PARAMETER
JURNAL MATEMATIKA, STATISTIKA DAN TERAPANNYA

e-ISSN. 2829-6303
Vol. 04 No. 2, August 2025
Pages 339-352

Categorical Boosting and Bayesian Optimization in Natural Disaster Tweet Classification

Enzelica Vica Christina[1], Wahyu S. J. Saputra[2]*, Kartika Maulida Hindrayani[3]

[1,2,3]Data Science Study Program, Faculty of Computer Science, Universitas Pembangunan Nasional Veteran Jawa Timur
Rungkut Madya St., Gn. Anyar, Dist. Gn. Anyar, Surabaya, 60294, East Java, Indonesia

E-mail Correspondence Author: *wahyu.s.j.saputra.if@upnjatim.ac.id*

Abstract

Multi-label classification is an important challenge in natural language processing, especially when a single text data point can have more than one label. This study applies a multi-label classification approach to group information in Twitter comments related to natural disasters in Indonesia. The data is categorized into six labels: disaster, location, damage, victims, aid, and others. To address the complexity of text data, the Categorical Boosting (CatBoost) algorithm is used, which is a decision tree-based boosting method that excels at handling categorical features and reducing overfitting. The model is built using the MultiOutputClassifier approach to handle multiple labels simultaneously. Additionally, Bayesian optimization is performed, which is a parameter search method that uses a probabilistic approach to select the best parameter combination based on previous evaluations. Optimization focused on four main parameters: number of iterations, learning rate, tree depth, and L2 regularization. The results showed that the model achieved an accuracy of 75.41% and a Hamming loss of 0.0520, demonstrating the effectiveness of this approach in handling multi-label classification on Twitter data.

Keywords: *Bayesian Optimization, Categorical Boosting, Multi-Label Classification, Text*

: https://doi.org/10.30598/parameterv4i1pp339-352

1. INTRODUCTION

Text classification is one of the important techniques in digital data management and belongs to the field of natural language processing (NLP). This technique allows the process of grouping text into certain categories automatically, thus reducing dependence on manual classification processes that tend to require more time and money [1]. Along with technological developments, text classification approaches have evolved with the application of ensemble algorithms, which are methods that combine several models, both conventional and modern, to improve prediction performance. One type of ensemble that is widely used is boosting, which aims to improve accuracy by combining some weak models into a more reliable model [2]. This boosting technique is also applicable to multi-label classification, where a single text instance can simultaneously belong to more than one category [3].

One boosting algorithm that shows superior performance in classification tasks is Categorical Boosting (CatBoost). This algorithm is known to be effective in handling various types of data, including text data such as comments [4]. CatBoost generates the final prediction by building and combining some Decision Tree models [5]. The performance of this algorithm is affected by some parameters, such as iterations, learning rate, and depth, which need to be adjusted to achieve the best possible output. To efficiently obtain the best combination of parameters, in this study, the Bayesian optimization method is used. This method utilizes the Gaussian process in exploring information from previous iterations, to find parameter values more purposefully and efficiently [4].

CatBoost's ability to manage text data becomes relevant when faced with the need to extract information from social media, especially in the context of disasters. This need arises due to the high intensity of natural disasters that occur every year in Indonesia. Based on the Indonesian Disaster Data and Information Management Database (DIBI), throughout 2023-2024, there were a total of 6,670 natural disasters with the most deaths occurring in floods (267 people) and landslides (256 people) [6]. The high frequency of disasters raises the need for fast and accurate information delivery. In this context, Twitter plays an important role as one of the most widely used social media platforms in Indonesia [7]. The information shared generally includes various aspects, such as the location of the incident, the number of victims, or the need for aid. Therefore, a multi-label classification approach is considered more suitable than a single classification in managing information from this social media.

Several classification approaches using CatBoost have been carried out to answer the challenges of managing information from social media, one of which is the classification of legal documents [8]. This research classifies legal texts such as petitions and judicial decisions with several scenarios combining them with numerical and categorical features. The results showed that the accuracy and f1-score had values above 0.7 by using the whole text or post tags. Other research on multi-label classification has also been done such as in the study of Bukhari hadith and article documents. One of the approaches used is Binary Relevance, which allows algorithms such as K-Nearest Neighbor and AdaBoost to be used on multi-label tasks [9], [10]. Other approaches for multi-label classification that can also be used are Classifier Chain and Multi-Output Classifier [11]. Previous research has shown that CatBoost performs better than other boosting algorithms, due to its symmetric tree structure which is more suitable for non-linear data [5]. Bayesian optimization has also been shown to improve the performance of CatBoost by finding the best parameters for the model [4].

