁵
Universitas Panca Budi, Medan

e-mail: ¹nauvalalfarizi026@gmail.com, ²satria.sinurat@yahoo.co.id,

³myreal.autobot@gmail.com, ⁴mhdamin9977@gmail.com, ⁵2anpu3@gmail.com

Abstract: *The development of digital marketplaces has led to an increasing number of user reviews, which can be used to understand consumer perceptions of products and services. However, sentiment analysis in marketplace reviews faces a major challenge: class imbalance, where positive sentiment often dominates to an extreme. This study aims to analyze the effects of various imbalanced data-handling techniques on the performance of machine-learning-based multiclass sentiment classification in Tokopedia marketplace reviews. The dataset used consists of 56,981 reviews with three sentiment classes, with more than 97% of them being positive. Feature extraction was performed using the TF-IDF method, resulting in 17,765 features. The handling of data imbalance was tested through four scenarios: class weighting, Random Oversampling, SMOTE, and ADASYN, with the Naive Bayes, Logistic Regression, and Random Forest algorithms. The experimental results show that Random Forest with SMOTE achieves the highest accuracy of 0.9749 but has limitations in recognizing minority classes, with a recall of 0.3786. In contrast, Logistic Regression with Random Oversampling provides the most balanced performance with the highest F1-score (macro) value of 0.4992 and recall of 0.5866.*

Keywords: *Analysis, Sentiment, Imbalanced Data, Multi-Class Classification F1-Score*

Abstrak: Perkembangan marketplace digital menyebabkan meningkatnya jumlah ulasan pengguna yang dapat dimanfaatkan untuk memahami persepsi konsumen terhadap produk dan layanan. Namun, analisis sentimen pada ulasan marketplace menghadapi tantangan utama berupa ketidakseimbangan distribusi kelas, di mana sentimen positif sering kali mendominasi secara ekstrem. Penelitian ini bertujuan untuk menganalisis pengaruh berbagai teknik penanganan data tidak seimbang terhadap kinerja klasifikasi sentimen multikelas pada ulasan marketplace Tokopedia berbasis machine learning. Dataset yang digunakan terdiri dari 56.981 ulasan dengan tiga kelas sentiment, di mana proporsi sentimen positif mencapai lebih dari 97%. Ekstraksi fitur dilakukan menggunakan metode TF-IDF yang menghasilkan 17.765 fitur. Penanganan ketidakseimbangan data diuji melalui empat skenario, yaitu class weighting, Random Oversampling, SMOTE, dan ADASYN, dengan algoritma Naive Bayes, Logistic Regression, dan Random Forest. Hasil eksperimen menunjukkan bahwa Random Forest dengan SMOTE menghasilkan akurasi tertinggi sebesar 0,9749, namun memiliki keterbatasan dalam mengenali kelas minoritas dengan nilai recall 0,3786. Sebaliknya, Logistic Regression dengan Random Oversampling memberikan performa paling seimbang dengan nilai F1-score (macro) tertinggi sebesar 0,4992 dan recall 0,5866.

Kata kunci: Analisis, Sentimen, Data Tidak Seimbang, Klasifikasi Multi Kelas F1-Score

PENDAHULUAN

Pesatnya perkembangan teknologi informasi mendorong industri untuk terus berinovasi dalam menghadirkan layanan

digital yang mendukung kemudahan aktivitas [1]. kemajuan teknologi dan investasi mendorong perubahan besar di pasar *e-commerce* Indonesia. Melalui sudut pandang inovasi disrupti, dapat melihat bahwa teknologi bukan sekadar alat, melainkan kekuatan yang mengubah ekosistem bisnis secara mendasar. [2]. Terkait Ulasan, terdapat ungkapan bahwa rating memiliki pengaruh yang signifikan terhadap keputusan pembelian. Rating yang tinggi merepresentasikan tingkat kepercayaan serta persepsi kualitas produk dari sudut pandang konsumen, sehingga mampu menekan keraguan dan memperkuat keyakinan dalam proses pengambilan keputusan pembelian [3]. analisis sentimen pada ulasan produk *e-commerce* dengan memanfaatkan algoritma *machine learning* berfokus pada beberapa aspek utama, antara lain perumusan pertanyaan terkait prediksi perilaku pengguna dengan manfaat ekonomi, serta pengembangan yang mengintegrasikan berbagai algoritma pembelajaran mesin dan metodologi teknik yang digunakan. Analisis sentimen juga berupaya agar ulasan produk *e-commerce* dapat untuk mengevaluasi efektivitas dari penerapan algoritma yang diimplementasikan [4].

