

Model *Library Support Vector Machine* (LibSVM) untuk *Sentiment Review* Penilaian Pesisir Pantai

Oman Somantri¹, Ratih HafSarah Maharrani², Santi Purwaningrum³

^{1,2,3}Jurusan Teknik Informatika Politeknik Negeri Cilacap, Jln. Dr. Soetomo No. 01 Sidakaya Karangcengis Cilacap 53212 INDONESIA, (tel.: 0282-537 992; email: ¹oman_mantri@yahoo.com, ²ratih.hafsarah@pnc.ac.id, ³santi.purwaningrum@pnc.ac.id)

[Diterima: 2 Januari 2023, Revisi: 27 Maret 2023]

Corresponding Author: Oman Somantri

INTISARI — Peningkatan pelayanan sebagai upaya untuk memberikan kenyamanan sebuah tempat wisata, khususnya pesisir pantai selatan pulau Jawa, bagi para pengunjung merupakan sebuah tuntutan bagi para pengelola wisata yang akan memberikan dampak positif di masa depan. Penilaian yang dilakukan untuk mengetahui respons pengunjung wisata pantai mengenai tempat tersebut, memberikan kesan positif atau tidak, menjadi kesulitan tertentu bagi pihak terkait, baik pemerintah maupun pengelola, untuk dapat meningkatkan pelayanan wisata, khususnya pesisir pantai di wilayah selatan pulau Jawa. Penerapan teknologi *text mining* berbasis *machine learning*, khususnya sebuah *sentiment review*, menjadi salah satu solusi yang diusulkan untuk mengatasi permasalahan tersebut, sehingga prediksi potensi wisata dapat diketahui sebelumnya. Pada makalah ini, diusulkan sebuah model *sentiment review* pesisir pantai dengan menggunakan metode *library support vector machine* (LibSVM). Proses optimalisasi model mengusulkan sebuah optimasi model yang berbasis *feature weights* menggunakan algoritma *particle swarm optimization* (PSO) sebagai optimasi model untuk peningkatan akurasi. Upaya peningkatan akurasi pada model yang diusulkan merupakan kontribusi utama pada makalah ini. Hasil penelitian dan eksperimen terhadap model yang diusulkan menghasilkan model terbaik yang diberi nama LibSVM_IG+PSO dengan menggunakan metode LibSVM berbasis *information gain* (IG) dan PSO, yang menghasilkan tingkat akurasi sebesar 88,97%. Model yang diusulkan ini diharapkan dapat menjadi pendukung keputusan dalam menilai sentimen terhadap pariwisata maritim pesisir pantai yang dapat dimanfaatkan oleh wisatawan, pemerintah, maupun pengelola wisata.

KATA KUNCI — *Sentiment Review*, Wisata, Pesisir Pantai, LibSVM, *Feature Weights*.

I. PENDAHULUAN

Potensi wisata pesisir pantai dan laut merupakan primadona bagi wisatawan, khususnya dari luar negeri ketika berkunjung ke Indonesia. Akan tetapi, wisatawan lokal yang datang untuk menikmati keindahannya juga tidak kalah banyak. Pengembangan potensi sektor ini menjadi salah satu fokus utama pemerintah dan seluruh dunia untuk dapat meningkatkan kunjungan wisatawan, terlebih setelah masa pandemi COVID-19 berangsur membaik sehingga membawa sedikit perubahan pada sektor wisata di bidang maritim ini. Pesisir pantai wilayah selatan pulau Jawa saat ini banyak diminati oleh para pengunjung, sehingga diperlukan upaya dalam peningkatan pengelolaannya agar menjadi lebih baik dan banyak pengunjungnya.

Sentiment review merupakan sebuah pengembangan teknologi dalam bidang pengolahan *text mining* yang berorientasi pada penilaian sentimen seseorang terhadap sebuah produk, tempat, berita, kebijakan, maupun lainnya yang sifatnya menimbulkan penilaian sentimen terhadap objek yang dinilai, yang kebanyakan diambil dari media sosial [1]. Penerapan sebuah *machine learning* pada teknologi *sentiment review* ini sangatlah diperlukan karena proses analisis yang dilakukan memerlukan sebuah kecerdasan tertentu. Beberapa algoritma yang sering dipergunakan pada *sentiment review*, khususnya untuk klasifikasi sentimen, adalah *decision tree*, *neural network*, *naïve Bayes*, *k-nearest neighbor* (k-NN), dan *support vector machine* (SVM).

Terdapat beberapa penelitian dan studi yang telah melakukan *sentiment review* terhadap potensi tempat wisata, baik yang ada di Indonesia maupun di luar negeri, dengan menggunakan algoritma *machine learning* dalam prosesnya. Sebuah penelitian dilakukan untuk mengetahui dan mendeteksi titik objek wisata menggunakan profil Instagram berdasarkan penilaian wisatawan [2]. Penelitian lain melakukan analisis

sentiment review pariwisata menggunakan pendekatan metode *latent Dirichlet allocation* (LDA) [3]. Penelitian lainnya menerapkan metode *semantic clustering* pada sentimen sistem rekomendasi tempat wisata [4], sedikit berbeda dengan penelitian mengenai prediksi permintaan wisata menggunakan data *review online* [5]. Beberapa peneliti melakukan analisis *sentiment review* pariwisata dengan menggunakan situs web seperti TripAdvisor [6]–[8]. Di samping itu, selain tempat wisata, *sentiment review* yang berhubungan dengan penunjang wisata, yaitu ulasan (*review*) hotel, juga dilakukan banyak peneliti [9]–[11].

Beberapa peneliti telah melakukan penelitian mengenai penilaian pesisir pantai. Penelitian dilakukan terhadap ulasan penilaian pantai dengan menggunakan Google Map dan media lainnya [12]–[14]. Pada [12] diusulkan metode *long-short term memory* (LSTM) dan *word2Vec* untuk *sentiment analysis* pada ulasan objek wisata pantai. Penelitian tersebut menghasilkan model dengan tingkat akurasi rata-rata sebesar 84%, presisi 76%, dan *recall* 0,73%. Penelitian tersebut cukup baik, tetapi belum menghasilkan tingkat akurasi yang sangat tinggi, sehingga memerlukan optimasi model. Penelitian selanjutnya melakukan analisis *sentiment review* terhadap berbagai tempat wisata di kota Garut, yang juga mencakup objek wisata pantai di dalamnya [13]. Penelitian ini memperlihatkan bahwa 80% orang melakukan ulasan dengan komentar positif. Pada penelitian ini hanya dilakukan proses analisis tanpa menggunakan metode *machine learning*. Dilakukan *fiber classification* sehingga nilai kinerja (*performance*) model yang didapatkan, seperti tingkat akurasi, presisi, dan *recall*, tidak diketahui.