This research aims to design and implement a multi-label classification algorithm on text data. While algorithms such as K-Nearest Neighbor (KNN) and AdaBoost have been widely applied for multi-label classification tasks, the use of Categorical Boosting (CatBoost) remains relatively limited. However, CatBoost has the structural advantage of using a uniform decision tree, which makes it more effective in dealing with data with non-linear patterns. In addition, to further optimize the performance of the model, this research also uses Bayesian optimization to find the best parameters that are expected to improve the evaluation and efficiency of the model. Along with advances in text classification, this research utilizes data from Twitter that contains various information related to natural disasters. The findings of this research are anticipated to support the development of text classification techniques, especially in the disaster domain, and become a useful source of information for those involved in disaster management. Thus, this research not only expands the utilization of CatBoost in multi-label classification, but also demonstrates its potential application in real contexts in the digital era.

2. METHODOLOGY

This research utilizes comment data obtained from Twitter social media, which is collected through a scraping process with a focus on flood and landslide disaster topics during the 2020-2024 period. The selection of these disaster types is based on the National Disaster Management Agency (BNPB)'s recapitulation data which shows that floods and landslides are the most dominant types of disasters that have occurred almost every year in the past decade [12].

Table 1. Label Description

Label	Description
Disaster	Comments containing the words 'flood' or 'landslide' that contextually refer to the natural disaster.
Location	Comments that mention the location of the disaster such as the name of the city, district, or village.
Damage	Comments stating the impact of disasters such as damaged roads, broken bridges, and collapsed houses.
Victim	Comments that have information about casualties such as deaths, drifting, and missing.
Aid Others	Comments that have information about the aid provided and needed. Comments that are not disaster labels.

In Table 1, presents a table containing descriptions of the six labels in this study. This description gives an overview and also serves as a reference for maintaining label consistency to minimize ambiguity.

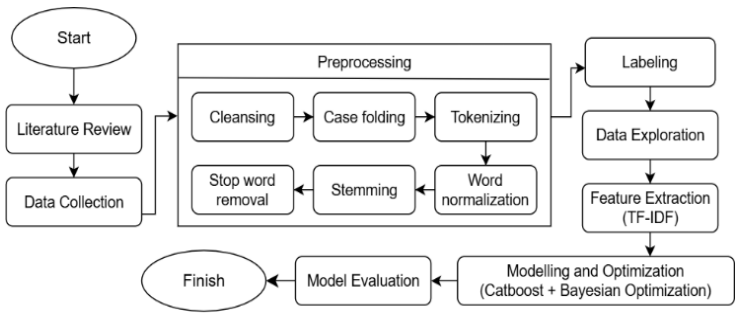

Figure 1. Research Methodology Flow

In Figure 1, the research is illustrated in detail, beginning with data collection and

concluding with model evaluation. This research begins with a literature study to deeply understand the problems raised and identify methods of solving them. The literature was obtained from credible sources such as scientific articles and books, which helped find gaps or topics that have not been widely studied in previous studies. After that, the research continued with the data collection stage until the evaluation.

2.1. Data Collection

Data was obtained through a scraping process from Twitter using API Key. Scraping is a technique to automatically obtain unstructured information from a website using a computer program or bot [13]. The data collected were user comments related to floods and landslides in Indonesia in 2020-2024. This entire process was run using Python with the help of the Tweet Harvest tool. As a result, a total of 4898 disaster-related comments were collected.

2.2. Preprocessing

The data that has been successfully collected is then processed through a series of processing stages consisting of six steps. The first step is cleansing, which is done to make the data clean by removing parts that are not relevant to the analysis. Case folding is required to convert the data into lowercase letters using modules available in Python [14]. Word normalization is done to correct the words in the text so that it has a format according to the language order to be used. Next, the stemming process converts words to their basic form or basic verbs. The last stage is stop word removal which removes words that are considered meaningless.

2.3. Labeling

Each piece of data is labeled based on the information contained in the comment text, with six possible categories: disaster, location, damage, victims, aid, and others. The labeling is binary, where each label is assigned a value of 0 (no) or 1 (yes) to each label. Two techniques are used in this process, namely Named Entity Recognition (NER) to detect location labels, and rule-based keyword matching for the remaining labels. The NER model used was developed by Cahya and is specifically trained for the Indonesian language, making it readily applicable. In the rule-based approach, a list of representative keywords is compiled for each label. If a data entry contains a word that matches a keyword in the list, the corresponding label is assigned.

2.4. Data Exploration

At this stage, visualization is done to get an overview of the distribution of labels and words. A bar chart is used to visualize label distribution by displaying the occurrence frequency of each label, allowing for an evaluation of distributional balance. In addition, visualization is also done in the form of bar charts and circles to display the most frequently occurring words in the comment data. The results of this visualization can provide initial insight into the dominance of certain topics in the data.

