The Effect of Pre-Processing on the Classification of Twitter's Flood Disaster Messages Using Support Vector Machine Algorithm

Mera Kartika Delimayanti
Department of Computer and Informatics Engineering
Politeknik Negeri Jakarta
Depok, Indonesia
mera.kartika@tik.pnj.ac.id

Risna Sari
Department of Computer and Informatics Engineering
Politeknik Negeri Jakarta
Depok, Indonesia
risna.sari@tik.pnj.ac.id

Mauldy Laya
Department of Computer and Informatics Engineering
Politeknik Negeri Jakarta
Depok, Indonesia
mauldy.laya@tik.pnj.ac.id

M.Reza Faisal
Department of Computer Science
Lambung Mangkurat University
Banjarbaru, Indonesia
reza.faisal@ulm.ac.id

Pahrul
Department of Computer Science
Lambung Mangkurat University
Banjarbaru, Indonesia
pahrul27@gmail.com

Rizqi Fitri Naryanto
Division of Mechanical Science and Engineering, Kanazawa University
Kanazawa, Japan
rizqi.fitri@stu.kanazawa-u.ac.jp

2020 3rd International Conference on Applied Engineering (ICAE) | 978-1-7281-9917-7/20/$31.00 ©2020 IEEE | DOI: 10.1109/ICAE50557.2020.9350387

Abstract— **The eyewitness message on twitter as a social network sensor aims to determine the classification process's performance. In the classification of flood disaster messages, preprocessing data is required before the classification process is carried out. Preprocessing affects the resulting level of accuracy in the classification process. Stopword removal is part of preprocessing so that the effect of stopword removal on classification performance will be examined. Support Vector Machine (SVM) is used to classify by weighting words using Term Frequency-Inverse Document Frequency (TF-IDF). The data taken from Twitter is 3000 data with 1000 data each for each label. The effect of the stopword on accuracy performance can be seen in several experiments that have been carried out. We had already conducted three different experiments, and the highest level of accuracy was 76.6%, 76.87%, and 77.87%. Based on the experiments that have been carried out, stopword is very influential on the accuracy generated by the classification of flood disaster messages on Twitter.**

Keywords—flood disaster messages, twitter, classification, stopword, support vector machine algorithm

I. Introduction

Social media had been proven to have an impact in accelerating the spread of news. Social media users tell what they see immediately after seeing the incident. This makes a lot of research conducted using social media data as Social Network Sensors or social media as sensors [1][2] and the use of social media as sensors for early warning and monitoring after natural disasters [3][4]. Natural disasters that can be handled by social network sensors include earthquakes, floods, tornados and forest fires.

A flood is an event when water immerses an area that is usually not flooded for a certain period of time. Flood usually occurs because rainfall continues to fall and result in overflowing river, lake, sea, or drainage water due to the amount of water that exceeds the capacity. The overflow of water caused by natural factors, namely high rainfall, and flooding also occurs due to human activity[5].

Flood disaster is one of the annual disasters, according to the National Disaster Management Agency (BNPB), there have been 111 flood events in Indonesia recorded during January 2020. Flooding can also have adverse impacts on the social and economic side of society. Therefore, almost everyone is looking for information of when and where the flooding occurred. People will find out about flooding in several ways, including listening to the news on television, in newspapers and on the radio. The research above has proven that social network sensors are a substitute for humans and conventional sensors to warn the public and monitor the situation when a natural disaster occurs.

The information from the social network sensor used in this study is from Twitter social media as the data source. Tweet data contain information on floods as information from observers and eyewitnesses. Information from eyewitness reports is preferred to other sources of information (for example, people outside the disaster area). Law enforcement agencies and first responders always look for first hand (eyewitness) and reliable information.

Previous studies have been conducted [6] to understand the various types of eyewitness reports consisting of three classes (i) eyewitnesses, (ii) non-eyewitnesses, and (iii) Do not know. The study classifies eyewitness messages on Twitter based on classes that have been determined. Meanwhile, Support Vector Machine (SVM) is a proper technique in the case of classification and regression. In the study [7], Support Vector Machine (SVM) was used for sentiment classification on Instagram comments, TF-IDF feature extraction and Support Vector Machine (SVM) classification algorithm. The highest accuracy of 90% was obtained from these tests. From various research references that have been carried out, the Support Vector Machine (SVM) algorithm is one of the classification algorithms that can be used to classify eyewitness messages during a flood disaster. Fig. 1 shows the flow diagram of this research.

