Scientific Journal of Informatics
Vol. 11, No. 2, May 2024
p-ISSN 2407-7658 https://journal.unnes.ac.id/journals/sji/index e-ISSN 2460-0040

Identifying Relevant Messages from Citizens in a Social Media Platform for Natural Disasters in Indonesia Using Histogram Gradient Boosting and Self-Training Classifier

Ariana Yunita[1*], Zhafran Ramadhan[2], Angelia Regina Dwi Kartika[3], Ade Irawan[4]
[1,2,3,4]Department of Computer Science, Faculty of Science & Computer, Universitas Pertamina, Indonesia

Abstract.
Purpose: This research aims to develop a classification model using histogram-based gradient boosting to identify relevant contextual tweets about disasters. This model can then be used for subsequent data cleaning stages.
Methods: This study uses a semi-supervised approach to develop a classification model using histogram-based gradient boosting. The model is trained to identify and remove irrelevant tweets that are related to disasters and gathered from Twitter. Optimization techniques, such as the AdaBoost classifier, calibrated classifier, and self-training classifier, are used to enhance the model's performance. The goal is to accurately recognize and categorize relevant tweets for additional data analysis and decision-making.
Result: The classification model that has been developed has achieved a high F1-score of 93.07%, which indicates its effectiveness in filtering disaster-related tweets that are relevant. This highlights the potential of the model to enable more precise aid distribution and faster decision-making in disaster response efforts. The successful implementation of the model also demonstrates its usefulness in utilizing social media data to enhance disaster management practices.
Novelty: This research contributes to the analysis of social media through machine learning algorithms. By utilizing social media, specifically Twitter, as a valuable resource for disaster response efforts, this study tackles challenges related to data collection and analysis in disaster management. The classification of relevant tweets into different types of natural disasters offers opportunities to enhance stakeholder decision-making processes in disaster scenarios.

Keywords: Text relevance, Histogram gradient boosting, Social media analysis, Natural disaster, Semi-supervised approach
Received March 2024 / **Revised** May 2024 / **Accepted** May 2024

INTRODUCTION

Indonesia, a nation comprised of islands with diverse terrains and located at the convergence of four tectonic plates, faces numerous natural disasters, such as volcanic eruptions, earthquakes, tsunamis, floods, and landslides. According to the World Bank, Indonesia ranks 12th out of 35 countries worldwide with a high risk of both human and economic loss due to various natural disasters [1]. These losses often occur because the government and donors lack comprehensive information when assisting regions affected by disasters, resulting in inaccuracies in aid distribution. Additionally, the slow decision-making process during natural disasters exacerbates economic losses [2], [3].

In the immediate aftermath of a disaster such as a flood, earthquake, or tsunami, individuals who are affected often need to contact their relatives in different areas to seek assistance. Similarly, both people and governments require up-to-date information from the area that has been affected by the disaster. In recent years, online social networks, especially X (formerly known as Twitter), have emerged as crucial platforms for accessing real-time updates during and immediately after natural disasters. X, which was originally a social media platform for expressing opinions [4]–[8], has now become instrumental in spreading news about the needs of communities in areas that have been struck by disasters [9]–[11]. Through its robust hashtag culture, X enables accurate and organized communication among users, making the process of collecting, categorizing, and expanding searches during data collection easier [12]. Furthermore, the volume of social media posts greatly increases during natural disasters. As a result, data from social media

*Corresponding author.
Email addresses: yunita.ariana@gmail.com (Yunita)
DOI: 10.15294/sji.v11i2.2287

platforms like X has been used as a valuable resource for decision-making during catastrophic events [10], [13]–[16]

Previous research [9] demonstrates that Twitter data can effectively identify the types of assistance needed by victims of natural disasters. The tweets are categorized into shelter assistance, food, clothing, health, or general assistance. Using a combination of the support vector machine (SVM) algorithm, classifier chain, and UniBiGram, the evaluation yields metrics of 82% precision, 70% recall, and 75% F1 score. Another study [17] focuses on identifying types of traffic events using tweet data from Twitter. The tweets are classified into six classes: accidents, damage, heavy traffic, light traffic, traffic lights, and others. Employing the decision tree and RAKEL algorithms, they achieve an average accuracy of 83%. Similarly, another research [18] utilizes tweet datasets to classify tweets based on whether they contain references to alcohol use or not. Various algorithms such as naive Bayes, SVM, Bayesian logistic regression, and random forest are employed. The random forest algorithm yields the best results with an area under the curve and a standard deviation value of (0.94+–0.02). These studies adopt a supervised approach, which has drawbacks including the time-intensive labelling of available data and significant computational requirements. Other studies [13], [19], [20] also leverage Twitter data within the context of natural disasters.

