

ANALISIS SENTIMEN TERHADAP MARKETPLACE MENGUNAKAN ALGORITMA NAÏVE BAYES (STUDI KASUS: LAZADA INDONESIA)

Fatihanursari Dikananda, Arif Rinaldi Dikananda, Saeful Anwar

Teknik Informatika, STMIK IKMI Cirebon

Jl. Perjuangan No.10B, Kota Cirebon

fatiha.dikananda@gmail.com

ABSTRAK

Penelitian ini menganalisis sentimen terhadap marketplace Lazada Indonesia menggunakan algoritma Naïve Bayes dengan teknik upsampling SMOTE. Marketplace menjadi platform penting dalam era digital, namun sering kali ulasan pengguna memiliki ketidakseimbangan dalam jumlah sentimen positif dan negatif, yang dapat mempengaruhi akurasi analisis. Data diambil dari Twitter API yang berisi 1.517 review. Algoritma Naïve Bayes diterapkan untuk analisis sentimen, dan teknik SMOTE digunakan untuk mengatasi ketidakseimbangan data. Hasil menunjukkan bahwa penerapan teknik SMOTE pada algoritma Naïve Bayes menghasilkan tingkat akurasi sebesar 81,89%, dibandingkan dengan 65,53% tanpa SMOTE. Penelitian ini menyimpulkan bahwa teknik SMOTE efektif dalam meningkatkan akurasi analisis sentimen, dan Lazada mendapatkan tanggapan positif dari penggunanya, meskipun ada beberapa kesalahan klasifikasi yang perlu diperbaiki. Hasil ini menunjukkan bahwa masyarakat memiliki respon yang sangat baik terhadap Lazada. Meskipun demikian, masih terdapat prediksi klasifikasi yang melenceng dari kenyataan.

Kata kunci: Analisis Sentimen, Lazada, Naïve Bayes, Synthetic Minority Oversampling Technique (SMOTE)

1. PENDAHULUAN

Covid-19 muncul pertama kali di Wuhan, China pada permulaan tahun 2020. Kehadiran pandemi tersebut mempengaruhi berbagai aspek kehidupan masyarakat secara langsung termasuk di Indonesia. Untuk mengurangi penyebaran virus, pemerintah membuat beberapa kebijakan, seperti menerapkan *social distancing* dan *physical distancing*. Pelaksanaan kebijakan tersebut berdampak pada segala aktivitas masyarakat, termasuk dalam bidang bisnis. Konsep bisnis di Indonesia mengalami perubahan yang sangat signifikan, dari transaksi jual-beli konvensional (*offline*) menjadi melalui berbagai metode transaksi secara *online*. Hal tersebut mempengaruhi perilaku konsumen dalam membeli produk, di mana salah satunya adalah dengan menggunakan transaksi *online* [1]

Marketplace merupakan sebuah elemen dari *e-commerce* yang memiliki perbedaan tertentu. Marketplace berfungsi sebagai sebuah *platform* pemasaran produk secara *online* yang mempertemukan banyak penjual dan pembeli untuk melakukan transaksi. Dengan adanya *marketplace*, para penjual tidak perlu repot mencari tempat berjualan *online* karena *platform* tersebut menyediakan layanan tersebut secara gratis. Hal ini memberikan peluang bagi pelaku usaha kecil untuk turut serta dalam *e-commerce* yang semakin berkembang. *Marketplace* juga menawarkan berbagai strategi pemasaran pada toko *online* yang dapat menjadi pertimbangan bagi konsumen dan calon konsumen untuk melakukan transaksi secara *online*. Dengan begitu, diharapkan konsumen menjadi lebih

mudah dalam bertransaksi melalui marketplace dibantu dengan adanya *review* dari para konsumen [2]

Salah satu *Marketplace* yang berkembang di Indonesia adalah Lazada. Pelayanan yang tidak memadai dari pihak marketplace dapat menyebabkan ketidakpuasan pelanggan dan berpotensi membuat pelanggan memberikan sentimen negatif terhadap Lazada. Lazada perlu membedakan masalah yang muncul dari survei yang diberikan oleh klien. Meskipun demikian, membaca dengan teliti dan mengelompokkan setiap survei yang diberikan oleh klien akan menghabiskan banyak waktu, dan dipandang tidak efektif. Untuk mengatasi masalah tersebut, diperlukan kerangka kerja yang dapat mengelompokkan sentimen ke dalam kelompok opini yang positif, negatif, dan netral yang biasanya disebut dengan analisis sentimen. Dengan menggunakan analisis sentimen ini dapat menghasilkan wawasan mengenai masalah-masalah yang dihadapi oleh para pelanggan, dan informasi ini dapat diberikan sebagai rekomendasi kepada Lazada untuk dijadikan saran dan pedoman dalam pengembangan produk, pelayanan, dan citra merek. Algoritma *Naïve Bayes* mampu melakukan klasifikasi secara efektif dan efisien. Bahkan ketika digunakan untuk menganalisis data dalam jumlah besar. Kata yang mewakili komentar berdasarkan kelas positif atau negatif akan dihitung oleh *Naïve Bayes* sehingga pengujian dilakukan untuk menentukan nilai akurasi hasil prediksi analisis sentimen berdasarkan hasil klasifikasi [3][4].