II. Material and Method

A. Social Network Sensor

Social Network Sensor is a new concept that comes from physical sensors. This concept is an attempt to bring the concept of real censorship to cyberspace through social media. Real censorship in the real world environment is the same as social censorship on social media, and they all

978-1-7281-4203-6/20/$31.00 ©2020 IEEE

produce large amounts of data. Physical sensors measure and produce physical or chemical data, such as light, heat, temperature, and humidity. Meanwhile, social network sensors collect data from social media in the form of tweets, users, locations, etc. More specifically, examples of physical sensors include luminosity sensors, temperature sensors, etc., while social sensors may include language sensors and text sensors [4].

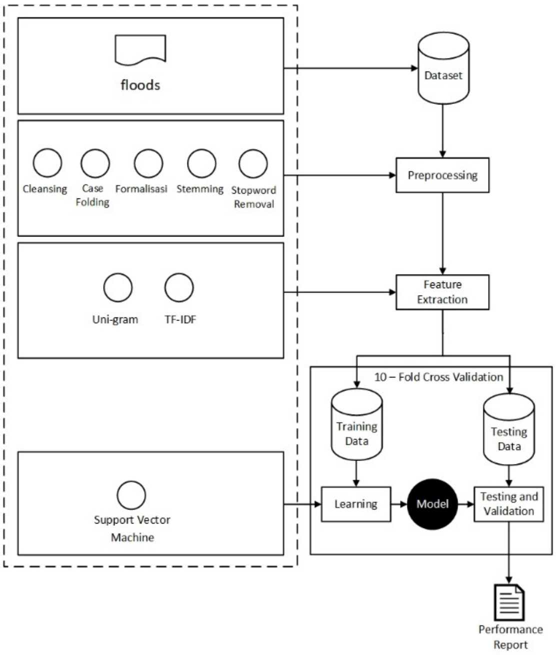


Fig. 1. The flow diagram of this research.

B. Text Mining

Text mining is a mining technique done by a computer to get something new, something that is not known before or rediscover information implicitly, derived from information extracted automatically from different text data sources. Text Mining explores to extract useful information from data sources through the identification and exploration of interesting patterns. Text Mining tends to lead to the field of data mining research. Therefore, it is not surprising that Text Mining and Data Mining are at the same architectural level [8].

C. Dataset

The dataset used in this research is tweet about floods taken from Twitter. The process of collecting the necessary tweets uses the Twitter Scraper package, one of the Python packages, in order to be able to scrap data from Twitter. The amount of data used is 3000 data. The total data came from the flood's disaster message from twitter located in Indonesia and using Bahasa Indonesia. The total data was exposed in Table I, and it is shown that there were 3300 data, and after the pre-processing, we used 3000 data for the next process.

D. Pre-Processing

Preprocessing is the initial stage of text mining to change data following the required format. This process is carried out to explore, process and manage information and analyze the textual relationship of structured and unstructured data [9]. The following processes had been executed in this research as follows:

- Data Labelling

The data scraped from Twitter was continued by doing data labeling. Additionally, those data were divided into three classes, namely (i) Eyewitnesses, (ii) non-eyewitnesses, and (iii) Do not know. Labeling is done manually using the characteristics.

- Remove Duplicate

Removing the duplicate data is the process of deletion of the same data. For example, in data retrieval, there are similarities of data taken in the scraping process

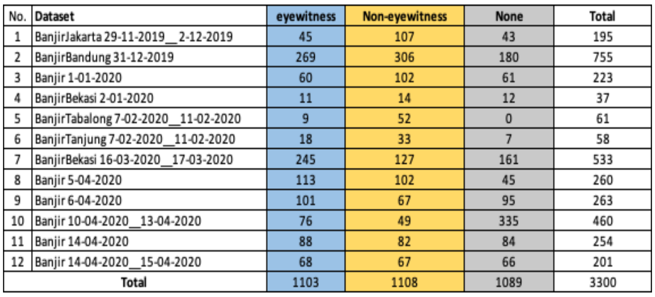
- Data Cleansing

Data cleansing removes characters that do not contribute to sentiment analysis, leaving only the alphabet characters. The cleansing process removes the username, hashtag, URL, RT, symbols / characters such as ("'+ =! &? * ^ ~ # -_), and numbers. Table II reveals the table of the example data from dataset after the data cleansing. The result of Data Cleansing was revealed in Table II.

