

Klasifikasi Emosional Ulasan Pelanggan dengan Pendekatan NLP menggunakan Metode Ensemble dan ROS

Emotional Classification of Customer Reviews with NLP Approach using Ensemble and ROS Methods

Adisaputra Zidha Noorizki¹, Heri Pratikno², Weny Indah Kusumawati^{3*}

^{1,2,3}Fakultas Teknologi dan Informatika, Universitas Dinamika

E-mail: ¹20410200018@dinamika.ac.id, ²heri@dinamika.ac.id, ^{3*}weny@dinamika.ac.id

**penulis korespondensi*

Abstrak

Konsep Orientasi Pelanggan sangat penting bagi perusahaan untuk berkembang di era saat ini, dengan memanfaatkan teknologi untuk mendapatkan wawasan yang mendalam tentang perilaku pelanggan mereka. Salah satu alat teknologi tersebut adalah pembelajaran mesin, khususnya yang menggunakan pendekatan pemrosesan bahasa alami (NLP). Penelitian ini menggunakan lima algoritma yang berbeda dan menggabungkan berbagai metode untuk meningkatkan kinerja model *machine learning*. Melalui penerapan teknik-teknik seperti *random over-sampling* (ROS) dan *ensemble learning*, akurasi prediksi keseluruhan untuk kelas minoritas meningkat secara signifikan. Model *ensemble* yang diintegrasikan dengan ROS mencapai akurasi 0,90 dan *mean square error* 0,91, mengungguli algoritma lain yang diuji dalam penelitian ini. Pendekatan yang dioptimalkan ini tidak hanya menunjukkan keefektifan pemanfaatan teknologi untuk sebuah perusahaan dapat menerapkan strategi yang berpusat pada pelanggan, tetapi juga menyoroti pentingnya peningkatan metodologi dalam pemodelan prediktif untuk keberlanjutan bisnis.

Kata kunci: Klasifikasi Emosi, Pembelajaran Mesin, Pemrosesan Bahasa Alami, *Hard Voting*, *Random Over Sampling*.

Abstract

Customer Orientation concept is crucial for companies to thrive in the current era, leveraging technology to gain deep insights into their customers' behaviors. One such technological tool is machine learning, particularly employing natural language processing (NLP) approaches. This research employs five distinct algorithms and incorporates various methods to enhance the performance of machine learning models. Through the implementation of techniques such as random over-sampling (ROS) and ensemble learning, the overall predictive accuracy for minority classes is significantly improved. The ensemble model, integrated with ROS, achieves an accuracy of 0.90 and a mean square error of 0.91, outperforming other algorithms tested in this study. This optimized approach not only showcases the effectiveness of leveraging technology for customer-centric strategies but also highlights the importance of methodological enhancements in predictive modeling for business sustainability.

Keywords: Emotional Classification, Ensemble Learning, Natural Language Processing, Hard Voting, Random Over Sampling.

1. PENDAHULUAN

Kepuasan pelanggan merupakan elemen kunci dalam kesuksesan sebuah bisnis atau perusahaan, terutama di era Industri 4.0 [1]. Dalam konteks ini, perusahaan dituntut untuk memiliki pemahaman yang mendalam mengenai kebutuhan dan preferensi target pasar atau pelanggan perusahaan. Penggunaan ulasan pelanggan menjadi sangat penting sebagai sumber umpan balik langsung yang membantu perusahaan mengukur tingkat kepuasan dan mengevaluasi

kualitas produk atau jasa yang ditawarkan perusahaan. Konsep *Customer Orientation* merupakan fondasi bagi perusahaan yang ingin sukses di era Industri 4.0. Hal ini melibatkan perubahan paradigma dari fokus pada efisiensi internal, menjadi memahami dan melayani pelanggan dengan lebih baik [2]. Perusahaan yang menerapkan pendekatan ini akan lebih responsif terhadap kebutuhan pelanggan, memastikan bahwa produk dan layanan yang dihasilkan sesuai dengan harapan pelanggan, dan membangun hubungan jangka panjang yang kuat dengan setiap pelanggannya. Pada Industri 4.0, banyak *startup* atau perusahaan yang telah menerapkan konsep *Customer Orientation* dalam pengembangan produknya [3]. Perusahaan telah memanfaatkan keberadaan teknologi terbaru yaitu *Big Data* dan analisis datanya untuk mendapatkan wawasan yang mendalam mengenai perilaku pelanggan. Dengan demikian, sebuah perusahaan dapat menciptakan produk yang lebih relevan dan memenuhi ekspektasi pelanggan, namun juga tidak lupa untuk memperhatikan efisiensi internal dalam perusahaan.

Setiap individu yang memiliki peran sebagai pelanggan suatu produk, pasti memiliki emosi yang terlibat dalam setiap pengalaman transaksi atau interaksi dengan suatu produk atau jasa. Emosi atau *emotion* merupakan sesuatu yang merujuk pada keadaan perasaan seseorang yang muncul sebagai respons terhadap suatu rangsangan atau pengalaman tertentu [4]. Emosi melibatkan reaksi psikologis dan fisik, dan sering kali dapat tercermin dalam perubahan suasana hati, ekspresi wajah, tingkat energi, atau bahkan respons fisik seperti peningkatan detak jantung. Emosi dapat mencakup berbagai macam perasaan, seperti kebahagiaan, kesedihan, kemarahan, kekecewaan, ketakutan, dan sebagainya. Setiap individu memiliki rentang emosi yang unik dan cara mereka bereaksi terhadap situasi tertentu juga dapat bervariasi [5]. Sementara itu, emosi pelanggan merupakan aspek yang sangat relevan dan penting dalam mengevaluasi kepuasan pelanggan. Emosi dapat mempengaruhi keputusan pembelian, loyalitas pelanggan, dan citra produk/jasa suatu perusahaan [6]. Oleh karena itu, menganalisis emosi melalui ulasan pelanggan merupakan kunci untuk memahami sejauh mana produk atau jasa perusahaan memenuhi atau bahkan melebihi ekspektasi pelanggan dari sisi emosional. Dengan memahami dan memetakan emosi yang terkandung dalam setiap ulasan pelanggan, perusahaan dapat lebih efektif dalam menanggapi kebutuhan pelanggan, meningkatkan kualitas produk atau layanan, dan membangun hubungan jangka panjang yang kuat dengan pelanggannya.