2.5. Feature Extraction (TF-IDF)

Term Frequency-Inverse Document Frequency (TF-IDF) is used to extract the data feature. The TF method transforms attributes into numerical form by measuring the importance of words in a sentence. TF expresses the frequency with which a specific word appears within a given sentence or document., while IDF shows how commonly the word appears in the entire set of documents [15].

$$Tf_{t,d}.Idf_t = Tf_{t,d} \times log\frac{N}{df_t} \quad (1)$$

Following **Equation (1)**, $\boldsymbol{Tf_{t,d}}$ is the Term Frequency, which states the number of $\boldsymbol{t}$ terms or words in $\boldsymbol{d}$ document. Meanwhile, $\boldsymbol{df_t}$ is Document Frequency, which reflects the frequency of documents that contain the word or t term. $\boldsymbol{N}$ is the total number of documents present within the dataset. The IDF part is expressed in the formula $\boldsymbol{log\frac{N}{df_t}}$, the value is calculated by dividing the N document by the number of documents that include the $\boldsymbol{df_t}$ term **[16]**.

2.6. Categorical Boosting

The Gradient Boosting method is constructed iteratively. Given a training dataset $\{\boldsymbol{x_i}, \boldsymbol{y_i}\}, \boldsymbol{i = 1, \dots, n}$, where $\boldsymbol{x_i}$ denotes the feature vector and $\boldsymbol{y_i}$ represents the target label, the loss function is defined as $\boldsymbol{L(y, F)}$, with $\boldsymbol{F}$ denoting the prediction model. The model update is expressed as $\boldsymbol{F_{t+1}(x) = F_t(x) + h_{t+1}(x)}$, where $\boldsymbol{F_{t+1}}$ is the updated model at iteration $\boldsymbol{t+1}$, $\boldsymbol{F_t}$ is the model from the previous iteration, and $\boldsymbol{h_{t+1}}$ is selected to minimize the loss function **[17]**.

$$\boldsymbol{h_{t+1} \approx arg\min_{h \in H}\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\partial L(y_i, F_t(x_i))}{\partial F_t(x_i)} - h(x_i)\right)^2} \quad (2)$$

Where $\boldsymbol{\frac{\partial L}{\partial F_t(x_i)}}$ denotes the partial derivative of the loss function with respect to the model output, which indicates both the direction and magnitude of the correction required to improve the prediction for the i-th data point. In contrast, H represents the set of candidate decision trees that serves as the search space for identifying the optimal tree at each iteration.

To enhance predictive performance, CatBoost introduces several key innovations. The first is Ordered Target Statistics (OTS). For categorical features, CatBoost does not necessarily employ one-hot encoding; instead, it applies target statistics, defined as **[17]**:

$$\hat{x}_{ik} = \frac{\sum_{x_j \in D_k} 1_{x_{ij}=x_{ik}} . y_j + ap}{\sum_{x_j \in D_k} 1_{x_{ij}=x_{ik}} + a} \quad (3)$$

In **Equation (3)**, $\hat{x}_{ik}$ denotes the transformed value of the i-th categorical feature for the k-th sample. D_k represents the subset of data preceding the k-th sample according to a random permutation. The indicator function $1_{x_{ij}=x_{ik}}$ equals 1 if the categorical values of samples j and k are identical, and 0 otherwise. y_j refers to the target label of sample j, while p is the prior, typically the mean of the target values in the dataset. The parameter a serves as a smoothing factor to prevent division by zero and to reduce estimation variance.

This approach mitigates target leakage, i.e., the unintended use of information from test data during training **[4]**. The second innovation is Ordered Boosting, which leverages multiple random permutations $(\sigma_1, \sigma_2, \dots, \sigma_s)$ to compute gradients, thereby yielding more stable estimates and improving generalization performance. The third innovation is the use of Oblivious Decision Trees (ODTs) for ensemble construction. ODTs are complete binary trees in which the same splitting criterion is applied at each level, ensuring balanced tree structures, reducing the risk of overfitting, and accelerating prediction **[5]**.

2.7. Bayesian Optimization

Bayesian optimization can find hyperparameter candidates based on historical data observations. Then find the hyperparameter value that optimizes the model performance [18]. When evaluating the function at various points, evidence about its performance is collected and can be used to update the initial belief. The updating process is caried out using Bayesian theorem, which integrates prior knowledge with new data to derive a posterior distribution. The acquisition function is used to determine the next sample point based on the new updated understanding. The cycle of repeated sampling, updating, and selection continues until it reaches a predetermined number of iterations or reaches a maximum value.