Masalah lain yang mungkin terjadi dalam melakukan sentimen analisis adalah ketidakseimbangan jumlah data antara sentimen positif dan negatif. Jika jumlah data positif jauh lebih

banyak daripada jumlah data negatif, maka hasil analisis dapat mengalami bias. Oleh karena itu, diperlukan teknik tambahan untuk mengatasi ketidakseimbangan data tersebut, seperti melakukan resampling pada data menggunakan teknik *Synthetic Minority Oversampling Technique (SMOTE)*.

Berdasarkan dari latar belakang masalah yang telah diuraikan dan pentingnya dilakukan penelitian lanjut yang berhubungan dengan analisis sentimen, maka peneliti mengajukan judul Analisis Sentimen Terhadap Marketplace Menggunakan Algoritma *Naïve Bayes* dengan studi kasus Lazada Indonesia dan dengan menerapkan Teknik *Synthetic Minority Oversampling Technique (SMOTE)*.

2. TINJAUAN PUSTAKA

Penelitian tentang analisis sentimen sebelumnya sudah pernah dilakukan, di antara penelitian-penelitian tersebut adalah:

Berdasarkan penelitian yang berjudul “Sentimen Analisis Customer Review Produk Shopee Indonesia Menggunakan Algoritma *Naïve Bayes Classifier*” yang terbit pada tahun 2021, jurnal tersebut menggunakan data sebanyak 3006 review salah satu produk smartphone dengan merek *Xiaomi Redmi Note 9*. Hasil klasifikasi menggunakan algoritma *Naïve Bayes* mendapati nilai akurasi sebesar 85% [3].

Berdasarkan penelitian yang berjudul “Analisis Sentimen Twitter Menggunakan *Ensemble Feature dan Metode Extreme Learning Machine (ELM)* (Studi Kasus: Samsung Indonesia)” yang terbit pada tahun 2018, jurnal tersebut menggunakan data twitter sebanyak 72 tweet dengan hasil klasifikasi menggunakan Metode *Extreme Learning Machine (ELM)* mendapatkan hasil tingkat akurasi hanya sebesar 42,857% [5].

Berdasarkan penelitian yang berjudul “Analisis Sentimen Review Produk Skincare Dengan *Naïve Bayes Classifier* Berbasis *Particle Swarm Optimization (PSO)*” yang terbit pada tahun 2022, jurnal tersebut menggunakan data sebanyak 800 data review dengan hasil klasifikasi menggunakan Metode *Naïve Bayes Classifier* dengan fitur *Particle Swarm Optimization*. Hasil penelitian menunjukkan peningkatan akurasi dari 77.96% menjadi 79.85% [5].

Berdasarkan penelitian yang berjudul “Analisis Sentimen Pada Ulasan Pembelian Produk di Marketplace Shopee Menggunakan Pendekatan *Natural Language Processing*” yang dilakukan pada tahun 2020, jurnal tersebut menggunakan 250 ulasan produk dengan menggunakan algoritma KNN. Hasil analisis pengujian menunjukkan tingkat akurasi 76,92%, presisi 80,00% dan *recall* 74,07%, sedangkan tanpa NLP hanya mendapat akurasi sebesar 69,23%, presisi 80,00% dan *recall* 64,52% [6].

Berdasarkan penelitian yang berjudul “*Sentiment Analysis of Twitter Data During Critical Events Through Bayesian Networks Classifiers*” yang terbit pada tahun 2020, jurnal tersebut menggunakan data twitter sebanyak 10222 opini masyarakat. Hasil

penelitian tersebut menghasilkan tingkat akurasi sebesar 78,1% dengan metode algoritma *naïve bayes* [7].

Kelima penelitian tersebut yang telah dilakukan oleh peneliti yang berbeda-beda membahas tentang analisis sentimen menggunakan algoritma *Naïve Bayes*, akan tetapi terdapat perbedaan dari sisi dataset yang digunakan dalam penelitian tersebut, serta terdapat perbedaan juga dari sisi hasil akurasi yang diperoleh dari masing-masing penelitian tersebut.

