JURNAL NASIONAL TEKNIK ELEKTRO DAN TEKNOLOGI INFORMASI
p-ISSN 2301–4156 | e-ISSN 2460–5719

Library Support Vector Machine (LibSVM) Model for Coastal Assessment Sentiment Review

Oman Somantri[1], Ratih Hafsarah Maharrani[2], Santi Purwaningrum[3]

[1,2,3] Department of Informatics, State Polytechnic of Cilacap, Jln. Dr. Soetomo No. 01 Sidakaya Karangcengis Cilacap 53212 INDONESIA (tel.: 0282-537 992; email: [1]oman_mantri@yahoo.com, [2]ratih.hafsarah@pnc.ac.id, [3]santi.purwaningrum@pnc.ac.id)

[Received: 2 January 2023, Revised: 27 March 2023]
Corresponding Author: Oman Somantri

ABSTRACT — Improving services as an effort to provide the convenience of tourist destinations, especially on the south coast of Java, is a demand placed on tourism managers, which in the long run will yield positive impacts. The assessment is conducted to determine whether the tourism destination give positive impressions to the tourists. The application of machine learning-based text mining technology, especially a sentiment review, is one of the solutions proposed to overcome this problem, therefore predictions of coastal tourism potential can be known beforehand. This research proposed a coastal sentiment review model using the library support vector machine (LibSVM) method. The process proposed a model optimization based on feature weights using the particle swarm optimization (PSO) algorithm as a model optimization to increase the accuracy level. Efforts to improve the accuracy of the proposed model are the main contribution of this research. The results of research and experiments on the proposed model produced the best model named LibSVM_IG+PSO using the information gain (IG). On the other hand, PSO-based LibSVM method generated an accuracy level of 88.97%. The model proposed in this research is expected to serve as a decision support for tourists, government, and tourism managers in assessing sentiment towards the coastal maritime tourism.

KEYWORDS — Sentiment Review, Tourism, Coastal, LibSVM, Feature Weights.

I. INTRODUCTION

The tourism potential of the coastal and marine areas is very promising, particularly for foreign tourists as well as local tourists who come to enjoy the beauty of the destination. The potential development of this sector is one of the main focuses of the government and the whole world in order to increase the number of tourist visits, particularly after the COVID-19 pandemic. It is expected that the maritime tourism sector will undergo a little change as the condition has gradually improved. The southern coast of Java is currently in great demand among tourists, meaning that efforts need to be made to improve its management and attract more tourists.

Sentiment reviews are a technological advancement in the field of text mining processing that is oriented toward assessing one's sentiments about products, places, news, policies, or other things and generates sentiment assessments of the objects, which are mostly taken from social media [1]. The application of machine learning to the sentiment review technology is necessary because the analysis process requires a certain intelligence. Some of the algorithms that are often used in sentiment reviews, especially for sentiment classification, include decision trees, neural networks, naïve Bayes, k-nearest neighbor (K-NN), support vector machines (SVM), and other algorithms.

Several studies have used machine learning algorithms to carry out sentiment reviews of potential tourist attractions in Indonesia and abroad. One of which is research that was conducted to find out and detect tourist attractions using Instagram profiles based on tourist ratings [2]. Meanwhile, another research was conducted to analyze tourism sentiment reviews using the latent Dirichlet allocation (LDA) method approach [3]. Another research applied the semantic clustering method to the sentiment of the tourist recommendation system [4], which slightly differs from that of predicting tourist demand using online review data [5]. Some researchers have carried out a sentiment analysis of tourism reviews by using websites such as TripAdvisor [6]–[8]. In addition, many researchers have performed sentiment reviews related to tourism support, namely hotel reviews [9]–[11].

There have been several researchers who have conduct research on the coastal assessments. Numerous research was conducted to review coastal assessments using Google Maps and other media [12]–[14]. In [12], the long-short term memory (LSTM) and Word2Vec methods were proposed for sentiment analysis in reviews of coastal tourism objects. This research produced a model with an average accuracy, precision, and recall rates of 84%, 76%, and 0.73%, respectively. This research is quite good, but it did not yield a high level of precision, so model optimization was required. In subsequent research, a sentiment review analysis was carried out on various tourist attractions in Garut as it has many beaches [13]. This study showed that 80% of people left reviews with positive comments. However, this research only carried out the analysis process without using the machine learning method. The fiber classification was carried out so that the performance model values obtained, such as the level of accuracy, precision, and recall, were unknown.