Referensi [14] mengusulkan metode *random forest classifier* dan *Bernoulli naïve Bayes* untuk *sentiment review* tempat wisata di Phuket. Hasil penelitian menghasilkan nilai kinerja *receiver operating characteristic* (ROC) *area under the*

ROC curve (AUC) menggunakan *random forest* terbaik adalah 0,89 dan ROC AUC menggunakan *Bernoulli naïve Bayes* sebesar 0,90. Namun, penelitian ini belum memperlihatkan nilai kinerja tingkat akurasi yang dihasilkan.

Sentiment analysis dengan menggunakan SVM dalam bidang wisata juga telah dilakukan [15]. Penelitian tersebut mengusulkan sebuah model untuk *sentiment analysis* yang dilakukan di Bangkok, Chiang Mai, dan Phuket. Model yang diusulkan menggunakan metode *decision tree*, *random forest*, dan SVM. Penelitian ini menghasilkan nilai kinerja akurasi sebesar 93,2% ketika menggunakan algoritma CART, 77,4% ketika menggunakan SVM, dan 95,54% ketika menggunakan *random forest*. Terlihat bahwa model SVM masih memiliki tingkat akurasi yang belum optimal, sehingga diperlukan cara untuk optimalisasi model SVM.

Berdasarkan fungsinya, pada dasarnya SVM merupakan sebuah *machine learning* yang hanya dapat mengklasifikasikan dua kelas saja. Akan tetapi, setelah dikembangkan, metode ini dapat juga digunakan untuk model yang mengklasifikasikan lebih dari dua kelas atau *multiclass*. *Library support vector machine* (LibSVM) merupakan sebuah metode yang dapat digunakan untuk melakukan klasifikasi dengan menggunakan parameter SVM yang telah dikembangkan [16], [17].

Particle swarm optimization (PSO) adalah salah satu algoritma optimasi untuk pengambilan keputusan berdasarkan populasi (*swarm*) yang mengeksplorasi individu (*particle*) dalam pencariannya. Secara garis besar, algoritma ini mengoptimasi permasalahan dengan cara menggerakkan partikel atau calon solusi yang dipengaruhi oleh solusi terbaik partikel tersebut, yang secara umum didapatkan dari partikel lain di dalam ruang permasalahan menggunakan fungsi tertentu untuk posisi dan kecepatan dari partikel. Tujuan utama pada penelitian dalam makalah ini adalah mencari model optimasi PSO pada algoritma SVM terbaik yang menghasilkan tingkat akurasi klasifikasi *sentiment review* pesisir pantai wilayah selatan pulau Jawa, khususnya pantai di wilayah Cilacap.

Meskipun tingkat akurasi yang dihasilkan pada model telah diteliti pada beberapa penelitian sebelumnya, tetapi masih terdapat permasalahan pada tingkat akurasi yang dihasilkan, sehingga perlu dilakukan upaya peningkatan tingkat akurasi. Berbeda dengan beberapa penelitian sebelumnya, proses pemilihan bobot terbaik pada model LibSVM yang diusulkan pada makalah ini merupakan tujuan utama dan kontribusi utama penelitian ini. Peningkatan akurasi pada model LibSVM dilakukan dengan menggunakan *feature weights* algoritma PSO. PSO merupakan salah satu algoritma optimasi yang dapat digunakan untuk optimasi pemilihan bobot terbaik.

II. METODOLOGI

A. DATASET

Dataset yang dipergunakan pada penelitian ini adalah hasil ulasan dari setiap orang yang telah mengunjungi pesisir pantai selatan pulau Jawa, yang diambil dari *website* <https://www.google.com/maps> dari tahun 2018 sampai 2022. Data penelitian adalah teks berbahasa Indonesia. Proses pelabelan data ditentukan oleh penilaian *rating* bintang yang diberikan oleh wisatawan.

Pada proses pelabelan data untuk menentukan klasifikasi sentimen, data teks yang telah didapatkan diklasifikasi menjadi dua label, yaitu kategori “positif” dan “negatif”. Proses penentuan label data dikategorikan berdasarkan hasil penilaian *sentiment review* yang telah dilakukan oleh orang yang

memberikan ulasan terhadap pantai tersebut, berdasarkan pengalaman orang tersebut, dilihat dari keadaan pantai, pelayanan tempat wisata, maupun penilaian lainnya. Penilaian ulasan tersebut tidak dibatasi pada batasan tertentu, sehingga data teks sentimen diambil berdasar pada semua aspek penilaian dari sisi penilai. Label data ditentukan berdasarkan kategori penilaian pemberian tanda bintang pada ulasan. *Rating* bintang 1 dan 2 dikategorikan pada label “negatif”, sedangkan bintang 4 dan 5 dikategorikan pada label “positif”. *Rating* bintang 3 tidak dimasukkan pada penelitian ini karena dibatasi dua label saja.

Dataset diambil dari ulasan penilaian orang pada pantai selatan Teluk Penyu wilayah Kabupaten Cilacap, Jawa Tengah, Indonesia sebanyak 390 data ulasan, yang terdiri atas 250 data berlabel “positif” dan 140 data berlabel “negatif”.

B. METODE YANG DIUSULKAN

Tahapan penelitian dilakukan dengan metode eksperimen. Eksperimen dilakukan kepada setiap model yang dianggap model terbaik, yang pada pelaksanaannya ditetapkan terlebih dahulu nilai parameter-parameter ada. Tahapan utama yang dilakukan adalah prapemrosesan (*pre-processing*) data, implementasi model, validasi data, optimasi model, dan terakhir adalah evaluasi model.

Setiap tahapan penelitian dilakukan berdasarkan uji coba yang dilakukan berkali-kali dalam mendapatkan model yang diinginkan. Tahapan prapemrosesan melaksanakan proses tokenisasi, penapisan (*filtering*), dan pembersihan (*cleansing*) data. Tahapan selanjutnya adalah pembobotan data dan variabel dengan menggunakan metode *term frequency - inverse document frequency* (TF-IDF) [18]. Tahapan yang dilakukan pada penelitian ini diperlihatkan pada Gambar 1.