Equation (4) is a formula used for optimization based on Bayesian theory which states that ***A*** is the model and ***B*** is the observation.

$$P(A \mid B) = \frac{\big(P(B|A)P(A)\big)}{P(B)} \tag{4}$$

Following **Equation (4)**, $P(A)$ is the initial probability before looking at the data. The value of $P(B)$ refers to the likelihood of variables A (model) and variable B (observation). The expressions $P(A|B)$ and $P(B|A)$ are conditional probabilities, which allows the equation to be simplified by omitting $P(B)$, leading to the formulation presented in **Equation (5)**. **[4]**

$$P(A \mid B) = P(B \mid A)P(A) \tag{5}$$

2.8. Model Evaluation

Hamming loss is one of the commonly used evaluation metrics in multi-label classification. It measures the proportion of prediction errors for each label individually, so that if a prediction contains a combination of correct and incorrect labels, only the incorrect labels are counted as errors. Hamming loss ranges between 0 and 1 with values nearer to 0 reflecting superior model performance. [19]

$$HL(x_i, y_i) = \frac{1}{|D|}\sum_{i=1}^{|D|}\frac{\Delta(x_i, y_i)}{|L|} \tag{6}$$

Following **Equation (6)**, $|D|$ is the total number of data, $|L|$ denotes the total number of labels, y_i corresponds to the ground truth label, and x_i refers to the predicted label.

The confusion matrix functions as a tabular representation that consolidates classification performances, employing different evaluation metrics to measure accuracy. It allows for the observation of both correct and incorrect classifications across all classes. The main evaluation metrics, such as accuracy, precision, recall, and f1-score are calculated based on equations 7, 8, 9, and 10. **[20]**

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{7}$$

$$Precision = \frac{TP}{TP + FP} \tag{8}$$

$$Recall \ = \frac{TP}{TP + FN} \tag{9}$$

$$F1 - score = \frac{(2 \ \times Recall \times Precision)}{(Recall + Precision)} \tag{10}$$

3. RESULT AND DISCUSSION

3.1. Data Collection

The data collection focused specifically on Indonesian-language tweets related to landslides and floods. This process was assisted by the Tweet Harvest tool with the input of authentication tokens and relevant keywords. The results of data collection are stored in the form of csv files, with only the comment column used.

Table 2. Data Sample

No.	Original Comments
1.	Dilanda Banjir dan Longsor 3.000 Rumah Warga Cianjur Rusak https://t.co/gaMmZzxMDe
2.	Perjuangan Mahasiswa Tasikmalaya Antarkan Langsung Donasi untuk Korban Banjir di Sukabumi https://t.co/5rVa7twhIV
3.	Banjir Melanda Gorontalo: 3520 Jiwa Terdampak Tanggul Jebol dan Infrastruktur Rusak https://t.co/x9fd90H0gR

3.2. Preprocessing

The purpose of this stage is to transform the raw data into a form that is ready to be modeled. This process is carried out through six stages in sequence, namely cleansing, followed by case folding, tokenizing, word normalization, stemming, and stop word removal.

Table 3. Preprocessing Result

Proses	Hasil
Original comments	Dilanda Banjir dan Longsor 3.000 Rumah Warga Cianjur Rusak https://t.co/gaMmZzxMDe
Cleansing	Dilanda Banjir dan Longsor Rumah Warga Cianjur Rusak
Case folding	dilanda banjir dan longsor rumah warga cianjur rusak
Tokenizing	['dilanda', 'banjir', 'dan', 'longsor', 'rumah', 'warga', 'cianjur', 'rusak']
Word normalization	['dilanda', 'banjir', 'dan', 'longsor', 'rumah', 'warga', 'cianjur', 'rusak']
Stemming	['landa', 'banjir', 'dan', 'longsor', 'rumah', 'warga', 'cianjur', 'rusak']
Stopword removal	landa banjir longsor rumah warga cianjur rusak

3.3. Labeling

After going through the cleaning stage, the data was then labeled. Labeling was done by utilizing two main approaches: the Named Entity Recognition (NER) method to identify the location entity, and rule-based keyword matching techniques for the other five labels.

Table 4. Labeling Result

Comment	Disaster	Location	Damage	Victim	Aid	Others
Rumah Budi Jebol Di Hantam Tanah Longsor Akibat Guyuran Hujan Deras Selama 3 Jam https://t.co/8MKi1j5LbC	1	0	1	0	0	0
Terdampak Banjir BNPB Bantu Pemkab Sergai 500 Juta https://t.co/MthKGZvTp2 via @idpontas	1	1	0	0	1	0

Comment	Disaster	Location	Damage	Victim	Aid	Others
@MelodyUnited Fix emyu banjir bandang teropi taun ini.	0	0	0	0	0	1

3.4. Data Exploration

The exploration was conducted using two types of visualizations, namely bar charts and pie charts. Bar charts were used to display the amount of data on each label as well as the frequency of word occurrence, while pie charts were used to show the proportion of the most frequently occurring words.