Sedangkan perbedaan dari penelitian-penelitian tersebut di atas dengan penelitian ini adalah dari sisi dataset yang digunakan, yaitu dalam penelitian ini menggunakan dataset tentang data review terhadap marketplace Lazada Indonesia <https://www.lazada.co.id/> melalui *twitter API*. Juga berbeda dalam hal penggunaan algoritma yang diterapkan yaitu algoritma *naïve bayes* dengan menerapkan *Synthetic Minority Oversampling Technique (SMOTE)* untuk mengatasi data imbalanced agar akurasi dari klasifikasi menjadi lebih meningkat. Ini disebabkan pada klasifikasi data imbalanced yang menghasilkan ketepatan akurasi yang tidak maksimal dan cenderung memprediksi ke kelas mayoritas, sehingga diperlukan upaya untuk menyeimbangkan proporsi kelas tersebut dengan cara melakukan resampling data minoritas dengan pendekatan *Synthetic Minority Oversampling Technique (SMOTE)*.

3. METODE PENELITIAN

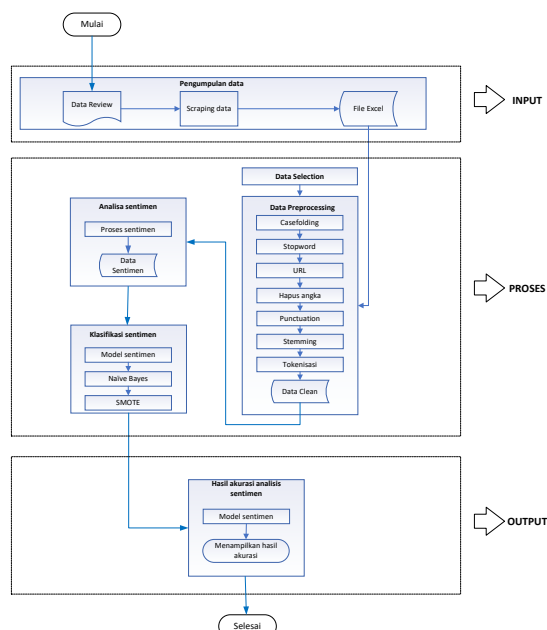
Jenis data yang digunakan dalam penelitian ini merupakan data terstruktur atau data teks. Data terstruktur adalah data kuantitatif yang teratur, umumnya data numerik atau berbasis teks. Sumber data merupakan data transaksional yang tercatat secara harian. Sumber data yang digunakan untuk penelitian ini adalah data analisis sentimen terhadap marketplace Lazada dengan metode *scraping* melalui *Twitter Application Program Interface (API)* sejumlah 1.517 dataset.

Penelitian ini didesain dengan melakukan adopsi terhadap tahapan dalam penelitian dengan menggunakan metode *Knowledge Discovery in Database (KDD)* yang merupakan tahapan proses dari sebuah kegiatan data mining [8]. Proses itu memuat beberapa iterasi terurut sebagai berikut:

1. *Data Selection*: Memilih kumpulan data target, menetapkan kumpulan data, atau memperhatikan pada subset variabel atau sampel data, di mana penemuan (*discovery*) akan dilakukan.
2. *Data Preprocessing/Cleansing*: Tahapan ini bermakna untuk menghilangkan noise dan data yang tidak konsisten.
3. *Data Integration*: Tahapan ini bermakna guna mengkombinasikan beberapa data.
4. *Data Transformation*: Tahapan ini bermakna guna mentransformasikan dan mengkonsolidasikan dataset menjadi sesuai dengan keperluan data mining.

5. *Data Mining*: Tahapan ini bermakna guna melakukan proses intelijen untuk mengekstrak pola data.
6. *Pattern Evaluation*: Tahapan ini berguna dalam mengidentifikasi pola yang merepresentasikan basis pengetahuan yang diperoleh.
7. *Knowledge Presentation*: Tahapan ini bermakna guna melakukan visualisasi dan representasi pengetahuan guna mendapatkan pengetahuan yang diperlukan pengguna.

Metode penelitian yang diperlihatkan dalam Gambar 1 terdiri dari beberapa langkah. Pertama-tama, data dikumpulkan melalui teknik scraping, yaitu teknik untuk mengambil data dari *twitter API*. Selanjutnya, data yang telah dikumpulkan akan diproses melalui tahap *pre-processing*, yaitu tahap di mana data diolah untuk memastikan data tersebut dapat dimasukkan ke dalam model algoritma *text mining* dengan baik. Setelah itu, data dimasukkan ke dalam model algoritma *text mining* untuk diproses lebih lanjut. Setelah data telah diproses, hasil yang diperoleh akan dievaluasi untuk menghasilkan laporan hasil pengujian.