Our primary research utilizes a semi-supervised learning (SSL) approach to classify the type of aid needed during natural disasters. This approach combines supervised and unsupervised learning methods [21]. SSL has several advantages over supervised learning (SL), including improved efficiency in modelling by utilizing both labelled and unlabelled data, reduced computational expenses, increased robustness to noise and label bias, and enhanced generalization capabilities compared to SL. Specifically, our self-training classifier utilizes unlabelled data to enhance classification performance by iteratively incorporating pseudo-labelled samples into the training dataset [21]. It is important to acknowledge that both SSL and SL approaches depend on high-quality input data to generate meaningful outputs (garbage in, garbage out). Therefore, preprocessing of tweet data is crucial to obtain relevant and reliable information.

When extracting data from Twitter using specific keywords, the process often retrieves unrelated tweets that simply include those keywords. As a result, there is a significant need for an efficient data cleaning step to remove irrelevant content [22]. Therefore, this study aims to develop a classification model using histogram-based gradient boosting to identify contextually relevant tweets about disasters, especially those seeking assistance for natural disasters. The selection of histogram-based gradient boosting is based on its capability to reduce memory usage and speed up computations [14]. Furthermore, these relevant tweets can subsequently be utilized in subsequent data cleaning stages.

The remainder of this paper is organized as follows. In the Methods section, we elaborate on the research methods employed. The Results and Discussion section presents, analyzes, and discusses the findings. The final section, Conclusion, provides a summary of the study and outlines our future endeavours.

METHODS

Our primary objective is to classify tweets according to the type of aid needed during natural disasters, employing a semi-supervised approach. This is illustrated in Figure 1.

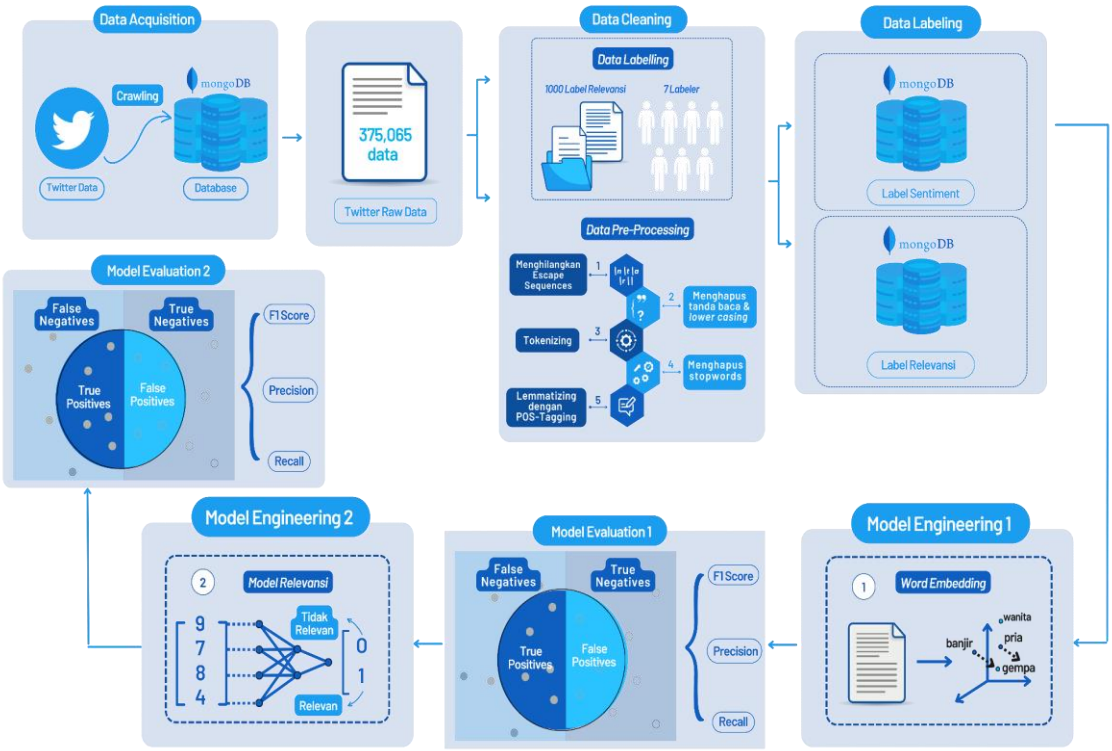

Figure 1. Outline of the research process

This study comprises six research stages: 1) data collection, 2) data labelling, 3) text preprocessing, 4) feature extraction, 5) training for tweet relevancy, and 6) model evaluation.

Data collection

This research gathers a dataset of tweets using Python libraries like SNScrape and Twint to facilitate the collection process. It is important to note that the data collection was conducted before Twitter transitioned to X. Initially, the tweet collection involved selecting keywords to crawl tweets. Subsequently, a timeframe is set to retrieve tweets. In this study, tweets spanning the past decade were obtained using keywords listed in Table 1.

