

Robust Optimization Model for Twitter Sentiment Analysis of *PeduliLindungi* Application

Alfia Azizah Fatimathuzahra¹, Diah Chaerani², Firdaniza³

^{1,2,3}Departement of Mathematics, Universitas Padjadjaran, Indonesia

alfia18001@mail.unpad.ac.id¹, d.chaerani@unpad.ac.id², firdaniza@unpad.ac.id³

ABSTRACT

Article History:

Received : 30-04-2022

Revised : 16-06-2022

Accepted: 22-06-2022

Online : 16-07-2022

Keywords:

PeduliLindungi;

Robust Optimization;

Multi-objective

Function;

Sentiment Analysis;

Support Vector

Machine.

Technological advances during the COVID-19 pandemic in Indonesia gave rise to the *PeduliLindungi* application which is developed by the government to prevent the spread of COVID-19. The advantages and disadvantages of developing *PeduliLindungi* can be seen from the responses and opinions from users, one of which is through the Twitter. A person's opinion about *PeduliLindungi* based on the tweet can be classified into positive, negative, or neutral categories using a Machine Learning approach with the Support Vector Machine (SVM) algorithm. In this paper, multi-objective optimization modeling is used to maximize the performance metrics, which are the value of Accuracy, Precision, Recall, and F1-Score. The value of the performance metrics is considered to contain uncertainty factors. Therefore, the optimization problem is solved by using Robust Optimization to handle the uncertainty factor. The data uncertainty is assumed to be belongs to polyhedral uncertainty set thus the resulted robust is computationally tractable. Numerical experiment is presented to complete the discussion.



<https://doi.org/10.31764/jtam.v6i3.8624>



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license

A. INTRODUCTION

The *PeduliLindungi* application is an application developed by the government in April 2020 to do digital tracking in order to prevent the spread of COVID-19 in Indonesia (Kurniawati et al., 2020). Every application which is still developed has advantages and disadvantages, including the *PeduliLindungi* application. The advantages and disadvantages can be seen from the responses as well as opinions from its users. People can write their opinions from various media, including social media. Social media user activity has been subjected to data collection, and considered as meaningful data sources in the industry and academia (Firdaniza et al., 2021). One popular social media that Indonesian people actively use is Twitter. Recorded as many as 18.45 million Twitter users in Indonesia and became the fifth largest country of Twitter users in the world as of January 2022 (Source: *Statista.com*). From Twitter, anyone can see information or even opinions from tweets that are shared by its users. Tweets about the *PeduliLindungi* application can contain positive or negative opinions.

Sentiment analysis or opinion mining is a study field that analyzes opinions, evaluations, attitudes, also emotions of someone from the written language (Liu, 2012). Sentiment analysis is a computational study in the field of Natural Language Processing (NLP). The sentiment analysis group sentiment in the data of text and divided into three classes, which are positive, negative, or neutral. The approaches used in the sentiment analysis are divided into three types, namely Machine Learning, Lexicon Based, and Hybrid Approach (Elsaid Moussa et al., 2021). This research uses Machine Learning Approach refers to Kumar et al., (2020), because on Twitter the are many non-standard languages, Lexicon Based Method cannot handle different dialects and informal words.

This research classifies opinion about *PeduliLindungi* application into categories of positive, negative, or neutral by using the approach of Machine Learning with the algorithm of Support Vector Machine (SVM). Information is retrieved by crawling Twitter data using the Twitter API Key with the Python programming language. The optimization is done in this research to maximize the performance metrics, which are the values of Accuracy, Precision, Recall and F1-Score from the sentiment analysis that uses the SVM algorithm. The state of the art in this research can be seen in Table 1.

Table 1. State of The Art about Sentiment Analysis Optimization

No	Author & Years	Title	Classifier	Optimization Method	Evaluation
1	Priya & Umamaheswari (2019)	Enhanced continuous and discrete multi objective particle swarm optimization for text summarization	Bag-of-Words model (BoW)	Particle Swarm Optimization (PSO)	ROUGE-n score
2	Noersasongko et al. (2019)	The Use of Particle Swarm Optimization to Obtain N-Gram Optimum Value for Movie Genre Classification	Naïve Bayes (NB)	Particle Swarm Optimization (PSO)	Accuracy (A), Precision (P), and Recall (R)
3	Janjanam & Reddy (2019)	Text Summarization: An Essential Study	Maximum Entropy (ME), Naïve Bayes (NB), Support Vector Machine (SVM)	Nature inspired, Fuzzy, Genetic, Differential Evolution	ROUGE-n score
4	Ramasamy, L. K. et al. (2020)	Selection of optimal hyper-parameter values of support vector machine for sentiment analysis tasks using nature-inspired optimization methods	Support Vector Machine (SVM) with sentiwordnet	Nature-inspired optimization techniques	Accuracy (A)
5	Kumar et al., (2020)	Aspect-based sentiment analysis using deep networks and stochastic optimization	Convolutional Neural Network (CNN) with Word2Vec	Particle Swarm Optimization (PSO)	Accuracy (A), Precision (P), Recall (R), and F-Measure (F)