Gambar 1. Metode Penelitian menggunakan Knowledge Discovery

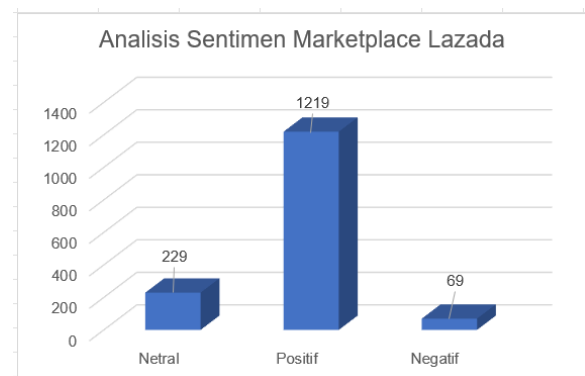
4. HASIL DAN PEMBAHASAN

Dataset analisis sentimen *marketplace* Lazada diproses dengan cara menggunakan metode *lexicon* menggunakan kamus bahasa Indonesia yang tersimpan dalam *file excel* untuk dilakukan klasifikasi. Selanjutnya setiap data teks pada analisis sentimen *marketplace* Lazada tersebut dihitung jumlah kata-kata yang termasuk kata positif dan dihitung jumlah kata-kata yang termasuk kata negatif berdasarkan yang termuat dalam kamus *lexicon* bahasa Indonesia yang sudah disimpan dalam *file excel*. Jumlah total kata-kata positif dibagi dengan seluruh kata dalam teks *review*

dikalikan 100% dinyatakan sebagai total positif. Jumlah kata-kata negatif dibagi dengan seluruh kata dalam teks *review* dikalikan 100% sebagai total negatif. Selanjutnya jika total positif dikurangi total negatif hasilnya lebih dari nol maka teks *review* diklasifikasikan sebagai klasifikasi sentimen positif. Hasil dari proses klasifikasi sentimen dengan menggunakan metode *lexicon* kamus Bahasa Indonesia terhadap dataset *review* analisis sentimen terhadap *marketplace* Lazada. Berikut perhitungan presentase klasifikasi sentimen *marketplace* Lazada:

Tabel 1. Perhitungan Presentase Klasifikasi Sentimen Marketplace Lazada

Netral	Positif	Negatif
(Jumlah review netral/ jumlah dataset) * 100%	(Jumlah review positif/ jumlah dataset) * 100%	(Jumlah review negatif/ jumlah dataset) * 100%
= (229/1.517) * 100%	= (1.219/1.517) * 100%	= (69/1.517) * 100%
= 15,096%	= 80,356%	= 4,548%



Gambar 2. Grafik Nilai Klasifikasi Sentimen Lazada

Berdasarkan temuan data penelitian tersebut dapat disampaikan bahwa opini/*review* yang disampaikan terkait analisis sentimen *marketplace* Lazada mendapatkan klasifikasi sentimen positif sebesar 80,356% yang berarti respon yang sangat baik dari masyarakat. Sentimen positif tersebut sangat dipengaruhi oleh 13 kata yaitu kata “banget”, “belanja”, “bonus”, “dadakan”, “gratis”, “habis”, “habisan”, “hemat”, “kejar”, “lazada”, “ongkir”, “promo”, “voucher” yang frekuensi kemunculannya dalam setiap *review* yang disampaikan oleh masyarakat sehingga memiliki kontribusi yang sangat besar terhadap sentimen positif dengan total jumlah sebanyak 9.945 kata. Kata-kata yang sering muncul dalam penelitian, hasil klasifikasi sentimen positif terhadap *marketplace* Lazada divisualisasikan dalam bentuk *wordcloud* sesuai dengan gambar 3.



Gambar 3. Wordcloud Kata-kata yang sering muncul

Pada gambar *wordcloud* tersebut besar kecilnya kata-kata tersebut menunjukkan frekuensi kemunculan kata tersebut. Semakin besar dan tebal kata-kata tersebut mengartikan bahwa kata tersebut memiliki frekuensi kemunculan kata yang tinggi. Semakin kecil frekuensi kemunculan kata tersebut ditunjukkan dengan kata-kata yang semakin kecil dan tipis. Temuan penelitian dari kata-kata yang sering muncul pada klasifikasi sentimen positif dalam *marketplace* Lazada tersebut termuat dalam tabel 2.