Table 1. Keywords for scraping based on types of assistance

Type of Assistance	Keywords
Evakuasi	Evakuasi AND Bencana
Makanan/minuman	(Butuh OR bantuan) AND (makanan OR minuman OR pangan) AND bencana
Pakaian	(Butuh OR bantuan) AND (baju OR celana OR pakaian) AND bencana
Kesehatan	(Butuh OR bantuan) AND (p3k OR medis OR dokter) AND bencana
Perlindungan	(Butuh OR bantuan) AND tempat perlindungan
Umum	(Tolong OR bantuan) AND bencana
	(Tolong OR bantuan) AND (banjir OR (gempa OR (gempa bumi)) OR tsunami OR longsor OR (gunung berapa OR gunung meletus) OR erupsi OR (angin puting beliung OR puting beliung) OR badai OR (angin kencang) OR kebakaran OR (gelombang pasang) OR abrasi)

The keywords chosen for different types of assistance, apart from general aid, are based on various search possibilities that users may use when requesting such assistance. In the case of general aid, the keyword selection is based on the most common types of natural disasters in Indonesia. The data collection process used AND and OR logic gates, as illustrated in Table 1. However, the use of logic gates has limitations when the collected text data is not relevant to the desired keywords. Thus, it is crucial to have a model that can accurately determine the relevance of text data to natural disasters for the purpose of classifying it.

The results of the crawling, based on the keywords listed in Table 1, yielded approximately 516,846 tweets. Before proceeding to the next stage, the tweets will be filtered to include only those in Indonesian. Out of the total, 71,561 tweets were found to be in languages other than Indonesian and will be discarded, leaving 445,285 tweets for further processing. Subsequently, the tweets will undergo further filtering to remove duplicate data. The dataset contains 70,220 duplicate tweets, which will all be discarded, resulting in 375,065 tweets remaining after filtering.

Data labelling

Labelling the data is a crucial step in this study. The tweet data collected in the previous step undergoes relevance labelling, which serves as the ground truth for creating a relevance model trained in a supervised manner. The relevance label is categorized as 1 (relevant) when the tweet contains elements or meaning related to natural disasters, and as 0 (not relevant) when the tweet lacks such elements entirely. Since our primary focus is on using SSL, data labelling is limited to the first 1,000 tweets. The labelling process is outsourced to seven individuals, who independently classify the tweets into the two classes of relevance to natural disasters or not. This independent labelling ensures that the process remains unbiased. Involving these seven individuals aims to enhance the quality of the labelling.

Text preprocessing

Data collected through the crawling process undergo further preprocessing to ensure cleanliness and organization, allowing for more accurate training of machine learning models. The text preprocessing stages employed in this research are outlined below.

1. Elimination of user tags, hashtags, and links
   The first step involves removing user tags, hashtags, and links from the tweet data, and keeping only the text content. Twitter data often contains user tags, hashtags, and links, which can introduce bias when identifying types of aid. An example of this elimination process is shown in Table 2.
2. Conversion of emojis and numbers
   The second stage involves converting emojis used by users into textual representations of the emojis and changing numeric data into text representations. Users often include emojis or numbers in their tweets. When preprocessing emojis or numbers, there are two options: removal or conversion into meaningful text. In this research, we choose the latter approach, converting emojis and numbers into text.
3. Removal of escape sequences
   Escape sequences, such as newline characters ("\n") or tab characters ("\t"), are frequently used for formatting text. While escape sequences are not common in tweet data, their presence can create inconsistencies and make it difficult to apply consistent text processing techniques. Removing escape sequences helps to ensure uniformity in the text data, making it easier to work with. Essentially, escape sequences can be considered as noise in the text data. By eliminating them during preprocessing, unnecessary noise is reduced, resulting in cleaner and more meaningful text. Removing escape sequences is essential for improving the quality and usability of text data.

Table 2. Example illustrating the elimination of user tags, hashtags, and links

Original Tweet	Preprocessed Tweet
[ELF CARE INDONESIA for BREBES] Kami menerima donasi berupa uang dan pakaian layak pakai untuk teman2 kita di Brebes yang terkena bencana banjir dan longsor. Donasi uang tanpa minimum donasi dan untuk alamat pengiriman pakaian akan kami DM. DM untuk join! deadline 5 Maret thanks	[ELF CARE INDONESIA for BREBES] Kami menerima donasi berupa uang dan pakaian layak pakai untuk teman2 kita di Brebes yang terkena bencana banjir dan longsor. Donasi uang tanpa minimum donasi dan untuk alamat pengiriman pakaian akan kami DM. DM untuk join! deadline 5 Maret thanks
Alhamdulillah, Bantuan tahap 2 dari para donatur & dermawan utk korban bencana gempa & tsunami di Sulawesi Tengah baru saja tiba di Posko Tim Relawan & Bantuan Kemanusiaan Mahasiswa Fakultas Teknologi Industri UMI di Jl S Parman (Rumah Bpk Julius Hoesan) Kota Palu https://t.co/Gab99nq4HK	Alhamdulillah, Bantuan tahap 2 dari para donatur & dermawan utk korban bencana gempa & tsunami di Sulawesi Tengah baru saja tiba di Posko Tim Relawan & Bantuan Kemanusiaan Mahasiswa Fakultas Teknologi Industri UMI di Jl S Parman (Rumah Bpk Julius Hoesan) Kota Palu

