JTAM (Jurnal Teori dan Aplikasi Matematika)
http://journal.ummat.ac.id/index.php/jtam
p-ISSN 2597-7512 | e-ISSN 2614-1175
Vol. 9, No. 3, July 2025, pp. 930-945

Robust Optimization Model Analysis for Online Sentiment Issues on Shopee using Support Vector Machine

Raisha Zuhaira Dongoran[1*], Hendra Cipta[1]
[1]Department of Mathematics, Universitas Islam Negeri Sumatera Utara, Indonesia
raisha0703212064@uinsu.ac.id

Article History:
Received : 16-05-2025
Revised : 25-06-2025
Accepted : 02-07-2025
Online : 03-07-2025

Keywords:
Polyhedral Uncertainty.
Robust Optimization;
Sentiment Analysis;
Lexicon-Based;
Shopee;
Support Vektor Machine (SVM).

ABSTRACT

In the digital economy era, e-commerce platforms like Shopee receive thousands of user reviews daily, which significantly influence customer perceptions and purchasing decisions. However, sentiment analysis of such reviews remains challenging due to the presence of noise, uncertainty, and dynamic data changes. This quantitative research aims to develop a more reliable sentiment classification model by integrating a Lexicon-Based labeling approach and Support Vector Machine (SVM) classification with a Robust Optimization framework. The labeling process uses a sentiment lexicon dictionary that assigns polarity values to words, classifying texts into positive, negative, or neutral categories. The classification process utilizes SVM to evaluate sentiment prediction based on key performance metrics: Accuracy, Precision, Recall, and F1-score. These performance metrics are treated as uncertain parameters in the optimization phase. The main contribution of this study is the formulation of a robust optimization model for sentiment analysis weighting problems, transforming a multi-criteria objective into a single-objective utility function. By applying polyhedral uncertainty modeling, the robust counterpart formulation accounts for worst-case scenarios in model evaluation. Numerical experiments using Python in Google Colab show that while the deterministic model achieves a higher performance score (0.865), the robust model yields a slightly lower score (0.825) but offers better stability under uncertainty. These results imply that robust optimization can enhance the reliability of sentiment classification systems in real-world e-commerce applications, providing more trustworthy insights for businesses in managing consumer feedback.

Crossref
https://doi.org/10.31764/jtam.v9i3.31555

———————————— ◆ ————————————

A. INTRODUCTION

E-commerce (electronic commerce) refers to trading activities conducted online via the internet, and it has played a major role in shaping modern consumer behavior. Online marketplaces have even contributed to the global spread of language use, including English, by facilitating international trade and communication (Wahyuningsih & Ziyana, 2023). Shopee, established by PT Shopee Indonesia in 2015, is one of the most widely used e-commerce platforms in Indonesia. It offers a seamless and secure shopping experience, connecting millions of buyers and sellers while managing logistics and order fulfillment efficiently (Khaw, 2023; Reza, 2022). The application serves as a digital marketplace for a wide variety of products, from fashion to beauty and personal care, aiming to meet daily needs in a practical way (Prayitno, 2023; Wulansari et al., 2024). Shopee's popularity is driven not only by its features but also by its perceived trustworthiness among Indonesian users (Sitorus & Zufria, 2024; Negm, 2024).

One unique aspect of Shopee's ecosystem is the volume of customer reviews it receives daily. These reviews are critical in shaping public perception, as positive comments tend to build trust, while negative ones may harm a product's reputation (Ventre & Kolbe, 2020). Given this reality, sentiment analysis becomes a valuable tool for interpreting user feedback. However, although numerous studies have employed lexicon-based or machine learning techniques to classify sentiments, few have considered the effects of uncertainty within performance metrics. Most models assume stable data conditions, which may not reflect the real-world volatility present in e-commerce platforms.

This research addresses that gap by proposing a hybrid sentiment analysis framework combining lexicon-based labelling and Support Vector Machine (SVM) classification, further enhanced through the application of a robust optimization model. The novelty lies in integrating performance uncertainty directly into the optimization process to improve model reliability. By doing so, the study contributes to the development of sentiment analysis methods that are not only accurate but also stable and adaptable in fluctuating data environments, particularly in dynamic platforms like Shopee.

Sentiment analysis is an area of research focused on examining human emotions, perspectives, opinions, and evaluations related to a wide range of subjects, including events, topics, products, services, individuals, organizations, and their associated characteristics (Khaw, 2023). Sentiment analysis is an important area within natural language processing that aims to classify text into three main categories: positive, negative, or neutral. As online platforms continue to grow, allowing people to freely share their thoughts and views, it has become essential for organizations to understand the emotions behind these expressions in order to make better-informed decisions (Tan et al., 2023). This study was conducted using the Lexicon-Based approach. Essentially, sentiment analysis can be carried out using two main approaches: a rule-based method or machine learning techniques (B & B, 2023).

The lexicon-based method assigns semantic orientation to words by utilizing either dictionary-driven techniques or approaches based on language corpora (Qi & Shabrina, 2023). Lexicon-based approaches assess the degree of emotion and subjectivity at the word level. The classification process focuses on identifying informative words within a text and assigns a quantitative score to each word's polarity. This study addresses a multi-class classification task, where texts are categorized as positive, negative, or neutral (Raees & Fazilat, 2024). The lexicon-based algorithm is capable of producing fairly accurate text classification results, especially when combined with complementary techniques like machine learning (Akbari et al., 2023).

