



Analisis Sentimen Ulasan Produk Perawatan Wajah menggunakan Algoritma CNN dan Seleksi Fitur PSO

Sentiment Analysis of Facial Care Product Reviews using CNN Algorithm and PSO Feature Selection

Rahma Ardianti, Prodi Matematika FMIPA UNY
Nur Hadi Waryanto*, Prodi Matematika FMIPA UNY
*e-mail: nur_hw@uny.ac.id

Abstrak

Penelitian ini bertujuan untuk menganalisis dan mengklasifikasi sentimen ulasan produk perawatan wajah merek AVS menggunakan *Convolutional Neural Network* (CNN) dan seleksi fitur *Particle Swarm Optimization* (PSO), serta meninjau pengaruh algoritma PSO, perbandingan rasio dataset, dan model *word embedding* terhadap hasil akurasi klasifikasi CNN. Penelitian dilakukan dengan mengumpulkan data melalui *web scraping*, pelabelan dan *preprocessing* data, *word embedding*, seleksi fitur dengan PSO, klasifikasi sentimen dengan CNN, dan evaluasi model klasifikasi. Model klasifikasi CNN memperoleh akurasi sebesar 61% hingga 83% untuk sembilan percobaan untuk tiga rasio dataset yang berbeda yaitu 70:30, 75:25, 80:20 dan tiga model *word embedding* yang berbeda yaitu Glove, Word2Vec CBOW, dan Word2Vec Skipgram. Sedangkan pengujian menggunakan PSO-CNN memilih 44-60 fitur atau kata-kata penting dari total 100 fitur, sehingga diperoleh nilai akurasi 77% hingga 86%. Rata-rata kenaikan akurasi yang diperoleh akibat penambahan seleksi fitur oleh PSO adalah 8,7%. Akurasi terbaik diperoleh oleh model PSO-CNN dengan rasio dataset 80:20 dan model *word embedding* Word2Vec Skipgram yaitu 86%.

Kata kunci: analisis sentimen, produk perawatan wajah, CNN, seleksi fitur, PSO.

Abstract

This study aims to analyze and classify review sentiments of AVS facial care products using the Convolutional Neural Network (CNN) and Particle Swarm Optimization (PSO) feature selection, as well as the effect of the PSO algorithm, dataset ratio, and word embedding models on the results of CNN classification accuracy. The research was conducted by collecting data through web scraping, labeling and data preprocessing, word embedding, feature selection with PSO, sentiment classification with CNN, and evaluation of classification models. The CNN classification model obtained an accuracy of 61% to 83% for nine trials for three different dataset ratios 70:30, 75:25, 80:20 and three different word embedding models Glove, Word2Vec CBOW, and Word2Vec Skipgram. Meanwhile, testing using PSO-CNN selects 44-60 important features or words from a total of 100 features, in order to obtain an accuracy value of 77% to 86%. The average increase in accuracy obtained due to the addition of feature selection by PSO is 8.7%. The best accuracy was obtained by the PSO-CNN model with a dataset ratio of 80:20 and the Word2Vec Skipgram word embedding model, namely 86%.

Keywords: sentiment analysis, skin care products, CNN, feature selection, PSO.

PENDAHULUAN

Produk kecantikan dan perawatan kulit wajah semakin dikenal dan diincar oleh konsumen Indonesia. Dikutip dari surat kabar Media Indonesia (2021), Analisis Kebijakan Kementerian Perindustrian memperkirakan Indonesia akan menjadi pasar kosmetik terbesar ke-5 di dunia dalam 10 hingga 15 tahun mendatang karena data tren belanja kosmetik yang mencakup segmen produk kosmetik, produk perawatan kulit serta wajah, dan *personal care* di Indonesia terus meningkat. Salah satu merek lokal yang tengah mendapat perhatian masyarakat adalah AVS yang awalnya berdiri pada tahun 2014 di kota Yogyakarta. Riset yang dilakukan oleh Kompas (2021) menyatakan bahwa selama bulan Juli 2021 saja AVS berhasil mencapai total penjualan sebesar Rp6,36 miliar dari kurang lebih 47 ribu transaksi di *marketplace* Shopee.

Ulasan (*review*) penggunaan produk dapat membantu calon konsumen menyimpulkan bagaimana kualitas produk dari suatu merek tertentu. Di Indonesia sendiri terdapat sebuah situs bernama Female Daily yang menyediakan laman khusus untuk ulasan dan deskripsi detail informasi produk perawatan wajah dan tubuh. Ketersediaan data yang banyak dan sentimen yang beragam mengakibatkan sulitnya mengolah data secara manual untuk memperoleh informasi. Oleh karena itu, untuk memperoleh hasil analisis dan informasi yang efisien maka dilakukan *text mining*. Analisis sentimen merupakan salah satu *text mining* yang diperlukan untuk mengekstrak sentimen dengan topik yang spesifik dalam jangka waktu singkat. Tidak hanya berguna bagi calon konsumen, sentimen masyarakat juga memiliki nilai bagi perusahaan sebagai bahan evaluasi untuk meningkatkan kualitas produk