- Case Folding

Case folding aims to change every word form to be he same form. The case folding was done by changing the word to a lower case or lowercase letters. The results of case-folding are shown in Table III.

TABLE I. TOTAL DATA



No.	Dataset	eyewitness	Non-eyewitness	None	Total
1	BanjirJakarta 29-11-2019__2-12-2019	45	107	43	195
2	BanjirBandung 31-12-2019	269	306	180	755
3	Banjir 1-01-2020	60	102	61	223
4	BanjirBekasi 2-01-2020	11	14	12	37
5	BanjirTabalong 7-02-2020__11-02-2020	9	52	0	61
6	BanjirTanjung 7-02-2020__11-02-2020	18	33	7	58
7	BanjirBekasi 16-03-2020__17-03-2020	245	127	161	533
8	Banjir 5-04-2020	113	102	45	260
9	Banjir 6-04-2020	101	67	95	263
10	Banjir 10-04-2020__13-04-2020	76	49	335	460
11	Banjir 14-04-2020	88	82	84	254
12	Banjir 14-04-2020__15-04-2020	68	67	66	201
Total		1103	1108	1089	3300

TABLE II. DATA CLEANSING

Text	Label
Depan rumah dh banjir Huhu	Eyewitness
Hujan lebat gais banjir disini huhu	Eyewitness
banjir dimanaa macet dimanaa hufff	Eyewitness
Takutnya banjir	Eyewitness
Semoga para korban banjir diberikan ketabahan	Non-Eyewitness
cm di Jl Gerbang Pemuda Senayan bagi kendaraan sejenis sedan dihimbau agar tidak melintas	Non-Eyewitness
Banjir Redam Ratusan Rumah di Karawang	Non-Eyewitness
Jangan sedih nanti banjir	None
Aku banjir keringet sampe ganti baju lho ini	None
Membongkar Habis Cara Banjir Orderan Dari Facebook	None

TABLE III. CASE FOLDING

Text	Label
depan rumah dh banjir huhu	Eyewitness
hujan lebat gais banjir disini huhu	Eyewitness
banjir dimanaa macet dimanaa hufff	Eyewitness
takutnya banjir	Eyewitness
semoga para korban banjir diberikan ketabahan	Non-Eyewitness
cm di jl gerbang pemuda senayan bagi kendaraan sejenis sedan dihimbau agar tidak melintas	Non-Eyewitness
banjir redam ratusan rumah di karawang	Non-Eyewitness
jangan sedih nanti banjir	None
aku banjir keringet sampe ganti baju lho ini	None
membongkar habis cara banjir orderan dari facebook	None

- Formalization

Formalization is a process of detecting and repairing or removing damaged data or inaccurate data to increase the accuracy of the classification process. To convert sentences that are not standard, the current use of 'alay' or 'slang' sentences causes non-standard use of Indonesian. Additionally, Stemming aims to transform words into their basic words (root words) by eliminating all word affixes. In Table IV, we can see an example of the results of formalization.

TABLE IV. DATA CLEANSING

Text	Label
depan rumah sudah banjir huhu	Eyewitness
hujan lebat teman banjir disini huhu	Eyewitness
banjir dimanaa macet dimana hufff	Eyewitness
takutnya banjir	Eyewitness
semoga para korban banjir diberikan ketabahan	Non-Eyewitness
cm di jalan gerbang pemuda senayan bagi kendaraan sejenis sedan dihimbau agar tidak melintas	Non-Eyewitness
banjir redam ratusan rumah di karawang	Non-Eyewitness
jangan sedih nanti banjir	None
saya banjir keringat sampai ganti baju lho ini	None
membongkar habis cara banjir orderan dari facebook	None

- Stopword Removal

Stopword Removal is filtering of words that appear in large numbers are familiar or words that are not standard and have no meaning (stopword). The removal process uses 2 types of stopword dictionaries. Stopword dictionary id.stopwords.02.01.2016 was downloaded at https://github.com/masdevid/ID-Stopwords.git. Meanwhile, the stopword two dictionary is designed to eliminate words that have no meaning, for example "awkokkk", "hufff", "heheheee" and "bahahahaha". In Table V, we can see an example of the results of stopword removal using dictionary 2.