Data ulasan yang bersifat tidak terstruktur dapat disajikan dalam bentuk teks bebas harus memerlukan teknik pengolahan yang tepat. Salah satu pendekatan yang banyak digunakan adalah analisis sentimen. Analisis sentimen merupakan bagian dari *Natural Language Processing* (NLP) yang bertujuan untuk mengidentifikasi serta mengklasifikasikan opini pengguna ke dalam kategori sentimen tertentu dengan memanfaatkan pendekatan berbasis *machine learning* [5]. *Supervised learning* merupakan pendekatan yang umum digunakan dalam klasifikasi teks karena memanfaatkan data berlabel untuk mempelajari pola dari fitur teks. Pada tugas klasifikasi sentimen, algoritma *supervised learning* seperti *Naive Bayes*, *Random Forest*, dan *Logistic Regression* banyak digunakan sebagai *model baseline*

untuk melakukan evaluasi dan perbandingan kinerja dalam proses klasifikasi sentiment [6]. *Machine learning* melakukan percobaan klasifikasi sentimen pada data komentar platform social menggunakan prinsip analisis untuk membandingkan kinerja algoritma *Random Forest* dan *AdaBoost*. Hasil penelitian menunjukkan bahwa *Random Forest* menghasilkan tingkat akurasi yang lebih tinggi dibandingkan dengan *AdaBoost* yang membuktikan bahwa peran *machine learning* sangat vital dalam melakukan sebuah uji data [7].

Salah satu penelitian menunjukkan bahwa pengambilan dan pengolahan data teks merupakan bagian dari kajian *pattern recognition* dan *artificial intelligence* yang telah lama diterapkan [12]. Pada klasifikasi multikelas dengan distribusi data tidak seimbang, model klasifikasi cenderung bias terhadap kelas mayoritas. Penelitian yang menunjukkan bahwa meskipun akurasi tinggi dapat dicapai pada kondisi tertentu, pendekatan GSVM-RU lebih mampu meningkatkan kemampuan model dalam mengenali kelas minoritas [13]. Berikutnya penelitian yang dilakukan membahas mengenai Penerapan teknik *Synthetic Minority Over-sampling Technique* (SMOTE) pada algoritma *Logistic Regression* dilaporkan mampu meningkatkan kinerja klasifikasi pada data tidak seimbang, khususnya berdasarkan metrik *F1-score*. Beberapa skenario penerapan SMOTE menunjukkan bahwa pendekatan ini memberikan performa yang lebih baik dibandingkan kombinasi dengan tahapan pra-proses tertentu, sehingga SMOTE dinilai efektif dalam meningkatkan kemampuan model dalam mengenali kelas minoritas [14]. Teknik PCA-KMeans dan ADASYN diterapkan untuk mengatasi ketidakseimbangan data, dengan pembagian dataset ke dalam data latih dan data uji pada rasio 80:20. Tiga skenario model kemudian dibangun, yaitu tanpa penyeimbangan data, dengan penyeimbangan PCA-KMeans, dan dengan penyeimbangan ADASYN, yang

selanjutnya digunakan sebagai masukan untuk pelatihan model DNN berbasis *focal loss* [15].

Berdasarkan latar belakang yang telah diuraikan sebelumnya, penelitian ini bertujuan untuk menganalisis pengaruh berbagai teknik penanganan data tidak seimbang terhadap kinerja klasifikasi sentimen multikelas pada data ulasan marketplace Tokopedia berbasis machine learning. Penelitian ini tidak hanya menerapkan satu metode penanganan data tidak seimbang, tetapi melakukan perbandingan beberapa teknik, yaitu *Random Oversampling*, *Synthetic Minority Over-sampling Technique* (SMOTE), ADASYN, serta pendekatan *class weighting*, guna memperoleh gambaran yang lebih komprehensif mengenai efektivitas masing-masing teknik.

Untuk mengevaluasi kinerja model secara lebih objektif pada kondisi data tidak seimbang, penelitian ini menggunakan beberapa algoritma *supervised learning*, yaitu *Naive Bayes*, *Logistic Regression*, dan *Random Forest*, dengan pengukuran performa yang tidak hanya berfokus pada nilai *accuracy*, tetapi juga mempertimbangkan metrik *precision*, *recall*, dan *F1-score*, khususnya *F1-score* (macro), serta analisis *confusion matrix* pada setiap kelas sentimen. Pendekatan evaluasi ini diharapkan mampu memberikan penilaian yang lebih adil terhadap kemampuan model dalam mengenali kelas minoritas.

