

# Ant Colony Algorithm for Determining Dynamic Travel Routes Based on Traffic Information from Twitter

Edy Subowo<sup>1,\*</sup>, Eko Sedyono<sup>2</sup>, and Farikhin<sup>3</sup>

<sup>1</sup> Magister Program of Information System, School of Postgraduate Studies, Diponegoro University, Semarang - Indonesia

<sup>2</sup> Faculty of Information Technology, Satya Wacana Christian University, Salatiga, Indonesia

<sup>3</sup> Department of Mathematics, Faculty of Science and Mathematics, Diponegoro University

**Abstract.** Combining the search method for fire suppression routes with ant colony algorithms and methods of analyzing twitter events on the highway is the basis of the problems to be studied. The results of the twitter data feature extraction are classified with Support Vector Machine after it is implemented in the Simple Additive Weighting method in calculating path weights with criteria of distance, congestion, multiple branching, and many holes. Line weights are also used as initial pheromone values. The C-means method is used to group the weights of each path and distance so that the path with the lowest weight and the shortest distance that will be simulated using the Ant Colony. The validation results with cross fold on SVM with linear kernels produce the greatest accuracy value is 97.93% for training data distribution: test data 6: 4. The simulation of the selection of the damkar car path from Feather to Pleburan with Ant Colony obtained 50 seconds of computation time, whereas with Ant Colony with Clustering the computation time was 15, resulting in a reduction in computing of 35. Ant colony with MinMax optimization gives the best computation time of 14.47 seconds with 100 iterations and 10 nodes.

**Keywords:** Data Mining; Ant Colony Clustering; Support Vector Machine.

## 1 Background

Highway is one of the infrastructures that connects one place to another. In the current modern era with the high demand for roads and the large number of road users, highway management is needed to respond and detect quickly and precisely the dynamic phenomena that occur [1]. Road management includes road maintenance management and traffic management for rapid and appropriate handling of traffic jam phenomena [2]. Roads are dynamic, because road events such as traffic jams caused by traffic signs, road construction, traffic accidents, driving style of drivers change with time, therefore traffic management is needed to solve these problems [3].

One of the traffic management arrangements is the management of fire engine routes. This management system must be able to present the fastest route from the location of the pool of fire engines based on distance, congestion value, number of road holes and branching, so that the calculation of the weight value is needed to obtain the optimal solution [4]. Algorithms for determining the fastest route include ant colony, djikstra, tabu search and warshall where from the four algorithms ant colony gets a near-optimal solution [5]. Because ant colony has quite a lot of complexity, computation time is quite long, so it requires path clustering for algorithm optimization [6].

Visibility values can be used to give ant directions for choosing the desired route [7]. The calculation results of SAW weight values can be used to determine the value of visibility. Therefore, accurate congestion data is needed so that firefighters can avoid traffic jams [8]. One provider of congestion data that is quite accurate and is real-time in nature is twitter data. In 2017, the number of active Twitter users reached 330 million and 80% of active users on mobile devices [9]. Twitter is flexible where Twitter users are able to create, consume, promote, distribute, get and share information with their communities so that large clusters of data are needed to reduce errors in data processing [10]. Previously, a classification analysis of twitter data with various methods such as naïve bayes, Maximum Entropy, or Support Vector Machine was conducted. It was concluded that the Support Vector Machine (SVM) method gave the best results compared to other methods, with accuracy up to 82, 2% [11]. SVM can be applied to handwriting recognition and text categorization, and SVM can avoid the difficulties of dimensionality problems [12].

The merger of the search methods for fire truck travel lines with the ant colony with clustering algorithm with the method of analyzing data on tweets about events on the road becomes the basis of the problems to be studied. Collection of twitter data detected based on predetermined keyword tables will be extracted based on syllables and the time the tweet was made. Furthermore,

\* Corresponding author: [edy.subowo@gmail.com](mailto:edy.subowo@gmail.com)

the system will classify the event with the Support Vector Machine algorithm after being implemented on the simple additive weighting method in calculating the criteria for distance, congestion, multiple branching, and holes. Ant colony with clustering algorithm is used to get the optimal path with twitter data in realtime traffic conditions and displayed with a mobile-based interface.

## 2 Literature review

The twitter data mining process consists of information extraction, information summarization, and document classification [13]. Previously, sentiment analysis studies on Twitter have been conducted with various methods such as Naïve Bayes Classification, Maximum Entropy, or Support Vector Machine. It was concluded that the use of the Support Vector Machine (SVM) method gave the best results compared to other methods, namely the accuracy of up to 82.2% [11]. Research related to traffic congestion in real-time in Chicago and New York (USA) is presented by Jorge et al [14] to extract information from the user's geotag location, time, date, and content of tweet text.

The use of twitter media as an ingredient for the case of land occupancy based on weather conditions with the algorithm naïve bayes classifier and SVM presented by Marchel et al [15], obtained results where the SVM method has a greater level of performance than naive bayes with a difference in accuracy of 0.0015 due to inability to naive bayes in overcoming data outlier problems. There are previous studies that also use 2 algorithms namely Naïve Bayes Classifier and SVM presented by Hassan et al [16]. This research was carried out under text enrichment through Wikitology using data from 20 newsgroups with 1000 categories, the results showed that the more training data, the greater the accuracy of SVM, the more features and external enrichment for labeling, the greater the accuracy of Naïve Bayes.

The ant colony with clustering algorithm is presented by Gao et al [17] where immigrant schemes are proposed to address dynamic location routing (DLRP) problems. DLRP is divided into two parts formed by location-allocation problems (LAP) and vehicle routing problems (VRP) in a dynamic environment. To handle LAP, the C-Means clustering algorithm is used to handle the location of nodes and surrounding cities in each class. Immigrant schemes are used to overcome dynamic location routing (DLRP) problems with ant colony with clustering (KACO), the results of the study show that compared to normal ACO algorithms such as US, EAS, MMAS, ACS, P-ACO, SA and GA, KACO produce better performance.

This study will use the approach taken by Gao et al. [17] to address the problem of dynamic travel route selection and in-depth learning to detect traffic accidents from social media data as done by Hasan et al. [16] using twitter data around events on the road with methods Support Vector Machine. The time-based data will be

weighted by the SAW method so that the visibility value is obtained as the basis for calculating the probability node  $P_{ij}^k(t)$  to determine the fastest route in real-time with the Ant Colony algorithm and applied to the problem of fire truck transportation.

### 2.1 Twitter Application Programming Interface

To facilitate interaction between developers and services, Twitter has provided an API that can be accessed publicly via the internet.

