PERCEIVING EMOTIONS IN TWEETS USING MULTI-LABEL CLASSIFICATION

Vaishali Tidake[1*], Balasaheb Tarle[2], Sopan Talekar[3]

Abstract
Nowadays, many people, especially youth, express their opinions regarding different topics using tweets. It has become essential to extract information from the tweets posted by users. This information is helpful for decision-making to increase business profit and to identify and stop the spreading of wrong or sensitive information in critical situations. It is observed that the majority of tweets express multiple emotions, thereby making use of multi-label (ML) classification crucial for processing data in them. MLCET, which takes preprocessed tweets as input and then uses prior and estimated probabilities computed from statistics of neighbors using feature similarities and label dissimilarities, performs better than three ML algorithms when evaluated for five metrics. The work also depicted the importance of stemming and lemmatization for enhancing performance on text data. Experimentation has shown 9%, 16%, 15%, 23%, and 16% improvements in MLCET for one error, accuracy, F1 score, macro, and micro F1, respectively.

Keywords: Multi-label classification, tweets, emotions, machine learning, preprocessing, natural language processing, sentiment

[1*,2,3]Department of Computer Engineering, MVPS's KBT College of Engineering, Savitribai Phule Pune University, Pune, India
E-mail: [1]tidake.vaishali@kbtcoe.org, [2]tarle.balasaheb@kbtcoe.org & [3]talekar.sopan@kbtcoe.org

***Corresponding Author:** Vaishali Tidake
*Department of Computer Engineering, MVPS's KBT College of Engineering, Savitribai Phule Pune University, Pune, India, E-mail: tidake.vaishali@kbtcoe.org

DOI: 10.48047/ecb/2023.12.si10.00212

I. Introduction

Nowadays, many people, especially youth, give their opinions regarding different topics using tweets. If these tweets are analyzed carefully, they show other emotions of people. Various organizations use these emotions for their benefit. For example, to add new features to electronic products, to make movies or serials, to prepare YouTube videos, and so on. There are different ways to deal with emotions in tweets.

Because of the increased social media usage, generating large amounts of content, it has become essential to perceive emotions expressed in customer tweets. It is termed sentiment analysis in natural language processing [1].
Analyzing information gathered from Twitter has become crucial, which can help make decisions in certain situations [2]. As social media like Twitter generates massive data at high speed, it is challenging and complex to understand shared information because users generally use abbreviations and variations of spellings [3].
Due to its growing community, Twitter is also used for emergency communication. When users express their opinions, it is possible to represent more than one emotion associated with that tweet. Most existing research treats these tweets as binary or multi-class [4].

The need for a multi-label (ML) classification model arises because of the observation that most of the tweets are associated with multiple emotions. An ML classifier can help better for information extraction [5]. Examples of multi-label tweets are shown in Table 1 [6].

Recognizing emotions in tweets is essential for manufacturers of electronic devices, home appliances, and movie makers. It is also crucial for society and the government to stop spreading fake news and wrong or sensitive information in situations like COVID-19 [7] [8].

Section II discusses multi-label classification and sentiment analysis of multi-label tweets. Section III describes the methodology for preprocessing and classifying ML tweets using neighbor statistics, feature similarity, and label dissimilarity. Section IV describes evaluation metrics, dataset characteristics, and performance comparison. Finally, the work is concluded in the last section.

II. Related Work

A. Multi-label Classification (MLC)

For an ML dataset E and label space S, let (x_q, L_q) denote q^{th} instance of E, where $L_q \subseteq S$.

Then the objective of $g_c(x_q)$ is to find P_q, a prediction of labels for an instance x_q. Table 2 differentiates between SLC and MLC [9] [22] [26].

Researchers have implemented MLC in three different ways: (i) transformation that uses

Table 1 Examples of multi-label tweets

ID	Tweet	anger	anticipation	Disgust	fear	joy	love	optimism	pessimism	sadness	surprise	trust
2017-En-31535	Whatever you decide to do make sure it makes you #happy.	0	0	0	0	1	1	1	0	0	0	0
2017-En-11034	Weak people revenge. Strong people forgive. Intelligent people ignore.	1	0	0	0	0	0	1	0	0	0	1
2017-En-31561	Sometimes I get to sit back and be proud of myself for pleasing Him so well.	0	0	0	0	1	1	1	0	0	0	0
2017-En-40175	It's way too hard not to get discouraged.	1	0	0	0	0	0	0	1	1	0	0