It can be seen from Table 1, there is no research that discusses the issue of sentiment analysis specifically to determine the weighting problem with Robust Optimization. Therefore, in this study the completion of the multi-objective model in the research of Kumar et al., (2020) is carried out to determine the weighting of the sentiment analysis problem using Robust Optimization. Robust Optimization model is used because it can handle uncertainty factors that represents the worst-case that can happen.

The main discussion in this research is the implementation of Robust Optimization to analyze sentiment analysis on Twitter with a case study of the *PeduliLindungi* application. There will be a reformulation of the multi-objective optimization model by Kumar et al., (2020) and considering the uncertainty of the data. It is assumed that the uncertainty factor of this problem is the value of Accuracy, Precision, Recall and F1-Score. The negative sentiment can be used for the government to evaluate the performance of *PeduliLindungi* application and the positive sentiment can be an appreciation for the government to *PeduliLindungi* application that is being developed. For the researchers, the result of Robust Optimization model for sentiment analysis in this research can be used to evaluate the performance of a classification model.

B. METHODS

In this section, the methods used in this research are discussed. This section provides the theories that are relevant to the research problem, including the theory of Robust Optimization, Multi-objective Optimization and Sentiment Analysis.

1. Research Stages

In this research, several methods is used to solve the problem. Figure 1 is a research flowchart with the methods that used in this research. As shown in Figure 1.