Another research proposed the random forest classifier method and Bernoulli naïve Bayes for sentiment reviews of Phuket tourist attractions [14]. The best performance values of the receiver operating characteristic (ROC) area under the ROC curve (AUC) using a random forest were 0.89 and 0.90 when using Bernoulli naïve Bayes. However, the performance values of the resulting accuracy level were not revealed in this research.

Sentiment analysis implementing SVM in the tourism sector has also been performed [15]. This research proposed a model for conducting sentiment analysis in Bangkok, Chiang Mai, and Phuket. The proposed model used the decision tree, random forest, and SVM. The classification and regression tree (CART) algorithm, SVM, and random forest, respectively,

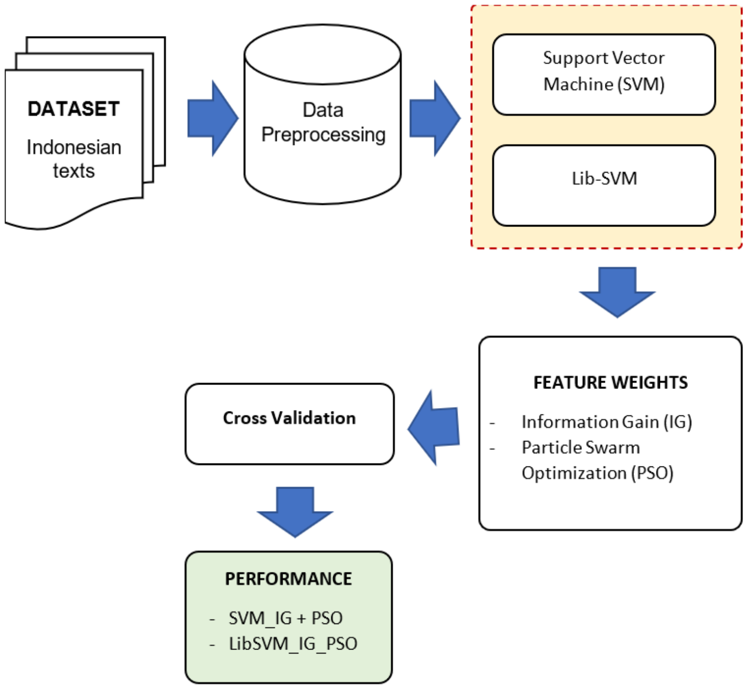

Figure 1. Framework and proposed model.

yielded accuracy performance values of 93.2%, 77.4%, and 95.54%. In this research, the SVM models' accuracy level was not quite optimal, so it was necessary to find a way to optimize the SVM model.

Based on its function, SVM is a machine learning that can only classify two classes; however, this technique, after being developed, can be applied to models that classify more than two classes, or multi classes. Library support vector machine (LibSVM) is a method that can be used for classification using SVM parameters that have been developed by [16], [17].

Particle swarm optimization (PSO) is an optimization algorithm for population-based decision-making (swarm) that exploits individuals (particles) in its search. Broadly speaking, this algorithm optimizes the problem by moving the particle or a candidate solution that is influenced by the best solution of the particle, which is generally obtained from other particles in the problem space using a certain function for the position and velocity of the particle. The main objective of this research is to find a PSO model for the best SVM that produces an accurate level of sentiment review classification accuracy on the coast of the southern region of Java, especially the Cilacap coastal area.

Even though the accuracy level produced by the model has been examined in several previous studies, however, there are still problems with the level of accuracy produced; therefore, efforts must be made to increase the level of accuracy. In contrast to several previous studies, the primary objective and contribution of this study is the process of selecting the optimal weights for the LibSVM model. Improved accuracy of the LibSVM model was carried out by using the feature weights of the PSO algorithm. PSO is one of the optimization algorithms that can be used to optimize the best weights selection.

II. METHOD

A. DATASET

This study's dataset was derived from a review of everyone who visited the southern coast of Java. The reviews were obtained from an open access website, namely, https://www.google.com/maps, between 2018 to 2022. The research data were texts in Indonesian. The data labeling process was determined by the star rating given by tourists.

To determine sentiment classification in the data labeling process, the obtained text data were then classified into two labels, namely the "positive and "negative" categories. The process of determining data labels is categorized based on the results of a sentiment review assessment conducted by the person who reviewed the beach based on the results of that person's experience ranging from the condition of the beach, tourist attractions services, and other assessments. The review was not constrained by any limitations so that the sentiment text data were taken based on all aspects from the tourist's point of view. Data labels were determined based on the stars given when reviewing, where 1 and 2 stars were classified as "negative" labels, while 4 and 5 stars were classified as "positive" labels. Meanwhile, reviews with 3 stars were excluded because they can only have two labels. Those with 3 stars are not included as the label was limited to only two labels.