Original Tweet	Preprocessed Tweet
pada hari Senin tgl 18 Maret 2019 mulai pukul 08.00 WIB s/d Selesai Babinsa Cibuniwangi Srd Aan Rukanto telah melaks pendampingan kegiatan Penyerahan BPNT (Bantuan Pangan Non Tunai) di… https://t.co/D0z9wa5OBu	pada hari Senin tgl 18 Maret 2019 mulai pukul 08.00 WIB s/d Selesai Babinsa Cibuniwangi Srd Aan Rukanto telah melaks pendampingan kegiatan Penyerahan BPNT (Bantuan Pangan Non Tunai) di…
RT muhammadiyah "Tim relawan Mapala Muhammadiyah yang terdiri dari Mapala UMRI, Mapala UMY, Stacia UMJ dan SARMMI, turut menyalirkan bantuan medis ke desa terdampak Karhutla Riau. Salah satu bantuan medis adalah Nebulizer bagi penderita ISPA … https://t.co/S0AKgrcqmB"	RT muhammadiyah "Tim relawan Mapala Muhammadiyah yang terdiri dari Mapala UMRI, Mapala UMY, Stacia UMJ dan SARMMI, turut menyalirkan bantuan medis ke desa terdampak Karhutla Riau. Salah satu bantuan medis adalah Nebulizer bagi penderita ISPA …
Bantuan itu akan diberikan kepada 20,5 juta keluarga yang termasuk dalam daftar Bantuan Pangan Non Tunai (BPNT) dan Program Keluarga Harapan (PKH), serta 2,5 juta PKL yang berjualan makanan gorengan," ungkap Presiden @jokowi https://t.co/w0WQXkFery	Bantuan itu akan diberikan kepada 20,5 juta keluarga yang termasuk dalam daftar Bantuan Pangan Non Tunai (BPNT) dan Program Keluarga Harapan (PKH), serta 2,5 juta PKL yang berjualan makanan gorengan," ungkap Presiden

4. Elimination of punctuation
   The fourth stage involves removing punctuation marks from the text and standardizing the text to all lowercase letters. Commonly used punctuation symbols include the full stop (.), comma (,), exclamation mark (!), question mark (?), semicolon (;), colon (:), hyphen (-), parentheses (()), quotation marks ("), and various other symbols.
5. Lowercasing or Case Folding
   In this phase, all capital letters were changed to lowercase. The aim is that because capital and lowercase letters differ from one another, the exact words are not recognized as having different meanings. This is a common text preprocessing step in natural language processing.

Table 3. Example illustrating the changes made during the preprocessing steps of converting emojis and numbers, eliminating escape sequences, and removing punctuation

Original Tweet	Preprocessed Tweet
[ELF CARE INDONESIA for BREBES] Kami menerima donasi berupa uang dan pakaian layak pakai untuk teman2 kita di Brebes yang terkena bencana banjir dan longsor. Donasi uang tanpa minimum donasi dan untuk alamat pengiriman pakaian akan kami DM. DM untuk join! deadline 5 Maret thanks	elf care indonesia for brebes kami menerima donasi berupa uang dan pakaian layak pakai untuk teman dua kita di brebes yang terkena bencana banjir dan longsor donasi uang tanpa minimum donasi dan untuk alamat pengiriman pakaian akan kami dm dm untuk join deadline lima maret thanks
Alhamdulillah, Bantuan tahap 2 dari para donatur & dermawan utk korban bencana gempa & tsunami di Sulawesi Tengah baru saja tiba di Posko Tim Relawan & Bantuan Kemanusiaan Mahasiswa Fakultas Teknologi Industri UMI di Jl S Parman (Rumah Bpk Julius Hoesan) Kota Palu https://t.co/Gab99nq4HK	alhamdulillah bantuan tahap dua dari para donatur amp dermawan utk korban bencana gempa amp tsunami di sulawesi tengah baru saja tiba di posko tim relawan amp bantuan kemanusiaan mahasiswa fakultas teknologi industri umi di jl s parman rumah bpk julius hoesan kota palu
pada hari Senin tgl 18 Maret 2019 mulai pukul 08.00 WIB s/d Selesai Babinsa Cibuniwangi Srd Aan Rukanto telah melaks pendampingan kegiatan Penyerahan BPNT (Bantuan Pangan Non Tunai) di… https://t.co/D0z9wa5OBu	pada hari senin tgl delapan belas maret dua ribu sembilan belas mulai pukul nol delapan nol nol wib sd selesai babinsa cibuniwangi srd aan rukanto telah melaks pendampingan kegiatan penyerahan bpnt bantuan pangan non tunai di…
RT muhammadiyah "Tim relawan Mapala Muhammadiyah yang terdiri dari Mapala UMRI, Mapala UMY, Stacia UMJ dan SARMMI, turut menyalirkan bantuan medis ke desa terdampak Karhutla Riau. Salah satu bantuan medis adalah Nebulizer bagi penderita ISPA … https://t.co/S0AKgrcqmB"	rt muhammadiyah tim relawan mapala muhammadiyah yang terdiri dari mapala umri mapala umy stacia umj dan sarmmi turut menyalirkan bantuan medis ke desa terdampak karhutla riau salah satu bantuan medis adalah nebulizer bagi penderita ispa …
Bantuan itu akan diberikan kepada 20,5 juta keluarga yang termasuk dalam daftar Bantuan Pangan Non Tunai (BPNT) dan Program Keluarga Harapan (PKH), serta 2,5 juta PKL yang berjualan makanan gorengan," ungkap Presiden @jokowi https://t.co/w0WQXkFery	bantuan itu akan diberikan kepada dua ratus lima juta keluarga yang termasuk dalam daftar bantuan pangan non tunai bpnt dan program keluarga harapan pkh serta dua puluh lima juta pkl yang berjualan makanan gorengan ungkap presiden