This study employs a machine learning algorithm, namely Support Vector Machine (SVM). Support Vector Machine (SVM) is a commonly applied classical supervised learning algorithm. It is built upon the structural risk minimization principle, enabling it to effectively reduce the likelihood of overfitting in various scenarios (Tang, 2024). support vector machine (SVM) is a supervised machine learning (ML) method capable of learning from data and making decisions based on the input (Valkenborg et al., 2023). In this study, Support Vector Machine is used to enhance model performance by optimizing values of Accuracy, Precision, Recall, and F1-Score. The values of the performance metric parameters are considered to involve uncertainty.

Therefore, this optimization problem is addressed using Robust Optimization to effectively manage the uncertainty factors.

Robust optimization is a method commonly applied when dealing with uncertain parameters. Unlike probabilistic approaches, it does not rely on predefined probability distributions; instead, it captures uncertainty by defining a specific range or set within which the parameters may vary (Syahputri & Cipta, 2024). Since these indicators involve uncertainty, a Robust Optimization approach is used a method capable of handling uncertainty without relying on specific probability distributions. This approach is expected to produce more accurate and stable sentiment predictions. Based on the literature review, no previous studies have specifically applied Robust Optimization for weighting in online sentiment analysis. Therefore, this study proposes the application of a robust optimization model for the case of reviews on the Shopee application, referring to the multi-objective model reformulation by (Kumar et al., 2020). In the context of sentiment analysis on the Shopee application, this approach is expected to improve sentiment classification accuracy by accommodating various variations in review data.

B. METHODS

1. Research Stages

This study employs several methods to address the problem. The research flowchart, presented in Figure 1, outlines the approaches utilized throughout the study. As illustrated in the figure, each step represents a stage in the research process.

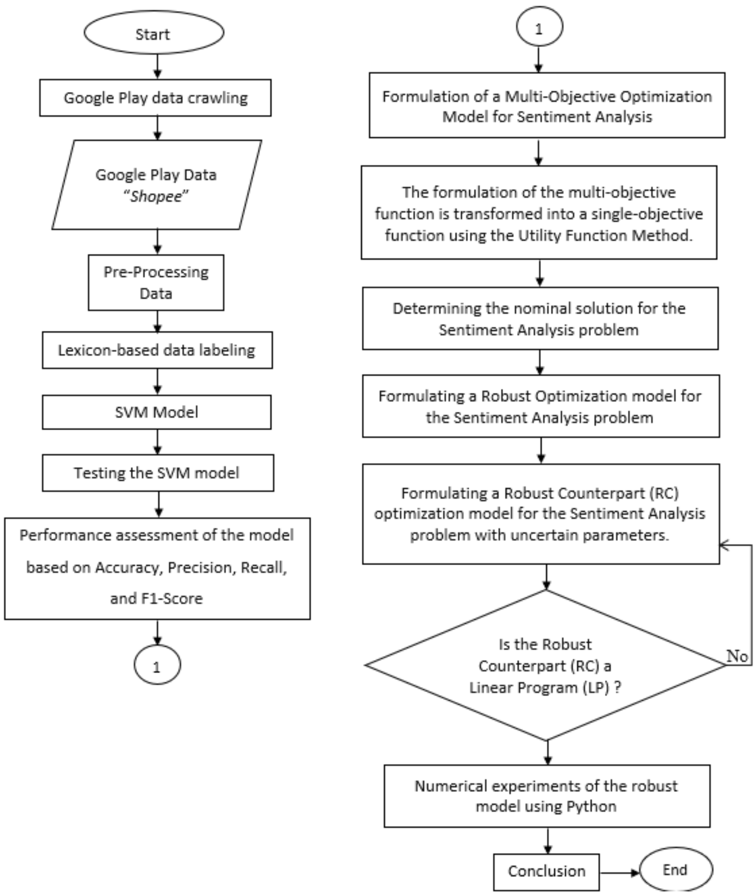

Figure 1. Flowchart of the Research

2. Crawling Data

Data crawling was performed to automatically and systematically collect relevant information from the Shopee platform. The data used in this study were obtained through a data crawling process targeting user reviews of the Shopee application available on the Google Play Store. The dataset consists of the most recent reviews collected in 2025, with a total of 1,500 data entries selected as the sample for analysis. These reviews were selected as a primary data source due to their relevance in reflecting real user experiences and opinions. Data crawling was done using Python in Google Colab, as shown in Figure 2.

```python
import csv

def export_to_csv(reviews, file_name='hasil_scraper_ulasan_app_shopee.csv'):
    if reviews:
        fieldnames = ['Review ID', 'Username', 'Rating', 'Review Text', 'Date']

        with open(file_name, mode='w', newline='', encoding='utf-8') as csv_file:
            writer = csv.DictWriter(csv_file, fieldnames=fieldnames)
            writer.writeheader()

            for review in reviews:
                writer.writerow({
                    'Review ID': review['reviewId'],
                    'Username': review['userName'],
                    'Rating': review['score'],
                    'Review Text': review['content'],
                    'Date': review['at']
```

Figure 2. Data crawling processed using Python libraries

3. Procesiing Data

Data processing is a processing and managing data into information that can be used for decision-making. Data is processed using Python in the Google Colab environment. Data pre-processing is performed with the following stages:"

a. Data cleaning refers to the process of detecting and correcting erroneous or inconsistent entries in a dataset, table, or database. (Borrohou et al., 2023).

b. Case folding is performed by converting all uppercase letters to lowercase. Only alphabetic characters from 'a' to 'z' are retained, while non-letter characters such as punctuation marks (e.g., &, #, :, -, ;) and digits (0–9) are treated as delimiters and therefore removed from the document (H, 2023).

c. Normalization becomes necessary when the feature values vary significantly in their ranges (Peshawa & Rezhna, 2014).

d. Tokenization serves as a fundamental process in natural language processing (NLP), acting as a bridge between raw text and computational language models (Schmidt et al., 2024).

e. Stopword removal eliminating stopwords allows the focus to shift toward the key terms that better reflect the underlying sentiment (Mostafa, 2025).

f. Stemming is defined as the process which produces variants of a root. In simple words,it reduces a base word to its stem word (Siddhartha et al., 2020).