Table 2 Single-label versus Multi-label Classification

Sr. No.	Single-label Classification (SLC)	Multi-label Classification (MLC)
1	One instance is associated with one label.	One instance is associated with a set of labels.
2	Also called Single-instance single-label learning (SISL).	Also called Single-instance multi-label learning (SIML).
3	$f_{SLC} : X \rightarrow L$	$f_{MLC} : X \rightarrow 2^L$
4	Every object represents only one semantic concept.	An object represents one or more semantic concepts.
5	Example of a single-label tweet: Tweet: My best friends driving for the first time with me in the car #terrifying — Emotion: fear	Example of multi-label tweet: Tweet: I'm doing all this to make sure you smiling down on me bro — Emotions: Joy, love, optimism

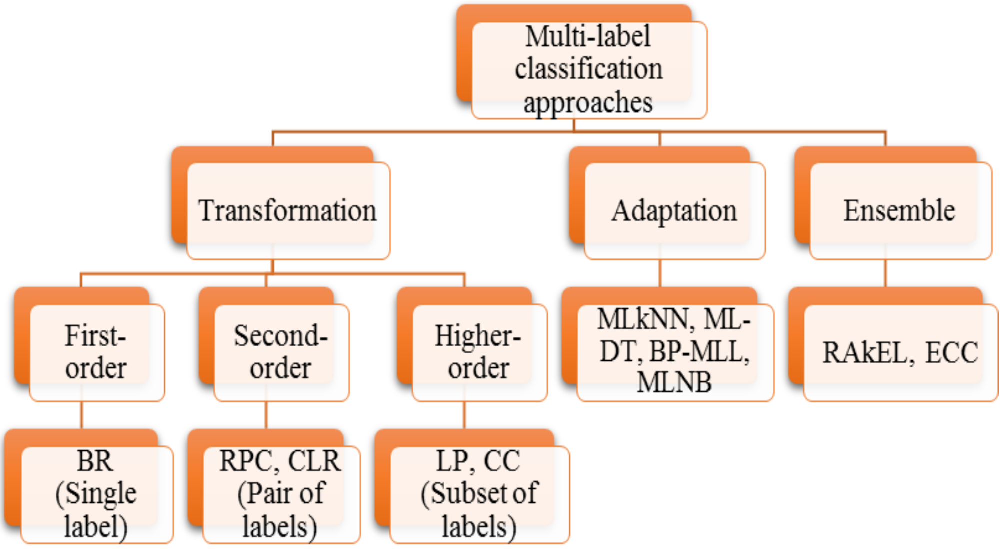


Fig. 1 Taxonomy of MLC approaches

single / pair / subset of labels, (ii) adaptation that alters conventional classifiers to deal with ML, and (iii) ensemble that groups multiple ML classifiers in different manners. Fig. 1 shows the taxonomy of MLC [9].

Methods that follow different MLC approaches are shown at the leaf nodes of Fig. 1 hierarchy.

B. Sentiment analysis

There are different ways to deal with emotions in tweets [6].

Alec Go et al. [10] used Twittratr, which lists positive and negative keywords. The authors have used the following parameters and compared their results:

- Unigrams: These represent the use of a single word.
- Bigrams: It represents the use of a pair of a word. It is helpful to understand emotions in the case of 'not easy,' for example.
- Use of both unigrams and bigrams: It has shown improved accuracy.
- POS: It uses Part of Speech tags because words may have different meanings per context. For example, the words 'book plays different roles in the following two sentences: 'book my ticket' and 'read a book.' It is a verb in the first sentence and a noun in the second.

Jishnu Ray Chowdhury et al. [11] combined various resources to form a multi-lingual database of tweets; constructing a dataset from different resources requires mapping existing classes to new ones. They used datasets of 4 languages, namely English, French, Italian, and Spanish, and used the remaining samples to form the fifth group.

Binary classifiers are used on the single vector representation of vectors that are obtained from a sentence encoder. The authors used the M-BERT model supporting Manifold Mixup to interpolate input data from different resources. The F1 score is used to test performance on four English datasets.

Alan Aipe et al. [5] proposed a Convolutional Neural Network (CNN) based deep learning architecture for MLC that categorizes tweets related to disaster into seven categories [5].

Twitter is used widely by politicians for promotions and communication with voters. Anuradha Surolia et al. [12] processed the PoliEMO dataset having 3.5K political tweets in the Indian context. They used five machine learning, three deep learning, and one transformer-based algorithm to understand six emotions in PoliEMO.

In the digital environment, online services must recognize the views expressed by their users. Xuan Liu et al. [13] designed sample-based and label-based MLkNN and compared their performance with MLkNN. Authors claim that basic MLkNN uses only features within the sentence. But accuracy can be improved if features of adjacent sentences are also used. So the probability of how much emotion is transferred between sentences and among all sentences of a tweet is calculated by authors, and pair-wise label correlation is also considered.