TABLE V. STOPWORD REMOVAL DATA

Text	Label
depan rumah sudah banjir	Eyewitness
hujan lebat teman banjir sini	Eyewitness
banjir dimanaa macet dimana	Eyewitness
takut banjir	Eyewitness
moga para korban banjir berik ketabahan	Non-Eyewitness
jalan gerbang muda senayan bagi kendara jenis sedan dihimbau agar tidak lintas	Non-Eyewitness
banjir redam ratus rumah karawang	Non-Eyewitness
jangan sedih nanti banjir	None
saya banjir keringat sampai ganti baju ini	None
bongkar habis cara banjir order dari facebook	None

E. Feature Extraction

Feature extraction is used to convert data into structured data so that the text mining process can be carried out. Uni-gram is used for feature extraction. Uni-gram is a single word feature extraction; in this process, the data will be extracted into a single word. Table VI shows how the uni-gram works to separate each text document into a single word. Classification performance can be affected by the feature extraction technique used[10].

TABLE VI. UNI-GRAM FEATURE EXTRACTION

depan rumah sudah banjir	"depan","rumah","sudah", "banjir"
hujan lebat teman banjir sini	"hujan","lebat","teman", "banjir","sini"
banjir dimanaa macet dimana	"banjir","dimanaa","macet", "dimana"
takut banjir	"takut","banjir"
moga para korban banjir berik ketabahan	"moga","para","korban", "banjir", "berik","ketabahan"
jalan gerbang muda senayan bagi kendara jenis sedan dihimbau agar tidak lintas	"jalan","gerbang","muda", "senayan","bagi","kendara", "jenis","sedan","dihimbau", "agar","tidak","lintas"
banjir redam ratus rumah karawang	"banjir","redam","ratus", "rumah","karawang"
jangan sedih nanti banjir	"jangan","sedih","nanti", "banjir"
saya banjir keringat sampai ganti baju ini	"saya","banjir","keringat", "sampai", "ganti","baju","ini"

bongkar habis cara banjir order dari facebook	"bongkar","habis","cara", "banjir","order","dari", "facebook"

The word weighting process is carried out after the data is extracted into single words. Tram Frequency Inverse Document Frequency (TF-IDF) is the weighting of a term which is taken from the number of occurrences of that term in a document. The feature extraction process can be seen in Table VII.

TABLE VII. TF-IDF WEIGHTING FOR FEATURE EXTRACTION

agar	baju	banjir	tidak	...	Label
0	0	0.038	0	...	Eyewitness
0	0	0.030	0	...	Eyewitness
0	0	0.030	0	...	Eyewitness
0	0	0.076	0	...	Eyewitness
0	0	0.025	0	...	Non-eyewitness
0.276	0	0	0.276	...	Non-eyewitness
0	0	0.030	0	...	Non-eyewitness
0	0	0.038	0	...	None
0	0.474	0.021	0	...	None
0	0	0.021	0	...	None

Term Frequency-Inverse Document Frequency (TF-IVDF) is a method used to calculate each word's weight that has been extracted. The TF-IDF weighting model is a method that integrates the Term Frequency (TF) and Inverse Document Frequency (IDF) models. Term Frequency (TF) is a process to count the number of occurrences of terms in one document and Inverse Document Frequency (IDF) is used to count terms that appear in various documents (comments) which are considered as general terms, which are considered not important [11].