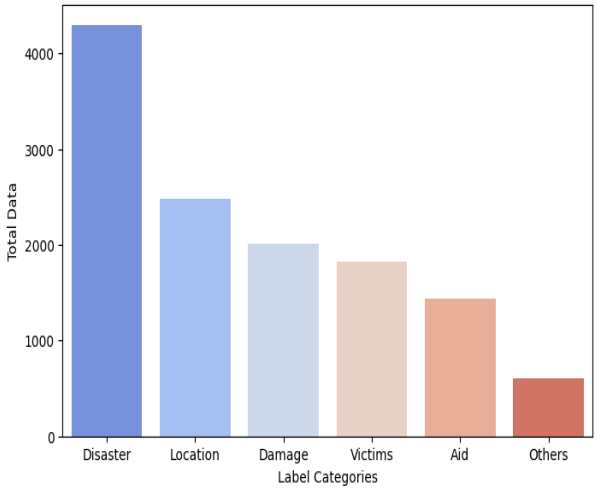

Figure 2. The Number of Labels

In **Figure 2**, the amount of comment data for each label category is shown. This visualization indicates that the distribution of labels is unbalanced, with the highest number being for the 'disaster' label, reaching more than 4,000, while the 'other' label has the lowest number, less than 1,000. This reflects that out of a total of 4,989 comment data, most of them discuss topics that are contextually directly related to flood or landslide events. In addition, the location and damage labels, which have around 2,000 records each, show that almost half of the retrieved data includes information about the place of occurrence as well as the impact of damage when reporting natural disasters.

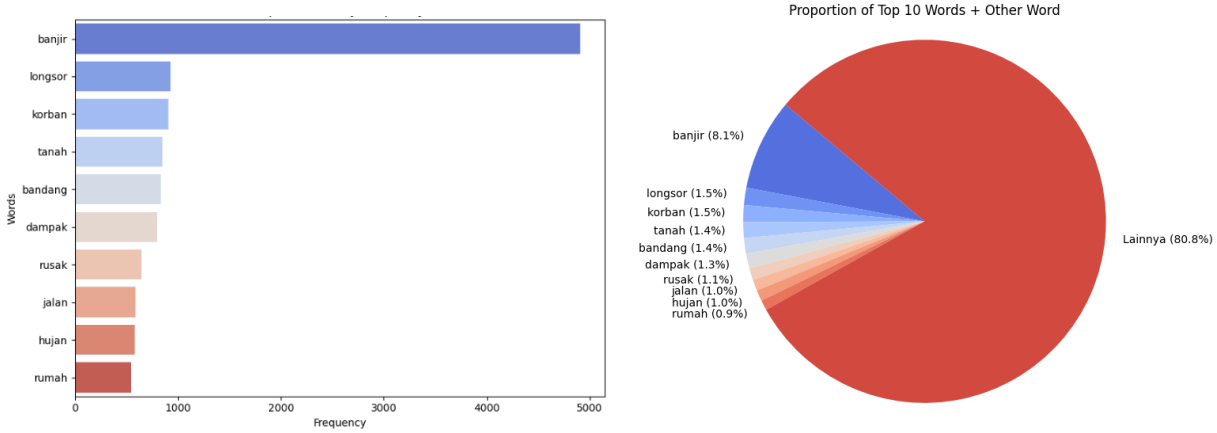

Figure 3. Word Frequency

In **Figure 3**, the relative frequency of words in the comment is shown. The bar chart reveals that the word 'banjir' is the most frequently occurring word, with a much higher number than other words. In comparison, the word 'longsor' only appears about 1,000 times, while other words have similar or lower frequencies. The pie chart shows that the word with the highest frequency only accounts for 8.1% of all words in the comments, while most of the other words only account for about 1% or even less.

3.5. Feature Extraction (TF-IDF)

Machines cannot process data in text form directly, so a technique is needed to convert the data into a numerical representation. In this research, the Term Frequency-Inverse Document Frequency (TF-IDF) method is used as a numerical feature extraction technique, which serves to measure the relative frequency of a word in a document compared to the entire corpus. This process plays an important role in extracting relevant features and helps in measuring the importance of a word to the whole corpus.

Table 5. TF-IDF

Id	term	tfidf
680	abad	0.379776
2121	abad	0.422498
1712	abadi	0.289922
76	abai	0.212866
1923	abai	0.258837
2162	abai	0.286026
2383	abai	0.241575
2739	abai	0.251517

In **Table 5**, the id column indicates the index or identity of the document where the word appears, the term indicates the word being analyzed, and tfidf is the TF-IDF weight value, which reflects how important the word is in the document compared to the whole corpus. The appearance of the same word in multiple rows, such as 'abai' or 'abad', indicates that the word appears in more than one document, and its TF-IDF value differs depending on the frequency of its appearance in each document as well as how common it is in the whole dataset.

