# Classification of Tweets Related to Natural Disasters Using Machine Learning Algorithms

https://doi.org/10.3991/ijim.v17i14.39907

Orlando Iparraguirre-Villanueva[1], Melquiades Melgarejo-Graciano[2], Gloria Castro-Leon[3], Sandro Olaya-Cotera[4], John Ruiz-Alvarado[5], Andrés Epifanía-Huerta[6], Michael Cabanillas-Carbonell[7], Joselyn Zapata-Paulini[8](✉)
[1] Universidad Autónoma del Perú, Lima, Perú
[2] Universidad Científica del Sur, Lima, Perú
[3] Universidad Nacional Tecnológica de Lima Sur, Lima, Perú
[4] Universidad San Ignacio de Loyola, Lima, Perú
[5] Universidad Tecnológica del Perú, Lima, Perú
[6] Universidad Católica los Ángeles de Chimbote, Lima, Perú
[7] Universidad Privada del Norte, Lima, Perú
[8] Universidad Continental, Lima, Perú
70994337@continental.edu.pe

**Abstract—**In recent years, computer science has advanced exponentially, helping significantly to identify and classify text extracted from social networks, specifically Twitter. This work identifies, classifies, and analyzes tweets related to real natural disasters through tweets with the hashtag #Nat-uralDisasters, using Machine learning (ML) algorithms, such as Bernoulli Naive Bayes (BNB), Multinomial Naive Bayes (MNB), Logistic Regression (LR), K-Nearest Neighbors (KNN), Decision Tree (DT), Random Forest (RF). First, tweets related to natural disasters were identified, creating a dataset of 122k geo-located tweets for training. Secondly, the data-cleaning process was carried out by applying stemming and lemmatization techniques. Third, exploratory data analysis (EDA) was performed to gain an initial understanding of the data. Fourth, the training and testing process of the BNB, MNB, L, KNN, DT, and RF models was initiated, using tools and libraries for this type of task. The results of the trained models demonstrated optimal performance: BNB, MNB, and LR models achieved a performance rate of 87% accuracy; and KNN, DT, and RF models achieved performances of 82%, 75%, and 86%, respectively. However, the BNB, MNB, and LR models performed better with respect to performance on their respective metrics, such as processing time, test accuracy, precision, and F1 score. Demonstrating, for this context and with the trained dataset that they are the best in terms of text classifiers.

**Keywords—**classification, tweets, disasters, machine learning, natural

## 1 Introduction

Natural disasters have increased the frequency of their manifestations in various parts of the world, with climate change as one of the main causes [1]. According to the United Nations Office for Disaster Risk Reduction (UNDRR) [2], this problem has generated an increase in the frequency and magnitude of natural disasters. Natural disasters are natural phenomena that generate great human and material losses [3]. There are several types of natural disasters, each of which can cause different damages, depending on the type and intensity of the phenomenon [4]. According to the UNDRR [5] between 2000 and 2019 there were more than 7,348 recorded natural disasters, which caused the death of approximately 1.23 million people, affected more than 4.2 billion people, and caused the loss of US$2.97 trillion in the world economy. The magnitude of the damage caused by these phenomena is of greater impact if there is no early warning system; risk indicators contribute to the reduction of human, material, financial and economic losses [6]. However, developing countries are the most affected since they do not have warning and prevention systems and are not prepared for disasters [7]. A clear example is what happened in Haiti in 2010 [8], where there was a seismic movement with a magnitude of 7 on the Richter scale, which caused multiple serious damages such as 200,000 human losses and 65% of buildings being destroyed.

Social networks as a means of information in the face of natural disasters are of great importance [9], [10] because they allow the sending of information content in real-time, and in many cases, they have been useful in various disasters that have occurred in recent years [11], [12]. Currently, the scope that technology has reached in society in terms of social networks, has become a means of information sources, which provides the reality of different events occurring in various parts of the world [13]. The processing and analysis of these social data in networks can provide specialists with a greater overview to understand the effectiveness of these situations [14]. One of the largest social networks in providing big data sources as a means of information is Twitter [15], in this social network daily millions of people around the world make updates of all the events happening in their environment [16]. However, not always the information provided by Twitter users is completely true [17], many times distorted, or false news has been made, which has only led to misinformation among other people [18]. Therefore, the objective of this work is to identify, classify and analyze the Tweets related to real natural disasters through publications with the hashtag #NaturalDisasters, using ML algorithms, such as BNB, MNB, LR, KNN, DT, and RF. These ML algorithms are used to classify the class or category to which a given tweet belongs, these algorithms are trained with previously labeled data[19], where each tweet is associated with a certain class or category.

This paper is organized as follows: section 2 describes the works related to the topic of study: BNB, MNB, LR, KNN, DT, RF, and natural disasters. Section 3 presents the work methodology and case implementation. Section 4 presents the results obtained and discusses them. Finally, Section 5 presents the conclusions of the work.

## 2 Related work

At present, natural language processing (NLP) has made great advances, because of which exceptional models have been developed [20], including the application of ML models in various prediction research, which has obtained efficient results.

In the research [21] the authors propose a computational model using ML for the purpose of predicting weekly rainfall, where they used recurrent neural networks (RNN) to perform the training process of the data, which obtained efficient results in the testing process. Similarly, researchers [22] authors developed a model to predict and label rainfall warnings by using classification models. In addition, the effectiveness of ML algorithms was compared, demonstrating higher effectiveness than MLP classifier. Also, in research [23], a prediction model was developed against emergency events of 4 types by employing the BERT-Att-BiLSTM model. Classification algorithms are a category of ML that learn from a training dataset containing labeled features and then use these learnings to make predictions on new data. For example, in [24], they implemented a sensor on Twitter to monitor natural disasters, where they tokenized the words in tweets to transform them into word embeddings, then used biLSTM and a conditional random field (CRF) output layer to increase the classification accuracy. Similarly, in the manuscript [25], the authors developed an automated system to monitor natural disasters using Twitter data, where they made a connection to the Twitter API and employed the NLTK tool for the filtering process, which had 93% algorithm efficiency. Likewise, in [26], an analysis was performed by extracting tweets for natural disaster behavior facilitation, in addition to comparing three classifiers: LSTM, BiLSTM, and Bert, to obtain the best extraction method. Similarly, in the following manuscript [27] where the authors performed an analysis of KNN, LR, RF, and DT Tweets to identify and categorize in greater detail the damage caused by disasters. Similarly, this research [28] developed a model to detect events in crisis situations through Twitter data, where they performed the combination of CNN and LSTM, in addition to performing a comparison with KNN and RF models among the results, it was obtained that the RF and KNN model has a better performance rate than LSTM, and even performs better than the traditional SVM model. Similarly, in paper [29] developed a model for event detection during a disaster situation by Twitter using ML algorithms, such as RF, DT, and perceptron. Finally, in this work [30] a model for Twitter text analysis for disaster resource management was performed, where they incorporated a hybrid model with ML and CNN algorithms, resulting in reasonable accuracy and proving to be useful during these natural crises.