1. Count Term Frequency (tft,d)
2. Count Weighting Term Frequency (Wtf t,d)

$$\mathrm{WTF}_{\mathrm{T,D}} = \begin{cases} 1 + \mathrm{LOG10\,TF}_{\mathrm{T,D}} & , \mathrm{IF\ TF}_{\mathrm{T,D}} > 0 \\ 0 & , \mathrm{IF\ TF}_{\mathrm{T,D}} = 0 \end{cases} \qquad (1)$$

3. Count document frequency (df)
4. Count Weight of Inverse Document Frequency (IDF)

$$idft = log10\frac{\mathrm{N}}{\mathrm{DF}} \qquad (2)$$

5. Count Weight Value of TF-IDF

$$Wt,d = Wtft,d \times idf\mathrm{T} \qquad (3)$$

F. *K-fold Cross Validation*

Different cross-validation methods are available in the literature for sample selection as a training data set. The k-fold cross-validation method divides the actual sample into k as samples of the same size. Each subsample is taken as validation data to test the classification model and the process is repeated k times. This method's advantage is more than the repetition of random samples as training and validation is conducted for each of them at least once. Here k is the variable parameter that will be selected by the user [12]. Sharing training data and testing data uses the K-fold Cross Validation method with a value of K = 10. 10-fold cross-validation divides the data into ten parts with nine parts as training data and 1 part as testing data. 10-fold cross-validation has ten repetitions with each repetition using different testing data.

G. *Support Vector Machine*

Support Vector Machine (SVM) is a proper technique in the case of classification and regression. Support Vector Machine (SVM) has a fundamental principle of a linear classifier. These classification cases can be separated linearly. However, the Support Vector Machine (SVM) has been developed to work on non-linear problems by incorporating the kernel concept in a high-dimensional workspace. A hyperplane will be sought in high-dimensional space that can maximize the distance (margin) between data classes.

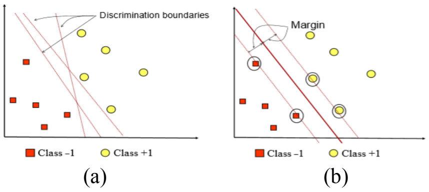


Fig. 2. The best Hyperplane of SVM that separates the two classes, –1 and +1.

SVM's concept can be explained simply as an effort to find the best hyperplane that functions as a separator of two classes. Figure 1-a shows several patterns that are members of two classes: +1 and –1. Patterns that are class members -1 are symbolized by red color (box), while patterns in class +1 are symbolized by yellow (circle)[13]. Classification problems can be translated by finding the line (hyperplane) that separates the two groups. The various alternative discrimination boundaries are shown in Fig. 2-a. The linear classification hyperplane Support Vector Machine (SVM) is notated on below:

$$f(x) = w^{T}. x + b \qquad (4)$$

with the equation as follows:

$$[(w^{T}. xi) + b\ ] \geq 1 \text{ untuk } yi = +1 \qquad (5)$$

$$[(w^{T}. xi) + b\ ] \leq 1 \text{ untuk } yi = -1 \qquad (6)$$

$$w^{T}. x + b = a^{T}.y^{T}.K(x_{i}\ .\ x_{j}) + b \qquad (7)$$

Initially, the Support Vector Machine (SVM) was developed for two-class classification problems, and then it was redeveloped for multi-class classification [13]. Two multi-class techniques are often used in SVM, namely One Versus One (OVO) and One Versus All (OVA). OVO is a classification technique that compares one class to another, while OVA compares one with all other than oneself who are considered one unit. This multi-class is used because SVM is machine learning that only classifies two classes linearly [14]. Support Vector Machine is used for message classification during a flood disaster based on the label (i) eyewitness, (ii) non-eyewitness, and (iii) do not know. The Support Vector Machine uses the OVO (One against One) method. OVO is a classification technique that compares one class to another for multiclass classification. In making the classification model, the Kernlab package is used and only Radial Basis Function (RBF) kernel is utilized.

H. Confusion Matrix

The confusion matrix is a method that can be implemented to measure the performance of a classification method. The confusion matrix contains information that compares the classification results carried out by the system with the classification of that should be[13]. Evaluation is carried out after performing the classification process, and the classification model is evaluated using a confusion matrix table. Testing is done by calculating the accuracy performance using the package caret.