3.6. Modeling and Optimization

Prior to the modeling stage, the data was divided into two parts with a proportion of 80% for training data and 20% for testing data. The model training process begins by building an initial model using a number of parameters set manually. At this stage, the parameter values used are default values or those commonly used in machine learning modeling practices.

Table 6. Catboost Base Model Parameters

Parameter	Value	Description
iteration	1000	Determine the number of trees to be formed in the model
learning_rate	0.08	How much does each new model affect the final model
Depth	10	Tree depth
l2_leaf_reg	5	L2 regulation parameter to measure overfitting
Verbose	500	Display logs every n iterations
loss_function	Logloss	A function to calculate the model error. Logloss is used for binary classification, the research is multi-label classification per label independently.

In **Table 6**, shows the summary of the parameter definitions used in the initial CatBoost model before optimization. The parameter selection is based on default values and common practices often used in similar studies to generate a sufficiently robust baseline before the optimization process is performed.

Table 7. The Optimal Parameters of Bayesian Optimization Results

Parameter	Value
iteration	811
learning_rate	0,0738400984694127
depth	9
l2_leaf_reg	8

In **Table 7**, the best parameter results based on Bayesian optimization with the MultiOutputClassifier approach are shown. This optimization considers the best combination of values based on f1-weighted performance evaluation with 10 training.

Table 8. CatBoost Model Result

Model	Evaluastion		
	Accuration	F1-weighted	Hamming Loss
CatBoost	75,20%	93,32%	0,0531
CatBoost + Bayesian Optimization	75,41%	91,66%	0,0520

In **Table 8**, the performance before and after applying Bayesian optimization to the CatBoost model is compared. The results show that the optimization improves accuracy and hamming loss, although it slightly decreases the f1-weighted score. This result indicates that the optimization results in a more balanced model in handling all labels. An accuracy value of 75.41% after applying Bayesian optimization shows that most of the data was correctly classified across all labels. Meanwhile, a hamming loss value of 5.2% indicates that, on average, there is one label error for every twenty possible labels. Based on these metrics, the model is estimated to have made approximately 307 label errors out of a total of 5880 labels in the test data, consisting of 980 data points.

Table 9. Classification Result

Original Comment	Actual	Prediction
Pemko Pekanbaru Waspada Bencana Banjir #BNetwork https://t.co/fCN5TQSRRk	disaster	disaster
bau bau yg bakal banjir air mata nih https://t.co/Kh1a6wdHRt	others	others
Ngeri Banjir Merangsek Masuk Pemukiman Padat Penduduk https://t.co/HT7qwaPyqa	disaster	disaster
Published on YouTube: Banjir Di Wajo Rusak Jembatan Gantung Warga Terisolir https://t.co/0YCskTxGQr	disaster, location, damage, victim	disaster, damage, victim
Banjir dan Longsor di Lebak Renggut 3 Korban Jiwa Status Tanggap Darurat Bencana Ditetapkan https://t.co/QYyOzQgZjI	disaster, location, victim	disaster, victim
The power of areng... 20 Jenazah Hanyut Dampak Banjir Bandang Blitar Tersisa Kain Kafan https://t.co/4Gr7IuZ50r	disaster, location, damage, victim	disaster, damage, victim

In **Table 9**, the classification results using the model after optimization are shown. These results are an example of model output that includes data with all correct label classifications as well as partially correct labels.

3.7. Evaluation

The evaluation results show that applying Bayesian optimization can improve the model's performance in parameter selection. This can be seen in Table 7, after optimization, the accuracy value increases from 75.20% to 75.41%, or an increase of 0.21%. In addition, the hamming loss value decreased from 5.31% to 5.20%, indicating that the optimized model produces more precise predictions overall. Although there is a slight decrease in the F1-weighted value, from 93.32% to 91.66%, this decrease is still within the acceptable range. This shows that Bayesian optimization can balance overall accuracy and prediction error per label in multi-label classification. In practical terms, this improvement supports the use of the model as a complementary tool to help organize large volumes of social media data related to disasters, particularly for filtering and grouping information by category before manual verification is performed.

In addition, the hamming loss value obtained is better than the previous research on Bukhari hadith text classification [9], which reached 0.0886 using the K-Nearest Neighbors (KNN) algorithm. This shows that CatBoost has superior performance in reducing prediction error in multi-label text classification. However, the accuracy of the model is still lower than the research on the CAPEC domain **[11]** which uses the MultiOutputClassifier approach. The difference can be caused by the characteristics of the data used, namely text data tends to be more complex than numerical data in CAPEC research. The results are also in line with research on lung cancer classification **[4]** which shows an increase in model performance after being optimized using Bayesian Optimization. Differences in the use of parameters such as subsample and colsample_bylevel in previous studies, which were not used in this study, may have influenced the final evaluation results.