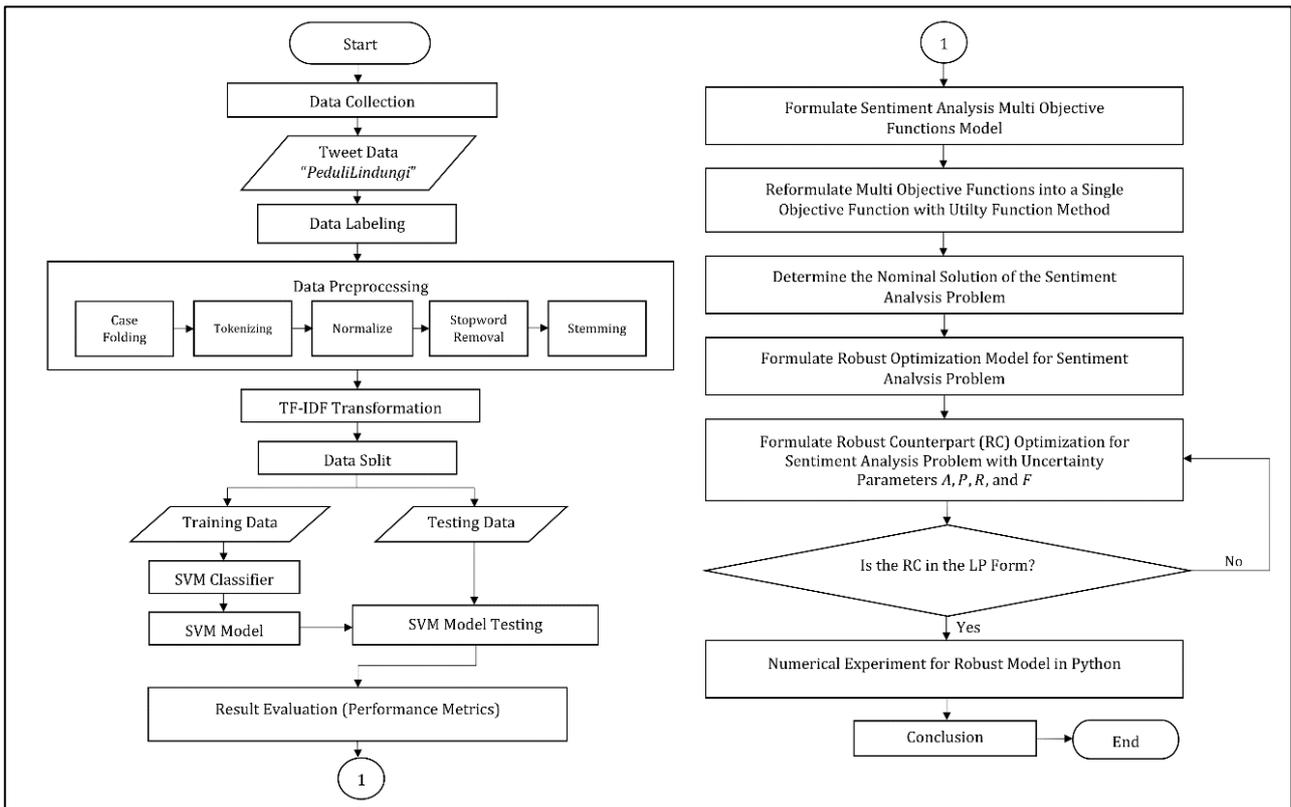


Figure 1. Flowchart of the Research

2. Robust Optimization

A problem relating to uncertainty can be solved by Robust Optimization. According to Ben-Tal & Nemirovski (2002), Robust Optimization is a mathematical methodology that is combined with computation tools in order to process the optimization problem with data uncertainty where the uncertainty can be found in an uncertainty set. This Robust Optimization aims to find a robust solution to the uncertain data on the parameter (Hertog, 2013). According to Ben-Tal and Nemirovski (2002) and discussed by Gorissen *et al.* (2015), the general model of the uncertain data can be formulated as problem (1)

$$\min_x \{c^T x : Ax \leq b \mid (c, A, b) \in \mathcal{U}\} \quad (1)$$

where c, A and b are uncertainty coefficients and \mathcal{U} is the uncertainty set. There are some basic assumptions of Robust Optimization (Ben-Tal *et al.*, 2009) which are as follows:

- a. All decision variables $x \in \mathbb{R}^n$ represent “here and now” decisions, meaning that specific numerical values must be obtained as a result of solving the problem before the actual data “reveals itself”.
- b. The decision-maker is fully responsible for the decision to be made if and only if the actual data has been determined in the uncertainty set \mathcal{U} .
- c. The constraints of an uncertain linear programming problem are “hard”, meaning that the decision-maker cannot tolerate a constraint violation when the data is in \mathcal{U} .

Besides the basic assumptions, it can be assume that the objective function is certain and the constraint on the right-hand side is certain (Gorissen *et al.*, 2015).

- a. The objective function $c^T x$ is certain.
- b. The constraint right-hand side b is certain
- c. *Robustness to \mathcal{U}* can be formulated in constraint-wise form, and the uncertainty set \mathcal{U} is a compact and convex.

Based on given assumptions, the Robust optimization approach changes the uncertain problem to become a single deterministic problem called Robust Counterpart (RC). If assumed $c \in \mathbb{R}^n$ and $b \in \mathbb{R}^m$ are certain, then the Robust Counterpart formulation of (1) is written as problem (2) (Gorissen *et al.*, 2015).

$$\begin{aligned} & \min_x \{c^T x : A(\zeta)x \leq b \mid \zeta \in Z\} \\ \Leftrightarrow & \min_x \{c^T x : a^T(\zeta)x \leq b \mid \zeta \in Z\} \end{aligned} \quad (2)$$

Define the parameter of uncertainty as equation (3)

$$a(\zeta) = a = \bar{a} + P\zeta \quad (3)$$

where $\bar{a} \in \mathbb{R}^n$ is a vector of nominal values and $P \in \mathbb{R}^{n \times L}$ is a perturbation matrix. Define the set \mathcal{U} as equation (4)

$$\mathcal{U} = \{a: a = \bar{a} + P\zeta, \zeta \in Z\}. \tag{4}$$

Substitute the parameter uncertainty equation (3) on the constraint (2)

$$(\bar{a} + P\zeta)^T x \leq b, \forall \zeta \in Z \tag{5}$$

where $(\bar{a} + P\zeta)$ is an affine function over the primitive uncertainty parameters $\zeta \in Z, a \in \mathbb{R}^n$, and $P \in M_{n,L}(\mathbb{R})$. The robust optimal solution is the optimal solution of the Robust Counterpart, and the optimal robust value of an uncertainty linear programming problem is the optimal value of the Robust Counterpart.

According to Ben-Tal & Nemirovski (2002) and Chaerani & Roos (2013), computationally tractable can be analyzed by representing Robust Counterpart in the form of Linear Programming (LP), Conic Quadratic Programming (CQP) and Semidefinite Programming (SDP). Table 2 shows the tractable formulation for constraints with uncertainty sets, as shown in Table 2.