## 3 Methodology

This chapter presents the theoretical basis of the BNB, MNB, LR, KNN, DT, and RF algorithms. The procedure to classify and analyze tweets related to real natural disasters through the publications with the hashtag #NaturalDisasters, using ML algorithms. For which the following process is followed: import and loading of the dataset, exploratory analysis of the dataset, data cleaning process, training of the models, and finally, testing of the algorithms.

### 3.1 Bernoulli Naive Bayes

It is a probabilistic classification model and uses the NB theorem to determine the probability that a record belongs to a given class. It is a variant of Naive Bayes that is used in cases where the predictor variable is binary[31]. The naive assumption of Naive Bayes refers to the assumption that the features are independent of each other. This algorithm is efficient in terms of training time and memory and is commonly used in text classification. This model is used to classify binary data into two categories [32]. The BNB model is represented in equation (1).

$$p(x) = [X = x] = \{q = 1 - p \qquad x - (0,1)\} \tag{1}$$

### 3.2 Multinomial Naive Bayes

It is a classification ML model that is generally used to predict the membership of an object to one of several categories based on its features. This model assumes that the features are independent and follow a multinomial distribution. It is mainly used in text analysis and document classification but can also be applied to other types of data [33]. The idea behind the MNB model is to calculate the a priori probability of each class and the conditional probability of each feature given a class, and then combine these probabilities to predict the most likely class for a given object. The MNB model is represented in logarithmic space with equation (2).

$$log\, \boldsymbol{P_r}(C_k|\boldsymbol{W}) \alpha\, log\, p(C_k) + \sum_{i=1}^{n} \boldsymbol{w}_i * log\, p(w_i|C_k) \tag{2}$$

### 3.3 Logistic regression

Is an ML model used for the binary classification of data. It is used to predict one of two possible categorical outcomes (e.g., yes/no, true/false) based on a set of predictors [34]. The output of RL is a probability that maps to a binary prediction (0 or 1). The LR is a form of generalized regression analysis that fits the data by applying a sigmoid logistic function on the linear combination of predictor variables [35]. The LR has a method with parameters for the distribution PY| X where Y is a discrete value and X=x1...xn is a vector with continuous values. The model is represented in equations (3) and (4).

$$PY = 1|X = \frac{1}{1+exp\, w_{o+\sum_{i=1}^{n} w_i x_i}} \tag{3}$$

y

$$PY = 0|X = \frac{exp\, w_{o+\sum_{l=1}^{n} w_{\mathrm{i}} x_i}}{1+exp\, w_{o+\sum_{l=1}^{n} w_{\mathrm{i}} x_i}} \quad (4)$$

The LR parameter W is selected by maximizing the conditional likelihood of the data. It is the likelihood of the Y values observed in the training data. The constraint is represented in equation (5).

$$w \Leftarrow argw \quad max\, \overline{\sum ln\ P}\, Y^l \big| X^1, W \quad (5)$$

### 3.4 K-Nearest neighbors

KNN is a model that relies on summary statistics such as median, and mode to make decisions, which means it is less sensitive to the shape of data distribution. This model works by comparing a new observation with the nearest observations in the training data set. Classification or prediction is performed by assigning the most common class or value among the K nearest neighbors. The quantity K is chosen beforehand and affects the accuracy and complexity of the model [36]. KNN is a simple and easy-to-implement algorithm, but it can be costly in terms of time and memory for large data sets. The model is represented in equation (6).

$$\mathrm{d}(x, y) = \sqrt{\sum_{i=1}^{n} (y_i - x_i)^2} \quad (6)$$

### 3.5 Decision tree

The DT is an ML model that represents a series of decisions based on certain conditions, in the form of a tree. Each internal node of the tree represents a test of a feature, and each leaf represents a class or an output value. Classification or prediction is performed by following the path through the tree from the root to a leaf based on the results of the tests at the internal nodes, as shown in Figure 1. The DT is an effective way to visualize and explain the logic behind decision making and is widely used in a wide variety of applications, including market research, risk management, and medical diagnosis [37].

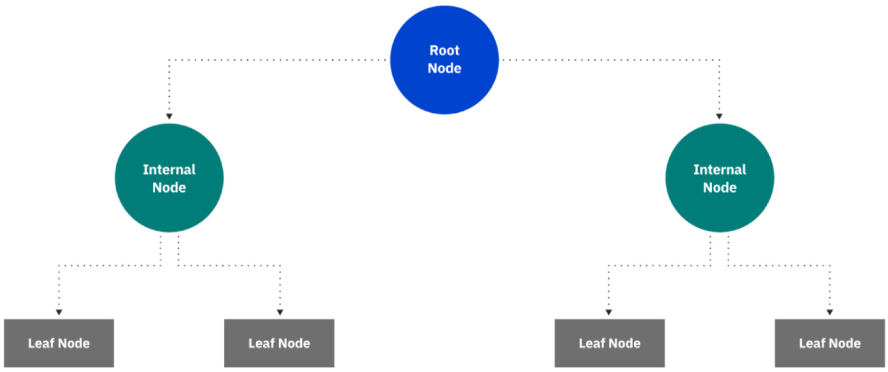


**Fig. 1.** Decision tree algorithm diagram

### 3.6 Random forest

RF is a supervised learning ML model based on decision trees. It works by creating a set of decision trees and combining them to produce a more accurate and robust prediction than a single decision tree. Instead of building a single complete decision tree, RF builds multiple trees from random subsamples of the training data set and performs voting to determine the final class or output value. RF is known for its ability to handle features with high dimensionality and correlation and is very effective for tackling classification and regression tasks [38]. In addition, RF can provide a measure of feature importance, which makes it useful for feature selection and interpretation of the results.