Salah satu pendekatan yang sangat relevan dan dapat diterapkan dalam menghadapi perkembangan teknologi yang pesat adalah analisis bahasa alami, yang lebih dikenal sebagai *Natural Language Processing* (NLP). Teknik ini merupakan bagian dari bidang ilmu kecerdasan buatan (*Artificial Intelligence*), terutama yang berfokus pada pembelajaran mesin (*machine Learning*). NLP telah terbukti efektif dalam mengidentifikasi kata kunci, frasa, dan sentimen dalam kalimat [7]. Sebagai contoh, dalam klasifikasi sentimen, dataset yang berasal dari pesan atau data yang dikirim oleh pengguna melalui media sosial populer seperti Twitter sering digunakan. Menurut penelitian [8], NLP dapat dioptimalkan untuk melakukan analisis sentimen terkait vaksin yang telah mendapatkan sertifikasi dari WHO, seperti Pfizer, Moderna, dan AstraZeneca, dengan menggunakan data dari Twitter. Hasil penelitian tersebut menyimpulkan bahwa sentimen positif tertinggi ditemukan pada vaksin Pfizer, diikuti oleh Moderna dengan persentase masing-masing 47,30% dan 46,20%. Sebaliknya, vaksin AstraZeneca menerima peringkat sentimen terendah sebesar 40,09%. Melalui pengetahuan ini, perusahaan dapat lebih memahami apakah pelanggan merasa puas, terkejut, kecewa, atau bahkan hanya netral terhadap produk atau layanan yang mereka tawarkan.

Di aspek lain, berbagai faktor dapat memengaruhi kinerja model dalam ruang lingkup pembelajaran mesin. Mulai dari ketidakseimbangan jumlah dataset dalam kelas atau kategori tertentu hingga pemilihan model yang tepat. Ketidakseimbangan jumlah data dalam dataset dapat mengakibatkan model yang dibangun cenderung lebih memihak pada salah satu kelas dengan jumlah data yang dominan. Oleh karena itu, diperlukan proses atau metode untuk menangani kondisi ini. Teknik *oversampling* adalah teknik yang bertujuan untuk menstabilkan ukuran atau jumlah data untuk setiap kelas dengan menduplikasi data sesuai dengan yang telah ditentukan. Teknik *oversampling* terdiri dari beberapa metode populer seperti *Random Oversampling* (ROS), *Adaptive Synthetic Sampling* (ADASYN), *Synthetic Minority Oversampling Technique*

(SMOTE), kombinasi SMOTE dan *Edited Nearest Neighbors* (SMOTE-ENN), SMOTE-Tomek, dan *Random Over Sampling Examples* (ROSE) [9]. Menurut penelitian [10], penggunaan teknik *oversampling* untuk menduplikasi data yang jumlahnya minoritas agar memiliki jumlah yang sama dengan kelas mayoritas, dapat menghasilkan model dengan kinerja prediksi yang lebih baik untuk kelas minoritas. Penelitian lainnya [11] mengungkapkan bahwa kombinasi dua metode, yaitu pendekatan *random over-sampling* (ROS) dan *neighborhood cleaning rule* (NCL) untuk mengatasi masalah ketidakseimbangan data, diuji pada tiga model sekaligus dengan hasil mencapai tingkat akurasi yang tinggi pada ketiga model klasifikasi, yaitu 97,71%, 98,01%, dan 98,11%.

Selain teknik *oversampling*, terdapat juga metode yang dikenal sebagai *ensemble learning* dalam pembelajaran mesin. Metode ini menjadi solusi ketika kinerja model yang dilatih secara individual tidak mencapai tingkat yang diinginkan. *Ensemble learning* memberikan alternatif untuk meningkatkan kinerja model dengan menggabungkan beberapa model untuk menghasilkan keputusan [12]. Sebagaimana telah dibuktikan dalam penelitian [13], hasil terbaik diperoleh oleh model yang telah diterapkan ensemble learning dengan menggabungkan beberapa model, yaitu *Bayesian Network*, *Random Forest*, *CART*, dan *Neural Network*.

Berdasarkan semua pertimbangan di atas, tujuan utama dari penelitian ini adalah menerapkan metode ROS dalam teknik *oversampling* untuk mengatasi ketidakseimbangan kelas dan melakukan proses klasifikasi emosi berdasarkan ulasan pelanggan melalui perbandingan berbagai *base learner*, termasuk *Multinomial Naïve Bayes*, *Support Vector Machine*, *Logistic Regression*, *K-Nearest Neighbors*, dan *Gradient Boosting*, serta menerapkan metode hard voting ensemble menggunakan pendekatan *Natural Language Processing* (NLP). Selain itu, penelitian ini juga mengevaluasi perbandingan hasil kinerja dari setiap model yang telah dilatih, menggunakan *confusion metrics*. Penelitian ini menawarkan pendekatan baru dalam mengoptimalkan klasifikasi emosi ulasan pelanggan dengan memanfaatkan metode ROS dalam *oversampling* serta kombinasi teknik *hard voting ensemble*, yang belum banyak diaplikasikan pada konteks serupa. Kontribusi penelitian ini terletak pada penerapan dan evaluasi metode *ensemble learning* berbasis ROS untuk meningkatkan akurasi model klasifikasi emosi pelanggan dalam dataset yang tidak seimbang, yang berbeda dari penelitian-penelitian terdahulu yang lebih umum menggunakan pendekatan tunggal. Dengan pendekatan ini, penelitian diharapkan dapat memperkaya analisis sentimen dalam *customer feedback* dan memberikan solusi yang lebih efektif untuk mengatasi keterbatasan pada dataset yang tidak seimbang.

2. METODE PENELITIAN

Penelitian ini terdiri dari serangkaian tahapan terstruktur yang dirancang dengan cermat untuk memfasilitasi pencapaian hasil yang diinginkan. Setiap tahap dalam kerangka kerja metodis ini memiliki tujuan yang berbeda, yang berkontribusi pada pemahaman dan analisis komprehensif dari pokok bahasan yang diteliti.

2.1 Pengumpulan Data

Dataset yang akan digunakan berasal dari dataset yang dapat diakses secara publik di halaman Kaggle dengan judul “Emotions dataset for NLP + Neutral Emotion” oleh Thuan Naheem Pakeer dan empat kolaborator lainnya. Dataset ini terdiri dari 26.800 baris data yang dikategorikan ke dalam tujuh label, yaitu ‘neutral’, ‘joy’, ‘sadness’, ‘anger’, ‘fear’, ‘love’, dan ‘surprise’. Di mana setiap kategori/label memiliki jumlah data yang berbeda.