METODE

Dalam penelitian ini digunakan algoritma klasifikasi, yaitu *Naive Bayes* sebagai *model baseline*, *Logistic Regression* sebagai model pembanding *linear*, serta *Random Forest* sebagai model pembanding *non-linear*. Setiap algoritma dievaluasi pada kondisi tanpa penanganan ketidakseimbangan kelas dan dengan masing-masing teknik penanganan *imbalanced* secara terpisah. Proses penelitian meliputi tahapan

pembersihan data, prapemrosesan teks, ekstraksi fitur menggunakan metode TF-IDF, pembagian data latih dan data uji secara terstratifikasi, serta pelatihan model pada berbagai skenario penanganan data tidak seimbang. Evaluasi kinerja model dilakukan menggunakan metrik akurasi, *precision*, *recall*, dan *F1-score*, dengan penekanan pada *F1-score* macro untuk menilai performa model secara adil terhadap kelas minoritas pada klasifikasi sentimen multikelas.

Pendekatan Penelitian

Pendekatan *machine learning* berbasis *supervised learning* diterapkan karena data yang digunakan telah memiliki label sentimen. Eksperimen dilakukan untuk menganalisis pengaruh penanganan data tidak seimbang terhadap performa model klasifikasi.

Dataset

Dataset yang digunakan dalam penelitian ini berupa data ulasan marketplace Tokopedia yang telah diberi label sentimen multikelas, yaitu positif, netral, dan negatif. Dataset memiliki distribusi kelas yang tidak seimbang, di mana kelas sentimen positif mendominasi dibandingkan kelas sentimen lainnya. Kondisi ini menjadi dasar penerapan teknik penanganan data tidak seimbang pada tahap pemodelan. Dataset diperoleh dari platform Kaggle dengan doi berikut : <https://doi.org/10.34740/kaggle/ds/8992580>

review_id	review_text	sentiment_label
11111111	positive
22222222	positive
33333333	positive
44444444	positive
55555555	positive
66666666	positive
77777777	positive
88888888	positive
99999999	positive
10101010	positive

Gambar 1 Hasil Data Cleaning

```

review_id      review_text
40  1137072591  Kualitas barang tidak rata, ada yg kecil kurang...
51  348885708   Datang pecah telur nya tadinya nya mau kompi...
54  958059464   Sudah berapa kali pecah di sini, , dulu oke bg...
77  309693973   pembelian 2 tray.. tray pertama pecah total 2...

...

57707  309694452   sudah sampai dg baik...terima kasih gan
57715  727351397   bagian belakang bengkok dan nyentak ban... ada...
57778  1137084854   Biar kan foto yg berbicara. Kudu dibotok dulu...
58348  1495294876   Belang bagian sisi
58731  1238801588   di mana pembelian tidak ada pilihan memilih si...

...

sentiment_label
40  negative
51  negative
54  negative
77  negative

...

57707  negative
57715  negative
57778  negative
58348  negative
58731  negative
    
```

Gambar 2 Sentimen multi kelas

Data tersebut didapatkan dari hasil extrasi dan proses *cleaning data* yang dilakukan dengan *Data Preparation Pipeline* terhadap dataset marketplace Tokopedia. Data tersebut sebanyak 58856 baris dan 4 kolom.

Data Cleaning

Dalam melakukan *data cleaning*, data dapat dilihat pada gambar 3 berikut dibawah ini dengan standar data preparation dalam bentuk *pipeline*.



Gambar 3 Alur Data Cleaning

Text Preprocessing

Tahapan prapemrosesan teks dilakukan sebelum proses penanganan data tidak seimbang. *Preprocessing* meliputi pembersihan teks awalan dengan mengubah seluruh karakter menjadi huruf kecil, menghapus URL, angka, karakter non-alfabet, serta simbol dan juga spasi. Tahapan ini bertujuan untuk mengurangi data kasar pada data teks dan memastikan kualitas fitur yang dihasilkan sebelum dilakukan ekstraksi fitur menggunakan TF-IDF. Dalam penelitian kali ini tahapan stemming

dilewatkan mengingat jumlah data yang cukup besar.