Tao and Fang [1] developed an AESA algorithm to capture implicit and explicit sentiments of multi-label tweets posted on social media.

Iqra Ameer et al. [14] used a combination of words 1-3 grams and characters 3-9 grams for emotion classification. They observed that character-n-

gram is inadequate, and word 1-gram is with BR and RFC works better on the SemEval-2018 dataset.
Rasha Obeidat et al. [15] noticed that misleading information caused severe issues, especially COVID-19. So it is critical to process such data for giving warnings or removing such posts immediately.

Mohammed Jabreel and Antonio Moreno [16] transformed a classification problem from multi-label to only one binary problem, unlike multiple binary problems as in BR. They designed a deep learning algorithm for solving this binary problem.
Semiu Salawu [17] introduced a dataset having information related to online abuse and cyberbullying detection. They also checked their dataset for the existence of gender bias.

Rohitash Chandra et al. [7] developed a framework that preprocessed extracted tweets and analyzed sentiments using LSTM and BERT models on COVID-19 data across different states in India.
Zahir Abbas Khan et al. [8] preprocessed collected news data from which features are extracted, followed by the application of CNN. Before CNN, the LISPW algorithm is applied to detect fake news. Based on linguistic features.

III. Methodology

The methodology for MLC of tweets adopted here uses a two-step process. To improve the quality of a dataset [18], step 1 uses the Algorithm PreMLT (Preprocessing of Multi-label Tweets) that takes Multi-label Dataset (MLDB) as input.

MLDB contains tweets and eleven associated emotions. PreMLT performs five operations on tweets (as shown in Fig. 2) to make it suitable for processing in step 2.

Machines cannot understand text data. Hence it is converted to number form using Count Vectorizer.

Table 3 Summary of related work for user emotion analysis

Reference	Methods	Datasets	Performance metrics
7	LSTM, BD-LSTM and BERT models	Senwave COVID-19 dataset	BCE Loss, Hamming Loss, Jaccard Score, LRAP Score, F1 Macro, F1 Micro
8	LISPW and CNN, LSTM and KNN for comparison	Gossipcop and PolitiFact dataset from FakeNewsNet Dataset	Accuracy, Precision, Recall, F1-score, Specificity, Kappa coefficient
9	MLC - BR, BPNN, CC, LP; SLC - Bayes Net, SGD, SMO, Voted Perceptron, AdaBoostM1, Attribute Selected Classifier, Bagging, Filtered Classifier, Decision Table, J48, and Random Forest	SemEval-2018	Accuracy, MicroF1, MacroF1, Exact Match, Hamming Loss
10	Twittratr, Naive Bayes, Maximum Entropy, Support Vector Machines	Collected data	Accuracy
12	5 MLC (BR, CC, LP, MLkNN with GB, LR) 3 deep learning (CNN, LSTM, and BiLSTM), and 1 transformer-based method (BERT)	PoliEMO	Micro-F1, Macro-F1, Accuracy
13	MLkNN, sample-based MLkNN (S-MLkNN), and label-based MLkNN (L-MLkNN)	Sentiment140 from www.kaggle.com	Subset Accuracy, Hamming Loss, One-Error, Ranking Loss, Average Precision, Accuracy, Precision, Recall, and F-score
15	5 transformer-based models: AraBERT , AraBERT-COV19, AraBERTv02-Twitter, MARBERTv02, XLM-R	ArCOV19-Rumors, Arabic COVID-19 binary misinformation dataset, Vaccine-related tweets	Macro-averaged and micro-averaged Precision, Recall, and F1 scores, accuracy
16	Binary Neural Network (BNet), transformation method called *xy*-pair-set	SemEval2018 Task1: Affect in Tweets	Accuracy (Jaccard), Precision, Recall, Micro F1, Macro F1
17	RoBERTa pre-trained model	Dataset created by Davidson, Kaggle Toxic Comments dataset	Macro ROC-AUC, Accuracy, Hamming Loss, Macro F1, Micro F1