A total of 390 reviews were obtained from reviews on the south coast of Teluk Penyu, Cilacap Regency, Central Java, Indonesia. These data comprised 250 data labeled as "positive" and 140 data labeled as "negative."

B. PROPOSED METHOD

Th e research stages were carried out using the experimental method. Experiments were carried out on each model that was considered the best model, whose parameter values were determined first. The main steps were data preprocessing, model implementation, data validation, model optimization, and the final stage is model evaluation.

Each research stage was carried out based on experiments that were conducted numerous times to acquire the desired model. The preprocessing stage performed tokenization, filtering, and data cleansing processes. The subsequent stage was data and variable weighting using the TF-IDF method [18]. The stages carried out in this research are shown in Figure 1.

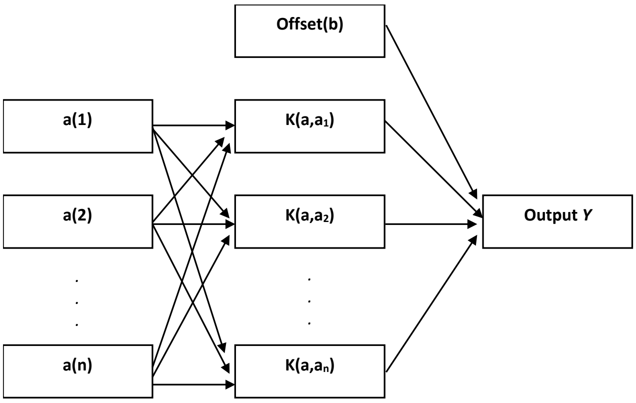

Figure 2. Architecture of LibSVM.

The data validation stage used was the cross-validation method. This method is considered one of the best methods in obtaining the results of the accuracy level of the resulting performance model. Determination of the number of folds used was carried out according to needs by prioritizing model's accuracy level in order to achieve best performance. At this stage, data was divided between training data and testing data: 90% was training data and 10% was testing data. The model validation process was carried out based on the number of folds determined for each model being tested so that the performance values of the proposed models had different levels of accuracy according to the fold parameters of each model.

Model evaluation process was part of the research stages as an assessment of the success rate indicators of the resulting model. This stage involved comparing the best overall models obtained so that the difference could be determined. In this stage, SVM_IG+PSO, a model applying an SVM algorithm based on information gain (IG) and PSO, was compared to LibSVM_IG_PSO, a model using the LibSVM algorithm based on IG and PSO. The data analysis process used for the model search was the RapidMiner Studio software.

Confusion matrix equation [19] to determine the performance of model accuracy as in (1).

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

In (1), *TP* denotes true positive, *TN* denotes true negative, *FP* denotes false positive, and *FN* denotes false negative.

C. LIBRARY SUPPORT VECTOR MACHINE (LIBSVM)

LibSVM is a library used to support vector classification, LibSVM is a library in the form of integrated software that is used to support vector classification, regression, and distribution estimation. The vector classifications include c-support vector classification (C-SVC) and nu-SVC; regressions include epsilon-support vector regression (SVR) and nu_SVR; whereas, the distribution estimation is one-class SVM [20]. LibSVM has features that make its implementation easy to do. It is an extension of SVM, in which LibSVM complements the original SVM parameters with additional parameters. LibSVM is currently used to solve the two-class problem by building hyper lanes and distinguishing between positive and negative as much as possible [17]. Figure 2 depicts the LibSVM architecture.

Several kernel functions were applied to the proposed model in LibSVM to determine its performance value. The kernel function is a method used in the process of receiving data as input and transforming it into the required form of data processing. In SVM, the kernel function is generally used to change the training dataset so that the nonlinear decision surface can be transformed into a linear equation in the highest dimensional spaces.

Some of the kernel functions used in this panel are the linear kernel in (2), the polynomial kernel in (3), the radial basis function (RBF) in (4), and the sigmoid kernel in (5).

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

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

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

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

K is the number of the class, *K(x_i,x_j)* is the kernel on the elements *x_i,x_j*, *β_0* is the influence on the classification of outcomes, *p* is the degree that governs the flexibility of the classifier.

III. RESULTS AND DISCUSSION

A. SUPPORT VECTOR MACHINE MODEL

The process of finding the best model at the experimental stage using an SVM in this research resulted in several combinations of different accuracy levels. The experimental results using SVM are shown in Figure 3. The test was performed utilizing a combination of 5- and 10-fold parameters, which resulted in two distinct accuracy values.