Tabel 2. Kata-kata Yang Sering Muncul dalam sentimen positif

No	Kata	Frekuensi (Positif)
1	banget	298
2	belanja	32
3	bonus	1510
4	dadakan	620
5	gratis	597
6	habis	923
7	habisan	894
8	hemat	514
9	kejar	344
10	lazada	1804
11	ongkir	609
12	promo	762
13	voucher	249

Berdasarkan tabel tersebut ditemukan bahwa kata-kata yang sering muncul dalam klasifikasi sentimen positif yang disampaikan pada penelitian ini terkait analisis sentimen tentang *marketplace* Lazada posisi pertama adalah kata “lazada” sebanyak 1.804, posisi kedua kata “bonus” sebanyak 1.510, posisi ketiga kata “habis” sebanyak 923, posisi keempat kata “habisan” sebanyak 894, posisi kelima kata “promo” sebanyak 762, posisi keenam kata “dadakan” sebanyak 620, posisi ketujuh kata “ongkir” sebanyak 609, posisi kedelapan kata “gratis” sebanyak 597, posisi kesembilan kata “hemat” sebanyak 514, posisi kesepuluh kata “kejar” sebanyak 344, posisi kesebelas kata “belanja” sebanyak 321, posisi keduabelas kata “banget” sebanyak 298, posisi ketigabelas kata “voucher” sebanyak 249. Berdasarkan data statistik dari ke-13 kata-kata yang sering muncul tersebut, kata

yang memiliki frekuensi tertinggi adalah kata “lazada” sebanyak 1.804. Sedangkan kata yang sering muncul dengan frekuensi kemunculan terendah adalah kata “voucher” sebanyak 249. Sedangkan klasifikasi sentimen negatif sebesar 4,548% yang berarti respon masyarakat yang negatif terhadap *marketplace* Lazada. Sentimen negatif tersebut sangat dipengaruhi oleh 11 kata yaitu kata “banget”, “belanja”, “bonus”, “barang”, “habis”, “habisan”, “hemat”, “kejar”, “lazada”, “ongkir”, “promo”, “voucher” seperti pada tabel 4.2 yang frekuensi kemunculannya dalam setiap review yang disampaikan oleh masyarakat sehingga memiliki kontribusi yang sangat besar terhadap sentimen negatif dengan total jumlah sebanyak 336 kata.

Berdasarkan pengujian penerapan algoritma *Naïve Bayes* dan algoritma *Naïve Bayes* dengan Teknik *SMOTE*, menggunakan *Cross Validation* dengan *nFold* sebesar 10 serta dengan teknik *stratified sampling*, yaitu sampling bertingkat membangun himpunan bagian acak. Ini memastikan bahwa distribusi kelas (didefinisikan oleh atribut label atau target) di subset dataset sama dengan di seluruh dataset. Misalnya dalam kasus klasifikasi binomial, *stratified sampling* membangun subset acak sehingga setiap subset berisi kira-kira proporsi yang sama dari dua nilai label tersebut. Berdasarkan eksperimen yang dilakukan dalam penelitian ini diperoleh *confusion matrix* untuk penerapan algoritma *Naïve Bayes* dengan Teknik *SMOTE* seperti terlihat pada tabel 3.

Tabel 3. Perhitungan Presentase Klasifikasi Sentimen *Marketplace* Lazada

Accuracy: 81,89%				
	true Netral	true Positif	true Negative	Class Precision
Prediksi Netral	82	202	0	28,87%
Prediksi positif	97	883	0	90,10%
Prediksi Negatif	50	134	1219	86,89%
Class Recall	35,81 %	72,44 %	100%	

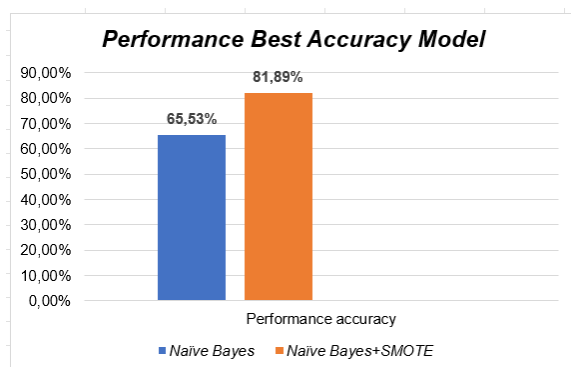
Berdasarkan *confusion matrix* tersebut dapat dijelaskan bahwa jumlah prediksi sentimen netral dan sentimen itu benar netral adalah berjumlah 82, sedangkan prediksi sentimen netral dan sentimen itu ternyata bukan netral akan tetapi sentimen itu ternyata positif berjumlah 202. Jumlah prediksi sentimen positif dan sentimen itu benar positif adalah berjumlah 883, sedangkan prediksi sentimen positif dan sentimen itu ternyata bukan positif akan tetapi sentimen itu ternyata netral berjumlah 97. Jumlah prediksi sentimen negatif dan sentimen itu benar negatif adalah berjumlah 1.219, sedangkan prediksi sentimen negatif dan sentimen itu ternyata bukan negatif akan tetapi sentimen itu ternyata netral berjumlah 50 dan sentimen positif berjumlah 134. Sehingga total jumlah prediksi sentimen netral, positif, negatif dan sentimen netral,

positif, negatif itu benar berjumlah 2.184 (82+883+1.219), sedangkan jumlah prediksi sentimen netral, positif, negatif akan tetapi sentimen tersebut bukan netral, positif dan negatif berjumlah 483 (202+97+50+134). Sedangkan eksperimen yang dilakukan dalam penelitian ini menggunakan algoritma *Naïve Bayes* tanpa menggunakan *SMOTE* diperoleh *confusion matrix* seperti terlihat pada Tabel 4.