Convolutional Neural Network (CNN) merupakan salah satu algoritma *deep learning* yang dapat digunakan untuk melakukan klasifikasi kalimat karena performanya yang baik dengan setiap *hyperparameter* CNN memberi pengaruh terhadap model klasifikasi untuk analisis sentimen (Zhang & Wallace, 2016). Penelitian dengan data Bahasa Indonesia oleh Afandika (2018) mengenai analisis sentimen operator telekomunikasi menggunakan CNN memberikan hasil akurasi terbaik sebesar 96,80%. Meskipun dapat menghasilkan akurasi yang cukup tinggi, proses klasifikasi data ulasan yang memiliki banyak fitur yang redundan dan tidak relevan berpotensi akan menghasilkan model klasifikasi yang *overfitting* (Kunnath et al., 2019), yaitu keadaan ketika model klasifikasi melatih data *training* dengan baik, tetapi tidak dengan data *testing*.

Menanggapi masalah tersebut, maka dinilai perlu untuk melakukan seleksi fitur menggunakan algoritma optimasi *Particle Swarm Optimization* (PSO). Penerapan PSO sebagai seleksi fitur akan mencegah terjadinya *overfitting* dengan mengidentifikasi subset fitur terbaik yang relevan yang paling berkontribusi terhadap akurasi dan efisiensi model klasifikasi. PSO juga merupakan salah satu algoritma yang berguna untuk mengatasi data berdimensi tinggi sehingga akan meningkatkan akurasi klasifikasi (Zahran & Kanaan, 2009). Menurut Xue et al., (2013) dibandingkan dengan algoritma pencarian global *Evolutionary Computation* lainnya seperti *Genetic Algorithm* (GA) dan *Genetic Programming* (GP), PSO memiliki kemampuan komputasi yang tidak mahal dan lebih cepat untuk mencapai nilai optimum. PSO telah digunakan sebagai teknik yang efektif di banyak bidang, termasuk pemilihan fitur. Penelitian oleh Kaur, et.al (2019) telah menunjukkan penggunaan algoritma klasifikasi CNN dan seleksi fitur PSO mampu meningkatkan akurasi deteksi sentimen konsumen *e-commerce* sebesar 8,99%.

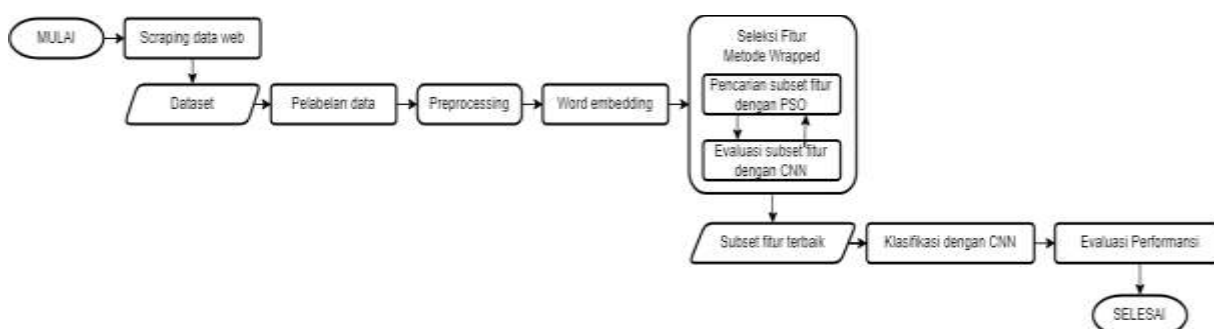
Proses klasifikasi sentimen dengan CNN dan PSO membutuhkan dataset yang nantinya akan dibagi ke dalam data latih dan data uji. Proses pemilihan rasio data diperlukan untuk mengetahui rasio mana yang akan memberi hasil kinerja terbaik dalam mengklasifikasikan sentimen. Penelitian oleh Muraina (2020) menunjukkan bahwa rasio dataset yang berbeda memiliki pengaruh pada model dan kinerja hasil klasifikasi. Selain itu, proses mengubah kata menjadi vektor (*word embedding*) dalam klasifikasi sentimen memiliki banyak pilihan model yang beragam. Dua model *word embedding* yang biasa digunakan adalah Glove dan Word2Vec.

Penelitian oleh Wang, et al., (2015), menunjukkan bahwa akurasi *word embedding* Glove-CNN dan Word2Vec-CNN mampu melebihi kombinasi model lainnya.

Berdasarkan permasalahan di atas, maka dalam penelitian ini akan melakukan analisis dan klasifikasi sentimen ulasan pengguna produk AVS untuk mengetahui respon dan penilaian kualitas produk AVS. Penelitian ini mengintegrasikan dua metode yaitu CNN sebagai metode klasifikasi dan PSO sebagai metode seleksi fitur untuk mengklasifikasikan ulasan berbahasa Indonesia terhadap produk perawatan wajah. Kemudian, digunakan pula tiga jenis rasio dataset yaitu 70:30, 75:25, 80:20 dan tiga *word embedding* yang berbeda yaitu Glove, Word2Vec CBOV, dan Word2Vec Skipgram untuk menguji model dengan rasio dan *word embedding* mana yang akan memberikan hasil yang terbaik.