In testing the model by implementing SVM using 5-fold, the results achieved an accuracy of 67.18%. The results obtained in this experiment were not so good and not as expected. The model obtained in this test still requires efforts to improve its accuracy, parameter values, and data preprocessing process. The following process was conducting an experiment utilizing different parameters. Based on the experimental results shown in Figure 3, an SVM model with the best accuracy value of 67.44% was obtained. The model with the highest level of accuracy was obtained using SVM, where the parameters set was radial kernel carried out using shuffled sampling. The accuracy obtained is the best of the proposed model in the range of predefined parameter values. The accuracy obtained in the experiments was still small, so another experiment was conducted in an attempt to optimize the results.

According to the dataset, each kernel type parameter used based on experimental results has different characteristics. It makes the performance results will likely be the same and/or different depending on the kernel type used for different datasets. Several kernel-type parameters exhibited slightly different performance model values; it can be caused by the characteristics of the text data used when preprocessing the data.

The subsequent experiment process was using the SVM. In this stage, the optimization process was conducted by applying feature weights, namely PSO algorithm. There is a difference in the resulting accuracy value compared to that of without PSO optimization. The experimental results of applying PSO to SVM in this experiment resulted in poor accuracy with the resulting accuracy being 70.77%; therefore, efforts to improve accuracy are still required. In this SVM+PSO experiment, the parameters applied were adjusted in order to achieve a different and optimal level of accuracy. The best accuracy level in this model used the 10-fold parameter with radial and shuffled

TABLE I
TESTING RESULTS OF THE LIBSVM METHOD WITH INFORMATION GAIN (IG) OPTIMIZATION

SVM type	Kernel type	Stratified Sampling	Shuffled Sampling	Linear Sampling
C-SVC	RBF	64.10%	64.10%	64.10%
C-SVC	Poly	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	Poly	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	Poly	52.05%	54.36%	49.49%
one-class	Linear	47.18%	44.87%	23.85%
one-class	Sigmoid	47.18%	44.62%	54.87%

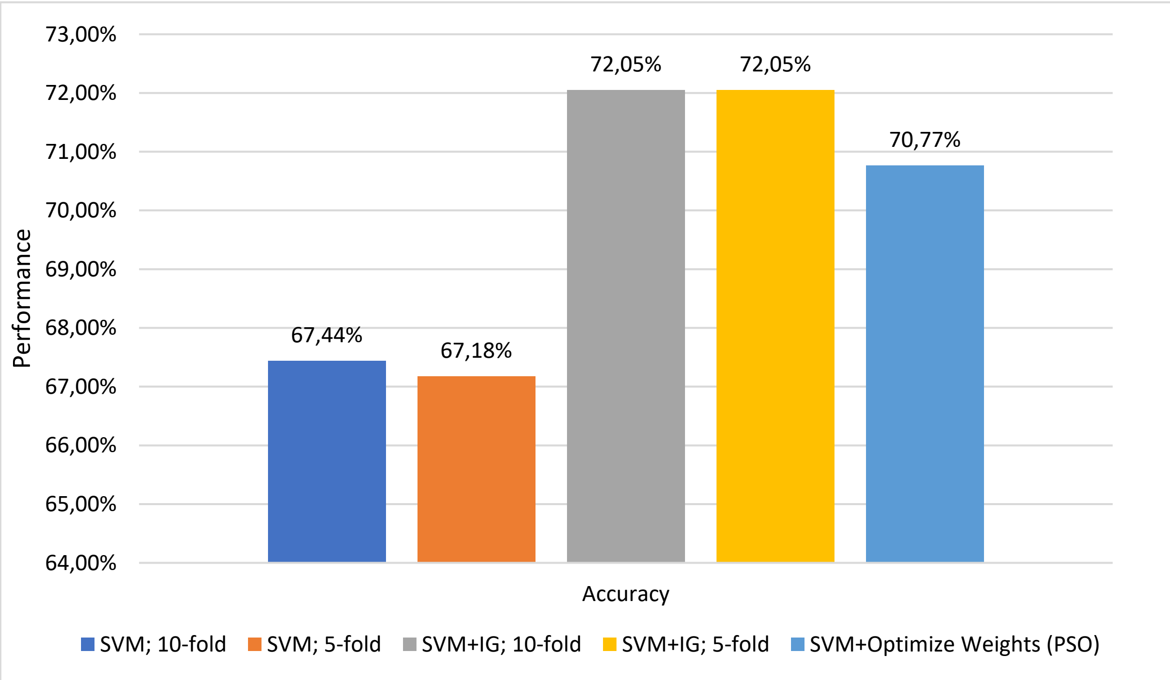

Figure 3. Comparison of model accuracy levels using SVM.