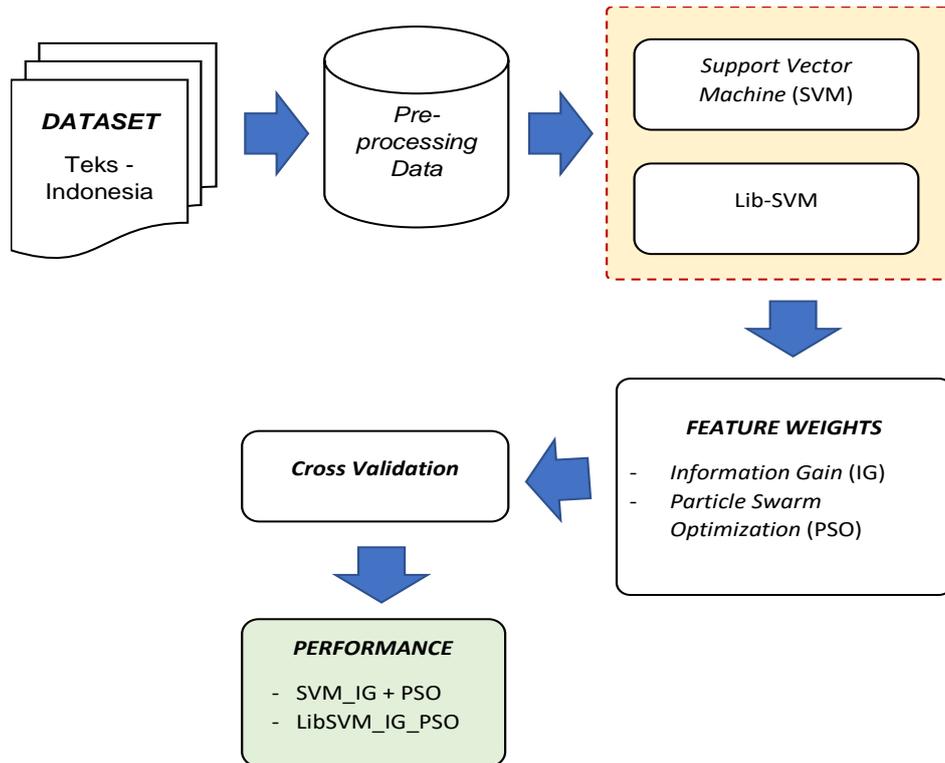
Tahapan validasi data yang dilakukan menggunakan metode *cross validation*. Metode ini dianggap sebagai salah satu metode terbaik dalam memperoleh tingkat akurasi kinerja model yang dihasilkan. Penentuan jumlah *fold* yang digunakan dilakukan sesuai kebutuhan, dengan mengedepankan cara memperoleh kinerja akurasi model terbaik. Pada tahapan ini, dilakukan pembagian data antara data latih dan data uji, yaitu ditentukan 90% data latih dan 10% data uji. Proses validasi model dilakukan berdasarkan jumlah *fold* yang ditentukan pada setiap model yang diujicobakan, sehingga model yang diusulkan memiliki tingkat akurasi yang berbeda-beda, sesuai dengan parameter *fold* masing-masing model.

Evaluasi model merupakan bagian dari tahapan proses penelitian sebagai penilaian indikator tingkat keberhasilan dari model yang dihasilkan. Pada proses evaluasi model, dilakukan komparasi keseluruhan model terbaik yang didapatkan, sehingga terlihat perbedaannya. Model-model yang dibandingkan adalah model yang menerapkan algoritma SVM berbasis *information gain* (IG) dan PSO (SVM_IG+PSO) dengan model yang menggunakan algoritma LibSVM berbasis IG dan PSO (LibSVM_IG+PSO). Analisis data yang digunakan untuk pencarian model dilakukan menggunakan perangkat lunak Rapidminer Studio.

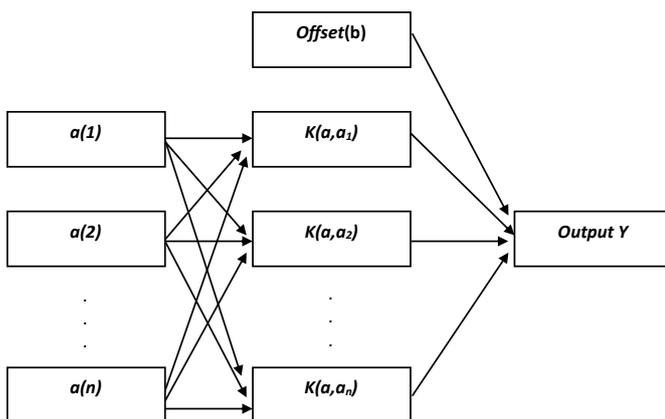
Rumus *confusion matrix* [19] untuk mengetahui kinerja akurasi model ditunjukkan pada (1).

$$Accuracy = \frac{\sum TP + TN}{\sum TP + TN + FP + FN} \quad (1)$$

dengan *TP* = *true positive*, *TN* = *true negative*, *FP* = *false positive*, dan *FN* = *false negative*.



Gambar 1. Framework dan metode yang diusulkan.



Gambar 2. Arsitektur LibSVM

C. LIBRARY SUPPORT VECTOR MACHINE (LIBSVM)

LibSVM merupakan sebuah *library* yang digunakan untuk klasifikasi *support vector*, regresi, dan estimasi distribusi, yang berbentuk sebuah perangkat lunak terintegrasi. Klasifikasi *vector* yang dimaksud adalah *c-support vector classification* (C-SVC) dan nu-SVC; *epsilon-support vector regression* (epsilon-SVR) dan nu-SVR untuk regresi; dan estimasi distribusi berupa *one-class SVM* [20]. LibSVM mempunyai fitur yang memungkinkan penerapannya mudah dilakukan. LibSVM merupakan pengembangan dari SVM; LibSVM melengkapi SVM asli dengan penambahan beberapa parameter. LibSVM saat ini digunakan untuk menyelesaikan masalah dua *class* dengan membangun *hyperlane* dan membedakan antara positif dan negatif sebanyak mungkin [17]. Gambar 2 menunjukkan sebuah arsitektur LibSVM.