METODE

Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari situs ulasan produk kecantikan yaitu Female Daily menggunakan metode *scraping* pada situs [website https://reviews.femaledaily.com](https://reviews.femaledaily.com). Data yang digunakan berupa ulasan konsumen produk perawatan wajah khususnya merek AVS yang dikumpulkan pada tanggal 11-12 Februari 2022. Jumlah data adalah sebanyak 10.056 ulasan yang disimpan dalam format *.xlsx*. Namun, setelah data dibersihkan dan melalui tahap *preprocessing*, data yang siap diproses untuk tahap klasifikasi adalah sebanyak 7.529 data yang terdiri dari 6.411 data ulasan positif dan 1118 data ulasan negatif.



Gambar 1. Diagram Alir Proses Klasifikasi Sentimen dengan CNN dan PSO

Scraping data ulasan di *website* Female Daily yaitu <https://reviews.femaledaily.com> adalah langkah pertama yang dilakukan dalam penelitian ini. *Scraping* data ulasan dilakukan dengan cara memanggil *library requests* yang berfungsi untuk mengirimkan permintaan (*request*) HTTP ke *server* situs web. *Library BeautifulSoup* akan dipanggil untuk *parsing* atau membaca kode HTML yang merupakan komponen utama dalam pembuatan struktur web.

Sebelum proses klasifikasi sentimen dijalankan, perlu dilakukan beberapa hal yang disebut *preprocessing* karena data ulasan yang dikumpulkan belum terstruktur dengan baik. Data teks yang terkumpul memiliki kata-kata dan karakter yang tidak informatif. Atribut-atribut ini tidak cukup berpengaruh untuk proses klasifikasi. Oleh karena itu, beberapa langkah *preprocessing* diperlukan untuk membuat data terstruktur dengan baik. *Preprocessing* pada penelitian ini dilakukan dengan beberapa langkah diantaranya: (1) membersihkan data dari data duplikat (*data cleaning*); (2) mengkonversi huruf kapital menjadi huruf kecil (*case folding*); (3) mengkonversi kata slang menjadi kata baku; (4) menyaring ulasan yang hanya menggunakan Bahasa Indonesia (*language filtering*); (5) membersihkan simbol/karakter dan angka (*punctuation and number removal*); (6) mengubah kata berimbuhan menjadi kata dasar (*stemming*); (7) dan menghapus kata-kata yang tidak penting (*stopword removal*).

Data ulasan dibagi menjadi dua jenis yaitu data latih dan data uji dalam tiga rasio yang berbeda yaitu 70:30, 75:25, dan 80:20. Tahap selanjutnya adalah mengubah kata-kata dari output *preprocessing* menjadi nilai vektor dengan *word embedding* Word2Vec arsitektur

Skipgram dan CBOW dan sebagai perbandingan juga dilakukan *word embedding* Glove. *Word embedding* dapat membantu untuk menentukan kemiripan kata berdasarkan konteks tertentu dilihat dari nilai vektor kata tersebut (Chatterjee, et.al, 2021). Glove adalah salah satu *word embedding* yang telah melatih 6 milyar kata di Wikipedia 2014 dan korpus Gigaword. Glove juga menggabungkan informasi *co-occurrence* kata atau statistik global untuk memperoleh hubungan semantik antarkata dalam korpus (Pennington et al., 2014). Sedangkan, Word2Vec adalah *word embedding* yang dikembangkan oleh Google di tahun 2013. Word2Vec mengimplementasikan dua arsitektur yang berbeda yaitu *Continuous Bag of Words* dan *Skip-gram*. Arsitektur *Continuous Bag of Words* (CBOW) membentuk urutan kata dan akan diprediksi kata mana yang lebih mungkin menjadi kata berikutnya dalam urutan ini. Sedangkan dalam arsitektur skip-gram, setiap kata akan menemukan kumpulan kata lainnya sesuai nilai probabilitas kata, sehingga dalam ruang vektor kata-kata dengan kesamaan semantik akan berdekatan atau dengan kata lain suatu kata akan berperan sebagai input. Hal ini menyebabkan skip-gram lebih cocok untuk penelitian pada dataset kecil dan perannya lebih sesuai untuk kata-kata yang memiliki frekuensi kemunculan sangat sedikit.