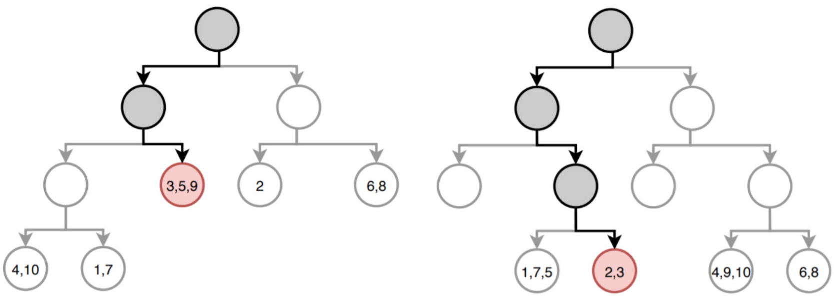


**Fig. 2.** Random Forest algorithm diagram

### 3.7 Understanding data

Natural disasters are natural phenomena that generate great human and material losses. There are several types of natural disasters, each of which can cause different damages, depending on the type and intensity of the phenomenon. However, understanding the data from these types of events is an important process in data analysis and decision making. It consists of examining, cleaning, transforming, modeling, and visualizing the data to obtain valuable and useful information. It is important to understand the data to identify and classify patterns, trends and relationships that can be useful for analysis and decision making. For this work, a total of 122k tweets related to natural disasters were used, extracted from Twitter with the hashtag #NaturalDisasters. This with the purpose of analyzing and classifying the key words in the tweets without disasters and the tweets with disasters. For which ML algorithms are used, such as: BNB, MNB, LR, KNN, DT, RF, to determine which of these are more accurate classification and present better performance. The extracted data set is composed of the following attributes, as shown in Table 1.

**Table 1.** Dataset attributes

| # | attribute | Non-Null Count | Dtype |
|---|---|---|---|
| 0 | Id | 7613 non-null | Int64 |
| 1 | Keyword | 7552 non-null | object |
| Test3 | 28 | 30 | 41 |
| 2 | Location | 5080 non-null | object |
| 3 | Text | 7613 non-null | object |
| 4 | target | 7613 non-null | Int64 |
| Dtypes: int64(2), object(3) | | | |

To understand the data, it is very important to perform a keyword analysis, as this allows to identify keywords that describe important tweets and trends. Also, this analysis allows to improve the classification efficiency to identify patterns and doing a keyword analysis on the Tweets is a crucial step to obtain understandable information from the dataset. As shown in Figure 3 the keyword distribution of non-disaster related Tweets, and in Figure 4 the keyword distribution of disaster related Tweets.

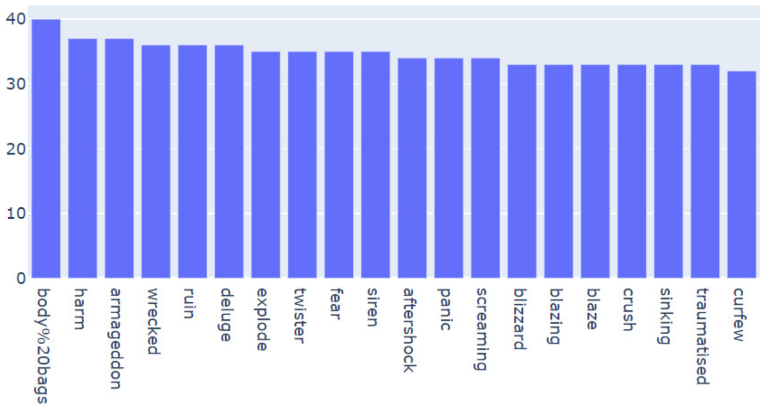


**Fig. 3.** Keyword distribution in non-disaster-related tweets

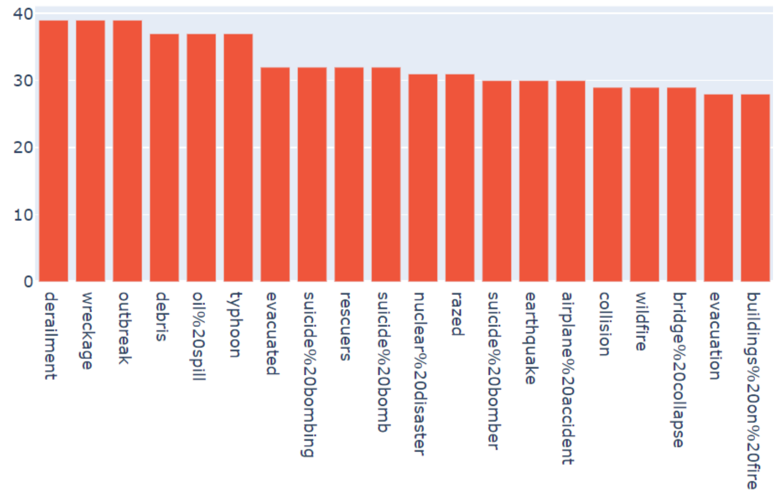


**Fig. 4.** Distribution of keywords in Tweets about disasters

The locations of disaster Tweets depend on several factors, such as magnitude, accessibility to the area, and availability of technology. In general, most Tweets are expected to come from urban areas, where technology and connectivity are available. Therefore, it is important to know the location of the Tweets to decide on the inclusion/exclusion of this attribute. Figure 5 shows the locations that are mostly recorded in the Tweets.

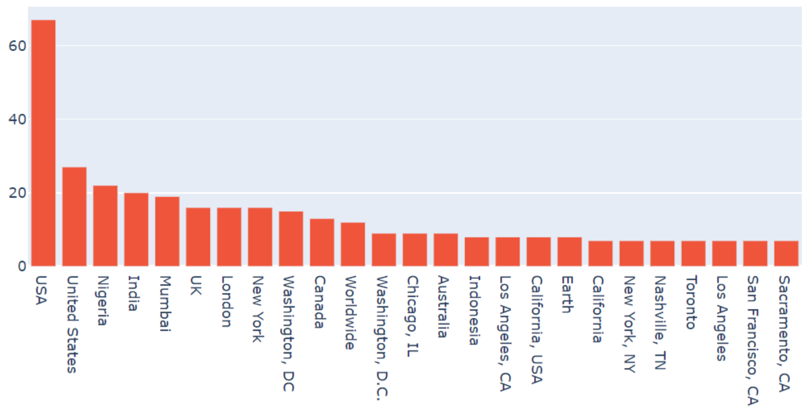


**Fig. 5.** Distribution of locations in disaster tweets

### 3.8 Data cleaning and processing

This section proceeds with data cleaning, a crucial step in data analysis, as it aims to prepare the data for analysis. This process involves correcting errors, removing duplicate data, removing missing values, converting to uppercase, cleaning special characters, removing punctuation, removing empty words, then applying stemming and lemmatization techniques, this process helps to find the base form of each word in the dictionary and reduce it to its lemma form. For example, "running" would be reduced to "run" or incomplete words such as "corr" would be reduced to "run".