4. CONCLUSION

This research discusses the application of the CatBoost model for multi-label classification on Twitter comment data with six label categories. According to the evaluation results, the CatBoost model optimized through the Bayesian algorithm demonstrates enhanced performance relative to the initial model, particularly in terms of accuracy and hamming loss metrics. This indicates that parameter optimization can improve model performance. However, the overall accuracy value is still at 75.41%, which indicates that the model is not fully optimized. Therefore, future research is recommended to consider more accurate labeling methods to improve the quality of the training data. Furthermore, exploring more advanced feature representation methods, such as word embeddings or transformer-based models, along with the application of alternative multi-label classification algorithms, such as Binary Relevance or Classifier Chains, may offer potential avenues for enhancing classification performance.

Acknowledgments

The authors declare that there are no acknowledgments to be made.

Funding Information

Author state no funding involved

Author Contributions Statement

Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Enzelica Vica Christina	✓	✓	✓		✓	✓	✓	✓			✓		✓	
Wahyu Syaifullah Jauharis Saputra				✓						✓		✓		
Kartika Maulida Hindrayani				✓						✓		✓		

C	: Conceptual	I	: Investigation	Vi	: Visualization
M	: Methodology	R	: Resources	Su	: Supervision
So	: Software	D	: Data Curation	P	: Project administration
Va	: Validation	O	: Writing – Original Draft	Fu	: Funding acquisition
Fo	: Formal analysis	E	: Writing – Review & Editing		

Conflict Of Interest Statement

The authors confirm that there are no financial, personal, or professional relationships that could be perceived as potential conflicts of interest in relation to this work.

Informed Consent

This study did not involve human participants in a clinical or personal context. Therefore, informed consent was not required for this research.

Ethical Approval

This study did not involve human participants or animals. Therefore, ethical approval was not required for this research.

Data Availability

The data supporting the findings of this study were collected from public tweets via the Twitter API. In accordance with Twitter's data sharing policy, raw tweet content cannot be shared openly. However, the dataset containing tweet IDs and labels used in this study is available from the corresponding author, Enzelica Vica Christina, upon reasonable request.

REFERENCES

[1] Q. Li *et al.*, "A Survey on Text Classification: From Traditional to Deep Learning," *ACM Trans Intell Syst Technol*, vol. 13, no. 2, pp. 31:1-31:41, 2022.

[2] K. Kowsari, K. J. Meimandi, M. Heidarysafa, S. Mendu, L. Barnes, and D. Brown, "Text classification algorithms: A survey," *Information (Switzerland)*, vol. 10, no. 4, pp. 1–68, 2019.

[3] A. Muhaimin, W. Wibowo, and P. A. Riyantoko, "Multi-label Classification Using Vector Generalized Additive Model via Cross-Validation," *Journal of Information and Communication Technology*, vol. 22, no. 4, pp. 657–673, 2023, doi: 10.32890/jict2023.22.4.5.

[4] Y. F. Zamzam, T. H. Saragih, R. Herteno, Muliadi, D. T. Nugrahadi, and P. H. Huynh, "Comparison of CatBoost and Random Forest Methods for Lung Cancer Classification using Hyperparameter Tuning Bayesian Optimization-based," *Journal of Electronics, Electromedical Engineering, and Medical Informatics*, vol. 6, no. 2, pp. 125–136, Apr. 2024, doi: 10.35882/jeeemi.v6i2.382.

[5] M. T. Syamkalla, S. Khomsah, and Y. S. R. Nur, "Implementasi Algoritma Catboost Dan Shapley Additive Explanations (SHAP) Dalam Memprediksi Popularitas Game Indie Pada Platform Steam," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 11, no. 4, pp. 777–786, Aug. 2024, doi: 10.25126/jtiik.1148503.

[6] BNPB, "Data informasi Bencana Indonesia (DIBI)." Accessed: Oct. 26, 2024. [Online].