### 2.2 Support Vector Machine

Support Vector Machine (SVM) is one of the classification methods using machine learning (supervised learning) that predicts classes based on models or patterns from the results of the training process. Classification is done by finding a decision boundary that separates a class from another class, in this case the line acts to separate positive-sent tweet (labeled +) with negative-sent tweet (labeled -) [18] In this SVM the first is training (learning) which is to study the text patterns on tweets into a model, according to the reference given, then the second step is prediction (testing) which uses a model to classify text and the results are predictions of classification. In the design, the tweets and labels will be divided into training and testing data as shown in Figure 1. Then after training, proceed with a prediction process that will result in system accuracy. Figure 2 shows the Hyperplane is formed between class-1 and +1, where

$$\text{Margin} = \frac{2}{\|w\|} = \frac{2}{\sqrt{w_1^2 + w_2^2}} \quad (1)$$

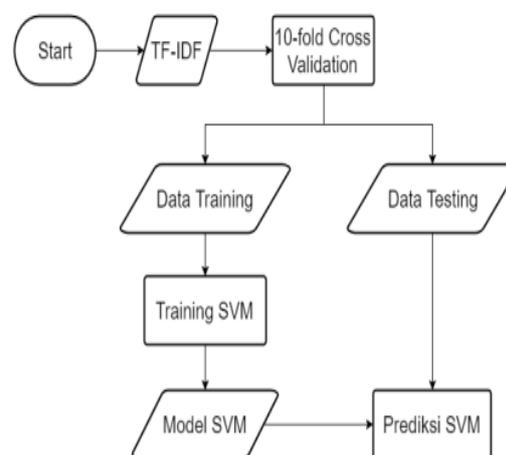
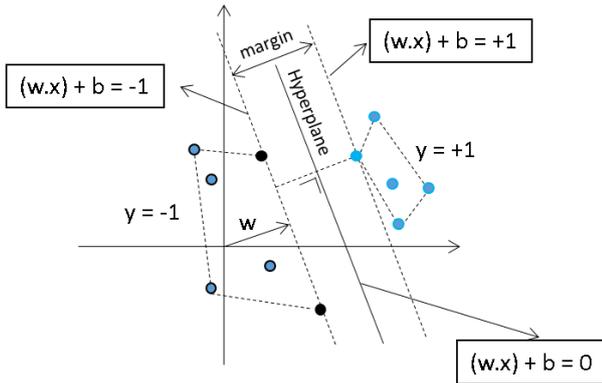


Fig. 1. Classification with Support Vector Machine [18]

### 2.3 SAW Method

The Simple Additive Weighting (SAW) method is often also known as the weighted sum method. The basic concept of the SAW method is to find a weighted sum of performance ratings on each alternative on all attributes. The SAW method requires the process of normalizing

the decision matrix (X) to a scale that can be compared with all available alternative ratings. The formula for normalizing is shown in equation 2 as follows:



**Fig. 2.** Hyperplane is formed between class -1 and +1 [19]

$$r_{ij} = \frac{x_{ij}}{\text{Max}_i x_{ij}} \quad (2)$$

Where  $r_{ij}$  is the normalized performance rating of the alternative  $A_i$  in the  $C_j$  attribute;  $i = 1, 2, \dots, m$  and  $j = 1, 2, \dots, n$ . The preference value for each alternative ( $V_i$ ) is given by equation (3). A larger  $V_i$  value indicates that the alternative is more chosen.

$$V_i = \sum_{j=1}^n w_j r_{ij} \quad (3)$$

## 2.4 Ant Colony algorithm

Dorigo and Colorni, called the ant system, were first applied to the problem of travel salesman (TSP) in 1991 [20] In general, the ACO algorithm for TSP follows the algorithm scheme [21] as follows:

procedure ACO Algorithm for TSPs

Determining parameters, initialize Pheromone trails

**while** (termination condition not met) do

ConstructAntSolutions

ApplyLocalSearch (optional)

UpdatePheromones

**Endwhile**

**End procedure**

Ant (k) will move from node (i) to (j) with the probability  $P_{ij}^k(t)$  namely:

$$P_{ij}^k(t) = \frac{[\tau_{ij}(t)]^\alpha [\eta_{ij}]^\beta}{\sum [\tau_{ik}(t)]^\alpha [\eta_{ik}]^\beta} \quad (4)$$

Where  $\tau_{ij}$  is the number of pheromones found at the edge between point i and point j,  $\eta_{ij}$  is visibility,  $\alpha$  is a parameter that controls the relative weight of Pheromone and  $\beta$  is the distance controller parameter ( $\alpha > 0$  and  $\beta > 0$ ). In selecting the next node (s) the equation (5) is used

$$s = \text{argmax}_{u \in J_i^k} \{ [\tau_{ij}]^\alpha [\eta_{ij}]^\beta \} \quad (5)$$

Furthermore, the global pheromone intensity in all arcs is updated using equation (6):

$$\tau_{ij}(t+n) = (1-\xi)\tau_{ij} + \xi\tau_{ij}^k \quad (6)$$

with  $\xi$  is the rate of local pheromone evaporation, and n is the ant cycle. Furthermore, the intensity of the local pheromone in all arcs is updated using equation (7):

$$\tau_{ij}(t+n) = \rho\tau_{ij}(t) + \Delta\tau_{ij}^k \quad (7)$$

With  $\rho$  is the rate of evaporation of the global pheromone, and  $\Delta\tau_{ij}^k = Q/L_k$  is the pheromone issued by ant k with  $L_k$  is the length of the ant line k and Q is a constant.

## 2.5 Elitist Ant System (EAS)

The additional intensity of Pheromone from Tbs tour is

$$\frac{e}{C^{bs}}$$

with gives additional quantity for each edge, where e parameters given to define the best tour value (Tbs) and  $C^{bs}$  is the best tour length. Pheromone changes are defined

$$\tau_{ij}(t+n) = \rho\tau_{ij}(t) + \Delta\tau_{ij}^k + \Delta\tau_{ij}^{bs} \quad (8)$$

The advantages of EAS are as follows:

- The number of iterations is done less, this is because the edges on the best tour get more Pheromone additions and cause the search space to be narrower.
- The space for finding solutions is narrower, making it faster to find solutions. Although the search space is more perfect but EAS does not show stagnation, it means that it continues to look for the possibility of a better new tour.

## 2.6 MAX – MIN Ant System (MMAS)

The MAX-MIN Ant System (MMAS) is a development of the next US algorithm, with several main changes as follows

- First, adding Pheromone can be done on the edges which are part of the best tour found at the start of the algorithm (best so-far-tour) or on the best tour found in the iteration (best-tour iteration). It could also be the addition of Pheromone to both, the best so-far-tour and the best-tour iteration at the same time.
- Second, to overcome the problem of the first change, the MMAS gives a limit in giving Pheromone values with  $[\tau \text{ min}, \tau \text{ max}]$  intervals.
- Third, initialize Pheromone with the upper limit of Pheromone's value, which together with Pheromone's small evaporation rate, has increased tour exploration since the start of the search.
- Fourth, Pheromone is re-initialized when there is stagnation or when no tour is found that matches the desired iteration.