2.2 Pra-pemrosesan Data

Dataset yang tersedia di Kaggle telah dibagi menjadi dua bagian, yaitu data pelatihan dan data pengujian. Sebelum melanjutkan ke proses utama pada tahap pra-pemrosesan data, peneliti akan menggabungkan kedua bagian data menjadi satu dokumen terlebih dahulu untuk melakukan perlakuan khusus dalam mengoptimalkan data pada dataset menggunakan pendekatan Pemrosesan Bahasa Alami (NLP).

Tabel 1 Jumlah data untuk setiap dokumen

Nama dokumen	Jumlah data untuk setiap dokumen	
	Sebelum	Sesudah
dataTrain.csv	21400	0
dataTest.csv	5400	0
emotionDataset.csv	0	26800

2.2.1 Pembersihan Data

Pada tahap pembersihan data, setiap baris data akan melalui serangkaian proses untuk memastikan konsistensi dan kebersihan informasi. Langkah-langkah yang dilakukan antara lain mengubah semua teks menjadi huruf kecil, menghilangkan kata-kata yang mengandung simbol '@', menghilangkan tanda baca, dan menghilangkan angka. Selain itu, langkah-langkah tambahan seperti penghapusan karakter tunggal, penghapusan spasi ganda yang berlebihan, dan penghapusan hyperlink atau URL dari setiap kalimat dalam dataset juga dilakukan.

2.2.2 Tokenisasi

Tokenisasi merupakan salah satu tahapan yang harus dilakukan jika sebuah penelitian menggunakan pendekatan *Natural Language Processing*. Tokenisasi merupakan langkah yang diperlukan untuk memecah teks menjadi unit-unit yang lebih kecil, yang kemudian disebut token. Token dapat berupa kata, kalimat, atau bahkan karakter, tergantung dari tingkat tokenisasi yang diterapkan [14]. Implementasi tokenisasi sangat penting karena memungkinkan pemrosesan teks lebih lanjut dengan memperlakukan setiap token sebagai unit yang terpisah. Dengan memecah teks menjadi bagian-bagian yang lebih kecil, analisis NLP dapat lebih efektif dalam mengidentifikasi pola linguistik, arti kata, atau struktur kalimat.

2.2.3 Stopwords

Stopwords merupakan tahap penting dalam pendekatan *Natural Language Processing* (NLP) yang bertujuan untuk menghilangkan berbagai kata umum yang sering diabaikan karena kurang memberikan informasi yang spesifik mengenai konten teks dan cenderung muncul secara umum dalam banyak kalimat atau dokumen [15]. Proses *stopword* ini bertujuan untuk meningkatkan kualitas analisis dengan berfokus pada kata-kata yang lebih informatif. Dalam teks bahasa Inggris, beberapa contoh kata yang sering dihilangkan selama proses *stopword* antara lain 'the', 'and', 'is', 'in', dan masih banyak lagi.

2.2.4 Lemmatization

Lemmatization adalah tahap yang melibatkan perubahan kata menjadi bentuk dasarnya atau kata benda menjadi kata kerja dasarnya. Tujuannya adalah untuk mengurangi variasi kata yang memiliki akar kata yang sama menjadi bentuk yang lebih standar dan seragam. Sebagai contoh, 'running' menjadi 'run', 'years' menjadi 'year', 'believes' menjadi 'believe'. *Lemmatization* juga bertujuan untuk meningkatkan pemahaman komputer terhadap teks yang diproses dengan mengurangi variasi kata yang sama ke dalam bentuk yang lebih standar [16].

Tabel 2 Kondisi data dalam prapemrosesan

Sebelum prapemrosesan	Sesudah prapemrosesan
['@qclindalou Go log in to your...', 'neutral']	[[feel, smart, figure, thing, ...]]
['Im grabbing a minute to post i feel...', 'anger']	[[graduate, two, day, feel, ...]]
['...Balisage Markup Conference 2009 Program online at http://tr.im/mL6Z...', 'neutral']	[[..., want, write, one, vain, hobby, makeup, ...]]

2.3 Pembagian Dataset

Setelah dataset melalui berbagai tahap dalam proses pra-pemrosesan, dataset tersebut telah mencapai tahap kesiapan yang memungkinkan untuk dibagi menjadi beberapa subset. Proses ini merupakan prasyarat penting sebelum melanjutkan ke tahap pembobotan, *random oversampling*, dan pelatihan oleh berbagai model yang telah dipilih.

Tabel 3 Kondisi data dalam proses pemisahan

Subset data	Jumlah data
Training data	21440
Testing data	5360

2.4 Pembobotan TF-IDF

Proses pembobotan pada dasarnya bertujuan untuk memberikan nilai bobot pada setiap kata dalam sebuah dokumen atau teks, yang menggambarkan tingkat kepentingan kata tersebut dalam konteks keseluruhan korpus atau koleksi dokumen. Dengan melakukan pembobotan, sistem dapat mengidentifikasi kata kunci yang memiliki kontribusi signifikan dalam representasi dan pemahaman suatu dokumen [17]. Penelitian ini menggunakan pembobotan TF/IDF (*Term Frequency-Inverse Document Frequency*), dimana dataset yang telah dibagi sebelumnya akan dilakukan pembobotan untuk mencari nilai TF (*Term Frequency*) dan nilai IDF (*Inverse Document Frequency*). Secara matematis, pembobotan TF/IDF dapat dituliskan:

$$TF\ IDF = \frac{f_d(t)}{\max f_t(d)} \times \log\left(\frac{n}{df_t}\right) \quad (1)$$