sampling. The PSO algorithm in the model during the optimization process used population parameters 5 and 10, but the best accuracy used population = 5. The lowest performance model was generated by using 5-fold with dot kernel and population = 5 resulting in an accuracy rate of 68.97%. SVM model optimization using PSO still had a small accuracy level so it required another model that could be applied. The accuracy of the SVM+PSO models is significantly different from one another due to differences in the values of the population parameter and fold parameter used during model validation.

The next effort to increase accuracy involved optimizing the model by applying the IG method to the SVM. Efforts to implement IG were expected to result in a higher accuracy value compared to the previously applied PSO method, which still exhibited a poor performance. The application of IG produced a significant change. It is evident from the increased level of accuracy produced. The accuracy level of the SVM_IG model had slightly better results and were above 70%. Using the polynomial kernel parameter in SVM and the stratified sampling method in the model validation process, the lowest model performance produced in the experiment was 70.26%. This result showed a better increase in accuracy during the testing. Despite being deemed quite good, the SVM_IG model trial yielded the lowest accuracy rate among sampling methods, namely 58.72%. Consequently, the model was still far from what is desired. This model employed the polynomial kernel and used the linear sampling method.

After the experiments were carried out, the performance model with the highest accuracy value for optimization using SVM_IG was 72.05%. The highest performance of the model was obtained by using 10- and 5-fold cross validation, and dot kernel parameters in the SVM.

B. LIBSVM MODEL RESULTS

The next experimental stage was to apply LibSVM model. The LibSVM model was used to classify coastal assessment review sentiments. The experimental results showed several combinations of accuracy levels that could be used, but the best accuracy level produced was the top priority in the model proposed. The experimental results of the model by implementing LibSVM yielded very different accuracy levels. In testing this model, several types of SVM were used, including C-SVC, nu-SVC, and one-class SVM. C-SVC is a

TABLE II
LIBSVM_IG MODEL EXPERIMENT RESULTS BASED ON PSO

SVM type	Kernel Type	Population	Sampling	Accuracy
nu-SVC	Poly	5	Linear	88.97%
nu-SVC	Poly	10	Linear	88.72%
nu-SVC	Poly	15	Linear	88.97%
C-SVC	Sigmoid	5	Stratified	75.64%
C-SVC	Sigmoid	10	Stratified	74.10%
C-SVC	Sigmoid	15	Stratified	75.13%

type of SVM that uses parameter C as the penalty parameter for the error term. Nu-SVC is almost similar to C-SVC, with the difference lying in the use of the nu parameter, which controls the number of support vectors and training errors. One-class SVM is a type of SVM used for distribution estimation, which is a form of unsupervised learning to study the ability to distinguish test samples from one class from those of other classes.

Moreover, the kernel type applied to the LibSVM model experiment at this stage used the RBF, polynomial, linear, and sigmoid kernel types. The obtained results demonstrated that the performance model is still below 80%, indicating that it still requires optimization.

The fold parameter used in the LibSVM model is determined by the value of the 10-fold parameter. The LibSVM model that produced the highest accuracy level used the SVM parameter type of nu-SVC, kernel type of sigmoid, and the stratified sampling method, which produced the best accuracy performance of 77.69%. Based on the test results, the model using sigmoid and linear type kernels and C-SVC and nu-SVC type SCVM produced the highest accuracy on average.

It can be seen that the model with the highest accuracy rate was 77.69%, as shown in the result of the model experiments. This high-accuracy model used the LibSVM method with SVM model parameters type of nu-SVC and kernel type of sigmoid. The accuracy level in the classical LibSVM model had a subpar accuracy level. It can be seen that the average level of accuracy produced is 34% - 70%, demonstrating the need to increase accuracy once more. The LibSVM model with one-class SVM and kernel polynomial type yielded a significant result with a very low accuracy rate between 34.62% and 35.64%.

C. IMPLEMENTATION OF LIBSVM & INFORMATION GAIN

The next optimization was to apply IG to the LibSVM model or LibSVM_IG in this research. The experimental results showed an increase in the accuracy level of the model, resulting a better model than the previous one. The results of implementing IG in LibSVM are shown in Table I. Another effort was to set different IG parameters. Hence, it can be seen that the model obtained was the best model with a high accuracy level.