Table 2. Tractable formulation for constraints with uncertainty sets

Uncertainty Set	Z	Robust Counterpart	Tractability
Box	$\ \zeta\ _\infty \leq 1$	$a^T x + \ P^T x\ _1 \leq b$	LP
Ellipsoidal	$\ \zeta\ _2 \leq 1$	$a^T x + \ P^T x\ _2 \leq b$	CQP
Polyhedral	$D\zeta + q \geq 0$	$\begin{cases} a^T x + q^T w \leq b \\ D^T w = -P^T x \\ w \geq 0 \end{cases}$	LP

3. Text Mining

Text mining is a technique that extracts information from structured and unstructured data, then finds the patterns (Kannan et al., 2015). Text preprocessing is an essential process in text mining. In this study, tweet data preprocessing is carried out with the following stages:

- a. Case folding is the process of transforming data into a uniform cases or letters. This process converts all letters to lowercase and removes numbers and punctuation marks.
- b. Tokenizing is splitting sentences into chunks of words called tokens.
- c. Normalization is the process of transforming non-standard text and slang into the standard text.
- d. Stopword removal is the removal of words that have no meaning (usually appear in large numbers).
- e. Stemming is the process of transforming a word into a root word by removing all affixes.

4. TF-IDF

Term Frequency-Inverse Document Frequency (TF-IDF) is a weighting method by combining the term frequency (TF) and Inverse Document Frequency (IDF) methods which illustrates how importance terms are in a corpus (Saadah et al., 2013). The equation for TF-IDF is written in equations (6) and (7).

$$IDF(t) = \log \frac{1 + n}{1 + df(t)} + 1 \tag{6}$$

$$TF - IDF(t, d) = TF(t, d) \cdot IDF(t) \quad (7)$$

where n is the total number of documents and $df(t)$ is the number of terms appearing in all documents. Therefore, TF-IDF is normalized using the Euclidean norm in equation (8)

$$v_{norm} = \frac{v}{\sqrt{v_1^2 + v_2^2 + \dots + v_n^2}} \quad (8)$$

where v_i is TF-IDF of each term in the entire documents.

5. Support Vector Machine (SVM)

Support Vector Machine (SVM) is one of the machine learning techniques used to do classification (Suthaharan, 2016). SVM aims to find the best hyperplane that separates two classes in the input space (Vapnik, 2000). In obtaining the hyperplane on the SVM, can use the following equation:

$$(w \cdot x_i) + b = 0 \quad (9)$$

In the data x_i , which belongs to class -1, is formulated as in equation (10)

$$(w \cdot x_i + b) \leq 1, y_i = -1, \quad (10)$$

while the data x_i which belongs to class +1, is formulated as in equation (11)

$$(w \cdot x_i + b) \geq 1, y_i = 1 \quad (11)$$

The optimal hyperplane search is done by finding the minimum point of the equation (12)

$$\begin{aligned} \min_{\vec{w}} \tau(w) &= \frac{1}{2} \|w\|^2 \\ \text{s. t } y_i(x_i \cdot w + b) - 1 &\geq 0, \forall i \end{aligned} \quad (12)$$

where x_i is input data, y_i is output value, w is a vector normal to hyperplane and b is an offset. Problem (12) can be solved with the Lagrange Multiplier. The optimal hyperplane search problem can be written as equation (13)

$$L(w, b, \alpha) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^l \alpha_i (y_i(x_i \cdot w + b) - 1), i = 1, 2, \dots, l. \quad (13)$$

where $\alpha_i \geq 0$ is the Lagrange multipliers. The optimal equation (13) value can be calculated by minimizing L to w and b while maximizing L to α_i . Since the optimal point of the gradient is $L = 0$, then equation (13) can be modified by maximizing equation (14)

$$\sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j x_i x_j \tag{14}$$

subject to:

$$\begin{aligned} \alpha_i &\geq 0 \quad (i = 1, 2, \dots, l) \\ \sum_{i=1}^l \alpha_i y_i &= 0. \end{aligned} \tag{15}$$

The maximization of (15) produces several α_i which has a positive value. The data associated with positive α_i are called Support Vector (SV) and can be used to find hyperplanes.

6. Performance Metrics

Performance metrics can be described as measuring tools to measure the performance of classifiers (M & M.N, 2015). To calculate performance metrics, the formula as shown in Table 3.