### 3.9 Exploratory data analysis

The EDA is a crucial process for data analysis that provides an initial in-depth understanding of the data and a solid foundation for further analysis. It includes data visualization before training models, descriptive statistics, and trend analysis. With this technique, it is possible to identify outliers, patterns, and relationships in the data. For this case, data such as frequency of Tweets, trends in Tweets, and sentiments in tweets are explored. Also, ML models are selected, such as BNB, MNB, LR, KNN, DT, and RF, which are the most suitable to solve the stated problem. Similarly, the data set for training and another set to evaluate the accuracy of the model are also divided. Also, in this section the metrics are selected, such as: Precision, accuracy, sensitivity or recall, specificity, F1 Score, ROC Curve and loss. These metrics are the most used to evaluate the performance of ML models, and the result will depend only on each model. A very important point before training is to identify the most relevant or representative words to better understand the context and content of the data, and for this purpose the Wordclouds are used, as shown in Figure 6. This helps to make informed decisions on how to classify the data and how to fit your model. Since Byesian algorithms such as BNB and MNB, are classification algorithms and assume their independence among features. Meanwhile, the LR algorithm is used to predict categorical variables. The DT algorithm is a supervised learning algorithm that models decisions and relationships between variables. The RF algorithm uses a combination of decision trees that improve the accuracy and stability of the model.

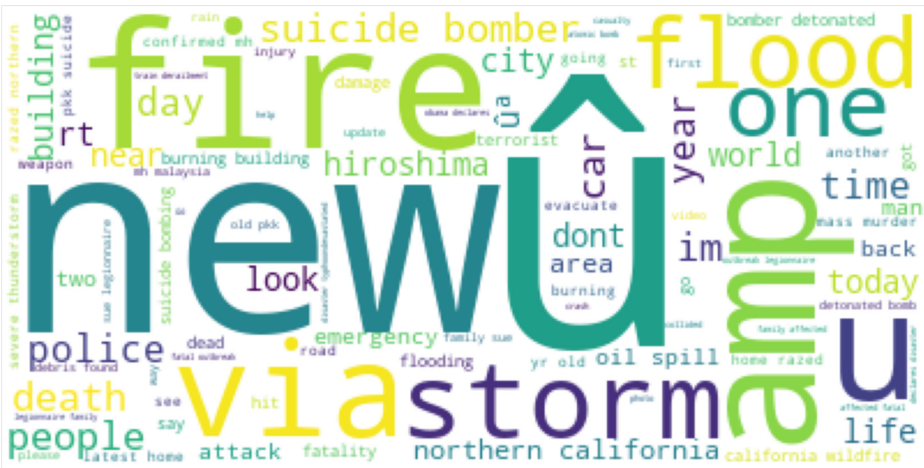


**Fig. 6.** Word cloud of Tweets on disasters

### 3.10 Model training and testing

The training process consists of feeding the model with a large dataset and adjusting the model parameters so that it can make accurate predictions on the input data. In this work, the following models will be trained: BNB, MNB, LR, KNN, DT, and RF, with the purpose of analyzing and classifying keywords in tweets without disasters and Tweets with disasters, using a total of 122k Tweets related to natural disasters extracted from Twitter with the hashtag #NaturalDisasters. Each of the six models to be trained has its own characteristics, settings, and performance. For example, the BNB model is based on calculating the probability of a given word appearing or not appearing in a document or category, and the a priori probability of each category in the training set, in this case, category, is associated with Tweets with disasters and Tweets without disasters. Meanwhile, the MNB model calculates the a priori probability of each category in the proportion of the dataset to be trained. Also, it stores the conditional probabilities for later use in classifying the Tweets. Training with the LR model aims to find the weights that best fit the training data and to be able to make accurate classifications with respect to the Tweets related to natural disasters. Training with the KNN model is simpler and consists of storing the training data in memory and calculating the distance between the stored features in each category. Training the DT model involves building a tree representing the decisions and rules used to classify the tweets, evaluating the quality of the splits, and cutting the tree to avoid overfitting. Training the RF model is very similar to the DT model since it also involves building several decision trees, combining their results, and evaluating the model to improve accuracy and stability.

After training, testing is performed to evaluate the accuracy of each of the models on data that was not used during training, for which the data was divided into 70% for training and 30% for testing. This helps to determine if the model is able to generalize well to new data and to evaluate its performance on each of the tasks. To evaluate the models, we will use the confusion matrix and the ROC curve, to indicate the quality of the models as a function of the four outcomes generated by the binary classification of the Tweets. Using a binary classifier, one can predict whether all data instances included in a test dataset will be positive or negative. This classification produces four results: true positive(VP) i.e. the model correctly predicts the presence of a condition and this condition actually exists, for work, it refers to real Tweets that are related to natural disasters; true negative(VN) i.e. the model correctly predicts the absence of a condition and this condition actually does not exist, in this case, it refers to real Tweets that exist, but are not related to natural disasters; false positive (FP) refers to a positive prediction that is incorrect, it can be said that the model predicts the presence of a condition, but this condition does not really exist, in this case, it can be stated the presence of Tweets with ambiguities or with special characters; and false negative (FN) i.e. the model predicts the absence of a condition or event, but this condition does not really exist. As shown in Figure 7.