2.5 Implementasi Random Oversampling (ROS)

Random Oversampling adalah teknik yang digunakan untuk menangani ketidakseimbangan kelas dalam dataset, terutama yang umum terjadi dalam masalah klasifikasi. Ketidakseimbangan kelas terjadi ketika jumlah sampel di satu kelas jauh lebih sedikit daripada jumlah sampel di kelas lain. ROS bertujuan untuk meningkatkan jumlah sampel di kelas minoritas dengan menduplikasi atau menggandakan sampel secara acak. Hal ini memastikan bahwa data yang digunakan untuk pelatihan model memiliki distribusi kelas yang seimbang, yang pada gilirannya dapat menghasilkan model yang lebih baik untuk mengklasifikasikan teks atau dokumen. Jika dituliskan secara matematis sebagai berikut:

$$\Delta N_{minority} = (N_{majority} - N_{minority}) \times r_{ros} \quad (2)$$

Tabel 4 Implementasi ROS

No	Kelas	Frekuensi sebelum ROS	Frekuensi sesudah ROS
1	Neutral	5458	5458
2	Joy	5426	5458
3	Sadness	4597	5458
4	Anger	2177	5458
5	Fear	1917	5458
6	Love	1284	5458
7	Surprise	581	5458

2.6 Implementasi Model Dasar

Pada tahap ini, peneliti akan mengimplementasikan data yang telah dipersiapkan sebelumnya dan telah melalui proses optimasi kualitas data. *Base Learners* yang akan dilatih terdiri dari lima jenis, yaitu *Multinomial Naïve Bayes*, *Support Vector Machine*, *Logistic Regression*, *Gradient Boosting*, dan *K-Nearest Neighbors*.

2.6.1 Algoritma Multinomial Naïve Bayes

Naïve Bayes merupakan salah satu algoritma klasifikasi yang umum digunakan dalam berbagai kasus atau penelitian. Terdapat beberapa varian algoritma *Naïve Bayes* yang biasa digunakan, seperti *Gaussian Naïve Bayes*, *Multinomial Naïve Bayes*, *Bernoulli Naïve Bayes*, *Complement Naïve Bayes*, dan *Categorical Naïve Bayes*. Penelitian ini akan berfokus pada penerapan *Multinomial Naïve Bayes* karena kesesuaiannya untuk klasifikasi teks pada beberapa penelitian sebelumnya. *Multinomial Naïve Bayes* dapat dituliskan sebagai berikut.

$$P(c, x) \propto P(c) \prod_{k=1}^n P(t_k | c) \quad (3)$$

2.6.2 Algoritma Support Vector Machine

Support Vector Machine (SVM) adalah metode yang umum digunakan untuk penelitian klasifikasi dan regresi. Fokus utama dari SVM adalah menemukan *hyperplane* terbaik yang memisahkan dua kelas dalam ruang fitur sedemikian rupa sehingga *margin* (jarak) antara *hyperplane* dan titik-titik terdekat dari kedua kelas (*support vector*) adalah maksimum. Secara matematis, *Support Vector Machine* dapat dituliskan sebagai berikut.

$$\hat{y} = \arg \max_c \sum (W_c \cdot x + b_c) \quad (4)$$

2.6.3 Algoritma Logistic Regression

Logistic Regression adalah salah satu teknik klasifikasi yang umum digunakan dalam pembelajaran mesin. Tujuannya adalah untuk memprediksi probabilitas bahwa sebuah data akan masuk ke dalam salah satu dari dua kelas yang ada. Regresi Logistik menggunakan fungsi logistik (juga dikenal sebagai *sigmoid*) untuk mengubah input kontinu menjadi output dalam rentang 0 hingga 1, yang dapat diartikan sebagai probabilitas.

$$P(y = C_n | x) = \frac{e^{-(w_n \cdot x + b_n)}}{1 + e^{-(w_n \cdot x + b_n)}} \quad (5)$$

2.6.4 Algoritma Gradient Boosting

Gradient Boosting adalah algoritma pembelajaran mesin yang kuat dan populer yang digunakan untuk masalah klasifikasi dan regresi. Algoritma ini bekerja dengan menggabungkan serangkaian model prediktif sederhana, seperti pohon keputusan, ke dalam model yang lebih kompleks dan kuat. Secara matematis, hal ini dapat dituliskan sebagai berikut.

$$P(x) = P_0(x) + \lambda \sum_{m=1}^M \gamma_m h_m(x) \quad (6)$$

2.6.5 Algoritma K-Nearest Neighbors

K-Nearest Neighbours (KNN) adalah algoritma klasifikasi dan regresi yang sederhana dan intuitif. KNN tidak memerlukan pembelajaran model sebelumnya dan hanya bergantung pada data yang diberikan untuk membuat prediksi. Algoritma ini bekerja dengan menemukan k tetangga terdekat dan menggunakan label mayoritas atau nilai target rata-rata dari tetangga tersebut.

$$y = \arg \max_c \sum_{i=1}^k 1(y_i = c) \quad (7)$$

2.7 Implementasi Model Hard Voting Ensemble

Hard Voting, umumnya dikenal sebagai voting mayoritas, adalah teknik pembelajaran mesin yang melibatkan tindakan sederhana untuk menghitung dari beberapa prediksi yang sebelumnya dihasilkan oleh model pembelajaran dasar. Dalam *hard voting*, setiap model dasar memberikan prediksinya pada data uji, dan hasil prediksi dari setiap model dijumlahkan. Prediksi akhir kemudian diambil berdasarkan suara terbanyak dari semua model yang terlibat. Jika dituliskan secara matematis, dapat dilihat pada persamaan 8.

$$H_{(x)} = \arg \max_x \{h_1(x), h_2(x), \dots, h_n(x)\} \quad (8)$$

3. HASIL DAN PEMBAHASAN

Hasil yang diperoleh dalam penelitian ini dilakukan dalam proses membangun beberapa model machine learning dan melakukan proses evaluasi dengan menggunakan metrik evaluasi sebagai berikut.

3.1 Hasil Implementasi Algoritma

Penelitian ini akan berfokus pada penggunaan 2 metrik evaluasi, yaitu *mean squared error* (mse) dan *confusion matrix* yang terdiri dari nilai akurasi, *presisi*, *recall*, dan *f1-score*. Hasil dari setiap implementasi algoritma adalah sebagai berikut:

3.1.1 Multinomial Naïve Bayes

Tabel 5 menunjukkan bahwa model *multinomial naive bayes* memiliki performa yang cukup baik karena mencapai nilai akurasi sebesar 0.75. Model ini juga memiliki kinerja yang cukup baik dalam mengklasifikasikan kelas ‘*fear*’, yang dibuktikan dengan nilai *precision*, *recall*, dan *f1-score* yang cukup stabil dan di atas 0.80. Namun, ia kesulitan dalam mengklasifikasikan kelas ‘*surprise*’.