Table 3. Performance metrics

Metric	Definition
Accuracy	$\frac{TP + TN}{TP + TN + FP + FN}$
Precision	$\frac{TP}{(TP + FP)}$
Recall	$\frac{TP}{(TP + FN)}$
F1-score	$\frac{2 \cdot \text{precision} \cdot \text{recall}}{(\text{recall} + \text{precision})}$

where TP, TN, FP and FN can be described as follows:

- a. TP (True Positive) is when the model correctly predicts the positive class.
- b. TN (True Negative) is when the model correctly predicts the negative class.
- c. FP (False Positive) is when the model incorrectly predicts the positive class.
- d. FN (False Negative) is when the model incorrectly predicts the negative class.

7. Multi-objective Optimization

Multi-objective optimization has several objective functions which are subject to some constraints. One of methods to solve multi-objective optimization problem is the Utility Function Method. Multi-objective decision-making problems can be considered as equation (16) (Rao, 2009).

$$\begin{aligned} \max & [f_1(x), f_2(x), f_3(x), \dots, f_k(x)] \\ \text{s. t} & g_i(x) \leq 0, \forall i = 1, 2, 3, \dots, m \end{aligned} \tag{16}$$

The problem in equation (16) has k goals that must be maximized. The utility function will be converted to a U function. U can be defined in many ways, one of the simplest ways is the Weighted Sum Method. The Weighted Sum Method combines all multi-objective functions into

one scalar, combining the objective function with the weighted sum (Yang, 2014). Problem (16) will turn into a problem (17) (Aliakbari & Seifbarghy, 2011).

$$\begin{aligned} & \max U(f_1, f_2, f_3, \dots, f_k) \\ & \text{s. t } g_i(x) \leq 0, \forall i = 1, 2, 3, \dots, m \\ & U(f_1, f_2, f_3, \dots, f_k) = \sum_{j=1}^k x_j f_j \end{aligned} \quad (17)$$

where x_j is a scalar weighting factor associated with the j -th objective function.

8. Optimization Model for Sentiment Analysis

Sentiment analysis as a classification problem uses performance metrics to evaluate performance (Vakili et al., 2020). Performance metrics can be determined by calculating the Accuracy, Precision, Recall and F1-Score values. The problem of sentiment analysis discussed in this study is to determine the weight of each performance metric that can maximize the performance metrics. This problem can be solved by using an optimization model to get optimal results.

The sentiment analysis weighting optimization problem is explained as follows.

- x_i represents the i -th scalar weighting factor, which can maximize the value of Accuracy, Precision, Recall and F1-Score.
- x_i is a weighting factor that shows the relative importance of each performance metric, so the total value of the weighting factor is 1.
- x_i has a minimum value and a maximum value at the closing interval [0,1].

The objective function of Kumar *et al.* (2020) is created to maximize the value of Accuracy, Precision, Recall and F1-Score. Several objective functions are combined into a single objective function by adding the objective function with scalar weights. This problem involves several parameters and variables as follows.

- a : Accuracy
- p : Precision
- r : Recall
- f : F1-Score

where x_i is a scalar weighting factor for each performance metric value, x_1 is weighting factor for Accuracy (a), x_2 is weighting factor for Precision (p), x_3 is weighting factor for Recall (r) and x_4 is weighting factor for F1- Score (f). The objective function is formulated as equation (18).

$$\begin{aligned} & \max a \\ & \max p \\ & \max r \\ & \max f \end{aligned} \quad (18)$$

The multi-objective function in equation (18) is formulated into a single objective function using the Utility Function Method, in order to obtain an optimization model for sentiment analysis as in equation (19) (Kumar et al., 2020)

$$\max g(x) = ax_1 + px_2 + rx_3 + fx_4 \quad (19)$$

$$\begin{aligned} \text{s. t } & \sum_{i=1}^N x_i = 1 \\ & x_i \in [0,1], i = 1, \dots, N. \end{aligned}$$

C. RESULT AND DISCUSSION

In this section the results are presented. Optimization Model Formulation for Sentiment Analysis Weighting Problem is discussed and also the robust version. Numerical experiment shows the validation for the case of *PeduliLindungi*.

1. Optimization Model Formulation for Sentiment Analysis Weighting Problem

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

a_1 : Accuracy

a_2 : Precision

a_3 : Recall

a_4 : F1-Score

A multi-objective function to maximize the value of performance metrics is formulated as equation (20) below.

$$\begin{aligned} & \max a_1 \\ & \max a_2 \\ & \max a_3 \\ & \max a_4 \end{aligned} \tag{20}$$

With the Utility Function Method as state in Rao (2009), the multi-objective function in (20) is reformulated into a single objective function. Then the optimization model formulation in this study is obtained as in equation (21) (Kumar *et al.*, 2020).

$$\begin{aligned} \max g(x) &= \sum_{i=1}^N a_i x_i & \forall a_i \in \mathcal{U} \\ \text{s. t } & \sum_{i=1}^N x_i = 1 \\ & x_i \in [0,1], i = 1, \dots, N. \end{aligned} \tag{21}$$

where x_i represents the i -th scalar weighting factor of each performance metric that indicates the level of importance.