4. Lexicon Based

Lexicon has been the most tested for Indonesian language sentiment analysis, however, there aren't many papers that discuss sentiment analysis in the Indonesian language using lexicon (Firdaus et al., 2021). Lexicon-based labeling is an approach in sentiment analysis that uses a predefined dictionary (lexicon) of words assigned with polarity values (positive, negative, or neutral) in order to assess the sentiment expressed in the text, as shown in Table 1.

Table 1. labeling results in three categories: Positive, Negative, and Neutral

No	Steming Data	Score	Sentiment
1	Item Selection and Others	-1	Negative
2	Price Applicator Based on Subscription and Shopping Frequency…	4	Positive
3	Okayy	0	Neutral
4	Normality	0	Neutral
5	good	0	Neutral

5. Suppport Vektor Mechine (SVM)

Support Vector Machine (SVM) is a supervised learning method introduced in the early 1990s, widely utilized for both classification and regression problems due to its strong performance and reliability, particularly when handling large-scale datasets (Khyathi et al., 2025). SVM is a popular machine learning technique known for its effectiveness in solving classification problems (Guido et al., 2024). This research involved the use of a Support Vector Machine by computing values from the Confusion Matrix. Support Vector Machine (SVM) is a supervised learning technique used to find the most effective hyperplane that separates data into two distinct classes for binary classification tasks (Nabilah et al., 2025).

Standard SVM operates under the assumption that all training data are provided at once for batch processing (Almaspoor et al., 2021). Classification is a fundamental task in machine learning that aims to develop a rule or model for assigning data into predefined categories, using information derived from a given set of training data. Here is an example of a classification matrix in Figure 3.

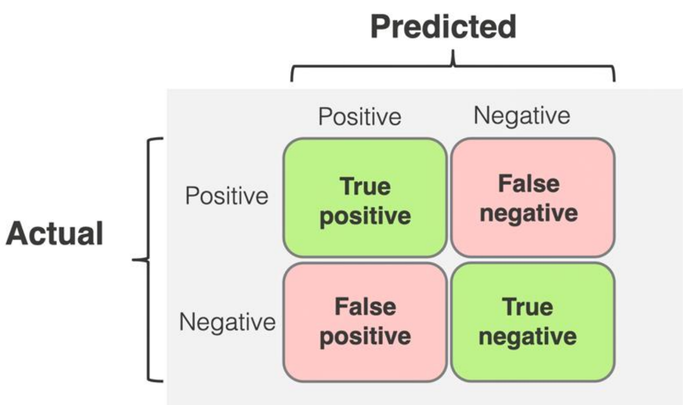

Figure 3. Classification Matrix

According to (Vakili et al., 2020), model evaluation can be performed using performance metrics, along with an analysis of the performance of the training process. To calculate the performance metrics, the formulas presented in Table 2 are used.

Table 2. Performance Matrix

Metric	**Definition**
Accuracy	$\frac{TP + TN}{number\ of\ data\ points}$
Precision	$\frac{TP}{TP + FP}$
Recall	$\frac{TP}{TP + FN}$
F1-score	$\frac{2\ .precission\ .recall}{(precission +\ recall)}$

where TP, TN, FP, and FN are defined as follows:

a. TP (True Positives)
b. TN (True Negatives)
c. FP (False Positives)
d. FN (False Negatives)

The values of TP, TN, FP, and FN are determined for each class (negative, neutral, and positive) in order to calculate the Precision, Recall, and F1-Score.

6. Optimization Model for Sentiment Analysis

Sentiment analysis, as a classification problem, uses performance metrics to evaluate model performance. This problem involves several parameters and decision variables x_i , which represent scalar weighting factors for each performance metric. Specifically, x_1 is the weight for Accuracy (a_2), x_2 is the weight for Precision (a_2), x_3 is the weight for Recall (a_3), and (a_4) is the weight for F1-Score $(a_4\ )$. The objective function is formulated in Eq. (1).

$$\max a_{1,2,3,4} \tag{1}$$

The multi-objective function in Eq. (1) is transformed into a single objective function using the Utility Function Method, resulting in the sentiment analysis optimization model as shown in Eq. (2) (Kumar et al., 2020).

$$\begin{gathered} max\ g\ (x) = a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4 \\ s.t. \sum_{i=1}^{n} x_i = 1 \\ x_i \in [0,1], i = 1, \dots, N \end{gathered} \tag{2}$$

7. Robust Optimization

Robust optimization is founded on the premise that uncertain data belongs to a convex and bounded set known as the uncertainty set. The approach begins by defining the robust counterpart, typically assuming that the uncertain data takes the form of boxes, intervals, box-interval combinations, ellipses, or polyhedral shapes. (Cipta et al., 2022). With robust optimization, the resulting robust solutions can help decision-makers avoid losses caused by such uncertainty with Polyhedral. Data uncertainty can be caused by measurement errors, modeling inaccuracies, or the unavailability of necessary information at the time decisions need to be made. According to (Cipta et al., 2022), the general model of uncertain data can be formulated as problem (3).

$$\min_{x}\{c^T x: Ax \leq b|(c, A, b)x \in u\} \tag{3}$$

Where c, A, and b are uncertain coefficients, and $\mathcal{U}$ is the uncertainty set. Robust Optimization is built upon several fundamental assumptions, which include:

a. All decision variables $x \in \mathbb{R}^n$ represent "here and now" decisions, meaning that specific numerical values must be determined as the solution before showing the data, here is a short explanation about how and why it was collected "reveals itself".
b. Responsibility for the decisions lies entirely with the decision-maker if, and only if, the observed data is contained within the uncertainty set $\mathcal{U}$.
c. The constraints in uncertain linear programming are considered rigid, meaning the decision-maker does not permit any violations when the data lies within the uncertainty set $\mathcal{U}$.