### 3 Research Methods

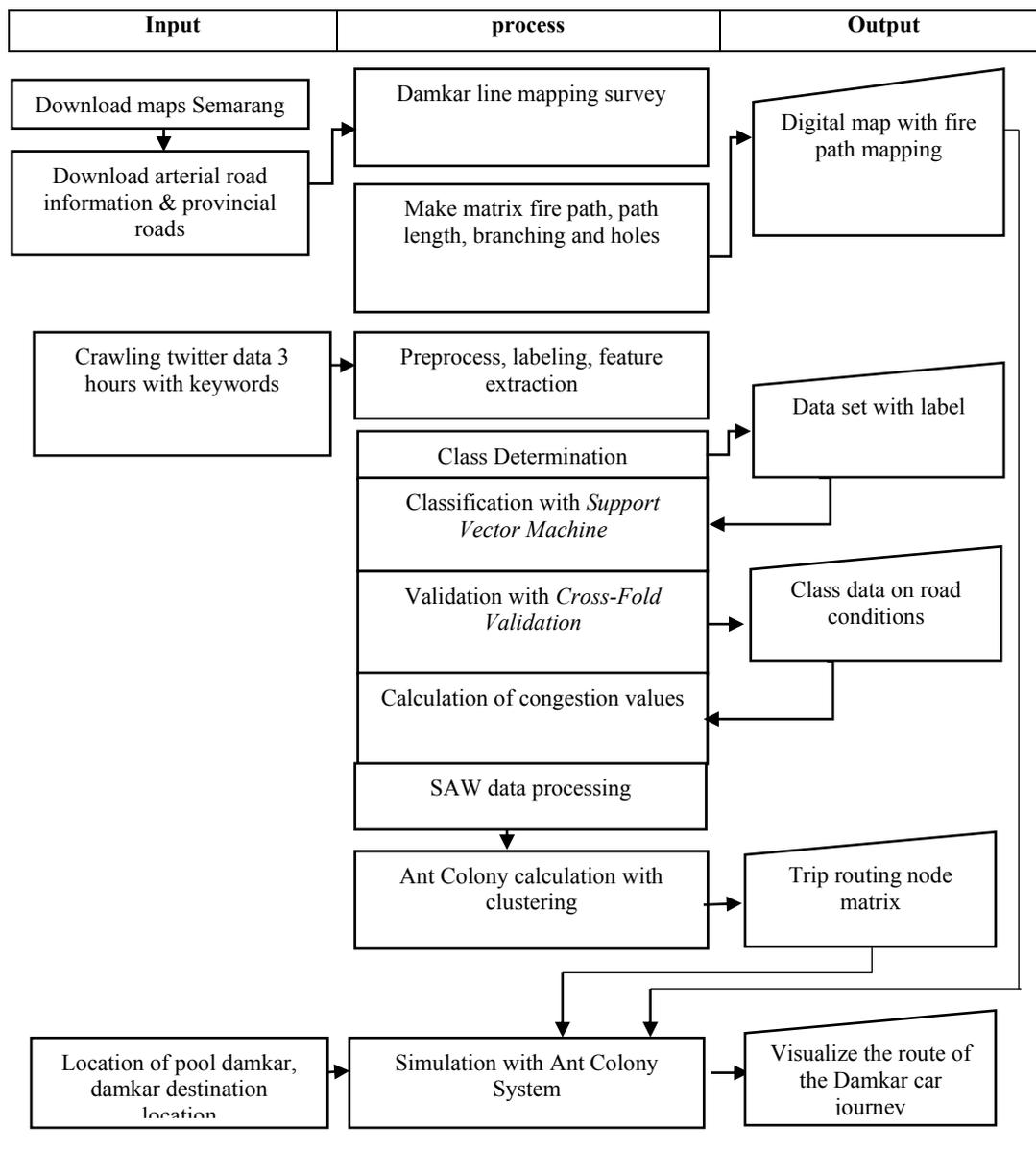


Fig. 3. Research procedure

#### 3.1 Railroad Car Route Routing

Data on road names and distance traveled by each road (km) that can be traversed by assumptions on arterial roads and provincial roads and survey of homebase damkar coordinates. Distance and one-way data for each street name were obtained from the Semarang City Transportation Office, while data on the number of Branches and holes were obtained from field observations on January 1, 2019. Table 1 shows examples of main and alternative fire trucks where the initial location is on Jalan Madukoro or the location of the pool fire engine and the destination is Jalan Imam Barjo, Pleburan Semarang.

Table 1. Description of road mapping

Rute	Node	Nama jalan	Jarak (km)	Percabangan	Lubang
A	1 → 2	Jl Raya Madukoro	0,35	1	2
	2 → 3	Jl Lingkar madukoro	1	1	0
	3 → 4	Tugu Muda	0,3	2	0
	4 → 5	Jl Pandanaran	0,75	1	1
	6 → 7	Simpang Lima	0,5	3	0
	5 → 6	Jl pahlawan	0,5	2	1
	8 → 9	Jl Imam Barjo	0,3	2	0
	B	1 → 2	Jl Raya Madukoro	0,35	1
2 → 3		Jl Lingkar madukoro	1	1	0
3 → 4		Tugu Muda	0,3	2	0
4 → 5		Jl Sutomo	1,4	1	3
5 → 6		Jl Veteran	1,3	1	2

### 3.2 Research Procedure for Twitter Data Classification

Tweet data obtained through the Twitter API with information tokens and APIkey. Data taken based on keywords is used to provide information about the traffic of the City of Semarang, namely #LalinSMG. The training tweet data taken is 500 tweets while the test data is 50 tweets automatically when reloading the page. Data validation is done by displaying location data (geotag) where only the location of Semarang City is taken for further data stored in the database. Pre-processing steps can be seen in Figure 4 as follows

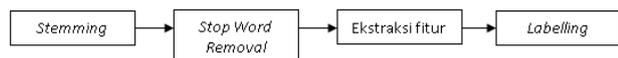


Fig. 4. Text Pre-processing

Twitter API is used for the twitter data retrieval process. The results of twitter data collection can be seen in Figure 5

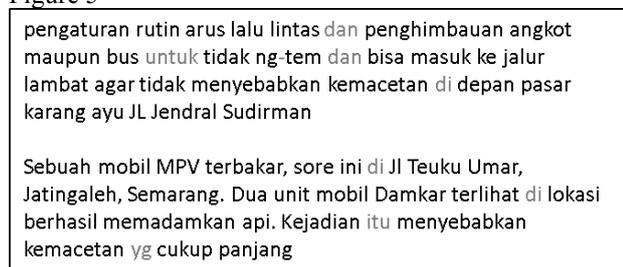


Fig. 5. Twitter data

Then the stemming process is carried out by breaking each word and removing the @ symbol, #, number. The stop word removal process is then carried out to eliminate the word stone, conjunctions so that feature extraction can be performed. Next is the feature extraction process to get the data results needed for classification.