TF-IDF

Hasil ekstraksi fitur menggunakan TF-IDF menunjukkan bahwa dari 56.981 data ulasan yang telah melalui tahap prapemrosesan, diperoleh matriks fitur berdimensi 17.765 kata unik. Tingginya dimensi fitur ini mencerminkan keragaman kosakata yang digunakan pelanggan dalam memberikan ulasan produk di marketplace Tokopedia. Selanjutnya, data dibagi menjadi data latih dan data uji dengan rasio 80:20 menggunakan teknik stratified split untuk menjaga proporsi kelas sentiment

Tabel 1 Hasil Ekstraksi fitur TF-IDF

Tahapan Proses	Jumlah Data / Dimensi
Total data ulasan setelah prapemrosesan	56.981
Dimensi fitur TF-IDF	17.765
Jumlah data latih (80%)	45.584
Jumlah data uji (20%)	11.397
Metode pembagian data	Stratified Train-Test Split

Imbalance Data

Untuk mengatasi permasalahan ketidakseimbangan kelas, penelitian ini menerapkan beberapa skenario penanganan data tidak seimbang yang diuji secara terpisah pada data latih. Pendekatan yang digunakan meliputi tanpa penanganan data tidak seimbang sebagai skenario dasar, pembobotan kelas (*class weighting*) pada tahap pemodelan, serta teknik resampling berbasis *oversampling*, yaitu *Random Oversampling*, *Synthetic Minority Oversampling Technique* (SMOTE), dan *Adaptive Synthetic Sampling* (ADASYN). Seluruh teknik penanganan *imbalance* diterapkan hanya pada data latih setelah proses pembagian data secara *stratified* untuk menghindari terjadinya data leakage.

HASIL DAN PEMBAHASAN

Pembagian data latih dan data uji dilakukan menggunakan stratified split dengan rasio 80:20 untuk menjaga proporsi kelas sentimen. Hasil pembagian menunjukkan bahwa distribusi kelas pada data latih dan data uji tetap konsisten, meskipun dataset memiliki tingkat ketidakseimbangan yang sangat tinggi. Hal ini dibuktikan dengan X shape : (56981, 17765) Dimana jumlah ulasan memiliki korelasi hubungan dengan jumlah fitur TF-IDF dengan hasil interpretasi yang dapat dilihat pada tabel 2 dibawah.

Tabel 2 Tabel Imbalanced data

Dataset	Jumlah (Baris)	Data Jumlah (Kolom)	Fitur
X total	56.981	17.765	
X_train	45.584	17.765	
X_test	11.397	17.765	

Juga pada y variabel, dimana y_train merupakan hasil interpretasi yang menyebutkan bahwa y memiliki hasil distribusi data hasil dari *split training* yang dapat dilihat pada tabel 3 dan tabel 4 menjelaskan bahwa terdapat proposional data yang terklarifikasi sebagai *imbalanced data* (data tidak seimbang) yang digunakan sebagai *baseline* awal dalam penanganan data tidak seimbang dikarenakan mayoritas dari *sentiment* terklasifikasi sebagai *sentiment positive*.

Tabel 3 Distribusi Data Latih (y_train)

Kelas Sentimen	Proporsi
Positive	0.974618
Negative	0.012921
Neutral	0.012461

Tabel 4 Distribusi Data Latih (y_test)

Kelas Sentimen	Proporsi
Positive	0.974642
Negative	0.012898
Neutral	0.012459

Hasil pembagian dari data latih dan data uji merupakan Gambaran , Dimana data dengan sentimen *positive* memiliki bias yang cukup tinggi dibandingkan

dengan *negative* dan *neutral* pada *dataset*. Dalam proporsi sentiment kelas cenderung konsisten dengan dominasi positive yang cukup tinggi dan bias terhadap kelas lain. Dikarenakan hal ini penggunaan matrix evaluasi diperlukan agar memperlihatkan hasil komparasi yang lebih presisi. Dengan mempertimbangkan penggunaan yang lebih representative seperti *precision*, *recall*, dan *f-1 score(macro)* agar dapat melihat hasil yang lebih jelas dan meminimalisir efek bias yang terjadi.

Baseline Model

Hasil dari model awal memperlihatkan terkait dari interpretasi yang dilakukan yaitu data memiliki Tingkat akurasi sebesar 97% dari model awal hal ini membuktikan adanya data hasil yaitu dominasi kelas *positive*. Hal ini dapat dilihat pada gambar 4 dibawah berikut ini.

Model	Imbalance Handling	Accuracy	Precision	Recall	F1_Macro
0	Naive Bayes	Baseline 0.974467	0.650268	0.337908	0.338281
1	Logistic Regression	Baseline 0.975695	0.645036	0.371892	0.396832
2	Random Forest	Baseline 0.975081	0.758525	0.353941	0.368526

Gambar 4 Evaluasi Model Baseline

Hal ini berbanding terbalik dengan nilai *f1-macro* yang relatif rendah , menunjukkan performa dalam mengenali dan mempelajari sentiment kelas neutral dan negative belum optimal mengingat data lebih dari 97% data terkategori dalam positive. Namun pada model baseline didapat hasil berupa algortima logistic regression menunjukkan nilai f1-macro paling tinggi diantara model algortima lainnya berupa 0.396832 dibandingkan dengan 0.338281 dan 0.368526.