Untuk memperoleh nilai kinerja model yang diusulkan, pada LibSVM diterapkan beberapa *kernel function*. *Kernel function* merupakan metode yang digunakan dalam proses pengambilan data sebagai masukan dan mengubahnya menjadi bentuk pemrosesan data yang diperlukan. Pada SVM, *kernel*

function secara umum berfungsi untuk mengubah kumpulan data latih sehingga permukaan keputusan nonlinier dapat diubah menjadi persamaan linier dalam jumlah ruang dimensi yang paling tinggi.

Beberapa *kernel function* yang digunakan pada penelitian ini yaitu kernel linier pada (2), *kernel* polinomial pada (3), *radial-basis function* (rbf) pada (4), dan *kernel sigmoid* pada (5).

$$K(x_i, x_j) = x_i^T x_j \quad (2)$$

$$K(x_i, x_j) = (1 + x_i^T x_j)^p \quad (3)$$

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \quad (4)$$

$$K(x_i, x_j) = \tanh(\beta_0 x_i^T x_j + \beta_1) \quad (5)$$

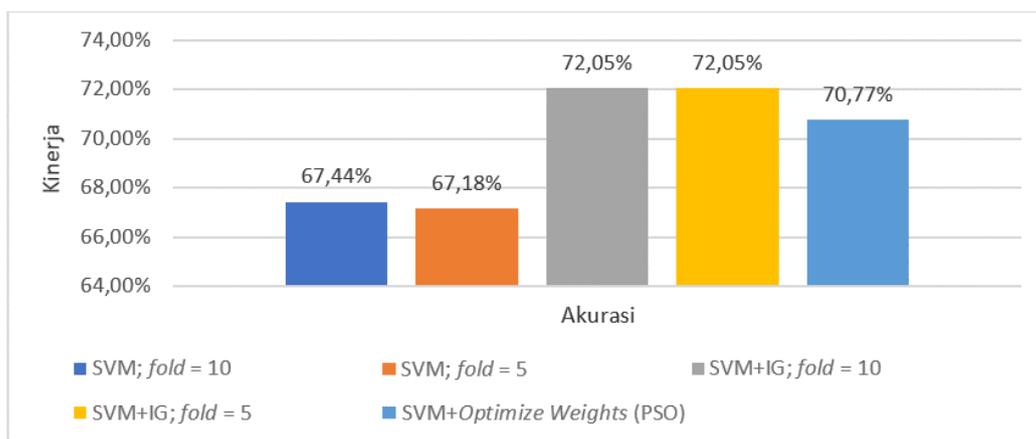
dengan K adalah nomor dari *class*, $K(x_i, x_j)$ adalah *kernel* pada elemen x_i, x_j , β_0 adalah pengaruh pada klasifikasi *out-comes*, dan p adalah derajat yang mengatur fleksibilitas *classifier*.

III. HASIL DAN PEMBAHASAN

A. MODEL SUPPORT VECTOR MACHINE (SVM)

Proses pencarian model terbaik pada tahapan eksperimen menggunakan SVM pada penelitian ini menghasilkan beberapa kombinasi tingkat akurasi yang berbeda. Hasil eksperimen menggunakan SVM diperlihatkan pada Gambar 3. Pengujian dilakukan dengan menggunakan kombinasi parameter *5-fold*, dan *10-fold*, yang kemudian memberikan sebuah nilai akurasi yang berbeda satu sama lain.

Pada pengujian model dengan menerapkan SVM menggunakan *5-fold*, diperoleh hasil akurasi sebesar 67,18%. Pada eksperimen ini, hasil yang didapatkan masih kurang begitu baik dan belum sesuai dengan harapan. Model yang didapatkan pada percobaan ini masih memerlukan upaya peningkatan akurasi dan upaya perbaikan, baik pada nilai



Gambar 3. Komparasi tingkat akurasi model menggunakan SVM.

TABEL I
 HASIL PENGUJIAN METODE LIBSVM DENGAN OPTIMASI INFORMATION GAIN (IG)

Tipe SVM	Tipe Kernel	Stratified Sampling	Shuffled Sampling	Linear Sampling
C-SVC	rbf	64,10%	64,10%	64,10%
C-SVC	polinomial	67,69%	67,95%	67,44%
C-SVC	linear	69,49%	73,08%	69,49%
C-SVC	sigmoid	73,33%	73,08%	69,49%
nu-SVC	rbf	64,36%	69,23%	54,87%
nu-SVC	polinomial	65,13%	67,95%	85,38%
nu-SVC	linear	73,33%	72,82%	65,90%
nu-SVC	sigmoid	67,95%	65,90%	35,13%
one-class	rbf	47,18%	44,87%	23,85%
one-class	polinomial	52,05%	54,36%	49,49%
one-class	linear	47,18%	44,87%	23,85%
one-class	sigmoid	47,18%	44,62%	54,87%

parameter yang digunakan maupun pada prapemrosesan data. Proses yang dilakukan selanjutnya adalah percobaan dengan menerapkan parameter yang berbeda. Berdasarkan hasil eksperimen, yang diperlihatkan pada Gambar 3, diperoleh sebuah model SVM dengan nilai akurasi terbaik sebesar 67,44%. Model dengan tingkat akurasi tertinggi didapatkan menggunakan SVM, dengan parameter yang ditetapkan adalah tipe *kernel radial* dan dilakukan dengan proses *shuffled sampling*. Akurasi yang didapatkan merupakan yang terbaik dari model yang diusulkan pada rentang nilai parameter yang sudah ditetapkan. Akurasi yang didapatkan pada percobaan yang masih dirasa kecil, sehingga dilakukan kembali eksperimen dengan optimasi.

Setiap parameter tipe *kernel* yang digunakan berdasarkan hasil eksperimen memiliki karakteristik yang berbeda, sesuai dengan *dataset* yang dipergunakan. Hal ini menyebabkan hasil kinerja akan sama dan/atau berbeda, bergantung pada tipe *kernel* yang digunakan untuk *dataset* yang berbeda-beda. Beberapa parameter tipe *kernel* yang digunakan memperlihatkan nilai kinerja model yang sedikit berbeda setiap *kernel*-nya, yang dapat diakibatkan oleh karakteristik data teks yang digunakan ketika dilakukan prapemrosesan data.