2. Uncertainty Optimization Model Formulation for Sentiment Analysis Weighting Problem

In this study, it is assumed that the uncertainty factors of this problem are a_i which is obtained from the classification process. As it is stated by (Ben-Tal & Nemirovski, 2002), the procedure on implementing Robust Optimization can be done by firstly, assume that the objective function is certain and bounded below by a single variable function

t . This implies that the problem (21) can be formulated such that the uncertainty only appears in the constraint function. The single variable $t \in \mathbb{R}$ is added instead of the objective function.

$$\begin{aligned} \max \quad & t \\ \text{s.t.} \quad & \sum_i a_i x_i \geq t \\ & \sum_i x_i = 1 \\ & x_i \in [0,1]. \end{aligned} \quad (22)$$

Thus, it can be seen in equation (22) that the uncertainty only appears in the constraint function. Defined uncertainty parameter as equation (23)

$$a_i(\zeta) = a_i = \bar{a}_i + P_i \zeta, \forall \zeta \in Z \quad (23)$$

where $\bar{a}_i \in \mathbb{R}^n$ is a nominal value, and $P \in \mathbb{R}^{n \times L}$ is a perturbation matrix. Define the set \mathcal{U} as equation (24) below

$$\mathcal{U} = \{a_i = \bar{a}_i + P_i \zeta \mid \forall \zeta \in Z\}. \quad (24)$$

Substitute the uncertainty of equation (24) on the constraint (22),

$$\begin{aligned} & \sum_i (\bar{a}_i + P_i \zeta) x_i, \quad \forall \zeta \in Z \\ & \sum_i \bar{a}_i x_i + P_i \zeta x_i, \quad \forall \zeta \in Z. \end{aligned} \quad (25)$$

The formulation of the Robust Optimization model for the weighting problem of sentiment analysis is obtained as the following equation (26)

$$\begin{aligned} \max \quad & t \\ \text{s.t.} \quad & \sum_i \bar{a}_i x_i + P_i \zeta x_i \geq t, \forall \zeta \in Z \\ & \sum_i x_i = 1 \\ & x_i \in [0,1]. \end{aligned} \quad (26)$$

3. Robust Counterpart Formulation for Sentiment Analysis Weighting Problem with Polyhedral Uncertainty Sets

Using the steps for defining polyhedral uncertainty set as introduce in Gorissen et al. (2015). Formulation of Robust Counterpart for Sentiment Analysis Weighting Problem with Polyhedral Uncertainty Sets is obtained in three steps:

a. Step 1

The constraint reformulation in equation (25) is equivalent to the worst-case formulation as in equation (27) below.

$$\begin{aligned}
& \sum_i a_i x_i \geq t \\
& \equiv \sum_i (\bar{a}_i + P_i \zeta) x_i \geq t \\
& \equiv \sum_i \bar{a}_i x_i + P_i \zeta x_i \geq t \\
& \equiv \sum_i \bar{a}_i x_i + \max_{\zeta: d-D \geq 0} P_i x_i \zeta \geq t
\end{aligned} \tag{27}$$

b. Step 2

Formulate the dual form of the maximization problem in equation (27). The reformulated form of equation (27) is obtained to be of the following form.

$$\begin{aligned}
& \sum_i \bar{a}_i x_i + \max_{\zeta: d-D \geq 0} P_i x_i \zeta \geq t \\
& \equiv \sum_i \bar{a}_i x_i + \min_{y_i} \{d_i y_i : D_i y_i = P_i x_i, y_i \geq 0\} \geq t \\
& \equiv \sum_i \bar{a}_i x_i + d_i y_i \geq t, D_i y_i = P_i x_i, y_i \geq 0
\end{aligned} \tag{28}$$

c. Step 3

The formulation of the Robust Counterpart model with polyhedral uncertainty set for the weighting problem of sentiment analysis is as equation (29) below.

$$\begin{aligned}
& \max t \\
& \text{s.t. } \sum_i \bar{a}_i x_i + d_i y_i \geq t \\
& \quad D_i y_i = P_i x_i, y_i \geq 0 \\
& \quad \sum_i x_i = 1 \\
& \quad x_i \in [0,1]
\end{aligned} \tag{29}$$

The Robust Counterpart of uncertain Linear Programming (LP) problem is computationally tractable (Ben-Tal et al., 2009). It can be seen that equation (29) is in the form of LP, so that the formulation of the Robust Counterpart model with a polyhedral uncertainty set is a computationally tractable problem, meaning that it can be solved computationally in polynomial time.