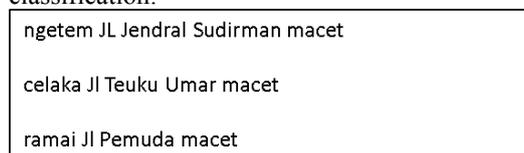


Fig. 6. Results of the feature extraction process

Labeling is then done to describe the classes. The first label shows the occurrence of a highway, the second label shows the condition of the highway. Figure 3.7 shows the labeling criteria obtained from the extraction process of 500 training tweets, values (+) show obstacles in the highway, while values (-) have no obstacles, while Table 2 shows an example of the results of labeling as follows:

Table 2. Labeling Criteria

Incident	Condition	Class	Info
ngetem	lancar	-1	tidak macet
ngetem	macet	1	macet
ngetem	padat	1	macet
celaka	lancar	-1	tidak macet
celaka	macet	1	macet
celaka	padat	1	macet
bakar	lancar	-1	tidak macet
bakar	macet	1	macet
bakar	padat	1	macet
mogok	lancar	-1	tidak macet
mogok	macet	1	macet
mogok	padat	1	macet
lengang	lancar	-1	tidak macet
lengang	macet	-1	tidak macet
lengang	padat	-1	tidak macet
merayap	lancar	-1	tidak macet
merayap	macet	-1	tidak macet
merayap	padat	-1	tidak macet

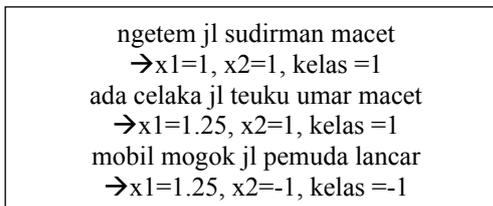
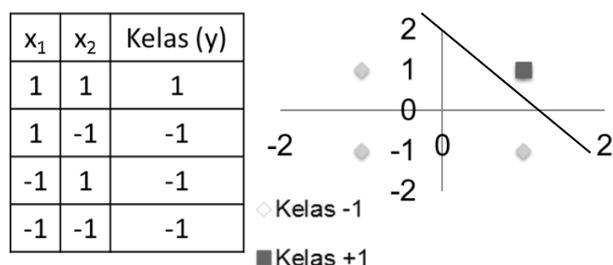


Fig. 7. Labeling results

### 3.3 Classification with Support Vector Machine and Validation with Cross Fold Validation

The classification process with linear SVM can be seen in Figure 8 as follows:



**Fig. 8.** Linear classification with Support Vector Machine

The purpose of the validation is to obtain the composition of training data and test data that provide the most optimal accuracy results, called the fold value. In Table 3 above, true positive (TP) is the number of positive records that are classified as positive, false positive (FP) is the number of negative records classified as positive, false negatives (FN) are the number of positive records classified as negative, true negatives (TN) is the number of negative records classified as negative.

**Table 3.** Examples of validation results

Classification	Predicted class	
	Class=Yes	Class=No
Class=Yes	a (true positive)	b (false negative)
Class=No	c (false positive)	d (true negative)

### 3.4 Determination of Congestion Values and Weighting with SAW

Table 1 shows the congestion value (nk) will be calculated as the number of tweets in class + in the specified data set (TPx) compared to the number of tweets + on all data sets ( $\sum TP_x$ ). x is the name of the street

$$nk = \frac{TP_x}{\sum TP_x} \quad (9)$$

Table 4 shows the value of the NK as an input for the calculation of SAW using data from the example of the main and alternative fire trucks from Jalan Madukoro to Jalan Imam Barjo where there are two tweets (+) at Tugumuda and Jl Pandanaran. The road mapping scenario used by the SAW method to do road weights by calculating the distance, congestion and position criteria.

**Table 4.** Congestion Value (NK) Damkar Travel Path

Route	Node	street	Dist (km)	Street brances	Street holes	NK
A	1 → 2	Jl Raya Madukoro	0,35	1	2	0
	2 → 3	Jl Lingkar madukoro	1	1	0	0
	3 → 4	Tugu Muda	0,3	2	0	0,5
	4 → 5	Jl Pandanaran	0,75	1	1	0,5
	6 → 7	Simpang Lima	0,5	3	0	0
	5 → 6	Jl pahlawan	0,5	2	1	0

B	8 → 9	Jl Imam Barjo	0,3	2	0	0
	1 → 2	Jl Raya Madukoro	0,35	1	2	0
	2 → 3	Jl Lingkar madukoro	1	1	0	0
	3 → 4	Tugu Muda	0,3	2	0	0,5
	4 → 5	Jl Sutomo	1,4	1	3	0
	5 → 6	Jl Veteran	1,3	1	2	0
	6 → 7	Jl pahlawan	1,2	2	0	0
	7 → 8	Jl Imam Barjo	0,3	2	0	0
C	1 → 2	Jl Raya Madukoro	0,35	1	2	0
	2 → 3	Jl Basudewo	1,5	1	2	0
	3 → 4	Jl Kaligarang	0,5	1	1	0
	4 → 5	Jl Sutomo	0,4	1	2	0
	5 → 6	Jl Veteran	1,3	1	2	0
	6 → 7	Jl pahlawan	1,2	2	0	0
	7 → 8	Jl Imam Barjo	0,3	2	0	0
	D	1 → 2	Jl Raya Madukoro	0,35	1	2
2 → 3		Jl Indraprasta	1	1	1	0
3 → 4		Jl Pemuda	0,3	2	3	0
4 → 5		Jl Gajahmada	1,4	1	2	0
6 → 7		Jl pahlawan	1,2	2	0	0
7 → 8		Jl Imam Barjo	0,3	2	0	0

In general, the system made in this study to find the best route based on 4 criteria, namely distance, many holes, lots of branching, and congestion value (NK). The workings of this system in general are:

- the determination of the initial node is the DAMKAR homebase until the destination node
- The ranking of pathways with SAW is based on 4 criteria and C-Means 2 cluster centers are clustered with distance as X axis and weighted ranking with SAW as Y axis.
- The path chosen for the next process (Ant Colony) is a path that is close to the cluster of the shortest path and ranks the lowest weight.
- Weighting value with SAW will also be used as a visibility value as the basis for calculating the

probability node  $P_{ij}^k(t)$  to determine the fastest route in real-time with the Ant Colony algorithm.