Tahap berikutnya adalah seleksi fitur menggunakan metode *wrapper* dengan algoritma optimasi yaitu PSO dan *learning algorithm* berupa algoritma klasifikasi CNN. Klasifikasi data dengan CNN dilakukan pada subset fitur terpilih. *Particle Swarm Optimization* (PSO) adalah algoritma *swarm intelligence* yakni studi mengenai sistem komputasi yang terinspirasi oleh kecerdasan kolektif yang muncul berdasarkan populasi atau kerja sama homogen dalam suatu lingkungan yang biasa digunakan untuk memecahkan masalah optimasi dan seleksi fitur (Sasongko, 2016). PSO melakukan pencarian solusi menggunakan populasi atau yang disebut *swarm* dari individu yang disebut partikel yang diperbarui dari iterasi ke iterasi. Untuk menemukan solusi optimal, masing-masing partikel mengubah arah pencariannya menurut dua faktor yaitu posisi terbaik sebelumnya (*pbest*) dan posisi terbaik dari semua anggota lainnya (*global best/gbest*) (Lin, et al., 2008). *Convolutional Neural Network* adalah salah satu algoritma *deep learning* dan merupakan pengembangan dari *Multilayer Perceptron* (MLP) yang didesain untuk mengolah data dua dimensi. Terdapat dua tahap proses yang dilakukan oleh CNN yaitu *feature learning* dan *classification*. Hal inilah yang kemudian membuat algoritma CNN dapat mengenali objek tanpa memperhatikan dimana posisi fitur tersebut muncul atau biasa disebut *translation invariance* (Raviya & Vennila, 2020).

Evaluasi performa model dilakukan menggunakan metode *confusion matrix*, yaitu tabel matriks yang memberi informasi mengenai perbandingan hasil klasifikasi oleh model dengan hasil klasifikasi sebenarnya (Han et al., 2011). Hasil performa model kemudian diukur dengan menghitung nilai akurasi, presisi, recall, dan F-1 score. Nilai presisi (*precision score*) adalah rasio setiap kelas sentimen prediksi yang tepat dari seluruh kelas sentimen yang diprediksi, sehingga menggambarkan keakuratan data yang diminta dengan hasil prediksi yang diberikan oleh model. *Recall score* adalah rasio prediksi benar untuk suatu kelas dibandingkan dengan keseluruhan data benar untuk kelas tersebut, sehingga menggambarkan keberhasilan model dalam menemukan kembali sebuah informasi. Nilai akurasi adalah rasio prediksi benar dengan keseluruhan data, sehingga menggambarkan seberapa akurat model dalam mengklasifikasikan data dengan benar. Parameter tunggal ukuran keberhasilan retrieval yang menggabungkan recall dan precision disebut dengan F-1 score atau F-measure.

HASIL DAN PEMBAHASAN

Data yang dihasilkan melalui proses scraping adalah sebanyak 10.056 ulasan. Data ini dipilah dan diberi label secara manual dan kemudian tersisa menjadi 7.636 data untuk diproses dalam tahap *preprocessing*. Sebanyak 6.504 data teridentifikasi dalam kategori sentimen positif dan sebanyak 1.132 data teridentifikasi dalam kategori sentimen negatif. Pada tahap *preprocessing* bagian *data cleaning*, ditemukan 37 data duplikat dari 7.636 data pada dataset.

Masalah ini diatasi dengan menghapus data ulasan yang duplikat sehingga menyisakan satu data tunggal. Pada tahap *language filtering* ditemukan pula sebanyak 70 sentimen ditulis dalam bahasa asing yang kemudian dihapus dalam dataset.

Hasil visualisasi kata yang sering muncul dan ditulis oleh pengguna produk AVS dalam data ulasan bersentimen positif dan negatif dapat divisualisasikan dalam *word cloud* seperti yang ditunjukkan pada Gambar 2. Kata “kulit”, “jerawat”, dan “tekstur” adalah kata-kata yang menjadi topik pembicaraan dan mendominasi pada bahasan positif yang sering ditulis. Sedangkan, kata “tidak”, “pakai”, dan “cocok” menempati urutan tertinggi yang ditulis oleh pengguna ketika menunjukkan rasa kekecewaan dan ketidakpuasan dalam menggunakan produk AVS.



Gambar 2. *Word Cloud* Kata yang Sering Muncul pada Ulasan Positif dan Negatif

Berdasarkan kata-kata yang sering muncul dapat diperoleh hubungan antarkata dalam data ulasan positif dan negatif. Hubungan atau asosiasi antarkata ini berguna untuk memperoleh informasi dari kata-kata yang sering muncul yang telah disajikan dalam *word cloud*. Hubungan antarkata pada data dihitung dengan nilai korelasi. Nilai korelasi yang semakin besar menandakan kemungkinan kedua kata muncul bersama-sama semakin besar.