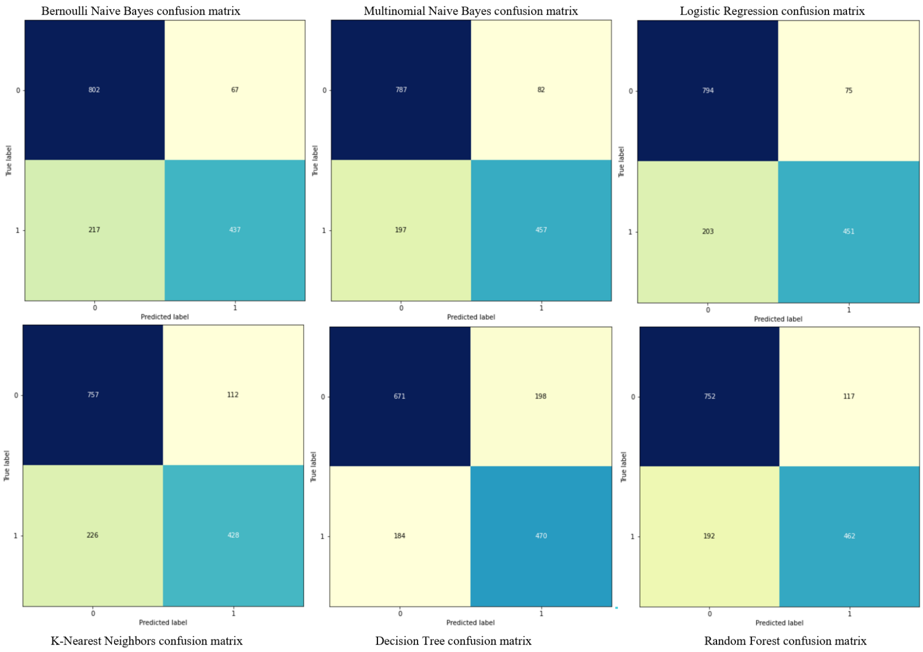


**Fig. 7.** Confusion matrix for BNB, MNB, LR, KNN, DT, and RF models

Figure 7. It indicates that the BNB model in the confusion matrix managed to correctly predict 802 Tweets in the positive class (VP); 437 Tweets in the negative class (VN), it also predicted 67 Tweets incorrectly in the positive class (false positive) and 217 Tweets in the negative class (FN). The MNB model in the confusion matrix managed to correctly predict 787 Tweets in the positive class (VP); 457 Tweets in the negative class (VN), it also predicted 82 Tweets incorrectly in the positive class (FP), and 197 Tweets in the negative class (FN). The LR model in the confusion matrix managed to correctly predict 794 Tweets in the positive class (VP); 451 Tweets in the negative class (VN), it also predicted 75 Tweets incorrectly in the positive class (FP) and 203 Tweets in the negative class (FN). The KNN model in the confusion matrix managed to correctly predict 757 Tweets in the positive class (VP); 428 Tweets in the negative class (VN), it also predicted 112 Tweets incorrectly in the positive class (FP), and 226 Tweets in the negative class (FN). The DT model in the confusion matrix managed to correctly predict 671 Tweets in the positive class (VP); 470 Tweets in the negative class (VN), it also predicted 198 Tweets incorrectly in the positive class (FP) and 184 Tweets in the negative class (FN) and The RF model in the confusion matrix managed to correctly predict 752 Tweets in the positive class (true positive); 462 Tweets in the negative class (true negative), also predicted 117 Tweets incorrectly in the positive class (FP) and 192 Tweets in the negative class (FN).

## 4 Results and discussion

This section presents the results obtained, although it is true that the six models used in this work have their strengths and weaknesses, the best option will depend on the nature of the problem and the characteristics of the data. The metrics used to evaluate the models are accuracy, recall, F1-Score and support, as shown in Table 2. Likewise, the ROC curve was used to compare the models directly, regardless of the rate of true positives and false positives in the data. In addition, it allows us to identify the point at which sensitivity and specificity are balanced, which is important to determine whether they are a priority or not. The results of the six models are compared below to determine which best fits this problem.

**Table 2.** Results of the training of ML models

| | | | | |
|---|---|---|---|---|
| **BNB** | | | | |
| | ***accuracy [%]*** | ***recall [%]*** | ***f1-score [%]*** | ***support*** |
| 0 | 79 | 92 | 85 | 869 |
| 1 | 87 | 67 | 75 | 654 |
| | | | | |
| accuracy | | | 81 | 1523 |
| macro avg | 83 | 80 | 80 | 1523 |
| weighted avg | 82 | 81 | 81 | 1523 |
| **MNB** | | | | |
| | ***accuracy [%]*** | ***recall [%]*** | ***f1-score [%]*** | ***support*** |
| 0 | 80 | 91 | 85 | 869 |
| 1 | 86 | 69 | 76 | 654 |
| | | | | |
| accuracy | | | 82 | 1523 |
| macro avg | 83 | 80 | 81 | 1523 |
| weighted avg | 82 | 82 | 81 | 1523 |
| **LR** | | | | |
| | ***accuracy [%]*** | ***recall [%]*** | ***f1-score [%]*** | ***support*** |
| 0 | 80 | 91 | 85 | 869 |
| 1 | 85 | 70 | 77 | 654 |
| | | | | |
| accuracy | | | 82 | 1523 |
| macro avg | 82 | 80 | 81 | 1523 |
| weighted avg | 82 | 82 | 81 | 1523 |
| **KNN** | | | | |
| | ***accuracy [%]*** | ***recall [%]*** | ***f1-score [%]*** | ***support*** |
| 0 | 77 | 87 | 82 | 869 |
| 1 | 79 | 65 | 72 | 654 |
| | | | | |
| accuracy | | | 78 | 1523 |

| | | | | |
|---|---|---|---|---|
| macro avg | 78 | 76 | 77 | 1523 |
| weighted avg | 78 | 78 | 77 | 1523 |
| **DT** | | | | |
| | ***accuracy [%]*** | ***recall [%]*** | ***f1-score [%]*** | ***support*** |
| 0 | 78 | 77 | 78 | 869 |
| 1 | 70 | 72 | 71 | 654 |
| | | | | |
| accuracy | | | 75 | 1523 |
| macro avg | 74 | 75 | 74 | 1523 |
| weighted avg | 75 | 75 | 75 | 1523 |
| **RF** | | | | |
| | ***accuracy [%]*** | ***recall [%]*** | ***f1-score [%]*** | ***support*** |
| 0 | 80 | 87 | 83 | 869 |
| 1 | 80 | 71 | 75 | 654 |
| | | | | |
| accuracy | | | 80 | 1523 |
| macro avg | 80 | 79 | 79 | 1523 |
| weighted avg | 80 | 80 | 80 | 1523 |

Table 2 shows the training results of the BNB, MNB, LR, KNN, DT, and RF models. It can be seen that the BNB, MNB, and LR models show better results in terms of processing time, test accuracy, precision and F1-Score, as shown in Table 3.