The results obtained in the LibSVM_IG model were slightly better than the classical type. It can be seen that the resulting accuracy was higher than the previous experiments without any IG optimization. The highest level of accuracy obtained from the LibSVM_IG model was 85.38% using the polynomial kernel type and nu-SVC SVM type. The lowest model accuracy rate was 23.85% using the one-class SVM and linear kernel. The best model in this experiment was generated using SVM parameters type of nu_SVC and kernel type of poly and was obtained through the linear sampling method. From Table I, it can be observed that the accuracy level of the proposed model has already increased significantly. However, it is dependent on the SVM type employed and the sampling method.

TABLE III
COMPARATION OF THE PERFORMANCE MODEL

No.	Model	Accuracy
1	LibSVM	77.69%
2	LibSVM+PSO	80.77%
3	LibSVM_IG	85.38%
4	LIBSVM_IG+PSO (proposed model)	88.97%

TABLE IV
ACCURACY VALUE BASED ON THE CONFUSION MATRIX

	True Negative	True Positive
Pred. Negative	102	5
Pred. Positive	38	245

D. LIBSVM_IG & FEATURE WEIGHTS PSO MODEL

The final stage of the experiment was to find a model by the application of feature weights, namely by applying the PSO algorithm. The proposed model was an optimization of the existing LibSVM model and was enhanced by implementing PSO or LibSVM+PSO in this research. The results of the experiments obtained in the proposed model are shown in Table II. In addition, the model optimized by PSO was a model that had previously implemented IG; consequently, at the stage of the model search process, it is an IG-based LibSVM optimization model or referred to as LibSVM_IG+PSO.

Two versions of SVM, namely nu-SVC and C-SVC, as well as two types of kernels, namely sigmoid and polynomial kernels, were utilized in earlier model experiments. Types of SVM and kernel were limited to determine good models and then optimized so that no additional experiments are necessitated to identify models that produced poor accuracy levels. The PSO parameters that were set and used in this model were determined by the range of population parameter values between 5 to 15.

The experiment process carried out on the LibSVM model using PSO without IG optimization showed quite good results; the level of accuracy produced on the LibSVM+PSO model resulted in the highest accuracy of 80.77% and the lowest of 78.46%. The LibSVM+PSO model was generated using nu-SVC SVM type and a sigmoid kernel with population = 10. The best model produced required around 6.15 minutes to produce the expected model using a predetermined application, namely RapidMiner Studio.

The model generated using LibSVM+PSO was 80.77%, which took 6.15 minutes, which was longer than other models; however, this result was quite good. The time required for other models was 2.16 minutes, but only produced an accuracy of

TABLE V
COMPARISON OF MODEL PERFORMANCE VALUE PROPOSED IN PREVIOUS RESEARCH

No.	Authors	Object of Research	Method	Performance (Accuracy/AUC)
1	N. Oktaviana, H.C. Rustamaji, and H. Sofyan [12]	Beach attractions	Long-short term memory (LSTM) + Word2Vec	Accuracy 84%, precision 76%, and recall 0.73%
2	D.T. Alamanda, A. Ramdhani, I. Kania, W. Susilawati, and E.S. Hadi [13]	Tourist attractions in the city of Garut	-	88% positive comments
3	Mathayomchan and K. Sripanidkulchai [14]	Phuket tourist attractions	Random forests classifier and Bernoulli naïve Bayes	• ROC AUC Random forest = 0.89 • ROC AUC Bernoulli naïve Bayes = 0.90
4	N. Leelawat *et al.* [15]	Bangkok, Chiang Mai, and Phuke	Decision tree, random forest, and support vector machines	• CART = 0.932 % • SVM = 77.4% • Random forest = 95.4%
5	Proposed model	Teluk Penyu beach, Cilacap, Indonesia	Library support vector machine (LibSVM) + IG + PSO	88.97%

78.46%. Similarly, the resulting model required less time, with 4.13 minutes, but produced an accuracy of 79.23%. The difference in time needed in the model search experiment was influenced by the type of SVM and the number of population parameters used, as well as the computer hardware's memory capacity, which is very influential.

Model optimization applying PSO to LibSVM_IG produces accuracy values, as seen in Table II. The results of the proposed model had a better level of accuracy and were higher than the model without prior optimization. The best level of accuracy for PSO-based LibSVM_IG produced an accuracy of 88.97% which was better than the previous model. The best model applied the nu-SVC SVM and polynomial kernel with population = 5 and population = 15.