Tabel 4. *Confusion Matrix* dari Algoritma *Naïve Bayes*

Accuracy: 65,53%				
	<i>true Netral</i>	<i>true Positif</i>	<i>true Negative</i>	<i>Class Precision</i>
Prediksi Netral	79	197	18	26,87%
Prediksi positif	100	890	26	87,60%
Prediksi Negatif	50	132	25	12,08%
<i>Class Recall</i>	34,50 %	73,01 %	36,23%	

Berdasarkan *confusion matrix* tersebut dapat dijelaskan bahwa jumlah prediksi sentimen netral dan sentimen itu benar netral adalah berjumlah 79, sedangkan prediksi sentimen netral dan sentimen itu ternyata positif berjumlah 197 dan sentimen negatif berjumlah 18. Jumlah prediksi sentimen positif dan sentimen itu benar positif adalah berjumlah 890, sedangkan prediksi sentimen positif dan sentimen itu ternyata bukan positif akan tetapi sentimen itu ternyata netral berjumlah 100 dan sentiment negative berjumlah 26. Jumlah prediksi sentimen negatif dan sentimen itu benar negatif adalah berjumlah 25, sedangkan prediksi sentimen negatif dan sentimen itu ternyata bukan negatif akan tetapi sentimen itu ternyata netral berjumlah 50 dan sentimen positif berjumlah 132. Sehingga total jumlah prediksi sentiment netral, positif, negatif dan sentimen netral, positif, negatif itu benar berjumlah 994 (79+890+25), sedangkan jumlah prediksi sentimen netral, positif, negatif akan tetapi sentimen tersebut bukan netral, positif dan negatif berjumlah 523 (197+18+26+100+50+132).



Gambar 4. Grafik Perbandingan Hasil Akurasi

Temuan hasil eksperimen dari hasil penelitian ini berdasarkan perbandingan nilai akurasi yang diperoleh melalui perhitungan *confusion matrix* tersebut di atas selanjutnya dapat divisualisasikan dalam bentuk tabel perbandingan akurasi seperti pada tabel 5 dan divisualisasikan dalam bentuk grafik seperti pada gambar 4.

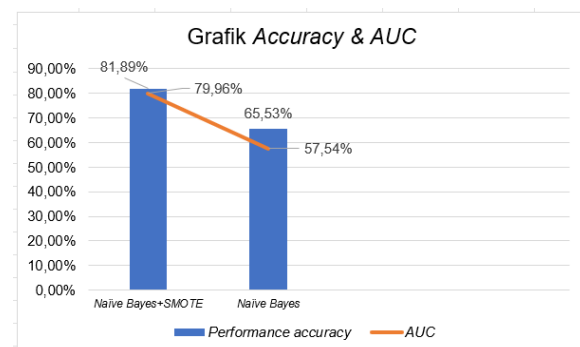
Tabel 5. Tabel Perbandingan Akurasi

No	Model	Performance Accuracy
1	<i>Naïve Bayes</i>	65,53%
2	<i>Naïve Bayes+SMOTE</i>	81,89%

Dalam penelitian ini juga dilakukan perhitungan nilai *AUC* berdasarkan *confusion matrix* yang diperoleh dari penerapan algoritma *Naïve Bayes* dan perhitungan nilai *AUC* berdasarkan *confusion matrix* yang diperoleh dari penerapan algoritma *Naïve Bayes* dengan Teknik *SMOTE*. Temuan hasil penelitian seperti ditunjukkan pada tabel 6 dan gambar 5 menunjukkan bahwa nilai *AUC* dari penerapan algoritma *Naïve Bayes* dengan Teknik *SMOTE* (79,96%) hasilnya lebih baik dari pada nilai *AUC* dari penerapan algoritma *Naïve Bayes* saja. Temuan ini menunjukkan bahwa penerapan Teknik *SMOTE* dalam model tersebut menghasilkan klasifikasi analisis sentimen sebagai *Fair Classification*.