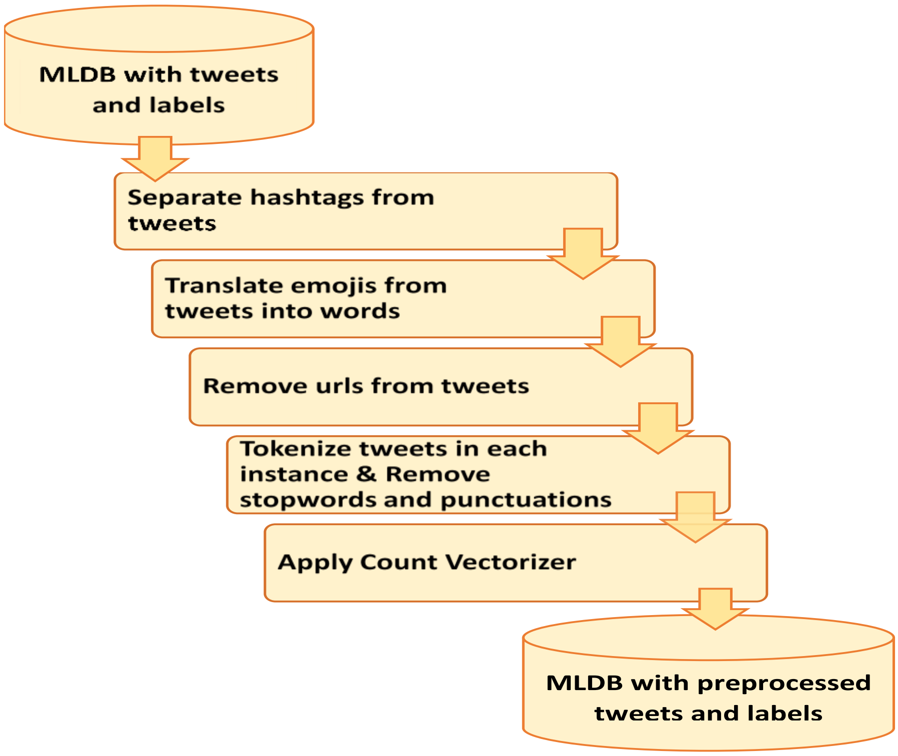


Fig. 2 PreMLT Algorithm

Step 2 uses Algorithm MLCET (*M*ulti-*l*abel *C*lassification of *E*motions in *T*weets) as shown in Fig. 3. It takes both train and test sets of MLDB with preprocessed tweets and associated emotions, #nearest neighbors k, threshold Th and a smoothing parameter p. Values 10 and 0.5 are used during experimentation for k and Th respectively. Pseudocode for MLCET is listed in Fig. 4.

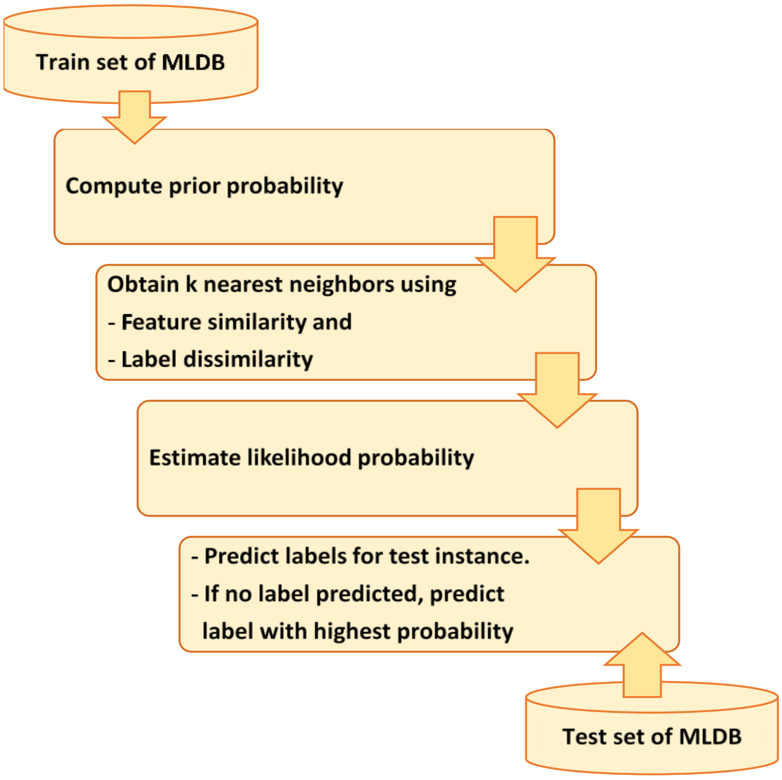


Fig. 3 MLCET Algorithm

Algorithm MLCET ($trainMLDB$, $testMLDB$, k, Th, p)

Begin

(1) Computation of prior probability distribution

a) For "a tweet T belongs to emotion e":

$P(H_e = 1) = (p + cnt^{(e)})\ /\ (2 \times p + q)$ *where q = |trainMLDB|*

b) For "a tweet T does not belong to emotion e":