Tabel 1. Asosiasi Kata pada Ulasan Positif

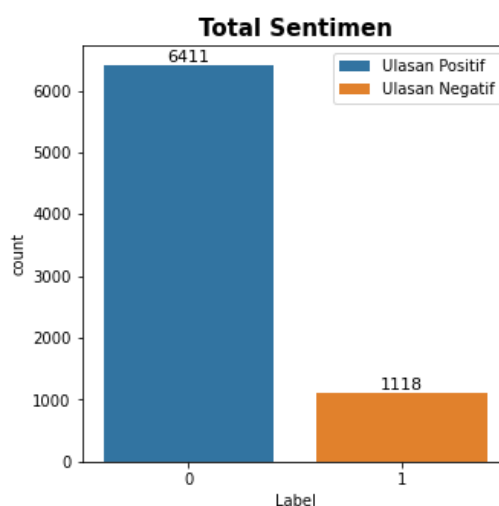
Kulit		Jerawat		Tekstur	
Kandungan	0.28	Pudar	0.29	Kental	0.36
Kering	0.24	Kempis	0.26	Resap	0.22
Sensitif	0.23	Hilang	0.22	Cair	0.21
Hidrasi	0.23	Mendem	0.22	Lengket	0.19
Antioksidan	0.18	Mateng	0.19	Mudah	0.18
Cerah		Bekas		Suka	
Niacinamide	0.16	Pudar	0.51	Wangi	0.08
Kusan	0.14	Hilang	0.25	Essence	0.07
Alpha	0.13	Hitam	0.22	Packaging	0.06
Arbutin	0.13	Arbutin	0.21	Tekstur	0.06
Grapeseed	0.13	Noda	0.13	Segar	0.06

Berdasarkan hasil analisis sentimen terhadap produk AVS, diperoleh beberapa hal yang menjadi keunggulan dari produk AVS yang dapat dipertahankan maupun ditingkatkan oleh AVS. Keunggulan tersebut terdapat pada kandungan produk, tekstur, aroma, dan kemasan yang disukai oleh pengguna. Kandungan produk AVS dinilai dapat menghidrasi kulit kering kering dan sensitif, memudahkan dan menghilangkan bekas jerawat, serta mencerahkan dan menyegarkan kulit wajah. Hal lain yang disukai pengguna adalah tekstur produk jenis essence dan serum juga dinilai mudah meresap ke dalam kulit. Selain itu, aroma produk yang tidak mengganggu dan kemasan yang sederhana juga menjadi nilai positif di mata pengguna.

Tabel 2. Asosiasi Kata pada Ulasan Negatif

Tidak		Pakai		Cocok	
Jenis	0.29	Purging	0.30	Tidak	0.25
Riwayat	0.27	Cekat-	0.30	Sedih	0.22
Khawatir	0.26	cekit		Kandungan	0.19
Bahan	0.25	Eksfoliasi	0.30	PHA	0.18
Cocok	0.25	Iritasi	0.30	Retinol	0.17
		Kering	0.30		
Kulit		Jerawat		Produk	
Sensitif	0.43	Bekas	0.41	<i>Pump</i>	0.24
Normal	0.43	Muncul	0.40	Sunscreen	0.20
Jenis	0.42	Radang	0.26	Klaim	0.19
Khawatir	0.29	Sakit	0.16	Kandungan	0.18
Rentan	0.29	Nanah	0.16	Ukurannya	0.18

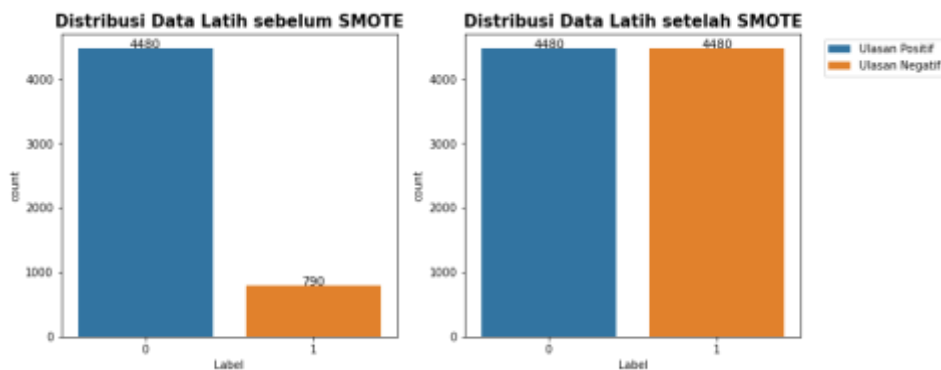
Alasan produk AVS memiliki ulasan negatif yang dinilai oleh pengguna adalah dari segi faktor kandungan, kemasan, dan ukuran produk. Hal ini dapat menjadi bahan masukan dan perbaikan bagi AVS untuk pengembangan produk selanjutnya. Adapun pemecahan masalah yang dapat dilakukan adalah dengan menyesuaikan kandungan produk AVS agar dapat digunakan dengan nyaman bagi pengguna kulit sensitif tanpa menimbulkan iritasi. Selain itu, kandungan yang membentuk tekstur pada produk *sunscreen* juga perlu disesuaikan agar mudah diaplikasikan pada kulit. Penggunaan kemasan produk AVS berbentuk botol *pump* juga perlu ditinjau lebih lanjut karena menjadi masalah bagi sebagian pengguna. Pengguna merasa jika isi produk susah dikeluarkan jika ingin digunakan dan pompa pun sering macet. Hal yang perlu ditinjau selanjutnya adalah sikap keberatan sebagian pengguna terhadap ukuran produk AVS yang kecil namun dijual dengan harga yang kurang terjangkau.