4. Support Vector Machine Classification

In this section, Support Vector Machine is classified. In this step, the tweets data are labeled manually. In this study, three classes of sentiment are used; positive, negative and neutral. The number of tweet data label can be seen in Table 4.

Table 4. Number of Tweets

Sentiment	Number of Tweets
Positive	630
Negative	578
Neutral	792
Total	2000

It can be seen from Table 4 that neutral sentiment has a greater number, followed by positive and negative sentiment. The criteria for each class are as follows:

- Positive class, contains tweet responses that support or praise the *PeduliLindungi* application.
- Negative class, contains tweet responses that contrast to the *PeduliLindungi* application.
- Neutral class, contains tweet responses that do not contain both positive or negative class.

The data classification process begins by dividing the data into training data and testing data. The ratio that commonly used is 80:20, which means that 80% of the data is for training and 20% is for testing (Joseph, 2022). Table 5 below shows the results of a 3x3 confusion matrix consisting of predict class and actual class.

Table 5. Confusion Matrix

		Predict Class		
		Negative	Neutral	Positive
Actual Class	Negative	61	47	12
	Neutral	21	113	20
	Positive	9	36	81

From Table 5, it can be calculated the value of the performance metrics from the SVM classification is obtained as shown in Table 6.

Table 6. Performance Metrics

Metric	Value
Accuracy	0,6375
Precision	0,66
Recall	0,63
F1-Score	0,64

5. Numerical Experiment

By using the Python programming language, the optimal solution of the deterministic model for the weighting problem of sentiment analysis is obtained, as shown in Table 7.

Table 7. Numerical Experiment for Deterministic Model

i	a_i	x_i	$a_i x_i$
1	0,6375	0	0
2	0,66	1	0,66
3	0,63	0	0
4	0,64	0	0
$g(x)$			0,66

After that, the optimal solution is sought from the Robust Counterpart model for the weighting problem of sentiment analysis with polyhedral uncertainty, as shown in Table 8.

Table 8. Numerical Experiment for Robust Model

i	a_i	x_i	$a_i x_i$
1	0,6375	0,3566	0,2273
2	0,66	0,3456	0,2281
3	0,63	0,2769	0,1744
4	0,64	0,02095	0,0134
$g(x)$			0,6433

The numerical experiment result with the deterministic model has a larger objective function value than the Robust Counterpart model with polyhedral uncertainty. The result difference is due to the Robust Counterpart Optimization model with polyhedral uncertainty considering the uncertainty factors of the performance metrics value.

D. CONCLUSION AND SUGGESTIONS

The form of Robust Counterpart formulation with polyhedral uncertainty set for the weighting problem of sentiment analysis can be seen as Linear Programming (LP). It can be concluded that Robust Counterpart formulation with polyhedral uncertainty set for the weighting problem of sentiment analysis is computationally tractable, i.e. the model can be solved computationally in polynomial time. The performance of the SVM model used has a value 0.6433 with the Robust Counterpart model with polyhedral uncertainty. The results of numerical experiment use the Robust Counterpart model with polyhedral uncertainty have a smaller objective function value than the deterministic model which has value 0.66, because the Robust Counterpart model maximizes the performance metrics by handling uncertainty factors that represents the worst that can happen. Further research can be carried out using the Robust Counterpart optimization model with other sets of uncertainties, the box uncertainty set and the ellipsoidal uncertainty set (Gorissen et al., 2015). Different classification methods can be used for sentiment analysis in order to know the comparison between the performance of other algorithms.

ACKNOWLEDGEMENT

This research is supported by Universitas Padjadjaran and the Indonesian Ministry of Education, Culture, Research, and Technology under project with Basic Research Scheme 2022 entitled "Adjustable Robust Counterpart Optimization Model and Social Media Analysis for Internet Shopping Online Problem" under contract number 2064/UN6.3.1/PT.00/2022.