Tabel 5 Hasil proses pelatihan algoritma *Multinomial Naïve Bayes*

kelas	<i>precision</i>	<i>recall</i>	<i>f1-score</i>	<i>accuracy</i>	<i>mse</i>
neutral	0.95	0.47	0.62	0.75	1.81
joy	0.89	0.40	0.55		
sadness	0.63	0.95	0.76		
anger	0.98	0.13	0.24		
fear	0.89	0.87	0.88		
love	0.73	0.92	0.82		
surprise	1.00	0.03	0.06		

3.1.2 Support Vector Machine

Tabel 6 menunjukkan bahwa model *Support Vector Machine* memiliki performa yang lebih baik dibandingkan model sebelumnya, meskipun nilai akurasi sama dengan model sebelumnya, nilai mse dari model ini lebih rendah dibandingkan model sebelumnya. Model ini juga memiliki performa yang baik dalam mengklasifikasikan kelas ‘*sadness*’, ‘*fear*’, dan ‘*love*’.

Tabel 6 Hasil proses pelatihan algoritma *Support Vector Machine*

kelas	<i>precision</i>	<i>recall</i>	<i>f1-score</i>	<i>accuracy</i>	<i>mse</i>
neutral	0.91	0.82	0.86	0.89	0.93
joy	0.86	0.78	0.82		
sadness	0.83	0.95	0.88		
anger	0.89	0.64	0.75		
fear	0.96	0.98	0.97		
love	0.91	0.93	0.92		
surprise	0.81	0.57	0.67		

3.1.3 Logistic Regression

Tabel 7 menunjukkan bahwa performa model *Logistic Regression* lebih baik dibandingkan dengan model *Multinomial Naïve Bayes* namun tidak lebih baik dibandingkan dengan model *Support Vector Machine*. Seperti halnya model SVM, model ini memiliki performa yang baik dalam mengklasifikasikan kelas ‘*sadness*’, ‘*fear*’, ‘*love*’, dan ‘*neutral*’.

Tabel 7 Hasil proses pelatihan algoritma *Logistic Regression*

kelas	<i>precision</i>	<i>recall</i>	<i>f1-score</i>	<i>accuracy</i>	<i>mse</i>
neutral	0.92	0.81	0.86	0.89	0.94
joy	0.87	0.77	0.82		
sadness	0.84	0.94	0.89		
anger	0.88	0.67	0.76		
fear	0.95	0.98	0.96		
love	0.90	0.94	0.92		
surprise	0.81	0.54	0.65		

3.1.4 Gradient Boosting

Tabel 8 menunjukkan bahwa model *Gradient Boosting* memiliki nilai akurasi yang lebih rendah dibandingkan dengan model *Logistic Regression* dan SVM, namun lebih unggul dibandingkan dengan model *Multinomial Naive Bayes*. Model ini menunjukkan nilai yang lebih stabil dibandingkan dengan beberapa model sebelumnya, meski belum optimal.

Tabel 8 Hasil proses pelatihan algoritma *Gradient Boosting*

kelas	precision	recall	f1-score	accuracy	mse
neutral	0.93	0.75	0.83	0.86	1.06
joy	0.91	0.74	0.82		
sadness	0.74	0.93	0.83		
anger	0.82	0.74	0.78		
fear	0.96	0.98	0.97		
love	0.95	0.82	0.88		
surprise	0.68	0.77	0.72		

3.1.5 K-Nearest Neighbours

Tabel 9 menunjukkan bahwa performa model KNN tidak terlalu baik pada penelitian ini, terlihat dari nilai setiap metrik evaluasi yang sangat berbeda dibandingkan dengan model lainnya. Selain itu, nilai mse juga sangat tinggi yaitu 3.45 yang dapat disimpulkan bahwa model memiliki tingkat kesalahan yang tinggi pada penelitian ini.

Tabel 9 Hasil proses pelatihan algoritma *K-Nearest Neighbours*

kelas	precision	recall	f1-score	accuracy	mse
neutral	0.82	0.09	0.16	0.32	3.45
joy	0.94	0.10	0.18		
sadness	0.87	0.08	0.15		
anger	0.84	0.04	0.09		
fear	0.27	0.99	0.42		
love	0.94	0.12	0.22		
surprise	0.73	0.08	0.14		

3.1.6 Hard Voting Ensemble

Metode ini diimplementasikan untuk melihat seberapa besar pengaruhnya dalam meningkatkan nilai dari beberapa metrik evaluasi yang digunakan. Beberapa model yang telah dilatih sebelumnya, pada model ini akan digunakan sebagai *base learner*, yang kemudian akan diterapkan *hard voting* untuk menentukan nilai akhir atau prediksi pada model ini.

Tabel 10 Hasil proses pelatihan algoritma *Hard Voting Ensemble*

kelas	precision	recall	f1-score	accuracy	Mse
neutral	0.82	0.09	0.16	0.32	3.45
joy	0.94	0.10	0.18		
sadness	0.87	0.08	0.15		
anger	0.84	0.04	0.09		
fear	0.27	0.99	0.42		
love	0.94	0.12	0.22		
surprise	0.73	0.08	0.14		

Tabel 10 menunjukkan bahwa performa model dengan mengimplementasikan model *ensemble* dan menggunakan pendekatan *hard voting*, menghasilkan nilai metrik evaluasi yang lebih baik dibandingkan dengan semua model sebelumnya. Hasil ini dapat membuktikan bahwa dengan menerapkan metode *ensemble* dapat meningkatkan performa dari suatu model, dan meminimalisir nilai *error rate* dari model tersebut dalam melakukan klasifikasi.

3.2 Hasil Evaluasi Model

Berdasarkan hasil implementasi dan/atau pengujian beberapa algoritma yang telah dilakukan, terlihat bahwa nilai metrik evaluasi secara keseluruhan memiliki perbedaan yang tidak

signifikan. Namun, jika diukur melalui nilai metrik evaluasi *mean squared error* (MSE).