Under the given assumptions, Robust Optimization converts the original uncertain problem into a single, equivalent deterministic problem called the Robust Counterpart (RC). The best robust solution corresponds to the optimal solution of the Robust Counterpart, and the robust optimal value of the uncertain linear program is equal to the optimal value derived from the Robust Counterpart. According to (Cipta et al., 2022), if it is assumed that $c \in \mathbb{R}^n$ if certification is confirmed, then the reformulated version of (3) is known as the Robust Counterpart problem, presented in (4).

$$\min_{x}\{c^T x: a(\zeta)x \leq b|\zeta \in Z\} \tag{4}$$

where $\mathbb{Z} \in \mathbb{R}^L$ shows the primitive uncertainty set. A solution $x \in \mathbb{R}^n$ a solution is said to be robust feasible if it complies with the uncertain constraint $C(\zeta) \leq q$ for all realization $\zeta \in \mathbb{Z}$. An uncertainty parameter defined as shown in Eq. (5)

$$a(\zeta) = a = \bar{a} + P\zeta \tag{5}$$

where $\bar{a} \in \mathbb{R}^n$ is the vector of nominal values and $P \in \mathbb{R}^{n \times L}$ is the perturbation matrix. The uncertainty set $\mathcal{U}$ is defined as shown in Eq. (6).

$$u = \{a: a = \bar{a} + P\zeta, \zeta \in Z\} \quad (6)$$

when $\bar{a} + P\zeta$ It can be expressed as an affine function depending on the primitive uncertain parameter $\zeta \in Z\}$, $a \in \mathbb{R}^n$ and $P \in M_{n,L}\{\mathbb{R}\}$. The solution that optimizes the Robust Counterpart is regarded as the robust optimal solution, and the corresponding value is known as the robust optimal value of the uncertain problem.

C. RESULT AND DISCUSSION

1. Lexicon-Based Labeling

Lexicon-based labeling is an approach in sentiment analysis that uses a predefined dictionary of words (lexicon), each assigned a polarity value, to assess the sentiment conveyed by a particular piece of text. Using the Python programming, the number of data points labelled as shown in Figure 4.

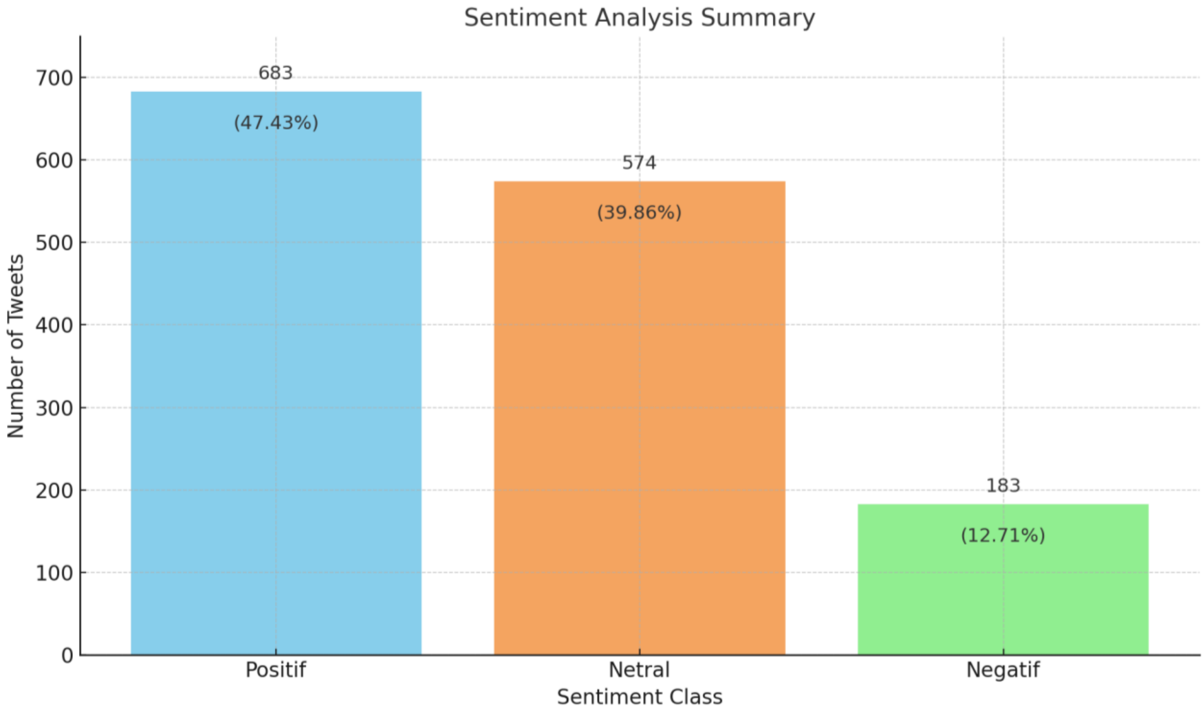

Figure 4. The count of data instances categorized as positive, negative, or neutral

Next, the dataset is divided into training and testing sets to train the Support SVM model, which will be used to evaluate its performance. There are 1,152 samples in the training set and 288 samples in the test set. The dataset was split using Python programming, with the data distribution shown in Figure 5.