Classweight Model

Dari hasil berikutnya didapatkan hasil berupa Tingkat akurasi tetap sama yaitu 97% namun pada *logistic regression* mengalami Tingkat penurunan akurasi sebesar 4% menjadi 93% diakibatkan model mulai dipaksa untuk menebak kelas minoritas secara lebih kolektif dan adil dibuktikan dengan Tingkat *f1-score*

yang mengalami peningkatan signifikan sebesar 0.485763. hal ini dapat dilihat pada gambar 5 terkait hasil evaluasi yang diterapkan dengan *class weight*.

	Model	Imbalance Handling	Accuracy	Precision	Recall	F1_Macro
0	Naive Bayes	Class Weight	0.974467	0.658268	0.337908	0.338281
1	Logistic Regression	Class Weight	0.936124	0.444599	0.601291	0.485763
2	Random Forest	Class Weight	0.973590	0.523301	0.348956	0.358550

Gambar 5 Evaluasi Class weight Random Oversampling

Dari hasil pengujian yang dilakukan, terdapat hasil penurunan bobot akurasi yang cukup sistematis. Dimana penurunan paling tinggi pada Tingkat akurasi didapat pada algoritma *naive bayes* dengan nilai akurasi 0,902255.

	Model	Imbalance Handling	Accuracy	Precision	Recall	F1_Macro
0	Naive Bayes	Random Oversampling	0.902255	0.412015	0.628465	0.446482
1	Logistic Regression	Random Oversampling	0.946477	0.459624	0.586612	0.489232
2	Random Forest	Random Oversampling	0.971308	0.486273	0.361760	0.376953

Gambar 6 Evaluasi Random oversampling

Hasil yang cukup signifikan juga dapat dilihat dengan Tingkat *recall* tinggi pada *naive bayes* sebesar 0,628465 hal ini membuktikan bahwa algoritma *naive bayes* lebih progresif terhadap Teknik penerapan *imbalanced data* meskipun data yang diperlihatkan model menebak dengan baik pada *logistic regression* dengan bobot *f1-macro* sebesar 0,499232 yang membuktikan model mulai kuat dalam menebak kelas minoritas dibandingkan sampling model lainnya dalam mengani data tidak seimbang.

Smote

Penerapan smote pada data latih menunjukkan kemampuan dalam mengenali sentiment multikelas. Hasil dari pengujian ini menunjukkan *logistic regression* memiliki bobot nilai paling tinggi dibandingkan dengan 2 algoritma lainnya dengan bobot *f1-score* sebesar 0.484180 menandakan bahwa model ini hanyalah interpretasi dengan bobot bias dibawah 50% namun memiliki hasil valuasi yang cukup dibandingkan dengan

ditandai hasil berupa random forest yang bekerja Bersama dalam melakukan variasi dengan nilai 0.405628.

	Model	Imbalance Handling	Accuracy	Precision	Recall	F1_Macro
0	Naive Bayes	SMOTE	0.905238	0.414384	0.629405	0.450025
1	Logistic Regression	SMOTE	0.940247	0.442752	0.597747	0.484180
2	Random Forest	SMOTE	0.974906	0.580500	0.378654	0.405628

Gambar 7 Evaluasi SMOTE

Adasyn

Hasil pengujian pada tiga model klasifikasi menunjukkan performa yang bervariasi. Random Forest mencatatkan akurasi tertinggi sebesar 97,48% dan presisi terbaik (0,588), namun memiliki *recall* paling rendah (0,374). Di sisi lain, Logistic Regression memberikan keseimbangan yang lebih baik dengan akurasi 93,96% dan skor F1_Macro tertinggi (0,480). Sementara itu, Naive Bayes memperoleh akurasi 90,55% dengan *recall* tertinggi di angka 0,627

	Model	Imbalance Handling	Accuracy	Precision	Recall	F1_Macro
0	Naive Bayes	ADASYN	0.905501	0.414442	0.627257	0.449923
1	Logistic Regression	ADASYN	0.939633	0.440334	0.593062	0.480928
2	Random Forest	ADASYN	0.974818	0.588093	0.374149	0.399673