Tabel 11 Hasil Evaluasi Model dengan Implementasi Metode ROS

Metrik Evaluasi		Model					
		MNB	SVM	LRG	GBS	KNN	HVE
accuracy		0.81	0.90	0.90	0.88	0.33	0.92
mean squared error		1.67	0.92	0.93	1.04	3.40	0.87
neutral	precision	0.72	0.90	0.88	0.88	0.68	0.88
	recall	0.84	0.84	0.88	0.85	0.15	0.89
	f1-score	0.77	0.87	0.88	0.87	0.25	0.89
joy	precision	0.73	0.86	0.86	0.86	0.74	0.86
	recall	0.79	0.81	0.82	0.76	0.15	0.83
	f1-score	0.75	0.84	0.84	0.81	0.25	0.84
sadness	precision	0.88	0.86	0.93	0.84	0.95	0.91
	recall	0.81	0.91	0.87	0.87	0.07	0.89
	f1-score	0.84	0.89	0.90	0.85	0.13	0.90

Metrik Evaluasi		Model					
		MNB	SVM	LRG	GBS	KNN	HVE
anger	precision	0.58	0.80	0.74	0.73	0.61	0.76
	recall	0.84	0.78	0.92	0.95	0.14	0.87
	f1-score	0.69	0.79	0.82	0.83	0.22	0.81
fear	precision	0.94	0.97	0.96	0.96	0.28	0.95
	recall	0.76	0.98	0.97	0.98	0.97	0.98
	f1-score	0.88	0.97	0.96	0.97	0.42	0.96
love	precision	0.73	0.91	0.90	0.95	0.94	0.92
	recall	0.92	0.93	0.94	0.82	0.12	0.91
	f1-score	0.82	0.92	0.92	0.88	0.22	0.91
surprise	precision	1.00	0.81	0.81	0.68	0.73	0.85
	recall	0.03	0.57	0.54	0.77	0.08	0.53
	f1-score	0.06	0.67	0.65	0.72	0.14	0.65

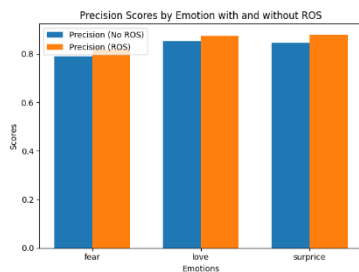
Keterangan:

- MNB : *Multinomial Naïve Bayes*
- SVM : *Support Vector Machine*
- LRG : *Logistic Regression*
- GBS : *Gradient Boosting*
- KNN : *K-Nearest Neighbors*
- HVE : *Hard Voting Ensemble*

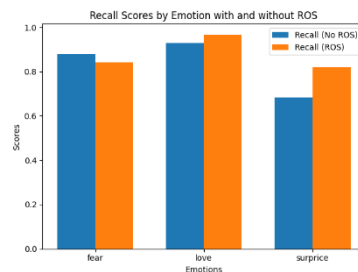
Tabel 12 menunjukkan hasil evaluasi dari model-model dengan penerapan metode *Random Over Sampling* (ROS). Penggunaan metode ini bertujuan untuk mengatasi ketidakseimbangan kelas dalam data. Hasilnya menunjukkan peningkatan kinerja keseluruhan model, namun KNN masih menunjukkan kinerja yang buruk dengan akurasi hanya sebesar 0,33. Terlihat bahwa model SVM dan *Multinomial Naïve Bayes* (MNB) berhasil mengurangi nilai MSE, dengan masing-masing 0,92 dan 0,87. Namun, KNN tetap memiliki MSE yang tinggi, yaitu mencapai 3,40, menunjukkan bahwa KNN tidak dapat memprediksi seakurat model lainnya, baik dengan atau tanpa penerapan metode ROS. Sementara itu, model MNB mencapai akurasi 0,81 dengan MSE 1,67, yang mengindikasikan kinerja yang baik dalam mengklasifikasikan data, terutama dalam memprediksi emosi ‘sadness’ dengan *precision*, *recall*, dan *f1-score* yang tinggi. Selain itu, model *Support Vector Machine* dan *Logistic Regression* menunjukkan kinerja keseluruhan yang lebih baik dengan akurasi masing-masing sebesar 0,90, meskipun tidak pada metrik evaluasi lainnya. SVM menonjol dalam hal stabilitas, sedangkan *Logistic Regression* stabil dalam memprediksi berbagai emosi. Namun, peningkatan kinerja yang konsisten pada kelas tertentu masih diperlukan untuk meningkatkan kemampuan model dalam memprediksi dengan lebih akurat.

Hasil implementasi dari semua algoritma yang digunakan, mengungkap beberapa poin penting. Model *Multinomial Naïve Bayes* (MNB) menunjukkan kinerja yang cukup baik dengan

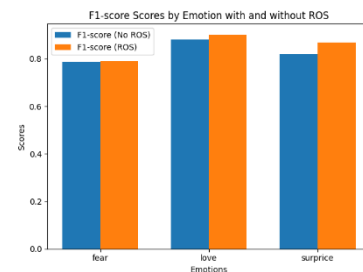
akurasi sebesar 0,89. Model ini menunjukkan kinerja yang baik dalam mengklasifikasikan kelas 'fear', namun berkinerja buruk dalam mengklasifikasikan kelas 'surprise' dengan nilai MSE sebesar 1,81. Model *Support Vector Machine* (SVM) mengungguli model MNB dengan akurasi sebesar 0,89 dan MSE yang lebih rendah, yaitu 0,93. Model SVM berkinerja baik dalam mengklasifikasikan beberapa kelas seperti 'sadness', 'fear', dan 'love'. Model *Logistic Regression* (LRG) juga lebih baik dari MNB tetapi tidak sebaik SVM, dengan akurasi 0,89 dan MSE sebesar 0,94. Model LRG berkinerja baik dalam mengklasifikasikan kelas 'sadness', 'fear', 'love', dan 'neutral'. Model *Gradient Boosting* (GBS) memiliki akurasi yang sedikit lebih rendah yaitu 0,86 dibandingkan SVM dan LRG. Namun, nilai *precision*, *recall*, dan *f1-score* yang dihasilkan oleh model GBS lebih stabil, dengan semua nilai di atas 0,70 dan MSE sebesar 1,06. Di sisi lain, model *K-Nearest Neighbours* (KNN) menunjukkan kinerja yang buruk dengan akurasi hanya 0,32 dan MSE yang tinggi sebesar 3,45, serta kinerja yang buruk pada semua metrik evaluasi yang digunakan. Model terakhir, *Hard Voting Ensemble* (HVE), menunjukkan bahwa metode *ensemble* menghasilkan nilai metrik evaluasi yang lebih baik daripada semua model yang disebutkan sebelumnya. Model HVE mencapai akurasi sebesar 0,90 dengan nilai MSE terendah, yaitu sebesar 0,91.