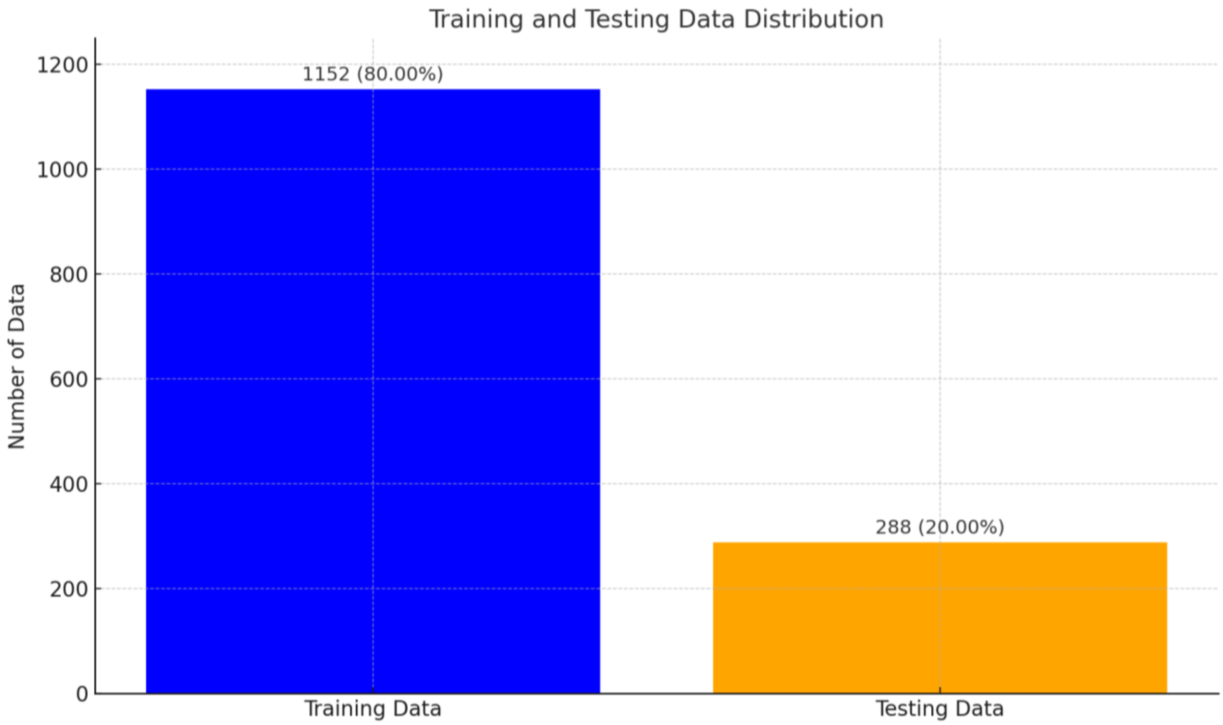

Figure 5. Data distribution

2. Support Vector Machine (SVM) Classification

Initially, the data is split into training and testing subsets. To measure the SVM model's effectiveness, a validity check is performed using a classification matrix. The 3x3 classification matrix, displaying predicted and true classes, is provided in Table 3.

Table 3. The Classification Matrix generated by the SVM model

		Estimated Class		
		Negative	**Neutral**	**Positive**
True Class	Negative	23	12	2
	Neutral	2	103	10
	Positive	3	10	123

with the classification matrix values already obtained, the next step is to calculate the overall performance metrics of the classification method using the formulas outlined in Table 2.

$$Accuracy = \frac{TP}{N} = 0{,}865$$
$$Precision = \frac{TP}{(TP+FN)} = 0{,}852$$
$$Recall = \frac{TP}{(TP+FP)} = 0{,}807$$
$$F1-Score = 0{,}825$$

The percentage values for performance metric are obtained as shown in Table 4.

Table 4. Performance Metrics

Metric	**Nilai**
Accuracy	0,865
Precision	0,852
Recall	0,807
F-1 Score	0,825

The performance metrics derived from the crawling, preprocessing, and SVM classification stages serve as sources of uncertainty when determining the weights in sentiment analysis through Robust Optimization. These performance metrics correspond to the parameters a_i listed in Table 5.

Table 5. Performance Metrics Data

	Value
a_1	0,865
a_2	0,852
a_3	0,807
a_4	0,825

3. Optimization Model Formulation for Sentimen Analysis Weighting Problem

The optimization model formulation in this study is obtained as in Eq. (16)(Kumar et al., 2020).

$$max\ g\ (x) = \textstyle\sum_{i=1}^{N} a_i x_i\ ,\ \forall a_i \in \mathcal{U} \tag{16}$$
$$s.t. \sum_{i=1}^{n} x_i = 1$$
$$x_i \in [0,1], i = 1, \dots, N$$

The Parameters used in this problem are a_i which are the performance metrics of sentiment analysis, where a_i are the following values.

a_1 : $Accuracy$
a_2 : $Precision$
a_3 : $Recall$
a_4 : $F1 - Score$

4. Uncertainty Optimization Model Formulation for Weighting Issue in Sentiment Analysis

In the present study, the uncertainty factor is represented by a_i, which is derived from the classification process. The objective function is considered deterministic, allowing the problem to be formulated with uncertainty present solely in the constraint functions. To facilitate this, a single variable $t \in \mathbb{R}$ is introduced as a replacement for the objective function, resulting in the form expressed in Eq. (17).