$P(H_e = 0) = 1 - P(H_e = 1)$

(2) Selection of k nearest neighbors for each tweet $T_i \in trainMLDB$ using feature similarity and label dissimilarity

(3) Estimation of a likelihood probability distribution:

a) When a tweet T has j neighbors having emotion e and "a tweet T belongs to the emotion e":

$$P(\mathrm{E} = \mathrm{j}|H_e = 1) = \frac{\mathrm{p}\ +\ F_1^{(e)}[j]}{p\ x\ (1 + \mathrm{k}) + \sum_{r=0}^{k} F_1^{(e)}[r]}, 0 \le j \le k$$

b) When a tweet T has j neighbors associated with emotion e and "a tweet T does not belong to the emotion e":

$$P(\mathrm{E} = \mathrm{j}|H_e = 0) = \frac{\mathrm{p}\ +\ F_0^{(e)}[j]}{p\ x\ (1 + \mathrm{k}) + \sum_{r=0}^{k} F_0^{(e)}[r]}, 0 \le j \le k$$

(4) Searching k nearest neighbors of an unlabeled tweet for a tweet $T_i \in testMLDB$ using feature similarity

(5) Predicting each emotion e for the unlabeled tweet t:

$$t_e = 1 \text{ if} \left(\frac{\mathrm{P}(H_e = 1) \times \mathrm{P}(\mathrm{E} = \mathrm{j}|H_e = 1)}{\mathrm{P}(H_e = 1) \times \mathrm{P}(\mathrm{E} = \mathrm{j}|H_e = 1) + \mathrm{P}(H_e = 0) \times \mathrm{P}(\mathrm{E} = \mathrm{j}|H_e = 0)} \right) \ge Th$$

(6) if $\forall_{e=1}^{\#emotions} t_e = 0$ then

$$x = \arg max_e \left(\frac{P(H_e=1) \times P(E=j|H_e=1)}{P(H_e=1) \times P(E=j|H_e=1) + P(H_e=0) \times P(E=j|H_e=0)} \right)$$

Set $t_x = 1$

End

Output: Prediction of emotions for unlabeled tweet *t*

Fig. 4 Pseudocode for Algorithm MLCET

MLCET performs operations in six steps, as shown in Fig. 4. $cnt^{(e)}$ in step 1 counts tweets in $trainMLDB$ associated with emotion *e*. In step 2, Feature similarity and label dissimilarity are calculated using Euclidian and Hamming distance, respectively. Thus the information obtained from features and labels is used to weigh neighbors.

In step 3, $F_1^{(e)}[j]$ denotes the total of tweets *x* belonging to emotion *e* and has *j* neighbors related to emotion *e*. $F_0^{(e)}[j]$ denotes total of tweets *x* not belonging to emotion *e* and has *j* neighbors related with emotion *e*. MLCET decides how many instances in *MLDB* have a total number of 0, 1…k neighbors where each neighbor is related with label *e*. This information is stored in $F_1^{(e)}[0 \dots k]$ and $F_0^{(e)}[0 \dots k]$ arrays, depending on whether the tweet under consideration whose neighbors are observed is related or not related to emotion *e*. This knowledge is utilized to estimate likelihood probabilities.

If no emotion is predicted for a tweet in $testMLDB$ in step 5, then MLCET predicts emotion with the highest probability in step 6. It is necessary because, in real scenarios, every tweet reveals at least one emotion to a certain extent in some context.

IV. Results and Discussion

A. Performance metrics

Let us denote an MLDB and label space by *E* and *S*, respectively. Let (x_q, AL_q) represent q^{th} instance of MLDB *E*, where x_q is a tweet having *f* features, q = 1…|*E*| and AL_q is a subset of *S*. AL_i and PL_i denote a set of actual and predicted labels by *g(.)* for instance x_i. MLC is evaluated using various metrics. The metrics used for experimentation in this work are listed in Fig. 5.