6. Replacement of slang words
   In the sixth stage, the language used in the tweets is standardized. Non-standard words, including slang terms, are replaced with uniform and standard equivalents. To convert slang words into formal language, we utilize the Colloquial Indonesian Lexicon, a lexicon compiled from Instagram by Salsabila et al. [23]. It is important to note that Twitter data often includes abbreviations due to its character limit.

Table 4. Example illustrating the replacement of slang words

Original Tweet	Preprocessed Tweet
met	selamat
kaka	kakak
slalu	selalu
bgt	banget
rb	ribu
mnis2nya	manis-manisnya
kaak	kak
baguuus	bagus
mantep	mantap
sampe	sampai
bnk	bang
abizzzz	abis
sdikit	sedikit

Word embedding

There are two types of embedding algorithms used in natural language processing: classical and contextual techniques [24]. For this study, the Word2Vec algorithm was chosen because it is the most suitable for our dataset. We compared Word2Vec, Doc2Vec, FastText, and Glove and found that Word2Vec outperformed the others. Word2Vec is a word embedding algorithm that converts words into numerical vectors while preserving their meaning. Words with similar meanings will have vector representations that are close to each other in a vector space with predefined dimensions. Word2Vec can learn linguistic patterns as linear relationships between word vectors. There are two Word2Vec algorithms: continuous bag of words and skip-gram. Word2Vec facilitates the capture of relationships between words in text, which are then represented as vectors. Moreover, this algorithm is suitable for large text datasets. However, its drawback is that it may produce inaccurate vector representations if the training data is too small.

Modelling tweet relevancy

The next step involves training the multidimensional vector using a supervised approach with the relevance labels obtained in the previous subchapter. The relevance model was constructed with the primary objective of determining the relevance of a corpus of tweets systematically acquired through web crawling. Examples of irrelevant tweets can be found in Table 5.

Table 5. Examples illustrating relevant and irrelevant text regarding requests for aid during natural disasters

	Tweet	Label (0 = irrelevant; 1 = relevant)
1.	bilas muka gosok gigi evakuasi :victoryhand:	0
2.	sedang dalam progress implementasi ide dan evakuasi alamat tinggal apa tantangan selanjutnya	0
3.	mungkin indonesia sama korea banyak yang rasis tapi di indonesia lebih manusiawi tragedi kanjuruhan pas kejadian banyak yang tolong menolong lah disitu malah asik joget sendiringalangin petugas buat evakuasi	0
4.	kisah di balik jembatan cinta selo boyolali jalur penting evakuasi Merapi	0
5.	gerombolan nasakom dan nol koma adalah produk berbahaya daya rusak bagi konoha sangat luar biasa segera evakuasi ke dasar Samudra	1
6.	relawan puan membagikan ribuan paket sembako untuk masyarakat kurang mampu dan memberi bantuan tunai untuk korban terdampak bencana angin putting beliung di sidoarjo	1
7.	bantuan sosial dari ibuibu relawan bandar lampung yang diwakili oleh ibu evi sukrana untuk korban bencana alam banjir desa banyumas kecamatan candipuro kabupaten lampung selatan	1

8.	polri peduli bencana polwan polres sidrap beri bantuan sembako kepada korban bencana kebakaran di kelurahan rijang pittu kec maritengngae kab sidrap	1
9.	tanggap peduli bencana polres sidrap serahkan bantuan sambako kepada korban kebakaran di keluarga rijang pittu kec maritengngae	1
10.	dilansir afp saksi mata yang berada di lokasi bercerita berebut keluar dari kerumunan orangorang menumpuk di atas satu sama lain paramedis yang melakukan evakuasi sampai kewalahan hingga meminta bantuan orang yang lewat untuk memberi pertolongan pertama kepada para korban	

The histogram gradient boosting (HGB) algorithm [18] is used to train the relevance model. In addition, we use a wrapper technique for feature selection [19]. Wrapper methods aim to improve classifier performance by optimizing feature selection. Other machine learning algorithms used for the wrapper technique include AdaBoost (AB), bagging classifier, balanced bagging classifier, and calibrated classifier.