VI. RESULT AND DISCUSSION

Flood disaster message classification is based on flood disaster tweet data using a support vector machine with the Radial Basis Function (RBF) kernel. The data used is 3000 data. The data used can be seen in table VIII. The flood disaster message classification on Twitter has the highest accuracy, with an accuracy of 77.87%. In some experiments using different data, the stopword can affect the level of accuracy. Table IX shows the effect of the data and the stopword dictionary used on the accuracy obtained.

The accuracy of each classification label in each experiment can also be seen in Fig.3. The effect of the stopword dictionary used can be seen in several experiments, the first experiment is on research data 1. In research data 1, the highest level of accuracy is obtained using the stopword two dictionary, which is 76.60 % using the Stopword 1 dictionary, the accuracy rate is 74.37%. The difference in accuracy can be seen in Fig.4.

TABLE VIII. THE TOTAL OF DATASET

Class data	**Total**
Eyewitness	1000
Non-eyewitness	1000
None	1000
Total	3000

TABLE IX. THE RESULT OF THE EXPERIMENTS

Code of Experiment	**Data**	**Using Stopword**	**Accuracy**
1	Experiment 1	Kamus 1	74,37%
2	Experiment 2	Kamus 2	76,60%
3	Experiment 1	Kamus 1	74,97%
4	Experiment 2	Kamus 2	76,87%
5	Experiment 1	Kamus 1	75,13%
6	Experiment 2	Kamus 2	77,87%

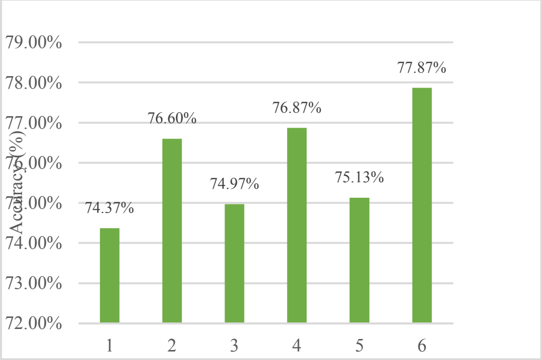


Fig. 3. The use of Stopword in preprocessing for the accuracy.

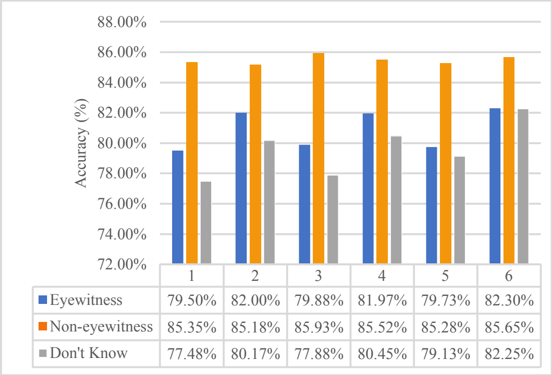


Fig. 4. The accuracy of every class label data.

In experiment 2, the stopword dictionary also affects the resulting level of accuracy. The level of accuracy using the stopword two dictionary is 76.87%, while using the stopword one dictionary, it is 74.97%. Data also affect the level of accuracy obtained. In experiment 1, the highest accuracy was 76.60%; in experiment 2, the highest level of accuracy was 76.87%, and in experiment 2, the highest level of accuracy was 77.87%. In spite of this research, the stopword doesn't affect the accuracy because accuracy is not substantially improved as the way it operates only cutting words into an important word. Often it has an incomprehension, while the sense of the term is important for sentiment analyzes position of an opinion judge[15].

VII. CONCLUSION

The preprocessing data before data mining is essential in the resulting performance level. One of the steps for preprocessing text data is stopword removal. In this study, using the stopword dictionary on accuracy performance can be seen in several experiments that have been carried out. In experiment 1, the highest level of accuracy was obtained using the Stopword 2 dictionary, namely 76.60%. In experiment 2, the highest level of accuracy was obtained by using the Stopword 2 dictionary, namely 76.87%. In experiment 3, the highest level of accuracy was obtained using the Stopword 2 dictionary, namely 77.87%.

A data-based stopword dictionary is very useful because it only removes meaningless words or noises. In sentiment analysis, words that express emotion, words that indicate perceptual senses, and adjectives are also very influential for increasing accuracy. Every tweet that contains adjectives is a characteristic of the tweet.