$$\max t \tag{17}$$

Next, an additional constraint function is introduced as shown in Eq. (18).

$$\sum_i a_i x_i \tag{18}$$

Thus, the formulation is obtained as shown in Eq. (19).

$$\begin{aligned} &\max t \\ &s.t \quad \sum_i a_i x_i \geq t \\ &\sum_i x_i = 1 \\ &x_i \in [0,1] \end{aligned} \tag{19}$$

It can be seen in Eq. (19) that the uncertainty only appears in the constraint functions. The uncertainty parameters are defined as in Eq. (20).

$$a_i(\zeta) = a_i = \bar{a}_i + P_i\zeta, \forall\zeta \in Z \tag{20}$$

when $\bar{a}_i \in \mathbb{R}^n$ is stated as the nominal value and $P \in \mathbb{R}^{nxl}$ is the perturbation matrix. The set $\mathcal{U}$ is defined as in Eq. (21).

$$\mathcal{U} = \{a_i = \bar{a}_i + P_i\zeta | \forall\zeta \in Z\} \tag{21}$$

Substituting the uncertainty from Eq. (21) into the constraint (19) results in Eq. (22) as follows.

$$\sum_i (\bar{a}_i + P_i\zeta)x_i, \quad \forall\zeta \in Z \tag{22}$$

$$\sum_i \bar{a}_i x_i + P_i\zeta x_i, \quad \forall\zeta \in Z \tag{23}$$

Robust Optimization model formulation for the sentiment analysis weighting problem is obtained as shown in Eq. (24)

$$\begin{aligned} &\max t \\ &s.t \quad \sum_i \bar{a}_i x_i + P_i\zeta x_i \geq t, \quad \forall\zeta \in Z \\ &\sum_i x_i = 1 \\ &x_i \in [0,1] \end{aligned} \tag{24}$$

5. Robust Counterpart Formulation for Sentimen Analisis Weighting Problem with Polyhedral uncertainty regions

To obtain the Robust Counterpart formulation of the sentiment analysis weighting problem with a polyhedral uncertainty regions, the process involves three key steps:

Stage 1

The constraint reformulation in the equation is equivalent to the worst-case formulation presented in Eq. (25)

$$\sum_i a_i x_i \geq t \equiv \sum_i (\bar{a}_i + P_i \zeta) x_i \geq t \equiv \sum_i \bar{a}_i x_i + P_i \zeta x_i \geq t$$
$$\equiv \textstyle\sum_i \bar{a}_i x_i + \max_{\zeta : d - D \geq 0} P_i x_i \zeta \geq t \tag{25}$$

Stage 2

The dual form of the maximization problem is formulated and can be expressed as the following problem in Eq. (26)

$$\begin{aligned} & max\ P_i x_i \zeta \\ & s.t\ d - D \geq 0 \end{aligned} \tag{26}$$

Move the variable ddd to the right-hand side and multiply both sides of the constraint by (−1), so that the problem in Eq. (26) is expressed as the inequality in Eq. (27)

$$\begin{aligned} & \max_{\zeta : d - D \geq 0} P_i x_i \zeta \\ & s.t\ D \leq d \end{aligned} \tag{27}$$

The problem in Eq. (27) has a maximization objective function with less-than-or-equal-to (≤) inequality constraints. In relation to the primal-dual construction, the dual form of the maximization problem (27) can be expressed as the problem in Eq. (28)

$$\begin{aligned} & \min d_i y_i \\ & s.t\ D_i y_i = P_i x_i, y_i \geq 0 \end{aligned} \tag{28}$$

Thus, the reformulation of the inequality in Eq. (28) is obtained in the form of Eq. (29)

$$\begin{aligned} & \sum_i \bar{a}_i x_i + \max_{\zeta : d - D \geq 0} P_i x_i \zeta \geq t \\ & \sum_i \bar{a}_i x_i + \min_{y_i} \{ d_i y_i : D_i y_i = P_i x_i, y_i \geq 0 \\ & \textstyle\sum_i \bar{a}_i x_i + d_i y_i \geq t, D_i y_i = P_i x_i, y_i \geq 0 \end{aligned} \tag{29}$$

Stage 3

Note that problem (29) is satisfied for a feasible solution y contained in the feasible set $\mathcal{F} = \{y : D_i y_i = P_i x_i, y_i \geq 0\}$, so the constraint function is guaranteed to be satisfied for the minimum value of y. The Robust Counterpart model formulation with a polyhedral uncertainty regions for the sentiment analysis weighting problem is as shown in Eq. (30) below.

$$\begin{aligned}
&\max t \\
&s.t \sum_i \bar{a}_i x_i + d_i y_i \geq t \\
&D_i y_i = P_i x_i, y_i \geq 0 \\
&\textstyle\sum_i x_i = 1 \\
&x_i \in [0,1]
\end{aligned} \tag{30}$$

Equation (30) is structured as a Linear Programming (LP) problem, LP involves a system of linear equations, with each equation containing one dependent variable and multiple independent variables (Hek et al., 2025). Which shows that the Robust Counterpart model with a polyhedral uncertainty set is computationally manageable and can be solved in polynomial time.

6. Computational Experiment

a. Deterministic Model

By using Google Colab with Python programming language, the optimal solution of the deterministic model for the weighting of sentiment analysis is obtained, as shown in Table 6.