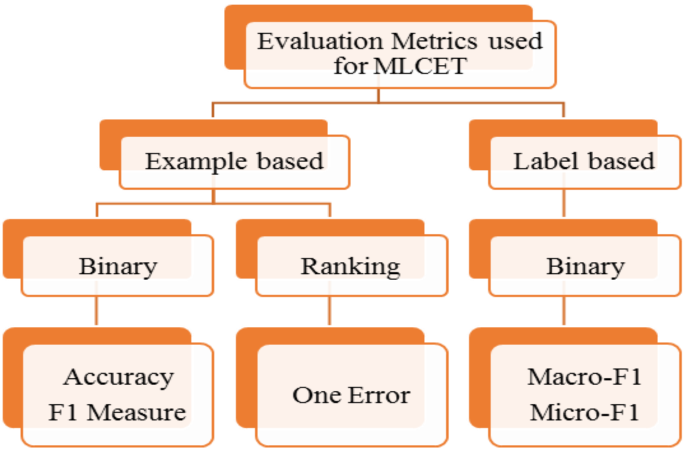


Fig. 5 Evaluation Metrics Used for MLCET

- **Example-based measures**

Performance measures that compute data from individual instances and average data obtained are called example-based measures. They can be grouped as binary and ranking. Example-based measures that predict whether an example is associated with a particular label are *binary*. *Ranking* measures are also example-based [19]. They are defined in terms of ranking function $\mu(l, i)$ that denotes relevance of label l with an instance i. F1-measure (Eq. 1) and accuracy (Eq. 2) [20] are binary, and one error (Eq. 3) is a ranking measure [19].

$$F1(g) = \frac{1}{|E|}\sum_{i=1}^{|E|}\frac{2|PL_i \cap AL_i|}{|AL_i|+|PL_i|} \quad (1)$$

$$Acc(g) = \frac{1}{|E|}\sum_{i=1}^{|E|}\frac{|PL_i \cap AL_i|}{|PL_i \cup AL_i|} \quad (2)$$

$$OE(g) = \frac{1}{|E|}\sum_{i=1}^{|E|} V((\arg\min_{y \in S} \mu(y, x_i)) \notin AL_i\ ) \quad (3)$$

V(.) in Eq. 3 returns 0 in case of a false condition, else it returns 1.

- **Label-based binary measures**

Measures that calculate average performance (macro or micro) from individual labels are called label-based measures. Macro (Eq. 4) and micro (Eq. 5) averaging are binary metrics. Macro (Micro) averaging gives equal importance to all the labels (instances). In other words, macro (micro) averaging finds an average across all the labels (example/label pairs) [21] [22].

$$MacroF1(g) = \frac{1}{|S|}\sum_{c=1}^{|S|}\frac{2 \times TP_c}{2 \times TP_c + FP_c + FN_c} \quad (4)$$

$$MicroF1(g) = \frac{2 \times \sum_{c=1}^{|S|} TP_c}{2 \times \sum_{c=1}^{|S|} TP_c + \sum_{c=1}^{|S|} FP_c + \sum_{c=1}^{|S|} FN_c} \quad (5)$$

B. Dataset

Saif M. Mohammad et al. [6] performed four tasks on tweets belonging to three languages English, Spanish, and Arabic. Some tasks are based on joy, anger, sadness, and fear, which are common emotions. But authors have considered sentiments also. Finally, they prepared a multi-label dataset from Tweets 2016 and 2017 datasets. This multi-label dataset is described as SemEval-2018 Affect in Tweets Dataset. Fig. 6 shows eleven emotions with their count in this dataset. Fig. 7 shows the percentage of emotion-wise distribution of tweets. After preprocessing, 6838 instances are used for training and 886 for testing in this work.