### 3.5 Clusterization calculation of the final matrix preference value

Clustering is based on the distance of each node to optimize computing time. The preference matrix calculation uses the C-Means algorithm (Gao, 2016) as follows:

#### C-Means Algorithm

- 01 enter the clustering value  $k = 1$  and  $n$  data
- 02: while  $k \leq \sqrt{n}$
- 03: for  $i = 1, 2, \dots$ , bro
- 04: random initiation of cluster centers
- 05: end for
- 06: while termination criteria not satisfied do
- 07: calculate distance and class, calculate the average distance in each class, value the new cluster center
- 08: end while
- 09: calculate the distance function F and save
- 10:  $k = k + 1$
- 11: end while specifying the optimal value  $k$  based on the minimum F value

**Ant Colony with C-Means Algorithm**

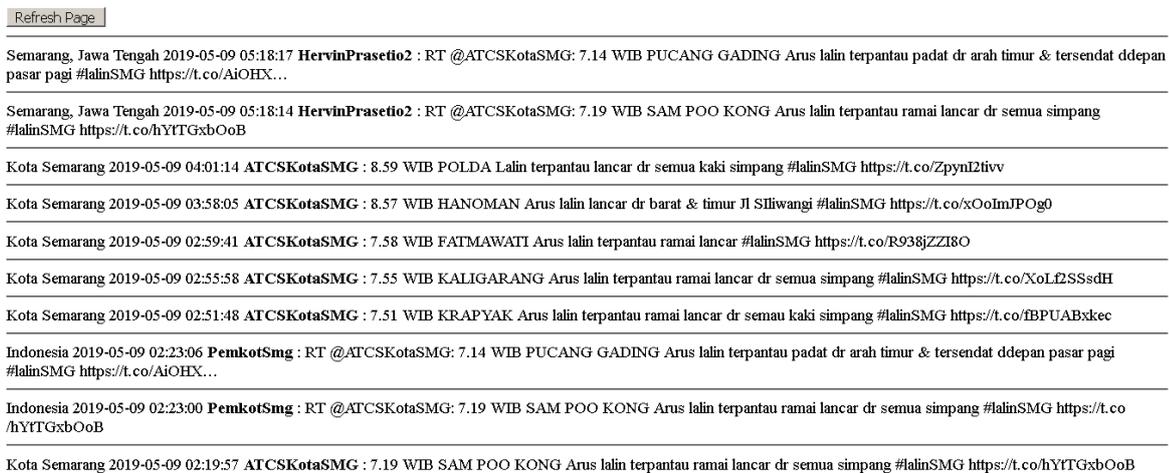
- 01: initiation of parameters  $\alpha$ ,  $\beta$ ,  $\tau_0$ ,  $K_s$ ,  $K_l$ ,  $N$ ,  $M$ ,  $C$ ,  $R$ ,  $r$ ,  $t$
- 02: specify  $k$  depot and class based on Algorithm 1
- 03: for  $i = 1, 2, \dots, bro$
- 04: while termination criteria not satisfied do
- 05: create a dynamic environment
- 06: clear routes,  $k_{short}$ ,  $k_{long}$ , Allocate lots of  $M$  ants at depot  $i$
- 07: if the city is not passed by then
- 08: if  $R \leq C$  then, create a new route with equation (2.1), update  $C$
- 09: else return to step 07
- 10: end if
- 11: end if
- 12: specify the best results iteration  $K_s$  as  $K_{short}$
- 13: rank  $r \times K_s$  with the SAW method for  $k_{short}$  updates

- 14: Update pheromone with equation (2.2)
- 15: compare the best solutions for the best iterations so far
- 16:  $t = t + 1$
- 17: end while
- 18: end for

**3.5 Research result**

Twitter data retrieval is done by initiating the twitteroAuth library, followed by determining the keywords and authentication of the twitter API. The training tweet data taken is 500 tweets while the test data is 20 tweets automatically when reloading the page as shown in Figure 9. The data stored in the database is extracted features for SVM classification

Tweet dengan Keyword #alainSMG



**Fig. 9.** Results of twitter data collection

The feature extraction process begins with retrieving data from the database and activating stemmer to break each word, change to lowercase letters and stopword removal to remove symbols, conjunctions and prepositions. The literary library is used to carry out the stemming process.

The process is continued with feature extraction and labeling. Feature extraction is done by calculating the frequency of occurrence of words related to highway events such as fire, accident, crowds and road conditions such as congestion, silence, etc. The words have been changed to become the basic words in the stemming process, making it easier to find the frequency of data. Table 5 shows the frequency of occurrence of words from 500 tweets.

**Table 5.** Frequency of Word Emergence Feature Extraction Results

Highway events		Highway condition	
word	frequency	word	frequency
Ngetem	3	lancar	215
Celaka	7	macet	144
Bakar	1	padat	27
Mogok	2		
Lengang	8		
Merayap	1		
Demo	1		
Ribut	2		
Perbaikan	55		

Labeling is done by searching for three different objects, including:

- a. Tweets relating to the name of the main street of Semarang city which number 23 street names are symbolized as  $j_1, j_2, \dots, j_{23}$ .

- b. Tweets related to highway events are symbolized as k1, k2, ..k6.
- c. Tweets related to road conditions are symbolized as ko1, ko2, ko3.

The extraction process of the feature is used to label using `preg_match()`

The data obtained in the labeling process shows the number of tweets related to congestion in 23 street names according to the criteria in Table 2. There is also a process to anticipate tweets related to the mention of places, for example, Sampookong's name is defined as Jl Kaligarang, krapyak with Jl Siliwangi, Fatmawati for Jl Majapahit, and Central Java Regional Police as Jl Pahlawan as shown in Table 6

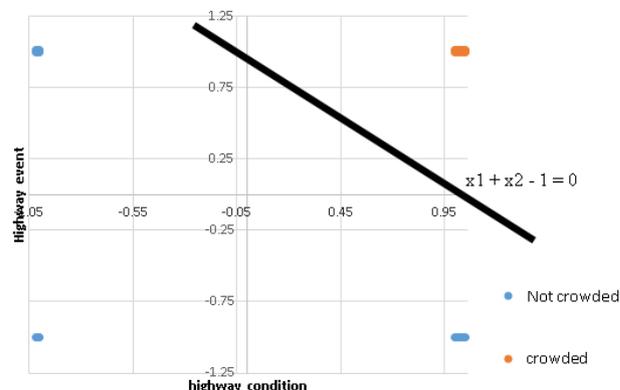
**Table 6.** Frequency of Labeling Values on Each Road

Street	Tweet +1	Tweet -1	Nilai Kemacetan
Jl Kaligarang	0	3	0
Jl Pengapon	0	1	0
Jl Soekarno	0	0	0
Jl Indraprasta	1	0	1/3 = 0.33
Jl Pemuda	1	0	1/3 = 0.33
Jl Pahlawan	1	1	1/3 = 0.33
Jl Majapahit	0	1	0

Table 6 shows that we can determine the congestion value (NK) where NK is based on the frequency of occurrence of tweets + (jammed). The classification with SVM starts with giving values to the event and road conditions as in Table 2, giving the value of each word using an interval of 0.01 to get a good data density as shown in Table 7. The table data is then plotted into a graph so that the data distribution is obtained with two classifications that have been determined in the labeling process, namely the clause is not jammed and the clause is jammed. The data distribution forms a data distribution pattern where data between classes has a distance that can be separated by a line. the line divides the hyperplane into two classes as shown in Figure 10.