Gambar 1 Nilai *precision*



Gambar 2 Nilai *recall*



Gambar 3 Nilai *f1-score*

Berdasarkan Gambar 1, terlihat bahwa nilai metrik evaluasi *precision* untuk tiga kelas minoritas mengalami peningkatan, terutama untuk kelas 'surprise', di mana perbedaannya sangat terlihat. Gambar 1 memvisualisasikan nilai dengan mengambil rata-rata dari semua model yang dilatih dengan dan tanpa metode ROS. Nilai *precision* untuk kelas 'fear' sebelum ROS adalah 0,79, dan saat menggunakan ROS adalah 0,82. Nilai *precision* untuk kelas 'love' sebelum ROS adalah 0,85, dan saat menggunakan ROS adalah 0,87. Nilai *precision* untuk kelas 'surprise' sebelum ROS adalah 0,85, dan saat menggunakan ROS adalah 0,88. Dari nilai-nilai tersebut, dapat disimpulkan bahwa penerapan metode ROS dapat meningkatkan nilai metrik evaluasi *precision*, meskipun peningkatan ini tidak terlalu signifikan, hanya sekitar 3,61% secara rata-rata.

Gambar 2 menunjukkan grafik nilai rata-rata metrik evaluasi *recall* untuk hasil pelatihan setiap model yang dikategorikan berdasarkan nama kelas minoritas. Terlihat bahwa nilai *recall* untuk kelas 'fear' sebelum ROS adalah 0,88, dan saat menggunakan ROS turun menjadi 0,84. Nilai *recall* untuk kelas 'love' sebelum ROS adalah 0,92, dan saat menggunakan ROS meningkat menjadi 0,96. Nilai *recall* untuk kelas 'surprise' sebelum ROS adalah 0,68, dan saat menggunakan ROS meningkat menjadi 0,82. Satu kelas menunjukkan peningkatan yang signifikan, yaitu kelas 'surprise', dengan peningkatan sebesar 20,59%.

Pada Gambar 3, peningkatan juga dapat dilihat pada beberapa kelas, sementara ada kelas yang tidak mengalami peningkatan dalam nilainya. Kelas yang tidak mengalami peningkatan adalah kelas 'fear', dengan nilai *f1-score* sebesar 0,79. Namun, untuk kelas lainnya, masih terdapat peningkatan. Untuk kelas 'love', nilai *f1-score* tanpa metode ROS adalah 0,88, dan saat menggunakan metode ROS, nilai tersebut meningkat menjadi 0,90. Sementara itu, untuk nilai *f1-score* kelas 'surprise' sebelum ROS adalah 0,81, dan saat menggunakan ROS, nilai tersebut meningkat menjadi 0,86. Secara rata-rata, kedua kelas ini mengalami peningkatan yang relatif tinggi sebesar 4,14%.

Selain ketiga nilai metrik evaluasi yang disebutkan di atas, nilai *mean squared error* (MSE) juga penting untuk diamati. Secara rata-rata, antara model yang tidak menerapkan ROS

dan yang telah menerapkan ROS, terdapat peningkatan sebesar -3,10%. Hal ini dapat disimpulkan bahwa dengan menerapkan metode ROS, kinerja model dapat ditingkatkan dengan mengurangi *mean squared error* (MSE) secara signifikan. MSE adalah salah satu metrik evaluasi penting dalam pemodelan statistik dan pembelajaran mesin, yang berfungsi untuk mengevaluasi kualitas atau akurasi model regresi atau prediksi dengan mengukur rata-rata kuadrat perbedaan antara nilai aktual (*ground truth*) dan nilai prediksi yang dihasilkan oleh model.

4. KESIMPULAN DAN SARAN

Berdasarkan hasil dan pembahasan penelitian yang telah disajikan sebelumnya, beberapa kesimpulan dapat diambil dari studi ini, antara lain:

1. Klasifikasi 5 model, yaitu Multinomial Naïve Bayes, Gradient Boosting, *Linear Regression*, *K-Nearest Neighbours*, dan *Support Vector Machine* dalam proses klasifikasi emosi ulasan pelanggan menggunakan pendekatan *Natural Language Processing* (NLP) telah berhasil dilakukan melalui beberapa tahapan sehingga model dapat secara optimal memahami pola klasifikasinya. Dari kelima model tersebut, model dengan kinerja terbaik memiliki nilai akurasi sebesar 0,89 dan nilai *mean squared error* sebesar 0,93. Sementara itu, model terburuk adalah *K-Nearest Neighbours* dengan nilai akurasi sebesar 0,32 dan nilai *mean squared error* sebesar 3,45. Hasil ini konsisten dengan penelitian sebelumnya yang menunjukkan bahwa model berbasis *boosting*, seperti *Gradient Boosting*, sering kali lebih unggul dalam klasifikasi berbasis teks jika dibandingkan dengan algoritma non-linear seperti *K-Nearest Neighbours*, yang ditunjukkan dalam analisis sentimen terkait vaksin COVID-19 [8].
2. Penerapan metode *Ensemble* dengan pendekatan *hard voting* dalam klasifikasi emosi ulasan pelanggan dapat meningkatkan nilai metrik evaluasi, terutama pada akurasi yang memperoleh nilai tertinggi sebesar 0,90 dengan nilai *mean squared error* sebesar 0,91. Kedua nilai ini menjadi yang tertinggi dibandingkan dengan hasil pelatihan dari 5 model lainnya, yaitu *Multinomial Naïve Bayes*, *Gradient Boosting*, *Linear Regression*, *K-Nearest Neighbours*, dan *Support Vector Machine*. Hasil ini mendukung temuan penelitian sebelumnya yang menunjukkan bahwa ensemble learning dapat meningkatkan performa model, seperti yang terlihat pada kombinasi model *Bayesian Network*, *Random Forest*, *CART*, dan *Neural Network* dalam studi serupa [13]. Namun, penelitian ini memperluas penerapannya ke dalam konteks klasifikasi emosi pelanggan yang lebih spesifik, memberikan kontribusi tambahan dalam kajian terkait.
3. Penerapan metode *Random Over Sampling* (ROS) dan *ensemble* dengan pendekatan *hard voting* berhasil mengatasi tantangan dalam ketidakseimbangan kelas dan secara keseluruhan meningkatkan kinerja prediksi kelas minoritas. Pendekatan ini memberikan pemahaman yang lebih baik pada model yang dilatih untuk mengklasifikasikan semua kelas secara adil dengan jumlah data yang setara. Pendekatan ini memberikan pemahaman yang lebih baik pada model yang dilatih untuk mengklasifikasikan semua kelas secara adil dengan jumlah data yang setara. Walaupun metode ROS juga diterapkan dalam beberapa penelitian sebelumnya [10, 11], kombinasinya dengan *ensemble hard voting* pada analisis sentimen ulasan pelanggan memberikan alternatif untuk menangani ketidakseimbangan kelas dalam konteks ini, berbeda dengan pendekatan-pendekatan lain seperti *SMOTE* atau *ADASYN* yang lebih umum digunakan dalam studi-studi serupa [9].