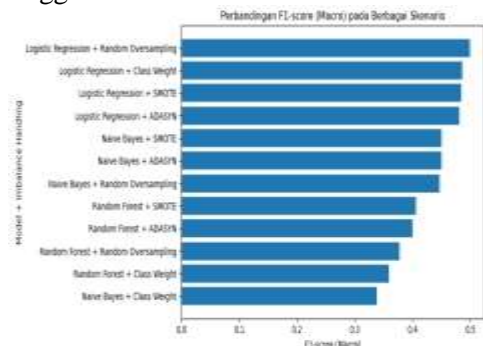
Gambar 8 Evaluasi ADASYN

	Model	Imbalance Handling	Accuracy	Precision	Recall	F1_Macro
0	Logistic Regression	Random Oversampling	0.946477	0.459624	0.586612	0.499232
1	Logistic Regression	Class Weight	0.936124	0.444599	0.601291	0.485763
2	Logistic Regression	SMOTE	0.940247	0.442752	0.597747	0.484180
3	Logistic Regression	ADASYN	0.939633	0.440334	0.593062	0.480928
4	Naive Bayes	SMOTE	0.905238	0.414384	0.629405	0.450025
5	Naive Bayes	ADASYN	0.905501	0.414442	0.627257	0.449923
6	Naive Bayes	Random Oversampling	0.902255	0.412015	0.628465	0.446482
7	Random Forest	SMOTE	0.974906	0.580500	0.378654	0.405628
8	Random Forest	ADASYN	0.974818	0.588093	0.374149	0.399673
9	Random Forest	Random Oversampling	0.971308	0.486273	0.361760	0.376953
10	Random Forest	Class Weight	0.973590	0.523301	0.348956	0.358550
11	Naive Bayes	Class Weight	0.974467	0.658268	0.337908	0.338281

Gambar 9 Performa Imbalanced data

Pada gambar 10 menunjukkan hasil Random Forest secara konsisten menghasilkan tingkat akurasi tertinggi, dengan puncaknya mencapai 0,974906 saat menggunakan teknik SMOTE, serta mencatatkan presisi yang relatif tinggi di atas 0,48 pada hampir semua skenario. Namun, meskipun akurasinya tinggi, model *Random Forest* memiliki nilai

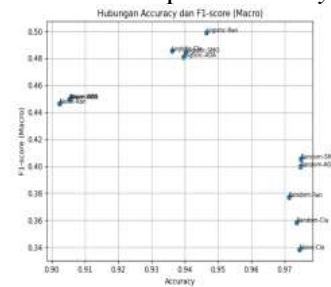
Recall yang rendah, berkisar antara 0,34 hingga 0,37, yang mengindikasikan kesulitan dalam mengidentifikasi kelas minoritas secara keseluruhan. Sebaliknya, model *Logistic Regression* dan *Naive Bayes* menunjukkan performa yang lebih seimbang dalam hal metrik Macro dan *recall*. Performa terbaik secara keseluruhan berdasarkan nilai macro (0,499232) diraih oleh *logistic regression* dengan teknik *random oversampling*, yang didukung oleh nilai *recall* sebesar 0,586612. Di sisi lain, *naive bayes* dengan teknik SMOTE mencatatkan nilai *recall* tertinggi di antara semua model, yaitu sebesar 0,629405, meskipun memiliki akurasi yang lebih rendah (sekitar 0,90) dibandingkan model lainnya. Secara umum, teknik penanganan ketidakseimbangan seperti SMOTE dan ADASYN cenderung meningkatkan kemampuan model dalam mengenali kelas minoritas (*recall*), sementara metode *class weight* pada *naive bayes* justru menghasilkan *recall* terendah yaitu 0,337908 meskipun akurasinya melonjak tinggi.



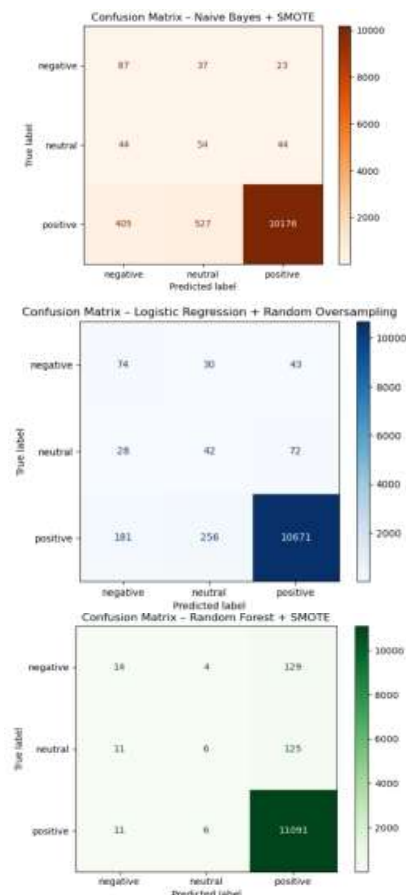
Gambar 10 Visualisasi Performa data

Jika meninjau keseimbangan antara akurasi dan kemampuan mendeteksi kelas minoritas melalui metrik *F1-score (Macro)*, *logistic regression* muncul sebagai model yang paling optimal. Nilai *F1_Macro* tertinggi (0,499232) dicapai oleh *Logistic Regression* dengan teknik *Random Oversampling*, yang juga memiliki nilai *Recall* yang cukup baik sebesar 0,586612. Grafik hubungan menunjukkan bahwa kelompok model *Logistic Regression* berada di posisi atas, menandakan performa *F1-score* yang

lebih unggul dibandingkan model *Random Forest* maupun *naive bayes*.