E. EVALUATION & RECOMMENDATIONS MODEL

The next step was to evaluate the model by comparing the experiment results to provide a better picture of the proposed model. Comparison results of all models that have been obtained with different accuracy levels are shown in Table III.

The model evaluation process was carried out by using the same dataset with a training data percentage of 90% and testing data of 10%. The best accuracy performance value for each model was determined using cross-validation with parameters that had been set based on the experiments conducted.

Table III shows that the proposed model, namely LibSVM_IG+PSO, is the model with the best and highest level of accuracy, namely 88.97%. Parameter values and types applied to the proposed model are one of the factors influencing the change in the accuracy value. Each parameter possesses a different characteristic that can significantly affect the classification accuracy. Based on the best accuracy level of the proposed model, the confusion matrix if the accuracy level, which generated based on (1), is shown in Table IV.

$$Accuracy = \frac{\sum TP + TN}{\sum TP + TN + FP + FN}$$

$$Accuracy = \frac{\sum 245 + 102}{\sum 245 + 102 + 38 + 5}$$

$$= 347/390$$

$$= 0.8897.$$

The accuracy level produced at this time is still possible to be optimized and, of course, requires much effort, especially in determining the model's parameters because it relies on intuition. It requires a suitable model to determine the parameters to achieve the expected results automatically. The application of other optimization methods can be considered, especially in optimizing the selection of parameter values used.

Table III shows an increase in the accuracy of the sentiment review model for the coastal assessment using the SVM and LibSVM models. The increase in the accuracy value of the model produced and proposed was continuously increasing. These results indicate that the optimization applied to the model can affect the increase in the accuracy level produced. When compared with previous studies, the proposed model has a fairly good accuracy level, as shown in Table V.

IV. CONCLUSION

The coastal review sentiment assessment has obtained the best model that can be used as decision support in predicting coastal assessment sentiment. This study proposed a LibSVM optimization model through optimization using PSO as a model for the sentiment. Compared to the existing LibSVM classic model, the proposed model boasts a higher level of accuracy. The best model obtained was LibSVM_IG+PSO, with an accuracy rate of 88.97%. This model used the LibSVM and IG based on PSO.

The drawback of this study is that the resulting accuracy level could be more optimal; one of the effects of this process is during data preprocessing, so further experiments are needed to find a more precise preprocessing model and increase the accuracy performance value of the resulting model. For future research, it is necessary to improve accuracy; therefore, it is better to get close to the best accuracy and employ other alternative methods that can be applied and redeveloped to find another appropriate model with a better level of accuracy. The simulated annealing (SA) and genetic algorithm (GA) methods can be proposed for further experimental research, and it is possible that later they will produce different performance values.

CONFLICT OF INTEREST

The authors state that there is no conflict of interest in the writing of the article or the research results.

AUTHOR CONTRIBUTION

Conceptualization, Oman Somantri, and Santi Purwaningrum; methodology, Oman Somantri; model analysis, Oman Somantri; validation, Oman Somantri, Santi Purwaningrum, and Ratih Hafsarah Maharrani; formal analysis, Oman Somantri; Model validation, Oman Somantri, Santi Puraningrum; data curation, Santi Purwaningrum, Ratoh Hafsarah Maharrani; writing—original drafting, Oman Somantri; writing—reviewing and editing, Oman Somantri, Ratih Hafsarah Maharrani; visualization, Oman Somantri.

ACKNOWLEDGMENT

This article is part of the output of the research grant funding for the Penelitian Dosen Pemula (PDP) scheme from the Academic Directorate of Vocational Education, the Director General of Vocational Education, Ministry of Education, Culture, Research, and Technology for the 2022 fiscal year with research contract number of 117/SPK/D4/PPK.01.APTV/VI/2022.

REFERENCES

[1] Y. Tao, W. Liu, Z. Huang, and C. Shi, "Thematic Analysis of Reviews on the Air Quality of Tourist Destinations from a Sentiment Analysis Perspective," *Tour. Manag. Perspect.*, Vol. 42, pp. 1–14, Apr. 2022, doi: 10.1016/j.tmp.2022.100969.

[2] K.D. Mukhina, S.V. Rakitin, and A.A. Visheratin, "Detection of Tourists Attraction Points Using Instagram Profiles," *Procedia Comput. Sci.*, Vol. 108, pp. 2378–2382, 2017, doi: 10.1016/j.procs.2017.05.131.