Saran yang dapat diberikan adalah agar melakukan penelitian dengan menguji lebih banyak metode klasifikasi, seperti *Neural Networks*, untuk meningkatkan kinerja model. Selain itu, disarankan menggunakan teknik NLP yang lebih canggih seperti *Word Embeddings*

(misalnya, Word2Vec atau GloVe) untuk menangkap makna kontekstual dari ulasan pelanggan dengan lebih baik. Penggunaan teknik peningkatan data lain, seperti SMOTE, juga bisa membantu dalam mengatasi ketidakseimbangan kelas, dan eksperimen dengan *hyperparameter tuning* dapat lebih meningkatkan akurasi model yang sudah ada.

DAFTAR PUSTAKA

- [1] K. Jermittiparsert, "Leadership and Industry 4.0 As A Tool to Enhance Organization Performance : Direct and Indirect Role of Job Satisfaction , Competitive Advantage and Business Sustainability," pp. 233–257, doi: 10.1108/978-1-80043-380-920201014.
- [2] P. Gaiardelli *et al.*, *Product-service systems evolution in the era of Industry 4.0*, vol. 15, no. 1. Springer Berlin Heidelberg, 2021. doi: 10.1007/s11628-021-00438-9.
- [3] K. Bruce, "Customer Orientation : Its Surprising Origins , Tumultuous Development and Place in the Future of Marketing thought and Practice and place in the future of marketing thought and practice," *Australas. Mark. J.*, no. May, 2020, doi: 10.1016/j.ausmj.2020.03.007.
- [4] T. Wharton, C. Bonard, D. Dukes, D. Sander, and S. Oswald, "Relevance and emotion," *J. Pragmat.*, vol. 181, pp. 259–269, 2021, doi: 10.1016/j.pragma.2021.06.001.
- [5] I. Masitoh, P. Supriadi, and R. Marliani, "Dampak Kepribadian Introvert dalam Interaksi Sosial," vol. 1, no. 2, pp. 245–249, 2023, doi: 10.59996/jurnalpelitanusantara.v1i2.203.
- [6] A. Torabi, H. Hamidi, and N. Safaie, "Effect of Sensory Experience on Customer Word-of-mouth Intention , Considering the Roles of Customer Emotions , Satisfaction , and Loyalty," vol. 34, no. 03, pp. 682–699, 2021, doi: 10.5829/ije.2021.34.03c.13.
- [7] T. Shaik, X. Tao, C. Dann, H. Xie, and Y. Li, "Sentiment Analysis and Opinion Mining on Educational Data : A Survey," *Nat. Lang. Process. J.*, vol. 2, no. Yan Li, p. 100003, 2023, doi: 10.1016/j.nlp.2022.100003.
- [8] P. Arsi, I. Prayoga, and M. H. Asyari, "Klasifikasi Sentimen Publik Terhadap Jenis Vaksin Covid-19 yang Tersertifikasi WHO Berbasis NLP dan KNN," *J. Media Inform. Budidarma*, vol. 7, no. 1, pp. 260–266, 2023, doi: 10.30865/mib.v7i1.5418.
- [9] T. Wongvorachan and S. He, "A Comparison of Undersampling , Oversampling , and SMOTE Methods for Dealing with Imbalanced Classification in Educational Data Mining," *Information*, vol. 14, no. 54, 2023, doi: 10.3390/info14010054.
- [10] P. Rana, A. Sowmya, E. Meijering, and Y. Song, "Imbalanced classification for protein subcellular localization with multilabel oversampling," *Bioinformatics*, vol. 39, no. 1, 2023, doi: 10.1093/bioinformatics/btac841.
- [11] U. Ependi, "A Hybrid Sampling Approach for Improving the Classification of Imbalanced Data Using ROS and NCL Methods," vol. 16, no. 3, pp. 345–361, 2023, doi: 10.22266/ijies2023.0630.28.
- [12] E. Widiputra, Harya; Mailangkay, Adele; Gautama, "Prediksi Indeks BEI dengan Ensemble Convolutional Neural Network dan Long Short-Term Memory," *Rekayasa Sist. dan Teknol. Inf.*, vol. 1, no. 10, pp. 456–465, 2021, doi: 10.29207/resti.v5i3.3111.
- [13] A. Malik, E. Michael, S. Dalal, and U. Kumar, "Forecasting students ' adaptability in online entrepreneurship education using modified ensemble machine learning model Forecasting students ' adaptability in online entrepreneurship education using modified ensemble machine learning model," *Array*, vol. 19, no. July, p. 100303, 2023, doi: 10.1016/j.array.2023.100303.
- [14] E. Salesky, C. Raffel, C. Si, and W. Y. Lee, "Between words and characters: A Brief History of Open-Vocabulary Modeling and Tokenization in NLP," *arXiv*, 2021, doi: 10.48550/arXiv.2112.10508.
- [15] L. Hickman, S. Thapa, and L. Tay, "Text Preprocessing for Text Mining in Organizational Research : Review and Recommendations," *Organ. Res. Methods*, vol. 25, no. 1, pp. 114–146, 2022, doi: 10.1177/1094428120971683.
- [16] K. Tuomo *et al.*, "Stemming and lemmatization in the clustering of finnish text conference

- on Information and knowledge management Editors of work : Pages : Stemming and Lemmatization in the Clustering of Finnish Text Documents,” 2004.
- [17] Z. Jiang, B. Gao, Y. He, Y. Han, P. Doyle, and Q. Zhu, “Text Classification Using Novel Term Weighting Scheme-Based Improved TF- IDF for Internet Media Reports Text Classification Using Novel Term Weighting Scheme-Based Improved TF-IDF for Internet Media Reports,” no. March, 2021, doi: 10.1155/2021/6619088.