Gambar 12 Hubungan Antar Data



Gambar 13 Confusion Matrix Performa komparasi Algoritma

Dalam hasil sebaran yang diperoleh pada gambar 13 menunjukkan bahwa data klasifikasi sentiment bias dikarenakan dataset yang ada terklasifikasi sebagai dataset imbalance data, hal ini dibuktikan dengan confusion matrix memiliki label positive dalam 3 model performa perbandingan. Hasil pengujian menunjukkan bahwa model *Random Forest* (khususnya dengan teknik

SMOTE) mendominasi dalam hal akurasi dengan nilai tertinggi mencapai 0,974906. Hal ini terlihat jelas pada grafik sebaran di mana kelompok Random Forest berada di posisi paling kanan, yang mengindikasikan kemampuan prediksi global yang sangat kuat. Melalui *confusion matrix*, dapat dilihat bahwa *Random Forest* dengan SMOTE berhasil memprediksi 11.091 data positif dengan benar, namun sangat kesulitan mendeteksi kelas minoritas, di mana hanya 14 data negatif dan 6 data netral yang terklasifikasi dengan tepat. Keterbatasan ini tercermin dari nilai *Recall* yang rendah di angka 0,378654.

Di sisi lain, *Logistic Regression* dengan *Random Oversampling* muncul sebagai model yang paling seimbang dan optimal berdasarkan metrik *F1_Macro* (0,499232). Grafik hubungan menempatkan pada titik tertinggi di sumbu Y, menandakan performa yang lebih baik dalam menangani ketidakseimbangan kelas dibandingkan model lainnya. *Confusion matrix* untuk model ini menunjukkan peningkatan deteksi kelas minoritas yang signifikan dibandingkan *Random Forest*, dengan keberhasilan memprediksi 74 data negatif dan 42 data netral secara akurat. Sementara itu, *Naive Bayes* dengan SMOTE mencatatkan *Recall* tertinggi (0,629405), yang dibuktikan dengan jumlah prediksi benar pada kelas negatif sebanyak 87 data. Namun, model *Naive Bayes* memiliki akurasi terendah di kisaran 0,90 dan cenderung lebih banyak melakukan kesalahan prediksi (*false positive*) pada kelas positif dibandingkan model lainnya.

SIMPULAN

Hasil penelitian menunjukkan bahwa meskipun Random Forest dengan SMOTE menghasilkan nilai akurasi tertinggi, model tersebut masih cenderung bias terhadap kelas mayoritas dan memiliki keterbatasan dalam mendeteksi kelas minoritas. Sebaliknya, *Logistic*

Regression yang dikombinasikan dengan *Random Oversampling* memberikan performa paling seimbang berdasarkan nilai *F1-score (macro)* tertinggi, yang menandakan peningkatan kemampuan model dalam mengenali seluruh kelas sentimen secara lebih adil. Dengan demikian, dapat disimpulkan bahwa pemilihan model terbaik pada klasifikasi sentimen multikelas dengan data tidak seimbang tidak cukup hanya mempertimbangkan akurasi, tetapi perlu fokus pada keseimbangan performa antar kelas

DAFTAR PUSTAKA

- U. F. Tanamal, “Analisis Kepuasan Pengguna Aplikasi Livin by Mandiri Menggunakan Metode E-Servqual dan Importance Performance Analysis (IPA),” *J. Pendidik. dan Teknol. Indones.*, vol. 5, no. 2, pp. 547–568, 2025, doi: 10.52436/1.jpti.664.
- G. D. P. Dewi and A. E. Lusikooy, “E-commerce Transformation in Indonesia,” *Nation State J. Int. Stud.*, vol. 6, no. 2, pp. 117–138, 2024, doi: 10.24076/nsjis.v6i2.1304.
- A. Syahbani, M. Fatchurrohman, S. Shobikin, and N. K. Kusmayati, “Pengaruh Ulasan Online dan Rating terhadap Keputusan Pembelian di Marketplace Tiktok Shop,” *RIGGS J. Artif. Intell. Digit. Bus.*, vol. 4, no. 3, pp. 7207–7215, 2025, doi: 10.31004/riggs.v4i3.3021.
- A. Daza, N. D. González Rueda, M. S. Aguilar Sánchez, W. F. Robles Espíritu, and M. E. Chauca Quiñones, “Sentiment Analysis on E-Commerce Product Reviews Using Machine Learning and Deep Learning Algorithms: A Bibliometric Analysis and Systematic Literature Review, Challenges and Future Works,” *Int. J. Inf. Manag. Data Insights*, vol. 4, no. 2, 2024, doi: 10.1016/j.jjime.2024.100267.
- M. R. Tanjung, M. Iqbal, and Z. Sitorus,