[3] T. Ali, B. Omar, and K. Soulaimane, "Analyzing Tourism Reviews Using an LDA Topic-Based Sentiment Analysis Approach," *MethodsX*, Vol. 9, pp. 1–10, Nov. 2022, doi: 10.1016/j.mex.2022.101894.

[4] Z. Abbasi-Moud, H. Vahdat-Nejad, and J. Sadri, "Tourism Recommendation System Based on Semantic Clustering and Sentiment Analysis," *Expert Syst. Appl.*, Vol. 167, pp. 1–10, Apr. 2021, doi: 10.1016/j.eswa.2020.114324.

[5] M. Hu *et al.*, "Tourism Demand Forecasting Using Tourist-Generated Online Review Data," *Tour. Manag.*, Vol. 90, pp. 1–19, Jun. 2022, doi: 10.1016/j.tourman.2022.104490.

[6] V. Taecharungroj and B. Mathayomchan, "Analysing TripAdvisor reviews of tourist attractions in Phuket, Thailand," *Tour. Manag.*, Vol. 75, pp. 550–568, Dec. 2019, doi: 10.1016/j.tourman.2019.06.020.

[7] F. Pollák, P. Dorčák, N. Račeta, and N. Svetozarovová, "Sustainable E-Marketing of Selected Tourism Subjects from the Mediterranean through Active Online Reputation Management," in *Smart City 360°*, A.L-. Gracia *et al.*, Eds., Cham, Switzearland: Springer, 2016, pp. 692–703, doi: 10.1007/978-3-319-33681-7_60.

[8] T. Ali *et al.*, "Exploring Destination's Negative E-Reputation Using Aspect Based Sentiment Analysis Approach: Case of Marrakech Destination on TripAdvisor," *Tour. Manag. Perspect.*, Vol. 40, pp. 1–15, Oct. 2021, doi: 10.1016/j.tmp.2021.100892.

[9] D. Apriliani *et al.*, "Sentiment Analysis for Indonesia Hotel Services Review Using Optimized Neural Network," *J. Phys. Conf. Ser.*, Vol. 1538, No. 1, pp. 1–8, May 2020, doi: 10.1088/1742-6596/1538/1/012060.

[10] P.J. Lee, Y.H. Hu, and K.T. Lu, "Assessing the Helpfulness of Online Hotel Reviews: A Classification-Based Approach," *Telemat., Inform.*, Vol. 35, No. 2, pp. 436–445, May 2018, doi: 10.1016/j.tele.2018.01.001.

[11] R.P. Nawangsari, R. Kusumaningrum, and A. Wibowo, "Word2Vec for Indonesian Sentiment Analysis towards Hotel Reviews: An Evaluation Study," *Procedia Comput. Sci.*, Vol. 157, pp. 360–366, Oct. 2019, doi: 10.1016/j.procs.2019.08.178.

[12] N. Oktaviana, H.C. Rustamaji, and 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, pp. 133–143.

[13] D.T. Alamanda *et al.*, "Sentiment Analysis Using Text Mining of Indonesia Tourism Reviews via Social Media," *Int. J. Humanit. Arts, Soc. Sci.*, Vol. 5, No. 2, pp. 72–82, Apr. 2019, doi: 10.20469/ijhss.5.10004-2.

[14] B. Mathayomchan and 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, pp. 260–265, doi: 10.1109/JCSSE.2019.8864150.

[15] N. Leelawat *et al.*, "Twitter Data Sentiment Analysis of Tourism in Thailand during the COVID-19 Pandemic Using Machine Learning," *Heliyon*, Vol. 8, No. 10, pp. 1–11, Oct. 2022, doi: 10.1016/j.heliyon.2022.e10894.

[16] D. Liu, Z. Wang, L. Wang, and L. Chen, "Multi-Modal Fusion Emotion Recognition Method of Speech Expression Based on Deep Learning," *Front. Neurorobot.*, Vol. 15, Jul. 2021, doi: 10.3389/fnbot.2021.697634.

[17] C.-C. Chang and C.-J. Lin, "LIBSVM," *ACM Trans. Intell. Syst., Technol.*, Vol. 2, No. 3, pp. 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, and D. Miao, "Three-Way Confusion Matrix for Classification: A Measure Driven View," *Inf. Sci.*, Vol. 507, pp. 772–794, Jan. 2020, doi: 10.1016/j.ins.2019.06.064.

[20] T. Chen *et al.*, "EEG Emotion Recognition Model Based on the LIBSVM Classifier," *Meas.*, Vol. 164, pp. 1–7, Nov. 2020, doi: 10.1016/j.measurement.2020.108047.