Tabel 6. Tabel Perbandingan Akurasi + *AUC*

No	Model	Performance Accuracy	<i>AUC</i>
1	<i>Naïve Bayes+SMOTE</i>	81,89%	79,96%
2	<i>Naïve Bayes</i>	65,53%	57,54%

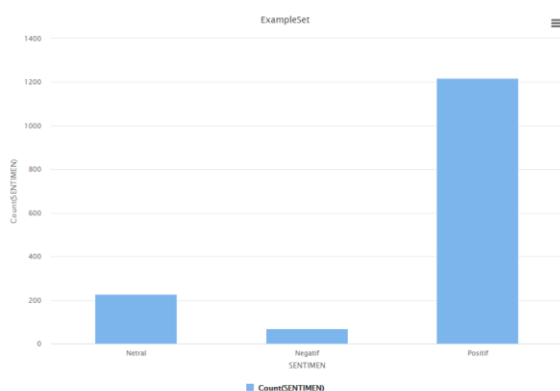


Gambar 5. Grafik Perbandingan Hasil Akurasi dan *AUC*

Penelitian yang dilakukan ini dibandingkan dengan penelitian terkait merupakan penelitian tentang analisis sentimen menggunakan algoritma *Naïve Bayes*, namun penelitian ini memiliki perbedaan dengan penelitian tersebut dalam hal dataset yang dipergunakan yaitu dataset analisis sentiment tentang *marketplace* Lazada yang diperoleh dengan Teknik *scraping* menggunakan *API*, metode yang dipergunakan adalah *KDD*.

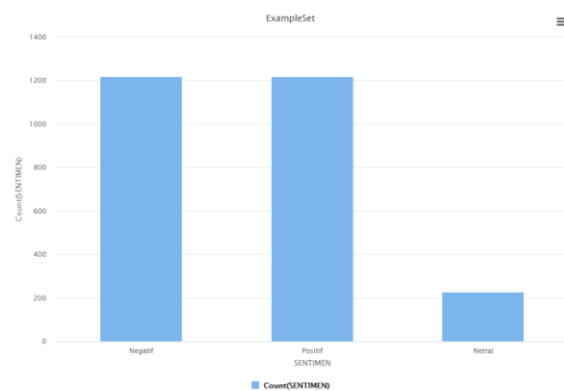
Hasil penelitian dengan menggunakan algoritma *Naïve Bayes* ini menunjukkan bahwa informasi tentang

analisis sentimen dilakukan sebanyak 1.517 dataset diperoleh hasil *review* prediksi sentimen netral dan sentimen itu benar netral adalah berjumlah 79, prediksi sentimen netral dan sentimen itu ternyata positif berjumlah 197 dan sentimen negatif berjumlah 18, prediksi sentimen positif dan sentimen itu benar positif adalah berjumlah 890, prediksi sentimen positif akan tetapi sentimen itu ternyata netral berjumlah 100 dan sentimen negatif berjumlah 26, prediksi sentimen negatif dan sentimen itu benar negatif adalah berjumlah 25, prediksi sentimen negatif dan sentimen itu ternyata bukan negatif akan tetapi sentimen tersebut ternyata netral berjumlah 50 dan sentimen positif berjumlah 132. Berdasarkan hasil tersebut, terjadi *over fitting* atau ketigakeseimbangan hasil pengujian analisis sentimen terhadap *marketplace* Lazada, di mana sentimen positif berjumlah 1.219, sedangkan sentimen negatif berjumlah 69, dan sentimen netral berjumlah 229. Rekomendasi yang dapat diberikan untuk mengatasi permasalahan tersebut adalah dengan menerapkan teknik *Synthetic Minority Over Sampling (SMOTE)* untuk digunakan sebagai metode yang efektif dalam mengatasi *over fitting*. Grafik Sentimen tanpa menggunakan *SMOTE* dapat dilihat pada gambar 6 berikut:



Gambar 6. Grafik Sentimen Tanpa SMOTE

Setelah menerapkan teknik *SMOTE*, hasil penelitian dengan menggunakan algoritma *Naïve Bayes* dengan *SMOTE* ini menunjukkan bahwa informasi tentang analisis sentimen dilakukan sebanyak 1.517 dataset diperoleh hasil *review* prediksi sentimen netral dan sentimen itu benar netral adalah berjumlah 82, prediksi sentimen netral dan sentimen itu ternyata bukan netral akan tetapi sentimen itu ternyata positif berjumlah 202, prediksi sentimen positif dan sentimen itu benar positif adalah berjumlah 883, prediksi sentimen positif dan sentimen itu ternyata bukan positif akan tetapi sentimen itu ternyata netral berjumlah 97, prediksi sentimen negatif dan sentimen itu benar negatif adalah berjumlah 1.219, sedangkan prediksi sentimen negatif dan sentimen itu ternyata bukan negatif akan tetapi sentimen itu ternyata netral berjumlah 50 dan sentimen positif berjumlah 134. Grafik Sentimen dengan menggunakan *SMOTE* dapat dilihat pada gambar 7 berikut:



Gambar 7. Grafik Sentimen Menggunakan SMOTE

5. KESIMPULAN DAN SARAN

Penelitian menunjukkan bahwa model analisis sentimen menggunakan algoritma *Naïve Bayes* dengan *SMOTE* memiliki akurasi 81,89%, lebih tinggi dibandingkan tanpa *SMOTE* yang hanya 65,53%. Penggunaan *SMOTE* menghasilkan dataset seimbang dengan 1.219 review positif, 1.219 review negatif, dan 229 review netral. Hasil analisis memperlihatkan 80,356% review bersentimen positif terhadap Lazada, yang dipengaruhi oleh kata-kata tertentu seperti "banget", "belanja", dan "promo". Namun, masih terdapat prediksi yang melenceng, dengan beberapa sentimen positif dan negatif yang sebenarnya netral. Nilai AUC juga lebih tinggi pada model dengan *SMOTE* (79,96%) dibandingkan tanpa *SMOTE* (57,54%), menunjukkan model dengan *SMOTE* lebih baik. Rekomendasi untuk Lazada adalah untuk terus memperhatikan sentimen pelanggan untuk meningkatkan layanan dan reputasi, serta melakukan evaluasi berkelanjutan pada model klasifikasi dan mempertimbangkan pendekatan *deep learning* untuk dataset yang lebih besar.

DAFTAR PUSTAKA

- [1] dan N. A. Arhadian Roliansyah, Amelia Putri Utami, Asnaini, "Analisis Transaksi Jual Beli Online di Marketplace Shopee Menurut Konsep Bisnis Islam Pada Masa Pandemi Covid-19," *Jurnal Riset Ekonomi*, vol. 1, no. 5, pp. 443–456, 2022, [Online]. Available: <https://www.bajangjournal.com/index.php/Juremi/article/view/1675/1173>
- [2] F. Alvi Herzegovino Sianipar and E. Dra Hj Yoestini, "ANALISIS PENGARUH CUSTOMER REVIEW DAN CUSTOMER RATING TERHADAP KEPUTUSAN PEMBELIAN PRODUK DI ONLINE MARKETPLACE (Studi Pada Mahasiswa Pengguna Tokopedia di Kota Semarang)," *Diponegoro Journal of Management*, vol. 10, no. 4, p. 1, 2021, [Online]. Available: <http://ejournal-s1.undip.ac.id/index.php/dbr>
- [3] L. O. Sihombing, H. Hannie, and B. A. Dermawan, "Sentimen Analisis Customer Review Produk Shopee Indonesia Menggunakan

- Algortima Naïve Bayes Classifier,” *Edumatic: Jurnal Pendidikan Informatika*, vol. 5, no. 2, pp. 233–242, 2021, doi: 10.29408/edumatic.v5i2.4089.
- [4] A. Rausanfita, P. P. Adikara, and S. Adinugroho, “Analisis Sentimen Twitter Menggunakan Ensemble Feature dan Metode Extreme Learning Machine (ELM) (Studi Kasus: Samsung Indonesia),” *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 2, no. 12, pp. 6409–6417, 2018, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [5] T. Astuti and Y. Astuti, “Analisis Sentimen Review Produk Skincare Dengan Naïve Bayes Classifier Berbasis Particle Swarm Optimization (PSO),” vol. 6, pp. 1806–1815, 2022, doi: 10.30865/mib.v6i4.4119.
- [6] E. H. Muktafin, K. Kusrini, and E. T. Luthfi, “Analisis Sentimen pada Ulasan Pembelian Produk di Marketplace Shopee Menggunakan Pendekatan Natural Language Processing,” *Jurnal Eksplora Informatika*, vol. 10, no. 1, pp. 32–42, 2020, doi: 10.30864/eksplora.v10i1.390.
- [7] G. A. Ruz, P. A. Henríquez, and A. Mascareño, “Sentiment analysis of Twitter data during critical events through Bayesian networks classifiers,” *Future Generation Computer Systems*, vol. 106, pp. 92–104, 2020, doi: 10.1016/j.future.2020.01.005.
- [8] Alfiqua and F. Y. Alfizi, “PENERAPAN MARKET BASKET ANALYSIS MENGGUNAKAN PROSES KDD (KNOWLEDGE DISCOVERY IN DATABASE) SEBAGAI STRATEGI PENJUALAN PRODUK SWALAYAN (STUDI KASUS : SWALAYAN X) Alfiqua,” pp. 509–516, 2018.