**Table 7.** Classification data with SVM

event	X value	condition	Y value	class	information
ngetem	1	lancar	-1	-1	tidak macet
ngetem	1	macet	1	1	macet
ngetem	1	padat	1.01	1	macet
celaka	1.01	lancar	-1	-1	tidak macet
celaka	1.01	macet	1	1	macet
celaka	1.01	padat	1.01	1	macet
bakar	1.02	lancar	-1	-1	tidak macet
bakar	1.02	macet	1	1	macet
bakar	1.02	padat	1.01	1	macet
mogok	1.03	lancar	-1	-1	tidak macet
mogok	1.03	macet	1	1	macet
mogok	1.03	padat	1.01	1	macet
lengang	-1	lancar	-1	-1	tidak macet
lengang	-1	macet	1	-1	tidak macet
lengang	-1	padat	1.01	-1	tidak macet
merayap	-1.01	lancar	-1	-1	tidak macet
merayap	-1.01	macet	1	-1	tidak macet
merayap	-1.01	padat	1.01	-1	tidak macet
demo	1.04	lancar	-1	-1	tidak macet
demo	1.04	macet	1	1	macet
demo	1.04	padat	1.01	1	macet
ribut	-1.02	lancar	-1	-1	tidak macet
ribut	-1.02	macet	1	-1	tidak macet
ribut	-1.02	padat	1.01	-1	tidak macet
perbaikan	1.05	lancar	-1	-1	tidak macet
perbaikan	1.05	macet	1	1	macet
perbaikan	1.05	padat	1.01	1	macet



**Fig. 10.** Classification with SVM Linear

Evaluation of the classification results is done using the cross fold method so that the confusion matrix is obtained. Confusion matrix is a tool used as an evaluation of classification models to estimate objects that are right or wrong. The process flow begins by reading the \*.csv data file then the data is divided between training data and test data to find predictive values and accuracy. In this study, there are 2 folds so that the 2x2 confusion matrix was obtained to obtain different accuracy values for each distribution of training data and test data in accordance with Table 8 as follows

**Table 8.** Value of Accuracy of Training Data Distribution and Test

Division	4:6	5:5	6:4	7:3	8:2
Accuracy	0.9452	0.9752	<b>0.9793</b>	0.9726	0.9591

Table 8 shows the greatest accuracy value is 97.93% for training data distribution: test data of 6: 4 and confusion matrix values as in Table 9 below:



**Debug Info**  
 Mula Hapus

**Konfigurasi**  
 Pilihan Mode Ant Colony Optimization: Max-Min

Max Pheromone: (1 \* Bobot Pheromone Deposit) / Jarak Terbaik

Min Pheromone: Max Pheromone \* Scaling Factor

Min Scaling Factor: 0.001

Ukuran Koloni Semut: 30

Banyak Iterasi: 200

Nilai Alpha: 1

Nilai Beta: 3

Nilai Rho: 0.1

Bobot Pheromone Deposit: 1

**Tweet dengan Keyword #lalinSMG**  
 Refresh Page

Kota Semarang 2019-05-13 07:53:55 ATCSKotaSMG : 12.51 WIB Arus lalin di simp. Thamin/ Pandanaran tepantau ramai bergerak lancar dr segala arah, patuhi rambu" lali... <https://t.co/Gsp5nCYBhg>

Kota Semarang 2019-05-13 06:26:45 ATCSKotaSMG : 11.24 WIB Arus lalin di Bundaran TUGUMUDA terantau ramai relatif lancar dari segala kaki simpang, patuhi rambu" la... <https://t.co/YQUv0Qyub>

Semarang, Indonesia 2019-05-13 04:53:57 Iwan\_madari : RT @ATCSKotaSMG: 09.43 WIB Arus lalin di simp. Krappyk terantau ramai lancar dr Jl. Siliwangi #lalinSMG <https://t.co/SD75pfd4jh>

Kota Semarang 2019-05-13 04:45:03 ATCSKotaSMG : 09.43 WIB Arus lalin di simp. Krappyk terantau ramai lancar dr Jl. Siliwangi #lalinSMG <https://t.co/SD75pfd4jh>

Kota Semarang 2019-05-13 03:16:00 ATCSKotaSMG : 08.13 WIB Arus lalin di simpang Exit TOL Gayamsari terantau ramai padat bergerak pelan dr timur menuju ke barat, a... <https://t.co/k6Q4NNyeR6>

Kota Semarang 2019-05-13 02:23:38 ATCSKotaSMG : 07.21 WIB Arus lalin di simp. Kalbanteng terantau ramai bergerak lancar KBM didominasi dari barat menuju Jl. Jend... <https://t.co/MZhuS21IzL>

Kota Semarang 2019-05-13 02:15:15 ATCSKotaSMG : 07.15 WIB Arus lalin di simpang Pucang Gading terantau ramai bergerak lancar, KBM di didominasi dr timur menuju pa... <https://t.co/bZQxQEXtRO>

Kota Semarang 2019-05-13 02:14:45 ATCSKotaSMG : 07.14 WIB Arus lalin di kawasan Terminal Mangkang pagi ini terantau ramai lancar dr barat menuju pusat kota, arah... <https://t.co/gaFivGYIX>

2019-05-12 03:00:34 tio89092746 : RT @ATCSKotaSMG: 08:56 WIB lalin simp Polda terantau lancar. Ngantuk? mensepilah... #LalinSMG <https://t.co/ZOP1hbnk0>

Kota Semarang 2019-05-11 04:06:16 ATCSKotaSMG : 08:56 WIB lalin simp Polda terantau lancar. Ngantuk? mensepilah... #LalinSMG <https://t.co/ZOP1hbnk0>

ramai gerak lancar kbm dominasi dari barat tuju jl jend <https://t.co/mzhuS21IzL>  
 08.13 wib arus lalin di simpang exit tol gayamsari pantau ramai padat gerak pelan dr timur tuju ke barat a <https://t.co/k6q4nnyer6>  
 09.43 wib arus lalin di simp krappyk pantau ramai lancar dr jl siliwangi lalin <https://t.co/5d75pfd4jh> rt atcskotasmg 09.43 wib arus lalin di simp krappyk pantau ramai lancar dr jl siliwangi lalin <https://t.co/5d75pfd4jh>  
 11.24 wib arus lalin di bundar tugumuda pantau ramai relatif lancar dari segala kaki simpang patuh rambu la <https://t.co/yquv0qyub>  
 12.51 wib arus lalin di simp thamin pandanaran tepantau ramai gerak lancar dr segala arah patuh rambu lali <https://t.co/gsp5nCyBhg>