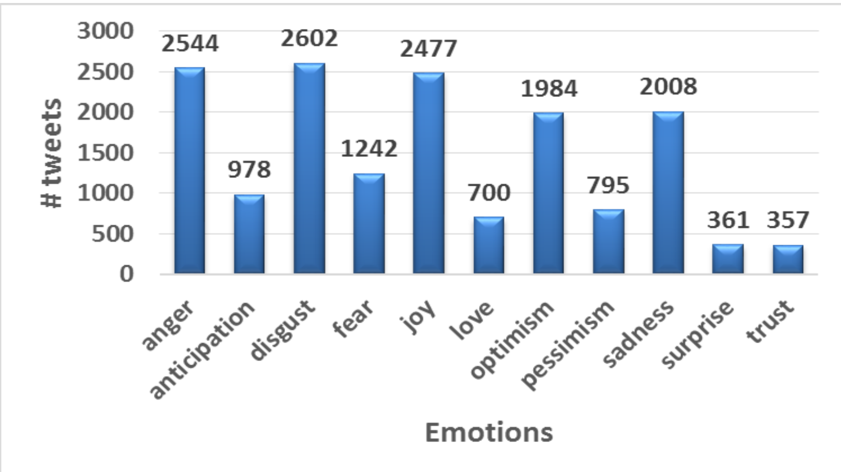


Fig. 6 Emotion-wise count of tweets

Fig. 7 Emotion-wise percentage of tweets

C. Experimentation

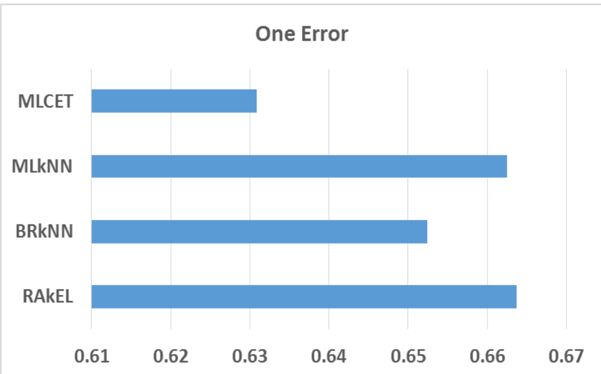
Three MLC, namely RAkEL [23], BRkNN [24], and MLkNN [25], are used for comparing the performance of MLCET. RAkEL uses LP and J48 as base classifiers in this work. For BRkNN, MLkNN, and MLCET, value 10 is used for k. Arrows (↓) and (↑) in Table 4 and Table 5 denote smaller and larger values expected for the corresponding metric. Also, bold values in both tables indicate e better result.

Table 4 and Fig. 8 show the results of four MLC methods. It indicates that MLCET is better than the three methods for all five metrics.

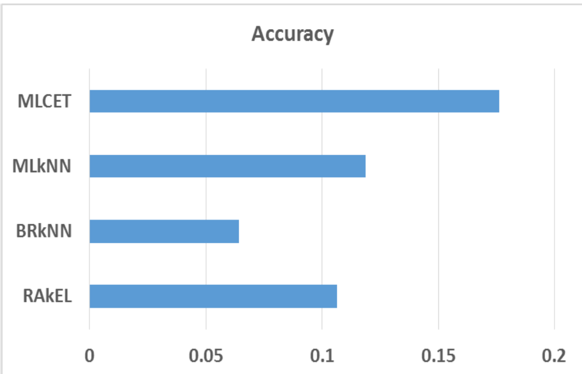
One error observes whether a predicted label at the top rank is not in an instance's list of relevant labels. It should have a smaller value. Fig. 8a shows that MLCET performed better among all, showing a 4% improvement. Accuracy is increased by 33% (Fig. 8b). Fig. 6 and 7 depict that # tweets per emotion vary. In such a scenario of unbalanced data, accuracy does not indicate improvement, if any. In such cases, the F1 score gives a better idea about the system. The F1 score is a metric that represents the harmonic mean of precision and recall. Macro-F1 is a class-wise average of the F1 score. But again, our dataset has imbalanced emotion distribution. Hence micro-F1 is also used in this work for evaluation, which observes the total values of TP, FN, and FP. Fig. 8c, 8d, and 8e show enhanced F1, Macro, and Micro F1 for MLCET by 37%, 29%, and 26%, respectively compared to other MLC methods.

Table 4 Performance comparison of MLCET

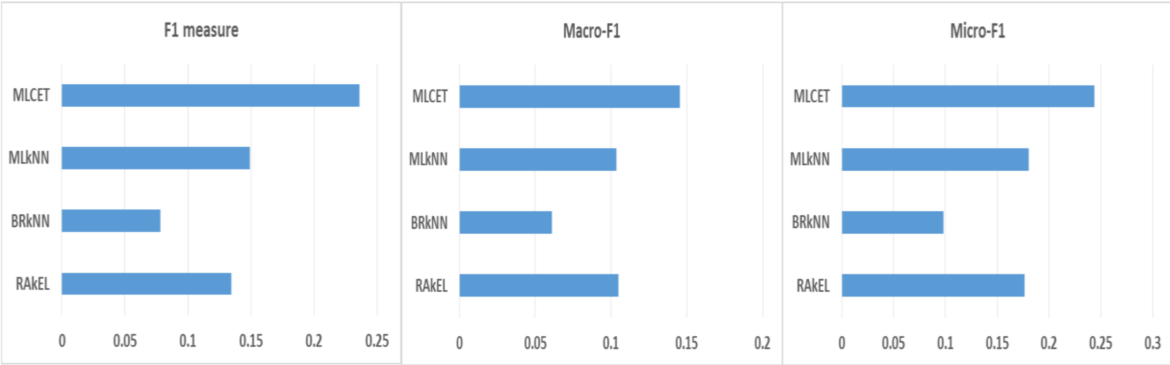
Performance Metric	RAkEL	BRkNN	MLkNN	MLCET
One Error (↓)	0.6637	0.6524	0.6625	**0.6309**
Accuracy (↑)	0.1066	0.0643	0.1187	**0.1763**
F1-Measure (↑)	0.1345	0.0781	0.1492	**0.2362**
Macro-F1 (↑)	0.1047	0.0610	0.1035	**0.1454**
Micro-F1 (↑)	0.1762	0.0980	0.1802	**0.2439**