- “Analisis Sentimen Google Review terhadap Mutu Kualitas Pendidikan pada Perguruan Tinggi STIE Al-Washliyah Sibolga dengan Metode Lexicon dan Algoritma Naive Bayes-Miftah Rusydi Tanjung et al Analisis Sentimen Google Review terhadap Mutu Kualitas Pendidikan,” vol. 07, no. 02, pp. 2721–1800, 2025, [Online]. Available: <https://journal.cattleyadf.org/index.php/jatilima/index>
- P. M. Susanti, M. Afdal, I. Permana, and A. Marsal, “Klasifikasi Sentimen Pengguna X Terhadap Pemboikotan Produk Pro Israel Menggunakan Algoritma Machine Learning,” *Technol. Sci.*, vol. 6, no. 4, pp. 2271–2280, 2025, doi: 10.47065/bits.v6i4.6533.
- A. Putri, M. Mustakim, R. Novita, and M. Afdal, “Analisis Sentimen Terhadap Publisher Rights Dalam Mengunggah Konten Digital Menggunakan Ensemble Learning,” *Build. Informatics, Technol. Sci.*, vol. 6, no. 1, pp. 64–73, 2024, doi: 10.47065/bits.v6i1.5179.
- A. Ichwani and R. Gantino, “Sentiment Analysis of Marketplace Application Reviews Using Support Vector Machine (SVM) and K-Nearest Neighbors (KNN),” *J. Technol. Open Source*, vol. 8, no. 2, 2024, [Online]. Available: <https://doi.org/10.36378/jtos.v7i2.4972>
- M. C. Untoro and M. A. N. M. Yusuf, “Evaluate of Random Undersampling Method and Majority Weighted Minority Oversampling Technique in Resolve Imbalanced Dataset,” *IT J. Res. Dev.*, vol. 8, no. 1, pp. 1–13, 2023,
- E. Erlin, Y. Desnelita, N. Nasution, L. Suryati, and F. Zoromi, “Dampak SMOTE terhadap Kinerja Random Forest Classifier berdasarkan Data Tidak seimbang,” *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 21, no. 3, pp. 677–690, 2022, doi: 10.30812/matrik.v21i3.1726.
- H. Cnn, “JURNAL RESTI Sentiment Analysis of ChatGPT on Indonesian Text,” vol. 5, no. 158, pp. 327–333, 2026.
- M. Amin, “Writing to Speech Conversion Application With Using an Android-Based Camera to Talk,” *Sci. Dev. Technol.*, vol. 3, no. 1, 2023, [Online]. Available: <http://creativecommons.org/licenses/by-sa/4.0/>
- Qorry Meidianingsih, D. E. Wardani, E. Salsabila, L. Nafisah, and A. N. Mutia, “Perbandingan Performa Metode Berbasis Support Vector Machine untuk Penanganan Klasifikasi Multi Kelas Tidak Seimbang,” *Stat. J. Theor. Stat. Its Appl.*, vol. 23, no. 1, pp. 8–18, 2023, doi: 10.29313/statistika.v23i1.1660.
- A. B. Putra Negara, “The Influence Of Applying Stopword Removal And Smote On Indonesian Sentiment Classification,” *Lontar Komput. J. Ilm. Teknol. Inf.*, vol. 14, no. 3, p. 172, 2023, doi: 10.24843/lkjiti.2023.v14.i03.p05.
- U. Ungkawa and M. A. Rafi, “Data Balancing Techniques Using the PCA-KMeans and ADASYN for Possible Stroke Disease Cases,” *J. Online Inform.*, vol. 9, no. 1, pp. 138–147, 2024, doi: 10.15575/join.v9i1.1293.
- Salman Abdurrahman. (2025). Tokopedia Product Reviews 2025 [Data set]. Kaggle. <https://doi.org/10.34740/KAGGLE/D S/8992580>