REFERENCES

- Aliakbari, A., & Seifbarghy, M. (2011). A Supplier Selection Model for Social Responsible Supply Chain. *Journal of Optimization in Industrial Engineering, Volume 4(8)*, 41–53. http://www.qjie.ir/article_86.html
- Ben-Tal, A., Ghaoui, L. El, & Nemirovski, A. (2009). *Robust Optimization: Princeton Series*. Princeton University Press.
- Ben-Tal, A., & Nemirovski, A. (2002). Robust Optimization-Methodology and Applications. *Mathematical Programming, 92(3)*, 453–480. <https://doi.org/10.1007/s101070100286>
- Chaerani, D., & Roos, C. (2013). Handling Optimization under Uncertainty Problem Using Robust Counterpart Methodology. *Jurnal Teknik Industri, 15(2)*. <https://doi.org/10.9744/jti.15.2.111-118>
- Dixon, S. (2022). *Leading countries based on number of Twitter users as of January 2022*. Statista. <https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/>
- Elsaid Moussa, M., Hussein Mohamed, E., & Hassan Haggag, M. (2021). Opinion mining: a hybrid framework based on lexicon and machine learning approaches. *International Journal of Computers and Applications, 43(8)*, 786–794. <https://doi.org/10.1080/1206212X.2019.1615250>
- Firdaniza, F., Ruchjana, B., Chaerani, D., & Radianti, J. (2021). Information Diffusion Model in Twitter: A Systematic Literature Review. *Information, 13(1)*, 13. <https://doi.org/10.3390/info13010013>
- Gorissen, B. L., Yanikoğlu, İ., & den Hertog, D. (2015). A practical guide to robust optimization. *Omega, 53*, 124–137. <https://doi.org/10.1016/j.omega.2014.12.006>
- Hertog, D. den. (2013). *Practical Robust Optimization: an Introduction*.
- Janjanam, P., & Reddy, C. H. P. (2019). Text summarization: An essential study. *ICCIDS 2019 - 2nd International Conference on Computational Intelligence in Data Science, Proceedings*, 1–6. <https://doi.org/10.1109/ICCIDS.2019.8862030>
- Joseph, V. R. (2022). Optimal ratio for data splitting. *Statistical Analysis and Data Mining: The ASA Data Science Journal, 1–8*. <https://doi.org/10.1002/sam.11583>
- Kannan, S., Gurusamy, V., Vijayarani, S., Ilamathi, J., & Nithya, M. (2015). Preprocessing Techniques for Text Mining Preprocessing Techniques for Text Mining. *International Journal of Computer Science & Communication Networks, 5(October 2014)*, 7–16. <https://doi.org/2249-5789>
- 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>
- Kurniawati, Khadapi, M., Riana, D., Arfian, A., Rahmawati, E., & Heriyanto. (2020). Public Acceptance Of Pedulilindungi Application In The Acceleration Of Corona Virus (Covid-19) Handling. *Journal of Physics: Conference Series, 1641(1)*, 012026. <https://doi.org/10.1088/1742-6596/1641/1/012026>
- Liu, B. (2012). *Sentiment Analysis and Opinion Mining*. Morgan & Claypool.
- M, H., & M.N, S. (2015). A Review on Evaluation Metrics for Data Classification Evaluations. *International Journal of Data Mining & Knowledge Management Process, 5(2)*, 01–11. <https://doi.org/10.5121/ijdkp.2015.5201>
- Priya, V., & Umamaheswari, K. (2019). Enhanced continuous and discrete multi objective particle swarm optimization for text summarization. *Cluster Computing, 22(S1)*, 229–240. <https://doi.org/10.1007/s10586-018-2674-1>
- Rao, S. S. (2009). *Engineering Optimization: Theory and Practice*. John Wiley & Sons, Inc.
- Saadah, M. N., Atmagi, R. W., Rahayu, D. S., & Arifin, A. Z. (2013). Information Retrieval Of Text Document With Weighting TF-IDF and LCS. *Jurnal Ilmu Komputer Dan Informasi, 6(1)*, 34.

<https://doi.org/10.21609/jiki.v6i1.216>

Suthaharan, S. (2016). *Support Vector Machine* (pp. 207–235). Springer.

https://doi.org/10.1007/978-1-4899-7641-3_9

Vakili, M., Ghamsari, M., & Rezaei, M. (2020). *Performance Analysis and Comparison of Machine and Deep Learning Algorithms for IoT Data Classification*.

<https://doi.org/https://doi.org/10.48550/>

Vapnik, V. N. (2000). *The Nature of Statistical Learning Theory*. Springer New York.

<https://doi.org/10.1007/978-1-4757-3264-1>

Yang, X.-S. (2014). Multi-Objective Optimization. In *Nature-Inspired Optimization Algorithms* (pp. 197–211). Elsevier. <https://doi.org/10.1016/B978-0-12-416743-8.00014-2>