Proses uji coba berikutnya adalah tahapan proses eksperimen dengan menggunakan SVM. Pada tahap ini, dilakukan optimasi dengan menerapkan *feature weights*, algoritma PSO. Pada hasil eksperimen ini terjadi perbedaan nilai akurasi jika dibandingkan dengan eksperimen tanpa optimasi PSO. Hasil eksperimen penerapan PSO pada SVM menghasilkan akurasi yang kurang begitu baik, yaitu sebesar

70,77%, sehingga masih diperlukan upaya peningkatan akurasi. Pada eksperimen SVM+PSO ini, parameter yang diterapkan disesuaikan dengan kebutuhan, dengan harapan diperoleh tingkat akurasi berbeda dan terbaik. Tingkat akurasi terbaik pada model ini tercapai menggunakan parameter 10-*fold* dengan tipe *kernel radial* dan proses *shuffled sampling*. Algoritma PSO pada model saat proses optimasi menggunakan parameter populasi 5 dan 10, tetapi akurasi terbaik diperoleh ketika digunakan populasi = 5. Kinerja model terendah dihasilkan dengan menggunakan 5-*fold* dengan tipe *kernel dot* dan populasi = 5, menghasilkan akurasi sebesar 68,97%. Optimasi model SVM dengan menggunakan PSO masih memiliki tingkat akurasi kecil, sehingga memerlukan model lain yang dapat diterapkan. Perbedaan tingkat akurasi pada model-model SVM+PSO ini cukup signifikan satu sama lainnya karena perbedaan nilai parameter populasi dan *fold* yang digunakan saat validasi model.

Upaya peningkatan akurasi selanjutnya adalah melalui proses optimasi model dengan menerapkan metode IG pada SVM. Upaya penerapan IG diharapkan dapat memberikan nilai akurasi yang lebih baik dibandingkan dengan metode PSO yang sudah diterapkan sebelumnya yang masih memiliki kinerja rendah. Berdasarkan hasil uji coba dengan penerapan IG pada SVM, diperoleh sebuah perubahan yang cukup signifikan. Hal ini terlihat dari peningkatan akurasi yang dihasilkan. Model SVM_IG memperlihatkan hasil akurasi yang sedikit lebih baik, yaitu di atas 70%. Terlihat bahwa kinerja model terendah yang dihasilkan pada eksperimen adalah sebesar 70,26%, dengan menggunakan parameter tipe *kernel*

TABEL II
HASIL EKSPERIMEN MODEL LIBSVM_IG BERBASIS PSO

Tipe SVM	Tipe Kernel	Populasi	Sampling	Akurasi
nu-SVC	polinomial	5	linear	88,97%
nu-SVC	polinomial	10	linear	88,72%
nu-SVC	polinomial	15	linear	88,97%
C-SVC	<i>sigmoid</i>	5	<i>stratified</i>	75,64%
C-SVC	<i>sigmoid</i>	10	<i>stratified</i>	74,10%
C-SVC	<i>sigmoid</i>	15	<i>stratified</i>	75,13%

polinomial pada SVM dan metode *stratified sampling* pada proses validasi model. Meskipun sudah cukup baik, pada uji coba model SVM_IG ini dihasilkan tingkat akurasi terendah dari keseluruhan metode *sampling* adalah sebesar 58,72%, sehingga model tersebut masih jauh dari yang diinginkan. Model tersebut menggunakan parameter tipe *kernel* polinomial serta menggunakan metode *linear sampling*.

Setelah dilakukan proses uji coba, kinerja model dengan akurasi tertinggi untuk optimasi menggunakan SVM_IG adalah 72,05%. Kinerja tertinggi model diperoleh menggunakan 10-fold dan 5-fold *cross validation*, serta parameter tipe *kernel dot* pada SVM.

B. HASIL MODEL LIBSVM

Tahapan eksperimen selanjutnya adalah menerapkan model LibSVM. Model LibSVM ini digunakan untuk dapat mengklasifikasikan *sentiment review* penilaian pesisir pantai. Hasil eksperimen menunjukkan beberapa kombinasi tingkat akurasi yang dapat digunakan. Namun, tingkat akurasi terbaik yang dihasilkan menjadi prioritas utama pada model yang diusulkan dalam makalah ini. Hasil eksperimen model dengan menerapkan LibSVM memperlihatkan hasil tingkat akurasi yang sangat beragam. Uji coba model ini menggunakan beberapa tipe SVM, di antaranya tiga tipe jenis SVM untuk klasifikasi data, yaitu C-SVC, nu-SVC, dan *one-class*. C-SVC merupakan jenis SVM yang menggunakan parameter C sebagai parameter penalti dari *error term*, sedangkan nu-SVC hampir mirip dengan C-SVC tetapi menggunakan parameter *nu* yang berfungsi untuk mengontrol jumlah *support vector* dan kesalahan pelatihan. *One-class SVM* merupakan tipe SVM yang dipergunakan untuk estimasi distribusi, yang merupakan jenis *unsupervised learning* untuk mempelajari kemampuan membedakan sampel uji kelas tertentu dari kelas lain.

Selain itu, tipe *kernel* yang diterapkan pada eksperimen model LibSVM tahapan ini adalah *rbf*, polinomial, linear, dan *sigmoid*. Berdasarkan hasil yang diperoleh, diketahui bahwa kinerja model masih di bawah 80%, sehingga masih memerlukan optimasi.

Parameter *fold* yang digunakan pada model LibSVM ditentukan bernilai 10-fold. Model LibSVM yang menghasilkan akurasi tertinggi adalah dengan model dengan parameter tipe SVM nu-SVC, tipe *kernel sigmoid*, dan menggunakan metode *stratified sampling*. Model ini menghasilkan kinerja akurasi terbaik sebesar 77,69%. Berdasarkan hasil uji coba, akurasi tertinggi kebanyakan dihasilkan oleh model dengan *kernel* tipe *sigmoid* dan linear menggunakan SCVM tipe C-SVC dan nu-SVC.

Berdasarkan hasil uji coba model yang diperoleh, terlihat bahwa tingkat akurasi tertinggi adalah sebesar 77,69%. Model dengan akurasi tertinggi ini menggunakan metode LibSVM dengan parameter SVM bertipe nu-SVC dan *kernel* bertipe *sigmoid*. Model LibSVM ini memiliki tingkat akurasi yang belum cukup baik, terlihat dari rata-rata tingkat akurasi yang

TABEL III
KOMPARASI KINERJA MODEL

No	Model	Akurasi
1	LibSVM	77,69%
2	LIBSVM+PSO	80,77%
3	LibSVM_IG	85,38%
4	LibSVM_IG+PSO (yang diusulkan)	88,97%

TABEL IV
NILAI AKURASI BERDASARKAN *CONFUSION MATRIX*

	<i>True Negative</i>	<i>True Positive</i>
<i>Prediksi Negative</i>	102	5
<i>Prediksi Positive</i>	38	245

dihasilkan, yaitu 34% hingga 70%, sehingga diperlukan upaya peningkatan akurasi kembali. Model LibSVM dengan tipe SVM *one-class* dan tipe *kernel* polinomial menghasilkan tingkat akurasi antara 34,62% sampai dengan 35,64%.