Available: https://dibi.bnpb.go.id/baru
[7] R. Stevany, "Indonesia Pengguna X atau Twitter Terbanyak Keempat di Dunia. ," RRI. Accessed: Oct. 26, 2024. [Online]. Available: https://www.rri.co.id/papua/lain-lain/859350/indonesia-pengguna-x-atau-twitter-terbanyak-keempat-di-dunia
[8] L. J. Gonçalves Freitas, P. S. D. Edokawa, T. Carvalho Valadares Rodrigues, A. H. Thomé de Farias, and E. Rodrigues de Alencar, "Catboost algorithm application in legal texts and UN 2030 Agenda," *Revista de Informatica Teorica e Aplicada*, vol. 30, no. 2, pp. 51–58, 2023, doi: 10.22456/2175-2745.128836.
[9] A. Hanafi, A. Adiwijaya, and W. Astuti, "Klasifikasi Multi Label pada Hadis Bukhari Terjemahan Bahasa Indonesia Menggunakan Mutual Information dan k-Nearest Neighbor," *Jurnal Sisfokom (Sistem Informasi dan Komputer)*, vol. 9, no. 3, pp. 357–364, 2020, doi: 10.32736/sisfokom.v9i3.980.
[10] I. G. A. P. Arimbawa and Dr. N. A. S. ER, "Penerapan Metode Adaboost Untuk Multi-Label Classification Pada Dokumen Teks," *Jurnal Elektronik Ilmu Komputer Udayana*, vol. 9, no. 1, pp. 127–140, 2020.
[11] T. S. Riera, J. R. B. Higuera, J. B. Higuera, J. J. M. Herraiz, and J. A. S. Montalvo, "A new multi-label dataset for Web attacks CAPEC classification using machine learning techniques," *Comput Secur*, vol. 120, no. 102788, 2022, doi: 10.1016/j.cose.2022.102788.
[12] T. Ridho Fariz, S. Suhardono, and S. Verdiana, "Pemanfaatan Data Twitter Dalam Penanggulangan Bencana Banjir dan Longsor Use of Twitter Data in Flood and Landslide Disaster Management," *Cogito Smart Journal |*, vol. 7, no. 1, 2021.
[13] T. M. Fahrudin, P. A. Riyantoko, and K. M. Hindrayani, "Implementation of Web Scraping on Google Search Engine for Text Collection Into Structured 2D List," *Telematika: Jurnal Informatika dan Teknologi Informasi*, vol. 20, no. 2, pp. 139–152, 2023, doi: 10.31515/telematika.v20i2.9575.
[14] T. M. Fahrudin, A. R. F. Sari, A. Lisanthoni, and A. A. Dewi, "Analisis Speech-to-Text pada Video Mengandung Kata Kasar dan Ujaran Kebencian dalam Ceramah Agama Islam Menggunakan Interpretasi Audiens dan Visualisasi Word Cloud," *SKANIKA: Sistem Komputer dan Teknik Informatika*, vol. 5, no. 2, pp. 190–202, 2022.
[15] P. A. Riyantoko, T. M. Fahrudin, D. A. Prasetya, Trimono, and T. D. Timur, "Analisis Sentimen Sederhana Menggunakan Algoritma LSTM dan BERT untuk Klasifikasi Data Spam dan Non-Spam, *Prosiding Seminar Nasional Sains Data*, vol. 2, pp. 103–111, 2022.
[16] V. R. Prasetyo, G. Erlangga, and D. A. Prima, "Analisis Sentimen Untuk Identifikasi Bantuan Korban Bencana Alam Berdasarkan Data Di Twitter Menggunakan Metode K-Means Dan Naïve Bayes," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*, vol. 10, no. 5, pp. 1055–1062, 2023, doi: 10.25126/jtiik.2023107077.
[17] J. T. Hancock and T. M. Khoshgoftaar, "CatBoost for big data: an interdisciplinary review," *J Big Data*, vol. 7, no. 94, pp. 1–45, 2020, doi: 10.1186/s40537-020-00369-8.
[18] Y. Wang, R. Wang, J. Wang, N. Li, and H. Cao, "A Rock Mass Strength Prediction Method Integrating Wave Velocity and Operational Parameters Based on the Bayesian Optimization Catboost Algorithm," *KSCE Journal of Civil Engineering*, vol. 27, no. 7, pp. 3148–3162, Jul. 2023, doi: 10.1007/s12205-023-2475-9.
[19] I. K. N. Ananda, N. P. N. P. Dewi, N. W. Marti, and L. J. E. Dewi, "Klasifikasi Multilabel Pada Gaya Belajar Siswa Sekolah Dasar Menggunakan Algoritma Machine Learning," *Journal of Applied Computer Science and Technology*, vol. 5, no. 2, pp. 144–154, 2024, doi: 10.52158/jacost.v5i2.940.
[20] M. Idhom, D. A. Prasetya, P. A. Riyantoko, T. M. Fahrudin, and A. P. Sari, "Pneumonia Classification Utilizing VGG-16 Architecture and Convolutional Neural Network Algorithm for Imbalanced Datasets," *TIERS Information Technology Journal*, vol. 4, no. 1, pp. 73–82, Jun. 2023, doi: 10.38043/tiers.v4i1.4380.