Table 6. Computational Experiment for Deterministic Model

i	a_i	x_i	$a_i x_i$
1	0,865	1	0,865
2	0,852	0	0
3	0,807	0	0
4	0,825	0	0
	$g(x)$		0,865

The deterministic model assigns the entire weight (x = 1) to the first component with the highest performance score (a_i = 0.865), while all other components are ignored (x = 0). This leads to an optimal solution of g(x) = 0.865, which represents the maximum performance under the assumption of no uncertainty. While this solution appears optimal in a stable environment, it may be overly sensitive to small perturbations or data noise.

b. Robust Counterpart Model

After that, the optimal solution is sought from the Robust Counterpart model for the weighting problem of sentiment analysis with polyhedral uncertainty, as shown Table 7.

Table 7. Computational Experiment for Robust Model

i	a_i	x_i	$a_i x_i$
1	0,865	0,01	0,009
2	0,852	0,845	0,720
3	0,807	0,866	0,699
4	0,825	0,0	0
	$g(x)$		0,825

Unlike the deterministic model, the robust model distributes the weights more evenly across multiple components. Although the resulting value of g(x) = 0.825 is slightly lower than that of the deterministic model, this solution provides greater resilience to uncertainty in the data. By not relying solely on a single input, the model becomes less sensitive to errors or fluctuations, making it more suitable for real-world applications such as sentiment analysis on dynamic e-commerce platforms like Shopee.

7. Analysis of Numerical Results for Sentiment Analysis Weighting Problems

This section presents the optimal solutions derived from numerical experiments conducted on both the deterministic model and the Robust Optimization Counterpart model incorporating polyhedral uncertainty regions in the sentiment analysis weighting problem. The outcomes of the numerical experiments can be noticed in Table 8.

Table 8. The Results of the Computational Experiments

Deterministic Model	RC Model with Polyhedral Uncertainty Regions
0,865	0,825

The comparison of the results in Table 9 shows a difference between the deterministic model and the RC model with polyhedral uncertainty regions. The SVM model under the RC approach achieved a performance score of 0.825, which is lower than the deterministic model's score of 0.865. This difference occurs because the Robust Counterpart model optimizes the performance metrics while explicitly considering uncertainty factors, thereby preparing for worst-case scenarios that might occur in practice. The lower score of the robust model reflects a conservative but more reliable estimate of performance under uncertain conditions, prioritizing stability over potentially optimistic results given by the deterministic model. Understanding this lower score is crucial because it highlights the trade-off between performance and reliability while the deterministic model might give a higher score assuming ideal data, the robust model offers a safeguard against variability and unexpected fluctuations in the input data or model parameters. To address the lower robust score, key strategies include refining data quality to reduce noise, tuning model parameters for optimal performance, and integrating adaptive methods to enhance resilience against data variability.

D. CONCLUSION AND SUGGESTIONS

This study highlights the main contribution of integrating Lexicon-Based sentiment labeling with the Support Vector Machine (SVM) classification method, enhanced by a Robust Optimization framework, to improve the reliability of sentiment classification on the Shopee platform. While the deterministic model achieved a slightly higher objective value (0.865), the robust model demonstrated superior stability and trustworthiness under uncertainty, with an objective value of 0.825. This finding emphasizes the practical value of incorporating uncertainty modeling in e-commerce sentiment analysis to ensure consistent performance. For future research, alternative robustness approaches such as ellipsoidal uncertainty sets or distributionally robust optimization can be explored. These methods could offer insights into the trade-offs between robustness, computational efficiency, and model interpretability, thereby advancing the development of more adaptive and resilient sentiment classification systems in dynamic online environments.

ACKNOWLEDGEMENT

Grateful appreciation is extended to the Department of Mathematics, Faculty of Sains and Tecnology, Universitas Islam Negeri Sumatera Utara for the support provided. It is hoped that the results of this study will make a meaningful contribution to the advancement of science.

REFERENCES

Akbari, W. A., Tukino, T., Huda, B., & Muslih, M. (2023). Sentiment Analysis of Twitter User Opinions Related to Metaverse Technology Using Lexicon Based Method. *Sinkron*, *8*(1), 195–201. https://doi.org/10.33395/sinkron.v8i1.11992

Almaspoor, M. H., Safaei, A., Salajegheh, A., Minaei-Bidgoli, B., & Safaei, A. A. (2021). *Support Vector Machines in Big Data Classiication: A Systematic Literature Review Support Vector Machines in Big Data Classification: 1 A Systematic Literature Review 2 3*. http://dx.doi.org/10.21203/rs.3.rs-663359/v1

B, H. G., & B, S. N. (2023). *Cryptocurrency Price Prediction using Twitter Sentiment Analysis*. 13–22. https://doi.org/10.5121/csit.2023.130302

Borrohou, S., Fissoune, R., & Badir, H. (2023). Data cleaning survey and challenges – improving outlier detection algorithm in machine learning. *Journal of Smart Cities and Society*, *2*(3), 125–140. https://doi.org/10.3233/scs-230008

Cipta, H., Suwilo, S., Sutarman, & Mawengkang, H. (2022). Improved Benders decomposition approach to complete robust optimization in box-interval. *Bulletin of Electrical Engineering and Informatics*, *11*(5), 2949–2957. https://doi.org/10.11591/eei.v11i5.4394

Firdaus, R., Asror, I., & Herdiani, A. (2021). Lexicon-Based Sentiment Analysis of Indonesian Language Student Feedback Evaluation. *Indonesia Journal of Computing*, *6*(1), 1–12. https://doi.org/10.34818/indojc.2021.6.1.408