The AB algorithm is a classification technique that uses an ensemble of weak learners to sequentially predict data. AB iteratively fits weak learners to the data, modifying the dataset with each iteration [25], [26]. Predictions from all weak learners are then combined through majority voting to generate the final prediction. The bagging classifier algorithm improves basic classification performance by introducing randomization in the classification process, which helps to reduce overfitting and is particularly effective for complex models [27], [28]. The balanced bagging classifier balances the training data while training using a given sampler. The calibrated classifier CV algorithm employs isotonic or logistic regression to enhance the probability estimate of a classification [29]. This algorithm fits a copy of the base estimator on the training subset and calibrates it using the test subset for each cross-validation split.

Evaluation

Evaluating a machine learning model is a crucial step in assessing its performance [30]. For classification tasks, a confusion matrix provides a detailed breakdown of true positives, true negatives, false positives, and false negatives, which offers insights into the model's performance. Furthermore, precision, recall, and F1 score can be calculated based on the confusion matrix, which further assists in evaluating the model's effectiveness.

Precision measures the accuracy of the model in making positive predictions. It indicates the proportion of items predicted as positive that are true positives. A high precision value indicates that the model is effective in generating positive predictions and has a low rate of false positives. High precision is especially important in situations where false positives are costly or undesirable, such as in medical diagnosis or fraud detection.

Recall, also known as sensitivity or the true positive rate, assesses the model's ability to correctly identify all relevant instances or true positives among the total actual positive instances. A high recall is preferred in situations where missing a positive instance is costly or unacceptable, as in medical diagnosis, where failing to detect a disease could have devastating consequences. However, prioritizing high recall may result in lower precision, as the model may generate more false positive predictions to ensure it captures all true positives. Therefore, there is often a trade-off between precision and recall, and the choice between the two depends on the specific requirements of the task at hand.

The F1 score is a commonly used metric in machine learning to evaluate the performance of classification models, including text classification [31], [32]. It gives a single numerical value that considers both precision and recall, providing a balanced measure of a model's accuracy. This is particularly useful when dealing with imbalanced class distributions or when the consequences of false positives and false negatives vary. A high F1 score indicates a model with both high precision and high recall [26].

RESULTS AND DISCUSSIONS

Out of the more than 350,000 available tweet data, only 1,000 tweets were labelled for relevance using an outsourcing approach. Since the majority of the data lacks relevance labels, we are employing a semi-supervised approach called the self-training classifier (STC) algorithm, combined with the supervised HGB algorithm as a base estimator. This approach allows us to train using tweet data that has both labelled and

unlabelled relevance information, making it possible to label a large portion of the unlabelled data. The parameters used for STC include the "k-best" criterion, with a value of 5,000 for k-best, and a maximum iteration (max iter) of 20.

STC is one of the semi-supervised algorithms available in the scikit-learn [X-sklearn] framework. The mechanism of STC involves training the base estimator on data with labels for the initial iteration. Subsequently, the 5,000 instances with the highest probabilities from both classes are selected and treated as labelled data for the subsequent iteration. This process is repeated for 20 iterations, as determined by the max_iter parameter. To develop a semi-supervised relevance model, the HGB and STC algorithms are combined, and the algorithm is further optimized with wrapper algorithms such as the AdaBoost classifier (ABC), bagging classifier (BC), balanced bagging classifier (BBG), and calibrated classifier-CV (CC-CV). Each algorithm combination is trained on the same dataset and evaluated using the same testing data for comparison. Performance is assessed using evaluation metrics such as precision, recall, and F1 score.

Table 6. Evaluation metrics results

Model	Precision	Recall	F1-Score
HGB + STC	87.35%	86.92%	86.65%
HGB + ABC + STC	92.46%	92.30%	92.30%
HGB + BC + STC	90.23%	90.00%	89.98%
HGB +BBC + STC	89.23%	89.23%	89.23%
HGB + ABS + CC-CV + STC	93.08%	93.07%	93.07%
HGB + BC + CC-CV + STC	88.01%	87.69%	87.66%
HGB + BBG + CC-CV + STC	83.87%	83.07%	82.97%

From the results shown in Table 6, it is clear that the HGB + ABC + CC-CV + STC algorithm model has the highest precision, recall, and F1 scores, achieving values of 93.08%, 93.07%, and 93.07%, respectively. Therefore, this model will be used as the classifier to determine whether a tweet is relevant or unrelated to natural disasters. The classified tweets will then be used for further stages of modelling.

In comparison, another study [33] aimed to compare four machine learning algorithms: AB, gradient boosting, extreme gradient boosting (XGB), and HGB. Their objective was to detect fake accounts using Twitter data during the Indonesian presidential election. However, the best scores achieved ranged from 60% to 68%. The research findings revealed that AB exhibited the highest accuracy and precision, reaching 62.3% and 61.3%, respectively, when utilizing 25 features. Additionally, XGB achieved a recall of 67.9%, which matched its recall result when using 20 features. On the other hand, Nurdeni et al. [8] classified types of assistance using SL, employing a combination of the SVM algorithm, classifier chain, and UniBiGram. Their evaluation results yielded 82% precision, 70% recall, and 75% F1 score. From these findings, it can be concluded that our proposed model outperforms previous models.