Nama Jalan	Tweet +	Tweet -
Jl. Kaligawe	0	0
Jl. Veteran	0	0
Jl. Mangkang	0	1
Jl. Pamulasih	0	0
Jl. Simpang Lima	0	0
Jl. MT Haryono	0	0
Jl. DI. Panjaitan	0	0
Jl. RA. Kartini	0	0
Jl. Tawangari	0	0
Jl. Mahakoro	0	0
Jl. Sudarso	0	0
Jl. Gayamsari	1	0
Jl. Mpu Tantular	0	0
Jl. Kaligarang	0	3
Jl. Pengapon	0	3
Jl. Soekarno	0	0
Jl. Indraprasta	0	0
Jl. Pemuda	0	1
Jl. Sutomo	0	0
Jl. Pahlawan	0	3
Jl. Majapahit	0	1

Fig. 12. User Interface

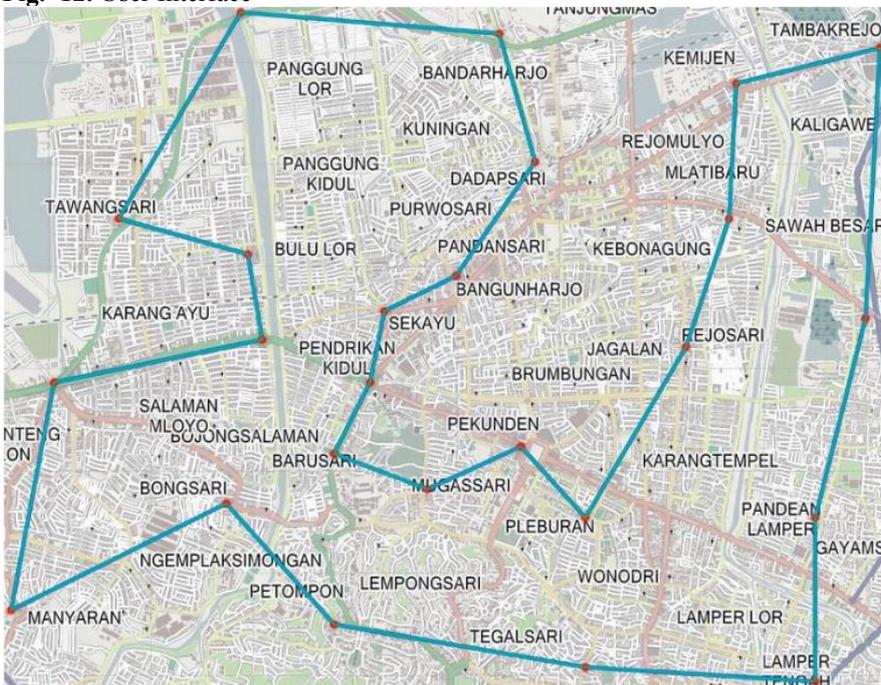
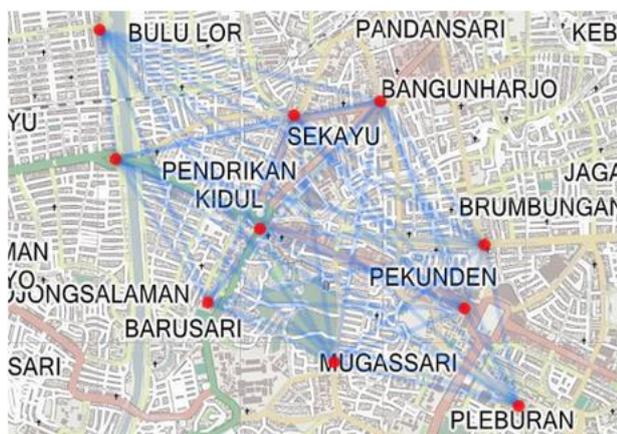


Fig. 13. Route and alternative routes



**Fig. 14.** Route with C-Means classification

Using C-Means Clustering the number of nodes is reduced from 25 nodes to 10 nodes and can provide a computation time of 15 seconds so that by using the ant colony path clustering it can reduce computing by 35 seconds or 51%. Here are the simulation parameters. Simulation has been continued with compare ant colony path clustering with MinMax and Elitis improvement.

- Many iterations = 100
- Size of ant colony / iteration = 30
- Alpha value = 1
- Beta value = 3
- rho value = 0.1
- Weight of pheromone deposit = 1
- Scale factor (MinMax) = 0.001
- elitist weight (elitist) = 2

Figure 15 shows the results of the simulation with ant colony, the blue line shows the best tour results while the green line shows the results of unfinished tours (there are remaining pheromones). With 100 iterations, it can be seen that in ACS there are still pheromones that connect Feather, Kiduldan Sekayu Education, whereas in Elitis there are still pheromones in Sekayu and Pendrikan kidul, while MinMax can produce the best tour. Figure



**Fig. 15.** Comparison of the results of ACS, Elitis and MinMax

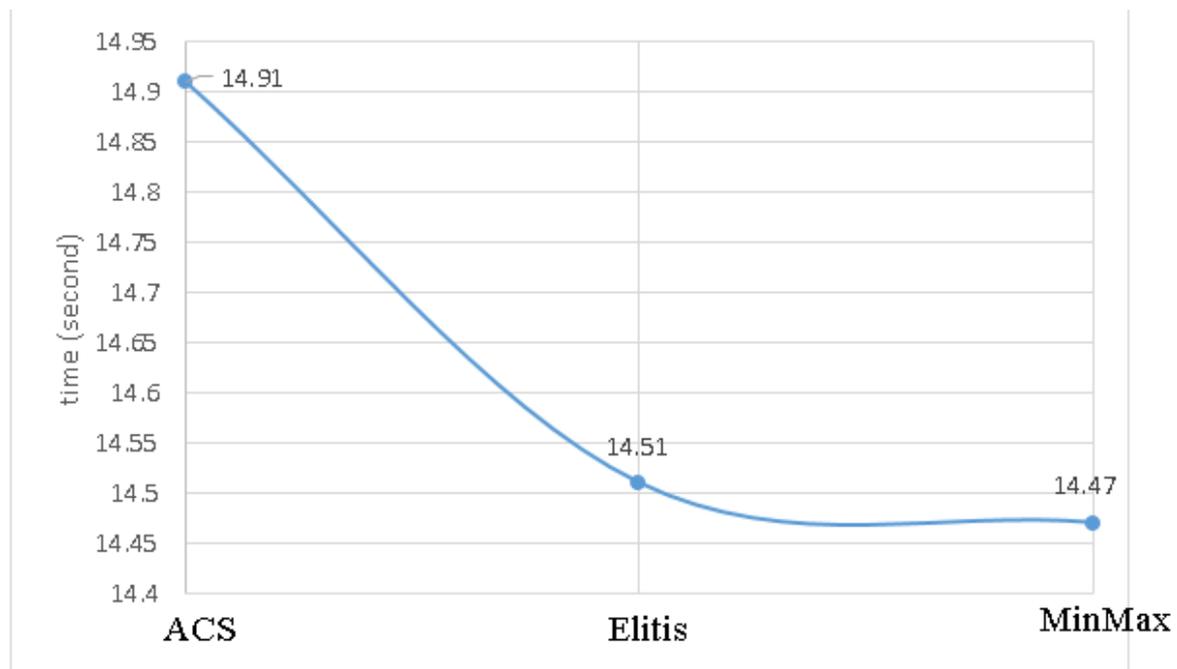
16 shows the time of the simulation process and it can be seen that the ant colony with MinMax optimization gives the best computation time of 14.47 seconds with 100 iterations and 10 nodes