Gambar 3. Jumlah Distribusi Ulasan berdasarkan Label Setelah Preprocessing

Data awal yang berjumlah 7.636 berkurang menjadi 7.529 dikarenakan penghapusan data duplikat dan penghapusan ulasan yang teridentifikasi ditulis dalam bahasa asing. Sebanyak 7.529 data ini adalah data final yang kemudian digunakan untuk proses klasifikasi. Sebanyak 6.411 data teridentifikasi dalam kategori sentimen positif dan sebanyak 1.118 data teridentifikasi dalam kategori sentimen negatif. Berdasarkan pelabelan ulasan, dapat ditunjukkan bahwa 85% ulasan terhadap produk perawatan wajah milik merek AVS memiliki sentimen positif dan hanya 15% ulasan bersentimen negatif. Artinya, konsumen telah merasa cukup puas dalam menggunakan produk perawatan wajah merek AVS.

Jumlah distribusi data ulasan untuk data berlabel 0 dan data berlabel 1 tidak seimbang. Proses klasifikasi menggunakan data latih dengan label tidak seimbang akan menghasilkan prediksi yang bias antara data berlabel yang berjumlah lebih banyak atau label mayoritas dan data berlabel yang berjumlah lebih sedikit atau label minoritas. Sistem akan sulit memprediksi data berlabel negatif karena jumlah data yang sedikit dan lebih fokus untuk mempelajari data berlabel positif yang berjumlah lebih banyak. SMOTE atau *Synthetic Minority Over-sampling Technique* diaplikasikan untuk menyeimbangkan data latih berlabel minoritas dan mayoritas. Berbeda dengan teknik *oversampling* yang dilakukan dengan menduplikasi data, SMOTE bekerja dengan cara membuat data sintetis atau data baru yang dibangkitkan dari label minoritas berdasarkan ketetanggaan data atau *k-nearest neighbour*.



Gambar 4. Distribusi Data Latih berdasarkan Label Sebelum dan Setelah SMOTE

Hasil dari pengujian terhadap pengaruh algoritma optimasi PSO, rasio dataset, dan model *word embedding* disajikan dalam tabel berikut.

Tabel 3. Hasil Pengujian Model Klasifikasi

Skenario	Rasio Data	Word Embedding	Accuracy	Recall	Precision	F1-Score
CNN	70:30	Glove	0,73	0,69	0,61	0,65
		W2V-Skipgram	0,78	0,67	0,62	0,64
		W2V-CBOW	0,61	0,69	0,60	0,64
	75:25	Glove	0,69	0,61	0,70	0,66
		W2V-Skipgram	0,74	0,68	0,61	0,64
		W2V-CBOW	0,70	0,73	0,62	0,67
	80:20	Glove	0,75	0,60	0,67	0,63
		W2V-Skipgram	0,83	0,63	0,63	0,63
		W2V-CBOW	0,72	0,68	0,60	0,64
PSO-CNN	70:30	Glove	0,78	0,58	0,64	0,61
		W2V-Skipgram	0,84	0,56	0,61	0,58
		W2V-CBOW	0,85	0,54	0,67	0,60

Skenario	Rasio Data	Word Embedding	Accuracy	Recall	Precision	F1-Score
	75:25	Glove	0,82	0,59	0,57	0,58
		W2V-Skipgram	0,85	0,55	0,66	0,60
		W2V-CBOW	0,78	0,60	0,59	0,59
	80:20	Glove	0,78	0,59	0,66	0,62
		W2V-Skipgram	0,86	0,71	0,71	0,71
		W2V-CBOW	0,77	0,59	0,57	0,58

Berdasarkan pengujian yang telah dilakukan terhadap klasifikasi ulasan produk perawatan wajah, diperoleh hasil akurasi sebagai nilai perbandingan untuk klasifikasi dengan *Convolutional Neural Network* tanpa seleksi fitur dan klasifikasi dengan *Convolutional Neural Network* berbasis seleksi fitur *Particle Swarm Optimization*. Model klasifikasi CNN mampu memberikan nilai akurasi dengan rentang sebesar 0,61 hingga 0,83 dari sembilan percobaan. Model ini bekerja dengan sangat baik pada dataset dengan rasio 80:20 dan *word embedding* Word2Vec Skipgram. Sedangkan, pada model PSO-CNN, nilai akurasi dari sembilan percobaan memiliki rentang sebesar 0,77 hingga 0,86.

Penggunaan Word2Vec Skipgram juga cocok digunakan pada model klasifikasi PSO-CNN. Berdasarkan pernyataan Mikolov (2013), pengembang *word embedding* Word2Vec, tidak seperti CBOW di mana setiap kata bersaing menjadi kata target terhadap suatu konteks kalimat, pada Skipgram, satu per satu kata menjadi kata input dan memprediksi konteks yang muncul di sekitar kata input. Hal ini menjadi kelebihan Skipgram yang dapat mempelajari kata yang jarang muncul pada dataset meskipun menghabiskan waktu *training* lebih lama. Sementara itu, penambahan rasio data latih tidak selalu meningkatkan performa klasifikasi meskipun rasio data latih terbesar yaitu 80% memberikan hasil terbaik pada model CNN dan PSO-CNN.