C. PENERAPAN LIBSVM DAN INFORMATION GAIN (IG)

Optimalisasi berikutnya adalah penerapan IG pada model LibSVM, yang dalam artikel ini diberi nama LibSVM_IG. Hasil eksperimen menunjukkan terjadinya peningkatan akurasi model, sehingga model yang dihasilkan lebih baik daripada sebelumnya. Hasil penerapan IG pada LibSVM ditunjukkan pada Tabel I. Upaya lain yang dilakukan adalah dengan menetapkan parameter-parameter IG dengan nilai berbeda-beda, sehingga dapat diketahui bahwa model yang didapatkan merupakan model terbaik dengan tingkat akurasi yang tinggi.

Hasil yang diperoleh pada model LibSVM_IG sedikit lebih baik dibandingkan dengan tipe klasik. Hal ini terlihat dari akurasi yang dihasilkan, yang lebih tinggi daripada percobaan sebelumnya, tanpa adanya optimasi IG. Tingkat akurasi tertinggi yang diperoleh dari model LibSVM_IG adalah sebesar 85,38%, dengan menggunakan *kernel* tipe polinomial dan SVM tipe nu-SVC. Tingkat akurasi model terendah adalah sebesar 23,85%, dengan menggunakan model SVM tipe *one-class* dan *kernel* tipe linear. Model terbaik pada percobaan ini dihasilkan menggunakan parameter SVM tipe nu_SVC dan *kernel* tipe polinomial dan didapatkan melalui metode *linear sampling*. Dari hasil yang ditampilkan pada Tabel I, terlihat bahwa model yang diusulkan sudah memiliki kenaikan tingkat akurasi yang cukup signifikan. Akan tetapi, hal tersebut dipengaruhi oleh tipe SVM yang digunakan serta metode *sampling* yang diterapkan.

D. MODEL LIBSVM_IG & FEATURE WEIGHTS PSO

Tahapan akhir dari eksperimen yang dilakukan adalah mencari model dengan menggunakan penerapan *feature weights*, yaitu dengan menerapkan algoritma PSO. Model yang diusulkan ini adalah optimasi dari model LibSVM yang telah ada, yang ditingkatkan kembali dengan menerapkan PSO. Pada makalah ini, model ini diberi nama LibSVM+PSO. Hasil eksperimen yang diperoleh pada model yang diusulkan

TABEL V
 PERBANDINGAN NILAI KINERJA MODEL YANG DIUSULKAN DENGAN PENELITIAN SEBELUMNYA

No.	Penulis	Objek Penelitian	Metode	Kinerja (Akurasi/AUC)
1.	N. Oktaviana, H.C. Rustamaji, dan H. Sofyan [12].	Objek wisata pantai	Long-short term memory (LSTM) + word2vec	Akurasi rata-rata 84%, presisi 76%, dan recall 0,73%
2.	D.T. Alamanda, A. Ramdhani, I. Kania, W. Susilawati, dan E.S. Hadi [13].	Tempat wisata di kota Garut	-	88% komentar positif
3.	Mathayomchan dan K. Sripanidkulchai, [14].	Tempat wisata Phuket	Random forests classifier dan Bernoulli naive Bayes	<ul style="list-style-type: none"> ROC AUC random forest = 0,89 ROC AUC Bernoulli naive Bayes = 0,90
4.	N. Leelawat dkk., [15].	Bangkok, Chiang Mai, dan Phuket	Decision tree, random forest, dan support vector machines	<ul style="list-style-type: none"> CART = 0,932% SVM = 77,4% Random forest = 95,4%
5	Metode yang diusulkan	Pantai Teluk Penyu, Cilacap, Indonesia	Library support vector machine (LibSVM) + IG + PSO	Akurasi 88,97%

diperlihatkan pada Tabel II. Selain itu, model yang dioptimasi oleh PSO adalah model yang sebelumnya telah menerapkan IG sebagai optimasi, sehingga pada tahapan proses pencarian model merupakan optimasi model LibSVM berbasis IG, yang diberi nama LibSVM_IG+PSO.

Berdasarkan hasil eksperimen sebelumnya, model yang telah didapatkan ini menggunakan dua tipe SVM, yaitu nu-SVC dan C-SVC, dengan dua tipe kernel, yaitu kernel sigmoid dan polinomial. Pembatasan tipe SVM dan kernel ini dilakukan untuk membatasi model yang sudah baik dan kemudian dioptimalkan, sehingga tidak memerlukan lagi uji coba tambahan untuk mencari model yang belum tentu baik akurasi. Jangkauan (range) nilai parameter populasi pada PSO yang diatur dan digunakan pada model ini ditetapkan antara 5 sampai 15.

Proses uji coba yang dilakukan pada model LibSVM dengan menerapkan PSO tanpa optimasi IG memperlihatkan hasil yang cukup baik. Tingkat akurasi tertinggi yang dihasilkan 80,77%, sedangkan yang terendah adalah 78,46%. Model LibSVM+PSO dihasilkan dengan menggunakan SVM tipe nu-SVC dan kernel tipe sigmoid dengan populasi = 10. Model terbaik yang dihasilkan membutuhkan waktu sekitar 6,15 menit untuk dapat menghasilkan model yang diharapkan dengan menggunakan aplikasi yang sudah ditetapkan, yaitu RapidMiner Studio.

Model yang dihasilkan menggunakan LibSVM+PSO memperoleh akurasi sebesar 80,77%. Diperlukan waktu 6,15 menit untuk mendapatkan model ini, yang lebih lama dibandingkan dengan model lainnya. Namun, hal ini sudah cukup baik. Waktu yang dibutuhkan model lainnya adalah 2,16 menit, dengan akurasi sebesar 78,46%. Selain itu, terdapat model yang dihasilkan dalam waktu 4,13 menit, tetapi akurasi sebesar 79,23%. Perbedaan waktu yang diperlukan dalam eksperimen pencarian model ini dipengaruhi oleh tipe SVM yang digunakan dan jumlah parameter populasi yang digunakan. Selain itu, kapasitas memori perangkat keras komputer yang digunakan untuk eksperimen sangat berpengaruh.