### 3.6 Conclusion

From the discussion described in the previous chapter, a conclusion can be drawn about the Ant Colony Clustering algorithm by weighting SAW on the twitter data. The extraction of 207 data twitter features taken from May 2, 2019 to June 30, 2019 with the #lalinSMG hashtag is the most "smooth" with 111 tweets. The SVM classification of twitter data generates liner Hyperplane, while the validation results with cross fold produce the greatest accuracy value is 97.9% for training data distribution: test data is 6: 4 with 2 prediction errors stuck (+1) in the class not jammed (-1) in the confusion matrix. The simulation of the selection of damkar car paths from Feather to Pleburan with Ant Colony obtained the shortest distance of 19.9 km and 50 seconds of computation time, while with Ant Colony with Clustering the computation time was 15 seconds and the number of nodes reduced from 25 nodes to 10 nodes computing for 35 seconds or 51%. Ant colony with MinMax optimization gives the best computation time of 14 seconds with 100 iterations and 10 nodes.

For the future works, the data obtained now is the sampling data provided by Twitter. To get all the national-scale tweets, you need access to the Firehose API that requires a lot of money so you can get more number of keyword events and road conditions. In addition, in this study the number of ant colony iterations is input from the system so it is expected that there are calculations to get the number of iterations automatically when the system reaches the best route.

\* Corresponding author: [edy.subowo@gmail.com](mailto:edy.subowo@gmail.com)



**Fig. 16.** Processing time from ACS, Elitis, and MaxMin with 100 number of iterations

## References

- Q. He, Y. Kamarianakis, K. Jintanakul, L. Wynter, *Incident duration prediction with hybrid tree-based quantile regression*, *Advances in Dynamic Network Modeling in Complex Transportation Systems*, Springer, 287–305. (2013)
- N. Sinsu, E. Sedyono, & A. P. Widodo, *Implementation of Web-based Analytic Network Process for Highway Maintenance Planning*, *International Journal of Computer Applications*, (0975 – 8887) Volume 181 – No.4 (2018)
- K. Fu, C.-T. Lu, R. Nune, & J. X. Tao, *Steds: Social media based transportation event detection with text summarization*, *IEEE 18th international conference on intelligent transportation systems (ITSC)* Las Palmas, Spain, 1952–1957 (2015)
- T. Pamula, *Road traffic parameters prediction in urban traffic management systems using neural network*, *Transport Problems*, 6, 123-128. (2011).
- E. Bonabeau, M. Dorigo, G. Theraulaz, *Inspiration for optimization from social insect behaviour*, *Nature* 406 (6791) 39–42 (2011)
- Gao S., Yirui W., Jiujun C., Yasuhiro I., Zheng T., *Ant colony optimization with clustering for solving the dynamic location routing problem*, *Applied Mathematics and Computation* 285 149–173 (2016)
- Z. C. S. S. Hlaing & M. A. Khine, *An ant colony optimization algorithm for solving traveling salesman problem*, In *International Conference on Information Communication and Management*, **16**, 54-59 (2011)
- C. O. Panyuwa, *Forecasting the Case of Traffic Accidents Through The Geographic Information System (GIS) Application Method with Double Exponential Smoothing and Analytical Hierarchy Process (AHP) in City of Jayapura Papua*, Doctoral dissertation, Magister Sistem Informasi Program Pascasarjana FTI-UKSW (2016)
- Twitter corporation, [www.twitter.com](http://www.twitter.com)
- A. Gal-Tzur, S. M. Grant-Muller, T. Kuflik, E. Minkov, S. Nocera, I. Shoor, *The potential of social media in delivering transport policy goals*, *Transport Policy* 32, 115–123 (2014)
- Go. A, R. Bhayani, and L. Huang, *Twitter Sentiment Classification using Distant Supervision*, Stanford, Project Report CS224N (2009)
- L. J. Wee, T. W. Tan, & S. Ranganathan, *SVM-based prediction of caspase substrate cleavage sites*, In *BMC bioinformatics*, BioMed Central, **7**, 5 p. S14. (2006)
- Han J, M. Kamber, *Data Mining : Concepts and Techniques.*: Series Editor Morgan Kaufmann (2000)
- Gonzalez Paule, Jorge David, et al., *Learning to geolocalise tweets at a fine-grained level*, *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*. ACM, (2018).
- Y. Marchel & J. Nasri, *Perbandingan Tingkat Akurasi Support Vector Machine Dengan Naive Bayes Pada Studi Kasus Okupansi Lahan Berdasarkan Kondisi Cuaca*, *eProceedings of Engineering*, 4(3). (2017)
- Hassan, S., Rafi, M., & Shaikh, M. S., *Comparing svm and naive bayes classifiers for text categorization with wikilogy as knowledge enrichment*. In *Multitopic Conference (INMIC)*, 2011 IEEE 14th International pp. 31-34 (2011)

17. Gao S., Yirui W., JiuJun C., Yasuhiro I., Zheng T., *Ant colony optimization with clustering for solving the dynamic location routing problem*, Applied Mathematics and Computation 285 149–173 (2016)
18. Han, Feng, et al., *A two-stage approach to people and vehicle detection with HOG-based SVM*, Performance Metrics for Intelligent Systems 2006 Workshop. (2006).
19. R. Feldman & J. Sanger, *The Text Mining Handbook: Advanced Approaches to Analyzing Unstructured Data*.
20. Dorigo, Marco, and Luca Maria Gambardella, *Ant colony system: a cooperative learning approach to the traveling salesman problem*, IEEE Transactions on evolutionary computation, 1.1 (1997): 53-66.
21. M. Dorigo, *Ant colony optimization*. Scholarpedia, 2(3), 1461. (2007)