**Table 3.** Summary of the three best performing models

| Model | Processing Time | Test accuracy | Precision | F1-score |
|---|---|---|---|---|
| Logistic Regression | 3.22 Secs | 0.82 | 0.85 | 0.77 |
| Multinomial Naïve Bayes | 0.22 Secs | 0.82 | 0.86 | 0.76 |
| Bernoulli Naïve Bayes | 0.92 Secs | 0.81 | 0.87 | 0.75 |

In general terms, LR, BNB, and MNB models are very good for binary classification and perform better in binary data problems, as is the case in this work. The BNB and MNB models are probabilistic classification models and are characterized by being very effective in text analysis with Twitter data. Also, the LR binary classification model is widely used in text analysis. It is important to note that, although it is true that in this work the BNB, MNB, and LR models obtained better results, this does not mean that they are the best classification models, but rather that it depends on the context and the data set. The next point to consider is that Twitter data is unstructured and usually has an informal and abbreviated language, which can affect the accuracy of the models. Figure 8 shows the ROC curve of the six models, though it allows evaluating the ability to discriminate the two classes (tweets related to natural disasters and unrelated tweets), in it, the relationship between the rate of true positives and the rate of true negatives is presented as the decision threshold of the model changes. For example, the BNB model in Figure 8, the ROC curve of class 0 as the iterations progress is approaching 1, this means that the accuracy of the optimal prediction, in this case the rate reached 87%, very similar are the results for the MNB and LR models, which reached 87%. With respect to the KNN model, the ROC curve in class 0 obtained a result of 82%, well below the BNB, MNB and LR models. This does not mean that the KNN model does not have an optimal rate of return, this depends on the data set and the context. The RF model obtained an optimal performance rate of 86%, placing it among the best classification and prediction models for this type of task. Finally, the DT model according to the ROC curve in class 0, obtained a performance rate of 75%, comparing with the other models, it ranks below the other trained models. However, it is important to keep in mind that the ROC curve only measures the discrimination capacity of the model and considers other important factors such as the complexity or interpretability of the decisions.

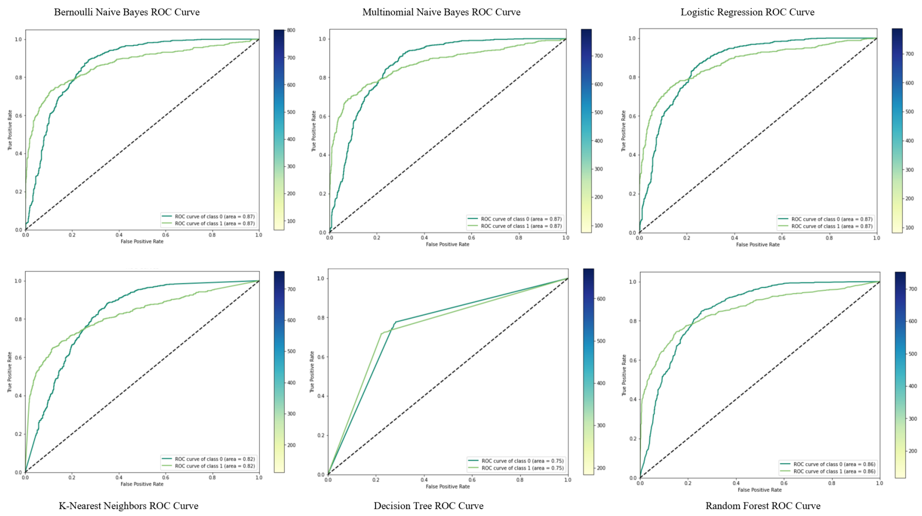


**Fig. 8.** ROC curve of the trained models

Knowing what information is available through social networks contributes significantly to humanitarian organizations and governments to prepare for and respond in a timely manner to natural disasters. In this sense, the results presented in Table 2, Table 3, and Figure 8, allow an analysis and a comparison with the results obtained in related works. The six models used have achieved a performance between 75% and 87% in classifying and analyzing tweets related to real natural disasters through posts with the hashtag #NaturalDisasters. For example, the BNB, MNB, and LR models achieved a performance rate on average with an accuracy of 86%, a recall of 91%, and an F1-Score of 85%, slightly superior to the results obtained in the work [24], where they used the RF, BNB, and Support Vector Machine (SVM) models to monitor natural disaster data on Twitter, managing to achieve an accuracy of 85%, a recall of 82%, and an F1-Score of 84%. However, if we compare with the work [23], where they created a clustering algorithm integrating BERT models and supervised logistic regression, the clustering accuracy obtained is 93.56%, much higher than that achieved in this work. The results of the models will depend on the context and the dataset. Lately, the identification of

Tweets related to natural disasters has been a focus of research where they aim to understand the scope and magnitude of the damage they can cause. For example, to determine the contribution of our work we compare with the research [26] and [28] unlike our work that focuses on identifying and classifying disaster-related Tweets, these two types of research [26] and [28], focus on analyzing the behavior of post-event tweet posts related to natural disasters, in order to gather information to monitor natural disasters using convolutional neural network (CNN) models. Although it is true that the two-research worked with different CNN architectures, they obtained very similar results in terms of F1-Score, the performance of the models used was on average 60%, a very low result to classify disaster-related information, this is because the models used are not optimal in classification tasks or present difficulties to manage multiple Twitter tags. While in this work the results of the trained models have shown efficiency to identify and classify disaster-related Tweets.

## 5 Conclusions

Social networks, specifically Twitter, are a medium through which large amounts of relevant information are shared regarding natural disasters, either announcing or requesting humanitarian aid. The identification and classification of disaster-related Tweets is of vital importance in crisis contexts. The methodology used in this work can be used during and after a disaster to identify and classify Tweets of real damage and significant information to make more informed decisions. The results obtained in this work generate valuable input in text classification to identify disaster-related Tweets. Six models (BNB, MNB, LR, KNN, DT, and RF) were trained to obtain the following results in performance: 87%, 87%, 87%, 87%, 82%, 75%, and 86%, respectively. All six models achieved very similar and acceptable results. However, the BNB, MNB, and LR models fared much better in terms of performance in their respective metrics, such as processing time, Test accuracy, precision, and F1-score, as can be seen in Table 3 and in the ROC curve in Figure 8. Therefore, in a disaster context, it is recommended to use one of the three models BNB, MNB or LR to classify Tweets since they have obtained the best results in this type of text classification tasks. A very important aspect of this work is that information was obtained from different types of very recent and geolocated disasters, which, usually, this type of data is lacking during the first hours of the event. To complement this work, future work can be carried out, such as the implementation of an ML model to assess material damage in natural disasters.