In future work, we suggest trying the SVM algorithm using One against All (OVA) for classification process to get the better performance in this data. Since this research is analysing data about flood, the method proposed and the research results will be very useful in the future to classify Twitter messages regarding information on any natural disaster.

ACKNOWLEDGMENT

The authors would like to gratefully acknowledge the research of Penelitian Produk Terapan from the Desentralisasi Scheme of Politeknik Negeri Jakarta in 2020. This research had been conducted in collaboration between the Department of Computer and Informatics Engineering, Politeknik Negeri Jakarta and Department of Computer Engineering, Lambung Mangkurat University in Indonesia. Additional computation time was provided by Division of Mechanical Science and Engineering, Kanazawa University, Japan.

REFERENCES

[1] N. A. Christakis and J. H. Fowler, "Social Network Sensors for Early Detection of Contagious Outbreaks," *PLoS ONE*, vol. 5, no. 9, p. e12948, Sep. 2010, doi: 10.1371/journal.pone.0012948.

[2] Y. Kryvasheyeu, H. Chen, E. Moro, P. Van Hentenryck, and M. Cebrian, "Performance of Social Network Sensors during Hurricane Sandy," *PLoS ONE*, vol. 10, no. 2, p. e0117288, Feb. 2015, doi: 10.1371/journal.pone.0117288.

[3] A. Hernandez-Suarez *et al.*, "Using Twitter Data to Monitor Natural Disaster Social Dynamics: A Recurrent Neural Network Approach with Word Embeddings and Kernel Density Estimation," *Sensors*, vol. 19, no. 7, p. 1746, Apr. 2019, doi: 10.3390/s19071746.

[4] C. H. Wu and T. Y. Li, "Social sensor: An analysis tool for social media," *International Journal of Electronic Commerce Studies*, vol. 7, no. 1, pp. 77–94, 2016, doi: 10.7903/ijecs.1411.

[5] BNPB, "Tanggap Tangkas Tangguh Menghadapi Bencana, Pedoman Penyusunan Rencana Penanggulangan Bencana." BNPB, 2008.

[6] K. Zahra, M. Imran, and F. O. Ostermann, "Automatic identification of eyewitness messages on twitter during disasters," *Information Processing & Management*, vol. 57, no. 1, p. 102107, Jan. 2020, doi: 10.1016/j.ipm.2019.102107.

[7] W. A. Luqyana, I. Cholissodin, and R. S. Perdana, "Analisis Sentimen Cyberbullying pada Komentar Instagram dengan Metode Klasifikasi Support Vector Machine," p. 10.

[8] D. D. A. Nur, "PENGEMBANGAN APLIKASI SENTIMENT ANALYSIS MENGGUNAKAN METODE NAÏVE BAYES," p. 6, 2015.

[9] G. A. P. Nugroho, *ANALISIS SENTIMEN DATA TWITTER MENGGUNAKAN K-MEANS CLUSTERING*. 2016.

[10] M. K. Delimayanti *et al.*, "Classification of Brainwaves for Sleep Stages by High-Dimensional FFT Features from EEG Signals," *Applied Sciences*, vol. 10, no. 5, p. 1797, Mar. 2020, doi: 10.3390/app10051797.

[11] M. Akbari, A. Novianty, and C. Setianingsih, "Analisis Sentimen Menggunakan Metode Learning Vector Quantization Sentiment Analysis Using Learning Vector Quantization Method," Aug. 2017.

[12] K. S. Raju, M. R. Murty, and M. V. Rao, "Support Vector Machine with K-fold Cross Validation Model for Software Fault Prediction," p. 15.

[13] B. Santosa, *Data Mining : Teknik Pemanfaatan Data untuk Keperluan Bisnis*. .

[14] A. S. Nugraha and K. K. Purnamasari, "PENERAPAN METODE SUPPORT VECTOR MACHINE PADA PART OF SPEECH TAG BAHASA INDONESIA," p. 8.

[15] A. W. Pradana and M. Hayaty, "The Effect of Stemming and Removal of Stopwords on the Accuracy of Sentiment Analysis on Indonesian-language Texts," *KINETIK*, vol. 4, no. 4, pp. 375–380, Oct. 2019, doi: 10.22219/kinetik.v4i4.912.