Optimasi model dengan menerapkan PSO pada LibSVM_IG menghasilkan nilai akurasi seperti disajikan pada Tabel II. Terlihat bahwa hasil dari model yang diusulkan memiliki tingkat akurasi yang lebih baik dan lebih tinggi daripada model tanpa dilakukan optimasi sebelumnya. Tingkat akurasi terbaik LibSVM_IG berbasis PSO adalah 88,97%, yang lebih baik dibandingkan model sebelumnya. Tipe SVM

yang diterapkan pada model terbaik ini adalah nu-SVC, dengan menggunakan kernel tipe polinomial dan populasi = 5 serta populasi = 15.

E. EVALUASI MODEL DAN REKOMENDASI

Pada tahapan berikutnya, dilakukan evaluasi model dengan melakukan komparasi hasil eksperimen untuk memberikan gambaran yang lebih baik pada model yang diusulkan. Hasil perbandingan dari seluruh model yang telah didapatkan dengan berbagai tingkat akurasi diperlihatkan pada Tabel III.

Proses evaluasi model dilakukan menggunakan dataset yang sama, dengan persentase data latih sebesar 90% dan data uji 10%. Pengujian model dilakukan menggunakan cross validation dengan parameter yang telah ditetapkan sesuai dengan percobaan yang dilakukan serta mendapatkan nilai kinerja akurasi terbaik setiap modelnya.

Berdasarkan Tabel III, terlihat bahwa model yang diusulkan, yaitu LibSVM_IG+PSO, merupakan model dengan tingkat akurasi terbaik dan tertinggi, yaitu sebesar 88,97%. Salah satu pengaruh perubahan nilai akurasi yang dihasilkan ini adalah ketika dilakukan penentuan nilai dan jenis parameter yang diterapkan dari model yang diusulkan, setiap parameter mempunyai karakteristik yang berbeda dan ini memberikan pengaruh yang signifikan terhadap tingkat akurasi klasifikasi yang dihasilkan. Berdasarkan Tabel III, confusion matrix nilai akurasi, yang dihasilkan menggunakan (1), ditunjukkan pada Tabel IV.

$$\begin{aligned}
 Accuracy &= \frac{\sum TP + TN}{\sum TP + TN + FP + FN} \\
 &= \frac{\sum 245 + 102}{\sum 245 + 102 + 38 + 5} \\
 &= 347/390 \\
 &= 0,8897.
 \end{aligned}$$

Tingkat akurasi yang dihasilkan saat ini masih dapat dioptimasi. Untuk itu, diperlukan sebuah upaya yang keras, terlebih dalam menentukan parameter yang ditentukan pada model tersebut, karena melibatkan intuisi. Diperlukan model yang tepat untuk dapat menentukan secara otomatis parameter yang digunakan agar hasil yang diharapkan tercapai. Penerapan metode optimasi lain dapat dipertimbangkan, terutama dalam optimasi pemilihan nilai parameter yang digunakan.

Tabel III menunjukkan peningkatan akurasi model sentiment review penilaian pesisir pantai dengan menggunakan model SVM dan LibSVM. Terlihat bahwa nilai akurasi model

yang dihasilkan dan diusulkan secara berkelanjutan mengalami kenaikan terus-menerus dan ini memperlihatkan bahwa optimasi yang diterapkan pada model dapat berpengaruh terhadap kenaikan tingkat akurasi yang dihasilkan. Apabila dibandingkan dengan penelitian-penelitian sebelumnya, model yang diusulkan memiliki tingkat akurasi cukup baik, seperti diperlihatkan pada Tabel V.

IV. KESIMPULAN

Penilaian *sentiment review* pesisir pantai telah mendapatkan model terbaik yang dapat dijadikan sebagai pendukung keputusan dalam memprediksi sentimen penilaian pesisir pantai. Penelitian ini mengusulkan model optimasi LibSVM melalui optimalisasi dengan menggunakan PSO sebagai model untuk sentimen. Model yang diusulkan memiliki keunggulan berupa tingkat akurasi yang lebih tinggi dibandingkan dengan model klasik LibSVM sebelumnya. Model terbaik yang didapatkan adalah model yang diberi nama LibSVM_IG+PSO, dengan tingkat akurasi sebesar 88,97%. Model ini menggunakan metode LibSVM dan IG berbasis PSO.

Kekurangan pada penelitian ini adalah masih belum optimalnya tingkat akurasi yang dihasilkan. Salah satu penyebabnya adalah prapemrosesan data, sehingga perlu dilakukan eksperimen lebih jauh untuk dapat menemukan model prapemrosesan yang lebih tepat agar nilai akurasi model yang dihasilkan dapat meningkat. Penelitian selanjutnya memerlukan upaya peningkatan akurasi yang lebih baik agar diperoleh akurasi terbaik dan menggunakan metode alternatif lainnya yang dapat diterapkan serta pengembangan kembali untuk mencari model lain yang tepat dengan tingkat akurasi lebih baik. Metode *simulate annealing* (SA) dan *genetic algorithm* (GA) dapat diusulkan untuk penelitian selanjutnya dan tidak menutup kemungkinan akan menghasilkan nilai kinerja yang berbeda dari PSO.

KONFLIK KEPENTINGAN

Penulis menyatakan bahwa tidak terdapat konflik kepentingan dalam penulisan artikel maupun hasil penelitian yang didapatkan.

KONTRIBUSI PENULIS

Konseptualisasi, Oman Somantri dan Santi Purwaningrum; metodologi, Oman Somantri; analisis model, Oman Somantri; validasi, Oman Somantri, Santi Purwaningrum, dan Ratih HafSarah Maharrani; analisis formal, Oman Somantri; validasi model, Oman Somantri, Santi Purwaningrum; kurasi data, Santi Purwaningrum, Ratih HafSarah Maharrani; penulisan—penyusunan draf asli, Oman Somantri; penulisan—peninjauan dan penyuntingan, Oman Somantri, Ratih HafSarah Maharrani; visualisasi, Oman Somantri.