CONCLUSION

The classification of types of natural disaster aid was developed to address a prevalent issue in Indonesia: inaccuracies in distributing aid to victims of natural disasters due to irrelevant data. This study aimed to classify relevant and irrelevant data using semi-supervised algorithms, specifically the combination of HGB and SCT, optimized with AB and calibrated classifier. The proposed algorithm achieved the highest accuracy of 93.08% among various combinations of HGB and other algorithms. However, a limitation of this research is the availability of labelled data; only 1,000 labels were applied due to the semi-supervised learning approach. For future studies, the proposed model can be extended to other contexts of social media data, such as stance detection or sentiment analysis. Additionally, our next agenda involves classifying relevant tweets into several types of natural disaster aid to facilitate faster decision-making for stakeholders.

ACKNOWLEDGEMENTS

This research is funded by Hibah Penelitian Dikti 2023 Skema Penelitian Dosen Pemula NO: 088/UP-WRP.1/PJN/VII/2023

REFERENCES

[1] World Bank, “World Bank Climate Change Knowledge Portal.”

[2] and M. H.-K. O. Zabihi, M. Siamaki, M. Gheibi, M. Akrami, “A smart sustainable system for flood damage management with the application of artificial intelligence and multi-criteria decision-making computations,” *Int. J. Disaster Risk Reduct.*, vol. 84, 2023, doi:

https://doi.org/10.1016/j.ijdrr.2022.103470.
[3] M. Tariq, I. Khan, S. Anwar, S. Asumadu, M. Rizwan, and A. Majeed, "Heliyon Do natural disasters affect economic growth ? The role of human capital , foreign direct investment , and infrastructure dynamics," *Heliyon*, vol. 9, no. 1, p. e12911, 2023, doi: 10.1016/j.heliyon.2023.e12911.
[4] M. R. Ningsih, K. A. H. Wibowo, A. U. Dullah, and J. Jumanto, "Global recession sentiment analysis utilizing VADER and ensemble learning method with word embedding," *J. Soft Comput. Explor.*, vol. 4, no. 3, 2023, doi: https://doi.org/10.52465/joscex.v4i3.193.
[5] W. B. Zulfikar, A. R. Atmadja, and S. F. Pratama, "Sentiment Analysis on Social Media Against Public Policy Using Multinomial Naive Bayes," *Sci. J. Informatics*, vol. 10, no. 1, pp. 25–34, Jan. 2023, doi: 10.15294/sji.v10i1.39952.
[6] N. A. Saputra, K. Aeni, and N. M. Saraswati, "Indonesian Hate Speech Text Classification Using Improved K-Nearest Neighbor with TF-IDF- ICSρF," *Sci. J. Informatics*, vol. 11, no. 1, pp. 21–30, 2024, doi: 10.15294/sji.v11i1.48085.
[7] A. Yunita, S. F. Telaumbanua, and A. Irawan, "The Tweetology of New and Renewable Energy in Indonesia," *IJCCS (Indonesian J. Comput. Cybern. Syst.*, vol. 17, no. 2, pp. 127–138, 2023, doi: 10.22146/ijccs.81397.
[8] D. A. Nurdeni, I. Budi, and A. ris B. Santoso, "Sentiment Analysis on Covid19 Vaccines in Indonesia : From The Perspective of Sinovac and Pfizer," *Indones. Conf. Comput. Inf. Technol. April 09-11. 2021, ISTTS Surabaya, Indones. Sentim.*, pp. 122–127, 2021.
[9] D. A. Nurdeni, I. Budi, and A. Yunita, "Extracting information from twitter data to identify types of assistance for victims of natural disasters: An Indonesian case study," *J. Manag. Inf. Decis. Sci.*, vol. 22, no. 1S, pp. 1–14, 2022.
[10] D. Reynard and M. Shirgaokar, "Harnessing the power of machine learning: Can Twitter data be useful in guiding resource allocation decisions during a natural disaster?," *Transp. Res. Part D Transp. Environ.*, vol. 77, pp. 449–463, 2019, doi: https://doi.org/10.1016/j.trd.2019.03.002.
[11] M. Karimiziarani, K. Jafarzadegan, P. Abbaszadeh, W. Shao, and H. Moradkhani, "Hazard risk awareness and disaster management: Extracting the information content of twitter data," *Sustain. Cities Soc.*, vol. 77, p. 103577, 2022, doi: https://doi.org/10.1016/j.scs.2021.103577.
[12] W. Ahmed, P. A. Bath, and G. Demartini, "Using Twitter as a Data Source: An Overview of Ethical, Legal, and Methodological Challenges," *Ethics Online Res. (Advances Res. Ethics Integr.*, vol. 2, pp. 79–107, 2017, doi: https://doi.org/10.1108/S2398-601820180000002004.
[13] S. Behl, A. Rao, S. Aggarwal, S. Chadha, and H. S. Pannu, "Twitter for disaster relief through sentiment analysis for COVID-19 and natural hazard crises," *Int. J. Disaster Risk Reduct.*, vol. 55, no. February, p. 102101, 2021, doi: 10.1016/j.ijdrr.2021.102101.
[14] E. Blomeier and S. Schmidt, "Drowning in the Information Flood : Machine-Learning-Based Relevance Classification of Flood-Related Tweets for Disaster Management," *information*, vol. 15, no. 3, p. 149, 2024, doi: 10.3390/info15030149.
[15] S. Mendon and P. Dutta, "A Hybrid Approach of Machine Learning and Lexicons to Sentiment Analysis : Enhanced Insights from Twitter Data of Natural Disasters," *Inf. Syst. Front.*, vol. 23, pp. 1145–1168, 2021, doi: https://doi.org/10.1007/s10796-021-10107-x.
[16] S. Auclair, R. Quique, and E. Soulier, "SURICATE-Nat : Innovative citizen centered platform for Twitter based natural disaster monitoring," *Int. Conf. Inf. Commun. Technol. Disaster Manag.*, 2019, doi: 10.1109/ICT-DM47966.2019.9032950.
[17] J. C. Chamby-diaz and A. L. C. Bazzan, "Identifying traffic event types from Twitter by Multi-label Classification," *Brazilian Conf. Intell. Syst.*, pp. 806–811, 2019, doi: 10.1109/BRACIS.2019.00144.
[18] P. K. Yin Aphinyanaphongs, Bisakha Ray, Alexander Statnikov, "Text Classification for Automatic Detection of Alcohol Use-Related Tweets A Feasibility Study," *Proc. 2014 IEEE 15th Int. Conf. Inf. Reuse Integr. (IEEE IRI 2014)*, pp. 93–97, 2014, doi: 10.1109/ICT-DM47966.2019.9032950.
[19] B. Kusumasari, N. Phydra, and A. Prabowo, "Scraping social media data for disaster communication : how the pattern of Twitter users affects disasters in Asia and the Pacific," *Nat. Hazards*, no. 0123456789, 2020, doi: 10.1007/s11069-020-04136-z.
[20] V. V Mihunov *et al.*, "Use of Twitter in disaster rescue : lessons learned from Hurricane Harvey Use of Twitter in disaster rescue : lessons learned from Hurricane," *Int. J. Digit. Earth ISSN*, 2020, doi: 10.1080/17538947.2020.1729879.
[21] J. E. van Engelen and H. H. Hoos, "A survey on semi-supervised learning," *Mach. Learn.*, 2019,