Guido, R., Ferrisi, S., Lofaro, D., & Conforti, D. (2024). An Overview on the Advancements of Support Vector Machine Models in Healthcare Applications: A Review. *Information (Switzerland)*, *15*(4). https://doi.org/10.3390/info15040235

Hek, T. K., Hou, A., Tinggi, S., Ekonomi, I., & Prasetya, E. (2025). Formation of Linear Programming Models of Water Price Compliant to the Regulation of Ministry of Home Affairs , Indonesia. *JTAM*. *9*(2), 555–567. https://doi.org/10.31764/jtam.v9i2.29484

Khaw, J. (2023). Shopee: How Does E-commerce Affect E-consumer Perception And Satisfaction? *International Journal of Tourism and Hospitality in Asia Pasific*, *6*(1), 1–13. https://doi.org/10.32535/ijthap.v6i1.2169

Khyathi, G., Indumathi, K. P., A, J. H., M, L. F. J., Siluvai, S., & Krishnaprakash, G. (2025). *Support Vector Machines : A Literature Review on Their Application in Analyzing Mass Data for Public Health*. *17*(1), 1–6. https://doi.org/10.7759/cureus.77169

Kumar, R., Pannu, H. S., & Malhi, A. K. (2020). Aspect-based sentiment analysis using deep networks and stochastic optimization. *Neural Computing and Applications*, *32*(8), 3221–3235. https://doi.org/10.1007/s00521-019-04105-z

Mostafa, G. (2025). *Improve the Sentiment of Bengali Language Texts with Stopword Removal Improve the Sentiment of Bengali Language Texts with Stopword Removal*. *March*. http://dx.doi.org/10.1145/3723178.3723233

Nabilah, M. F., Fauzan, A., & Indonesia, U. I. (2025). Analyzing Multiclass Land Cover and Spatial Point Patterns on Sentinel-2 Imagery Using Machine Learning and Deep Learning. *JTAM. 9*(2), 346–361. https://doi.org/10.31764/jtam.v9i2.29683

Peshawa, J. M. A., & Rezhna, H. F. (2014). Data Normalization and Standardization: A Technical Report. *Machine Learning Technical Reports*, *1*(1), 1–6. http://dx.doi.org/10.13140/RG.2.2.28948.04489

Prayitno, S. B. (2023). Generation Z perception of national online shopping day on Shopee e-commerce.

Journal of Management Science (JMAS, *6*(4), 596–604. https://doi.org/10.35335/jmas.v6i4.306

Qi, Y., & Shabrina, Z. (2023). Sentiment analysis using Twitter data: a comparative application of lexicon- and machine-learning-based approach. *Social Network Analysis and Mining*, *13*(1), 1–14. https://doi.org/10.1007/s13278-023-01030-x

Raees, M., & Fazilat, S. (2024). *Lexicon-Based Sentiment Analysis on Text Polarities with Evaluation of Classification Models*. 1–18. https://doi.org/10.48550/arXiv.2409.12840

Schmidt, C. W., Reddy, V., Zhang, H., Alameddine, A., Uzan, O., Pinter, Y., & Tanner, C. (2024). *Tokenization Is More Than Compression*. https://doi.org/10.48550/arXiv.2402.18376

Siddhartha, S. B., Khyani, D., Niveditha, M. N., & Divya, M. B. (2020). An Interpretation of Lemmatization and Stemming in Natural Language Processing. *Journal of University of Shanghai for Science and Technology*, *22*(10), 350–357. https://www.researchgate.net/publication/348306833

Sitorus, R. A., & Zufria, I. (2024). Application of the Naïve Bayes Algorithm in Sentiment Analysis of Using the Shopee Application on the Play Store. *Digital Zone*, *15*(1), 53–65. https://doi.org/10.31849/digitalzone.v15i1.19828

Syahputri, L., & Cipta, H. (2024). Implementation of Robust Optimization Model to Controlling the Inventory Costs of Consumable Medical Equipment at Malahayati Islamic Hospital. *Jurnal Matematika, Statistika Dan Komputasi*, *20*(3), 710–723. https://doi.org/10.20956/j.v20i3.34284

Tan, K. L., Lee, C. P., & Lim, K. M. (2023). A Survey of Sentiment Analysis: Approaches, Datasets, and Future Research. *Applied Sciences (Switzerland)*, *13*(7). https://doi.org/10.3390/app13074550

Tang, W. (2024). Application of support vector machine system introducing multiple submodels in data mining. *Systems and Soft Computing*, *6*(April), 200096. https://doi.org/10.1016/j.sasc.2024.200096

Vakili, M., Ghamsari, M., & Rezaei, M. (2020). *Performance Analysis and Comparison of Machine and Deep Learning Algorithms for IoT Data Classification*. http://dx.doi.org/10.48550/arXiv.2001.09636

Valkenborg, D., Rousseau, A. J., Geubbelmans, M., & Burzykowski, T. (2023). Support vector machines. *American Journal of Orthodontics and Dentofacial Orthopedics*, *164*(5), 754–757. https://doi.org/10.1016/j.ajodo.2023.08.003

Ventre, I., & Kolbe, D. (2020). The Impact of Perceived Usefulness of Online Reviews, Trust and Perceived Risk on Online Purchase Intention in Emerging Markets: A Mexican Perspective. *Journal of International Consumer Marketing*, *32*(4), 287–299. https://doi.org/10.1080/08961530.2020.1712293

Wahyuningsih, S., & Ziyana Untsa, F. (2023). English as Business Lingua Franca: Examining the Use of English in Indonesian Online Business. *ELT-Lectura*, *10*(2), 96–104. https://doi.org/10.31849/elt-lectura.v10i2.13699