Penggunaan algoritma optimasi seleksi fitur PSO memiliki pengaruh dalam proses klasifikasi ulasan positif dan negatif. Performa model klasifikasi dan seleksi fitur PSO-CNN lebih baik dibandingkan model CNN tanpa seleksi fitur. Hal ini ditunjukkan melalui nilai akurasi pada model PSO-CNN pada Tabel 3 yang secara keseluruhan memiliki akurasi yang lebih tinggi dibandingkan dengan model klasifikasi CNN tanpa seleksi fitur. Hasil akurasi tertinggi diperoleh dari model PSO-CNN dengan data latih 80% dan *word embedding* Skipgram yaitu sebesar 86%. Rata-rata peningkatan akurasi adalah sebesar 8,7% untuk sembilan percobaan dengan percobaan untuk tiga rasio dataset yang berbeda dan tiga model *word embedding* yang berbeda. Peningkatan akurasi disebabkan oleh algoritma PSO melakukan seleksi fitur di mana fitur atau kata-kata yang redundan dan kurang berperan dalam proses klasifikasi tidak akan digunakan dalam proses klasifikasi.

Hasil penelitian ini sejalan dengan penelitian yang dilakukan oleh Kaur, et.al (2019) yang menyatakan bahwa seleksi fitur PSO mampu meningkatkan akurasi klasifikasi sentimen dengan CNN sebesar 8,99%. Penelitian oleh Kaur, et.al (2019) berhasil mengklasifikasikan sentimen positif, negatif dan netral data sentimen pengguna *e-commerce* berbahasa Inggris dengan akurasi 98,32% setelah penambahan seleksi fitur PSO tetapi dengan hanya melibatkan empat langkah *text preprocessing* saja yaitu *case folding*, *punctuation removal*, *stopword*

removal, dan tokenisasi.

Meskipun akurasi yang dihasilkan meningkat, pemrosesan model PSO-CNN membutuhkan waktu yang lebih lama dibanding model CNN tanpa seleksi fitur karena meningkatnya proses komputasi yang harus dilakukan dalam memilih fitur-fitur penting. Selain itu, peningkatan waktu komputasi juga terjadi ketika rasio data latih semakin bertambah. Proses *training* dengan *word embedding* Word2Vec Skipgram bekerja dengan sangat baik pada model klasifikasi CNN maupun PSO-CNN, tetapi juga membutuhkan waktu komputasi yang lebih lama karena Skipgram dapat bekerja lebih baik untuk memprediksi kata yang jarang muncul pada dataset. Kekurangan lain pada penelitian ini adalah nilai *precision*, *recall*, dan F1-Score yang dihasilkan pada model secara keseluruhan masih rendah dibandingkan nilai akurasi yang dapat disebabkan karena distribusi label pada data uji yang tidak seimbang dan akurasi hanya berfokus kepada hasil prediksi yang benar.

Pada penelitian ini dapat diketahui pula bahwa algoritma optimasi PSO memilih untuk menggunakan 44 hingga 60 fitur atau kata-kata penting dari 100 fitur untuk setiap data ulasan. Pemilihan subset fitur dengan algoritma PSO mengakibatkan waktu proses komputasi yang lebih lama karena membutuhkan lebih banyak iterasi dalam memilih fitur-fitur penting. Waktu komputasi yang dibutuhkan ketika menggunakan model CNN membutuhkan waktu 1 jam 10 menit hingga 2 jam 5 menit, sedangkan ketika menggunakan model PSO-CNN membutuhkan waktu 3 jam 31 menit hingga 4 jam 20 menit.

Tabel 4. Hasil Confusion Matrix

Aktual	Prediksi	
	Negatif	Positif
Negatif	102	106
Positif	105	1193

Tabel 4 adalah *confusion matrix* untuk model klasifikasi PSO-CNN yang dilatih pada 80% dataset dengan *word embedding* Word2Vec Skipgram yang menghasilkan akurasi 0,86. Berdasarkan tabel dapat diketahui bahwa sebanyak 1.193 data ulasan berlabel positif berhasil diklasifikasi dengan benar ke dalam label positif oleh model (*true positive*). Sebanyak 102 data ulasan berlabel negatif juga berhasil diklasifikasi dengan benar ke dalam label negatif oleh model (*true negative*). Namun, sebanyak 106 data berlabel negatif diklasifikasi sebagai data berlabel positif oleh model (*false positive*). Hal yang serupa terjadi pada 105 data berlabel positif yang diklasifikasi sebagai data berlabel negatif oleh model (*false negative*). Melalui *confusion matrix* pula dapat dihitung nilai akurasi, recall, precision dan F1-Score secara manual.

SIMPULAN

Proses analisis sentimen terdiri atas beberapa langkah diantaranya: pengumpulan data dengan scraping, preprocessing, word embedding, seleksi fitur, klasifikasi, dan evaluasi kinerja model. Sentimen positif dan negatif yang ditulis oleh pengguna produk AVS memberikan informasi bahwa produk AVS disukai oleh pengguna karena kandungan, tekstur, dan kemasannya. Namun, ketiga hal ini juga masih perlu ditingkatkan karena ketiga hal ini beserta ukuran produk juga menjadi alasan pengguna memberi ulasan negatif pada produk AVS.