## 6 References

[1] T. Iancu *et al.*, "A Scientometric Analysis of Climate Change Adaptation Studies," *Sustainability 2022, Vol. 14, Page 12945*, vol. 14, no. 19, p. 12945, Oct. 2022. https://doi.org/10.3390/su141912945

[2] United Nations Office for Disaster Risk Reduction, "Statement on the establishment of a G20 Working Group on Disaster Risk Reduction," Aug. 2018, Accessed: Jan. 28, 2023. [Online]. Available: https://www.undrr.org/

[3] Global Network of Civil Society Organisations for Disaster Reduction, "Risk-Informed Development Guide," 2019. https://www.ifrc.org/es/nuestro-trabajo/desastres-clima-y-crisis/que-es-desastre (accessed Jan. 28, 2023).
[4] H. Y. Kwon and Y. O. Kang, "Risk analysis and visualization for detecting signs of flood disaster in Twitter," *Spatial Information Research*, vol. 24, no. 2, pp. 127–139, Apr. 2016. https://doi.org/10.1007/s41324-016-0014-1
[5] A. P. Rodrigues, R. Fernandes, A. Bhandary, A. C. Shenoy, A. Shetty, and M. Anisha, "Real-Time Twitter Trend Analysis Using Big Data Analytics and Machine Learning Techniques," *Wirel Commun Mob Comput*, vol. 2021, 2021. https://doi.org/10.1155/2021/3920325
[6] BID, "Natural disasters | IADB." https://www.iadb.org/es/desastresnaturales (accessed Jan. 28, 2023).
[7] D. Voloshchuk *et al.*, "Neural network modeling of seismic behaviour of the hellenic Arc: strengths and limitations," *International Journal of Advanced Technology and Engineering Exploration*, vol. 9, no. 95, pp. 1428–1447, Oct. 2022. https://doi.org/10.19101/IJATEE.2021.876293
[8] New company, "The persistent causes of a disaster that has not yet ended." https://nuso.org/articulo/terremoto-en-haiti-las-causas-persistentes-de-un-desastre-que-no-ha-terminado/ (accessed Jan. 28, 2023).
[9] M. Zhang, Q. Huang, and H. Liu, "A Multimodal Data Analysis Approach to Social Media during Natural Disasters," *Sustainability*, vol. 14, no. 9, p. 5536, May 2022. https://doi.org/10.3390/su14095536
[10] R. R. Arinta, S. Suyoto, and A. W. R. Emanuel, "Effectiveness of Gamification for Flood Emergency Planning in the Disaster Risk Reduction Area," *International Journal of Engineering Pedagogy (iJEP)*, vol. 10, no. 4, p. 108, Jul. 2020. https://doi.org/10.3991/ijep.v10i4.13145
[11] M. F. Omar, M. N. Mohd Nawi, J. M. Jamil, A. M. Mohamad, and S. Kamaruddin, "Research Design of Mobile Based Decision Support for Early Flood Warning System," *International Journal of Interactive Mobile Technologies (iJIM)*, vol. 14, no. 17, p. 130, Oct. 2020. https://doi.org/10.3991/ijim.v14i17.16557
[12] D. Rudrapal, S. Boro, J. Srivastava, and S. Singh, "A Deep Learning Approach to Predict Football Match Result," H. S. Behera, J. Nayak, B. Naik, and D. Pelusi, Eds., in Advances in Intelligent Systems and Computing, vol. 990. Singapore: Springer Singapore, 2020, pp. 93–99. https://doi.org/10.1007/978-981-13-8676-3_9
[13] A. Frifra, M. Maanan, H. Rhinane, and M. Maanan, "An artificial intelligence approach to prediction of extreme events: The case of storms in western france," *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, vol. 46, no. 4/W3-2021, pp. 115–122, Jan. 2022. https://doi.org/10.5194/isprs-archives-XLVI-4-W3-2021-115-2022
[14] J. Ren, S. Zhou, J. Wang, S. Yang, and C. Liu, "Research on Identification of Natural and Unnatural Earthquake Events Based on AlexNet Convolutional Neural Network," *Wirel Commun Mob Comput*, vol. 2022, 2022. https://doi.org/10.1155/2022/6782094
[15] J. Liu, F. Dong, and Z. Li, "Application of Convolutional Neural Network to GIS and Physics," *Comput Intell Neurosci*, vol. 2022, 2022. https://doi.org/10.1155/2022/8559343
[16] M. Avand, A. Kuriqi, M. Khazaei, and O. Ghorbanzadeh, "DEM resolution effects on machine learning performance for flood probability mapping," *Journal of Hydro-Environment Research*, vol. 40, pp. 1–16, Jan. 2022. https://doi.org/10.1016/j.jher.2021.10.002
[17] E. Koo and H. Kim, "Empirical strategy for stretching probability distribution in neural-network-based regression," *Neural Networks*, vol. 140, pp. 113–120, Aug. 2021. https://doi.org/10.1016/j.neunet.2021.02.030

[18] J. Won and J. Shin, “Machine learning-based approach for seismic damage prediction method of building structures considering soil-structure interaction,” *Sustainability (Switzerland)*, vol. 13, no. 8, Apr. 2021. https://doi.org/10.3390/su13084334
[19] M. Koivisto, “Gamified Learning of Project Business Skills,” *International Journal of Advanced Corporate Learning (iJAC)*, vol. 15, no. 1, pp. 45–60, Mar. 2022. https://doi.org/10.3991/ijac.v15i1.26689
[20] I. Dhall, S. Vashisth, and S. Saraswat, “Text Generation Using Long Short-Term Memory Networks,” in *Lecture Notes in Networks and Systems*, Springer, 2020, pp. 649–657. https://doi.org/10.1007/978-981-15-2329-8_66
[21] E. P. Prasetya and E. C. Djamal, “Rainfall Forecasting for the Natural Disasters Preparation Using Recurrent Neural Networks,” *Proceedings of the International Conference on Electrical Engineering and Informatics*, vol. 2019-July, pp. 52–57, Jul. 2019. https://doi.org/10.1109/ICEEI47359.2019.8988838
[22] C. Y. Tang, W. W. Cheng, T. Y. Hsu, C. J. Jeng, and Y. L. Wu, “Using Neural Networks to Label Rain Warning for Natural Hazard of Slope,” *Proc Int Conf Mach Learn Cybern*, vol. 2019-July, Jul. 2019. https://doi.org/10.1109/ICMLC48188.2019.8949267
[23] L. Huang, P. Shi, H. Zhu, and T. Chen, “Early detection of emergency events from social media: a new text clustering approach,” *Natural Hazards*, vol. 111, no. 1, pp. 851–875, Mar. 2022. https://doi.org/10.1007/s11069-021-05081-1
[24] 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 2019, Vol. 19, Page 1746*, vol. 19, no. 7, p. 1746, Apr. 2019. https://doi.org/10.3390/s19071746
[25] M. Maldonado, D. Alulema, D. Morocho, and M. Proano, “System for monitoring natural disasters using natural language processing in the social network Twitter,” *Proceedings - International Carnahan Conference on Security Technology*, vol. 0, Jul. 2016. https://doi.org/10.1109/CCST.2016.7815686
[26] Y. Suzuki, Y. Yoneda, and A. Nadamoto, “Analysis of Behavioral Facilitation Tweets for Large-Scale Natural Disasters Dataset Using Machine Learning,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 12924 LNCS, pp. 161–169, 2021. https://doi.org/10.1007/978-3-030-86475-0_16
[27] M. A. Sit, C. Koylu, and I. Demir, “Identifying disaster-related tweets and their semantic, spatial and temporal context using deep learning, natural language processing and spatial analysis: a case study of Hurricane Irma,” *Int J Digit Earth*, vol. 12, no. 11, pp. 1205–1229, Nov. 2019. https://doi.org/10.1080/17538947.2018.1563219
[28] N. R. Paul and R. C. Balabantaray, “Detecting Crisis Event on Twitter Using Combination of LSTM, CNN Model,” *Communications in Computer and Information Science*, vol. 1372 CCIS, pp. 71–80, 2021. https://doi.org/10.1007/978-981-16-2723-1_8
[29] Z. Jiang *et al.*, “Coupling machine learning and weather forecast to predict farmland flood disaster: A case study in Yangtze River basin,” *Environmental Modelling & Software*, vol. 155, p. 105436, Sep. 2022. https://doi.org/10.1016/j.envsoft.2022.105436
[30] S. Xie, C. Hou, H. Yu, Z. Zhang, X. Luo, and N. Zhu, “Multi-label disaster text classification via supervised contrastive learning for social media data,” *Computers and Electrical Engineering*, vol. 104, p. 108401, Dec. 2022. https://doi.org/10.1016/j.compeleceng.2022.108401
[31] F. Ouatik, M. Erritali, F. Ouatik, and M. Jourhmane, “Students’ Orientation Using Machine Learning and Big Data,” *International Journal of Online and Biomedical Engineering (iJOE)*, vol. 17, no. 01, p. 111, Jan. 2021. https://doi.org/10.3991/ijoe.v17i01.18037