UCAPAN TERIMA KASIH

Artikel ini merupakan bagian dari luaran hasil penelitian pendanaan hibah penelitian skema Penelitian Dosen Pemula (PDP) dari Direktorat Akademik Pendidikan Vokasi Dirjen Pendidikan Vokasi Kemendikbudristek tahun anggaran 2022 dengan nomor kontrak penelitian 117/SPK/D4/PPK.01.APTV/VI/2022.

REFERENSI

[1] Y. Tao, W. Liu, Z. Huang, dan C. Shi, "Thematic Analysis of Reviews

on the Air Quality of Tourist Destinations from a Sentiment Analysis Perspective," *Tour. Manag. Perspect.*, Vol. 42, hal. 1–14, Apr. 2022, doi: 10.1016/j.tmp.2022.100969.

- [2] K.D. Mukhina, S.V. Rakitin, dan A.A. Visheratin, "Detection of Tourists Attraction Points Using Instagram Profiles," *Procedia Comput. Sci.*, Vol. 108, hal. 2378–2382, 2017, doi: 10.1016/j.procs.2017.05.131.
- [3] T. Ali, B. Omar, dan K. Soulaïmane, "Analyzing Tourism Reviews Using an LDA Topic-Based Sentiment Analysis Approach," *MethodsX*, Vol. 9, hal. 1–10, Nov. 2022, doi: 10.1016/j.mex.2022.101894.
- [4] Z. Abbasi-Moud, H. Vahdat-Nejad, dan J. Sadri, "Tourism Recommendation System Based on Semantic Clustering and Sentiment Analysis," *Expert Syst. Appl.*, Vol. 167, hal. 1–10, Apr. 2021, doi: 10.1016/j.eswa.2020.114324.
- [5] M. Hu dkk., "Tourism Demand Forecasting Using Tourist-Generated Online Review Data," *Tour. Manag.*, Vol. 90, hal. 1–19, Jun. 2022, doi: 10.1016/j.tourman.2022.104490.
- [6] V. Taecharungroj dan B. Mathayomchan, "Analysing TripAdvisor reviews of tourist attractions in Phuket, Thailand," *Tour. Manag.*, Vol. 75, hal. 550–568, Des. 2019, doi: 10.1016/j.tourman.2019.06.020.
- [7] F. Pollák, P. Dorčák, N. Račeta, dan N. Svetožarovová, "Sustainable E-Marketing of Selected Tourism Subjects from the Mediterranean through Active Online Reputation Management," in *Smart City 360°*, A.L.-Gracia dkk., Eds., Cham, Swiss: Springer, 2016, hal. 692–703, doi: 10.1007/978-3-319-33681-7_60.
- [8] T. Ali dkk., "Exploring Destination's Negative E-Reputation Using Aspect Based Sentiment Analysis Approach: Case of Marrakech Destination on TripAdvisor," *Tour. Manag. Perspect.*, Vol. 40, hal. 1–15, Okt. 2021, doi: 10.1016/j.tmp.2021.100892.
- [9] D. Apriliani dkk., "Sentiment Analysis for Indonesia Hotel Services Review Using Optimized Neural Network," *J. Phys. Conf. Ser.*, Vol. 1538, No. 1, hal. 1–8, Mei 2020, doi: 10.1088/1742-6596/1538/1/012060.
- [10] P.J. Lee, Y.H. Hu, dan K.T. Lu, "Assessing the Helpfulness of Online Hotel Reviews: A Classification-Based Approach," *Telemat., Inform.*, Vol. 35, No. 2, hal. 436–445, Mei 2018, doi: 10.1016/j.tele.2018.01.001.
- [11] R.P. Nawangsari, R. Kusumaningrum, dan A. Wibowo, "Word2Vec for Indonesian Sentiment Analysis towards Hotel Reviews: An Evaluation Study," *Procedia Comput. Sci.*, Vol. 157, hal. 360–366, Okt. 2019, doi: 10.1016/j.procs.2019.08.178.
- [12] N. Oktaviana, H.C. Rustamaji, dan H. Sofyan, "Sentiment Analysis on Reviews of Beach Tourism Objects on Google Maps Using Long-Short Term Memory Method," *Pros. Semin. Nas. Inform.*, 2022, Vol. 1, No. 1, hal. 133–143.
- [13] D.T. Alamanda dkk., "Sentiment Analysis Using Text Mining of Indonesia Tourism Reviews via Social Media," *Int. J. Humanit. Arts. Soc. Sci.*, Vol. 5, No. 2, hal. 72–82, Apr. 2019, doi: 10.20469/ijhss.5.10004-2.
- [14] B. Mathayomchan dan K. Sripanidkulchai, "Utilizing Google Translated Reviews from Google Maps in Sentiment Analysis for Phuket Tourist Attractions," in *2019 16th Int. Joint Conf. Comput. Sci., Softw. Eng. (JCSSE)*, 2019, hal. 260–265, doi: 10.1109/JCSSE.2019.8864150.
- [15] N. Leelawat dkk., "Twitter Data Sentiment Analysis of Tourism in Thailand during the COVID-19 Pandemic Using Machine Learning," *Heliyon*, Vol. 8, No. 10, hal. 1–11, Okt. 2022, doi: 10.1016/j.heliyon.2022.e10894.
- [16] D. Liu, Z. Wang, L. Wang, dan L. Chen, "Multi-Modal Fusion Emotion Recognition Method of Speech Expression Based on Deep Learning," *Front. Neurobot.*, Vol. 15, Jul. 2021, doi: 10.3389/fnbot.2021.697634.
- [17] C.-C. Chang dan C.-J. Lin, "LIBSVM," *ACM Trans. Intell. Syst., Technol.*, Vol. 2, No. 3, hal. 1–27, Apr. 2011, doi: 10.1145/1961189.1961199.
- [18] C.W. Schmidt, "Improving a TF-IDF Weighted Document Vector Embedding," 2019, *arXiv:1902.09875*.
- [19] J. Xu, Y. Zhang, dan D. Miao, "Three-Way Confusion Matrix for Classification: A Measure Driven View," *Inf. Sci.*, Vol. 507, hal. 772–794, Jan. 2020, doi: 10.1016/j.ins.2019.06.064.
- [20] T. Chen dkk., "EEG Emotion Recognition Model Based on the LIBSVM Classifier," *Meas.*, Vol. 164, hal. 1–7, Nov. 2020, doi: 10.1016/j.measurement.2020.108047.