Model klasifikasi CNN memperoleh akurasi sebesar 61% hingga 83%. Sedangkan pengujian menggunakan PSO-CNN memilih 44-60 fitur atau kata-kata penting dari total 100 fitur, sehingga diperoleh nilai akurasi 77% hingga 86%. Pengujian analisis sentimen menggunakan algoritma Convolutional Neural Network berbasis seleksi fitur Particle Swarm Optimization dapat meningkatkan performa algoritma klasifikasi Convolutional Neural Network tanpa seleksi fitur. Pemilihan model word embedding Word2Vec Skipgram mampu

mengungguli word embedding Glove dan Word2Vec CBOW. Pemilihan rasio data latih dan data uji senilai 80:20 memberikan hasil akurasi yang paling optimal di antara keseluruhan uji coba yaitu 86% untuk model PSO-CNN. Rasio ini juga dapat direkomendasikan untuk model klasifikasi CNN tanpa seleksi fitur.

UCAPAN TERIMA KASIH

Terimakasih kepada koordinator Prodi Maatematika dan seluruh Dosen Prodi Matematika yang telah memberikan ilmu dan bimbingan hingga terselesaikannya artikel ini.

DAFTAR PUSTAKA

- Afnandika, A. (2018). Analisis Sentimen Teks Bahasa Indonesia Pada Media Sosial Menggunakan Convolutional Neural Network (Studi Kasus: Operator Telekomunikasi).
- Chatterjee, S., Chakrabarti, K., Garain, A., Schwenker, F., & Sarkar, R. (2021). JUMRv1: A Sentiment Analysis Dataset for Movie Recommendation. *Appl. Sci.* 2021, 11(20), 9381; <https://doi.org/10.3390/app11209381>.
- Compas. (2021). *Top 5 Produk AVS Terlaris di Shopee Bulan Juli 2021*. Jakarta: Kompas.
- Kaur, K., Kaur, C., & Bhatia, T. K. (2019). An optimized CNN based robust sentiment analysis system on big social data using text polarity feature. *International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-6, April 2019*, 1871-1877.
- Kunnath, S. N., Nikhil, A., & Poroli, H. (2019). Comparing the Wrapper Feature Selection Evaluators on Twitter Sentiment Classification. *Second International Conference on Computational Intelligence in Data Science (ICCIDS-2019)*.
- Lin, S.-W., Ying, K.-C., Chen, S.-C., & Lee, Z.-J. (2008). Particle swarm optimization for parameter determination and feature selection of support vector machines. *Expert Systems with Applications* 35, 1817-1824.
- Media Indonesia. (2021, November 9). *Trend Pengguna Kosmetik Meningkat, Indonesia Siap Ambil Peluang*. Retrieved from Media Indonesia: <https://mediaindonesia.com/ekonomi/409325/trend-pengguna-kosmetik-meningkat-indonesia-siap-ambil-peluang>
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space. *arXiv preprint*, arXiv:1301.3781.
- Muraina, I. O. (2022). Ideal Dataset Splitting Ratios in Machine Learning Algorithms: General Concerns for Data Scientists and Data Analysts. *7th International Mardin Artuklu Scientific Researchers Conference* (pp. 496-504). Mardin: Iksad Institute
- Pennington, J., Socher, R., & Manning, C. D. (2014). GloVe: Global Vectors for Word Representation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1532-1543.
- Raviya, & Vennila, D. M. (2020). A Hybrid Deep Learning Approach for Sentiment Analysis using CNN and Improved SVM with Multi Objective Swarm Optimization for Domain Independent Datasets. *International Journal of Advanced Trends in Computer Science and Engineering Volume 9, No.3, May-June2020*, 3200-3206.
- Sasongko, T. B. (2016). Komparasi dan Analisis Kinerja Model Algoritma SVM dan PSO-SVM (Studi Kasus Klasifikasi Jalur Minat SMA). *Jurnal Teknik Informatika dan Sistem Informasi*, 244-253.
- Wang, P., Xu, B., Xu, J., Tian, G., Liu, C.-L., & Hao, H. (2015). Semantic Expansion using Word Embedding Clustering and Convolutional Neural Network for Improving Short Text Classification. *Neurocomputing*, 806-814.

- Xue, B., Zhang, M., & Browne, W. N. (2013). Particle Swarm Optimization for Feature Selection in Classification: A Multi-Objective Approach. *IEEE TRANSACTIONS ON CYBERNETICS, VOL. 43, NO. 6, DECEMBER 2013*, 1656-1671.
- Zahran, B. M., & Kanaan, G. (2009). Text Feature Selection using Particle Swarm Optimization Algorithm. *World Applied Sciences Journal 7 (Special Issue of Computer & IT)*, 69-74.
- Zhang, Y., & Wallace, B. (2016). A Sensitivity Analysis of (and Practitioners' Guide to) Convolutional Neural Networks for Sentence Classification.