[32] A. O. Salau, T. A. Assegie, A. T. Akindadelo, and J. N. Eneh, "Evaluation of Bernoulli Naive Bayes model for detection of distributed denial of service attacks," *Bulletin of Electrical Engineering and Informatics*, vol. 12, no. 2, pp. 1203–1208, Apr. 2023. https://doi.org/10.11591/eei.v12i2.4020
[33] N. Matar, W. Matar, and T. Al Malahmeh, "Predictive Model for Students Admission Uncertainty Using Naïve Bayes Classifier and Kernel Density Estimation (KDE)," *International Journal of Emerging Technologies in Learning*, vol. 17, no. 8, pp. 75–96, 2022. https://doi.org/10.3991/ijet.v17i08.29827
[34] S. Justice, D. K. Sewell, H. Li, A. C. Miller, and P. M. Polgreen, "Evidence of within-facility patient-patient Clostridiodes difficile infection spread across diverse settings," *Epidemiol Infect*, vol. 151, p. e4, Dec. 2022. https://doi.org/10.1017/S0950268822001893
[35] F. E. Harrell, "Regression Modeling Strategies," 2015. https://doi.org/10.1007/978-3-319-19425-7
[36] O. Iparraguirre-Villanueva *et al.*, "Text prediction recurrent neural networks using long short-term memory-dropout," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 29, no. 3, pp. 1758–1768, Mar. 2023. https://doi.org/10.11591/ijeecs.v29.i3.pp1758-1768
[37] O. Iparraguirre-Villanueva, V. Guevara-Ponce, O. R. Paredes, F. Sierra-Liñan, J. Zapata-Paulini, and M. Cabanillas-Carbonell, "Convolutional Neural Networks with Transfer Learning for Pneumonia Detection," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 9, pp. 544–551, Dec. 2022. https://doi.org/10.14569/IJACSA.2022.0130963
[38] J. Yu, "Academic Performance Prediction Method of Online Education using Random Forest Algorithm and Artificial Intelligence Methods," *International Journal of Emerging Technologies in Learning*, vol. 16, no. 5, pp. 45–57, 2021. https://doi.org/10.3991/ijet.v16i05.20297

## 7 Authors

**Orlando Iparraguirre-Villanueva** Systems Engineer with a master's degree in Information Technology Management, Ph.D. in Systems Engineering from Universidad Nacional Federico Villarreal Peru. Undergraduate and postgraduate lecturer in different universities in the country. Specialist in software development, IoT, business intelligence, open-source software, augmented reality, machine learning, text mining, and virtual environments (Email: oiparraguirre@ieee.org).

**Efraín Melgarejo-Graciano** Systems Engineer with a master's degree in IT Management. Certified in Scrum Master, DevOps, DBA, and others. Specialist in technological innovation projects. Undergraduate and postgraduate teacher in different universities in Peru. Coordinator in entrepreneurship projects in universities developing software applications, IoT, Business Intelligence, free software, Augmented Reality, Machine Learning, text mining (Email: profesormelquiades@gmail.com).

**Gloria Castro-Leon** D. in Systems Engineering, Senior Lecturer, with more than 25 years of experience in undergraduate and postgraduate courses, both in national and private universities with extensive knowledge in project management and information technology, ITIL® Foundation Certificate in IT Service, with extensive experience in administrative management of the Peruvian University (Email: gcastrol@untels.edu.pe).

**Sandro Olaya Cotera** Certified Public Accountant in 2008 at the National University of San Marcos, with a master's degree in business administration and international Relations. Professor at various universities. He is an advisor and jury of research papers, co-author in various publications (Email: sandro.olaya@usil.pe).

**John Ruiz-Alvarado** D. in Education. Currently, assigned to the Faculty of Engineering of the Universidad Tecnológica del Perú (UTP). At the same time, I am a lecturer in Neuropedagogy, Management and Pedagogical Leadership (Email: C20326@utp.edu.pe).

**Andrés Epifanía-Huerta** master's in system engineering with mention in information and communication technologies. Currently, attached to the Faculty of Engineering of the Universidad Tecnológica del Perú (UTP). Speaker on issues of technological innovation, technologies applied to education, management of business technologies (Email: C25267@utp.edu.pe).

**Michael Cabanillas-Carbonell** Systems Engineer. Qualified researcher by the National Council of Science, Technology and Innovation - Peru. Professor conducting research at Universidad Privada del Norte. President of the IEEE-Peru Chapter of the Education Society. President of the EIRCON (Engineering International Research Conference) (Email: mcabanillas@ieee.org).

**Joselyn Zapata-Paulini** Systems Engineering. M.Sc. student in Environmental Management and Sustainable Development at Universidad Continental, author of scientific articles. Specialization in augmented reality, and software development (Email-mail: 70994337@continental.edu.pe).

Article submitted 2023-03-26. Resubmitted 2023-05-17. Final acceptance 2023-05-23. Final version published as submitted by the authors.