doi: 10.1007/s10994-019-05855-6.
[22] A. Cobo, D. Parra, and J. Navón, "Identifying Relevant Messages in a Twitter-based Citizen Channel for Natural Disaster Situations," *Proc. 24th Int. Conf. world wide web*, pp. 1189–1194, 2010, doi: 10.1145/2740908.2741719.
[23] N. A. Salsabila, Y. Ardhito, W. Ali, A. Septiandri, and A. Jamal, "Colloquial Indonesian Lexicon," *2018 Int. Conf. Asian Lang. Process.*, pp. 226–229, 2018, doi: 10.1109/IALP.2018.8629151.
[24] S. F. Sabbeh, "A Comparative Analysis of Word Embedding and Deep Learning for Arabic Sentiment Classification," *Electronics*, 2023, doi: https://doi.org/10.3390/electronics12061425.
[25] Y. Freund and R. E. Schapire, "A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting *," *J. Comput. Syst. Sci.*, vol. 139, pp. 119–139, 1997, doi: https://doi.org/10.1006/jcss.1997.1504.
[26] ScikitLearn, "ADA Boost Classifier."
[27] ScikitLearn, "Bagging Classifier."
[28] L. E. O. Bbeiman, "Bagging Predictors," *Mach. Learn.*, vol. 140, pp. 123–140, 1996.
[29] B. Zadrozny, "Obtaining Calibrated Probability Estimates from Decision Trees and Naive Bayesian Classifiers Obtaining calibrated probability estimates from decision trees and naive Bayesian classifiers," *Proc. Eighteenth Int. Conf. Mach. Learn.*, no. March, 2013, doi: https://doi.org/10.24963/ijcai.2023/475.
[30] A. Géron, "Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow," O'Reilly Media, Inc, 2019.
[31] M. Heikal, "ScienceDirect Sentiment Analysis of Arabic Tweets using Deep Learning Sentiment Analysis of ∗ Arabic Tweets using Deep Learning," *Procedia Comput. Sci.*, vol. 142, pp. 114–122, 2018, doi: 10.1016/j.procs.2018.10.466.
[32] E. A. Sukma *et al.*, "Sentiment Analysis of the New Indonesian Government Policy (Omnibus Law) on Social Media Twitter," *Int. Conf. Informatics, Multimedia, Cyber Inf. Syst.*, no. July 2017, pp. 153–158, 2021, doi: 10.1109/ICIMCIS51567.2020.9354287.
[33] A. Suciati, A. Wibisono, and P. Mursanto, "Twitter Buzzer Detection for Indonesian Presidential Election," *Int. Conf. Informatics Comput. Sci. Twitter*, 2019, doi: 10.1109/ICICoS48119.2019.8982529.