(a) One error



(b) Accuracy



(c) F1 measure (d) Macro-F1 (e) Micro-F1

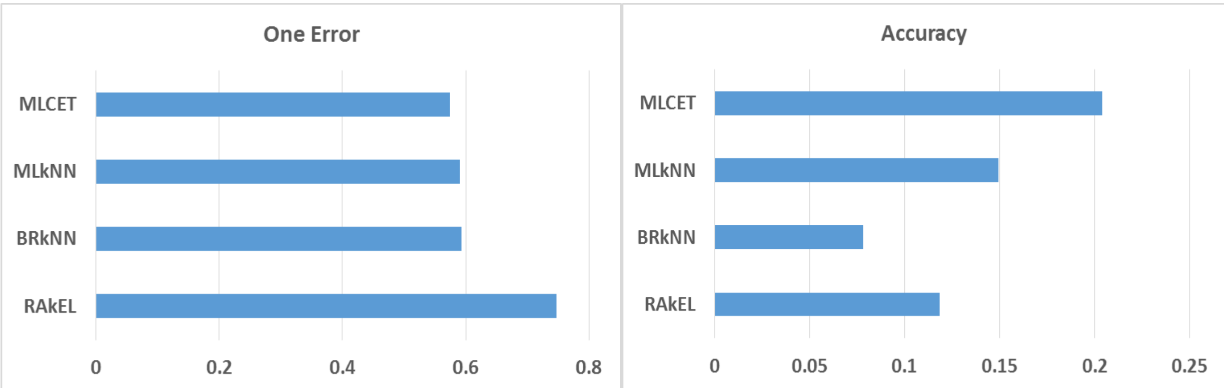
Fig. 8 Performance improvement of MLCET

- **Performance comparison after applying stemming and lemmatization on multi-label dataset**

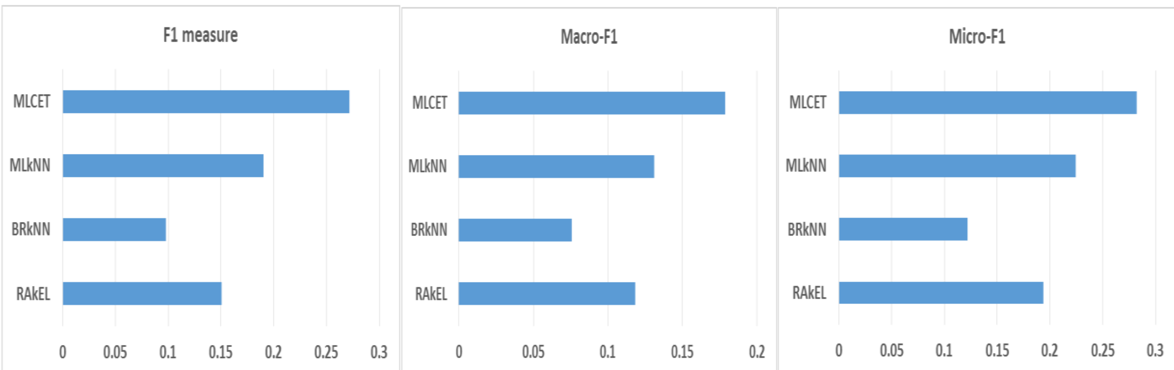
In English, words take different forms when used in sentences in different situations like play, played, playing. So it becomes difficult to compare them. The use of stemming and lemmatization brings all these words to their root base, say, play in the above example. It naturally leads to enhanced results of algorithms. The same is observed in this work and is shown in Table 5 and Fig. 9. Again, MLCET performed better with 23%, 27%, 30%, 27%, and 21% improvements in five metrics, respectively.

Table 5 Performance comparison on stemmed and lemmatized dataset

Performance Metric	RAkEL	BRkNN	MLkNN	MLCET
One Error (↓)	0.7472	0.5926	0.5903	**0.5745**
Accuracy (↑)	0.1187	0.0783	0.1495	**0.2042**
F1-Measure (↑)	0.1508	0.0981	0.1905	**0.2715**
Macro-F1 (↑)	0.1183	0.0757	0.1311	**0.1785**
Micro-F1 (↑)	0.1939	0.1221	0.2244	**0.2824**



(a) One error (b) Accuracy



(c) F1 measure (d) Macro-F1 (e) Micro-F1

Fig. 9 Performance improvement of MLCET on stemmed and lemmatized dataset

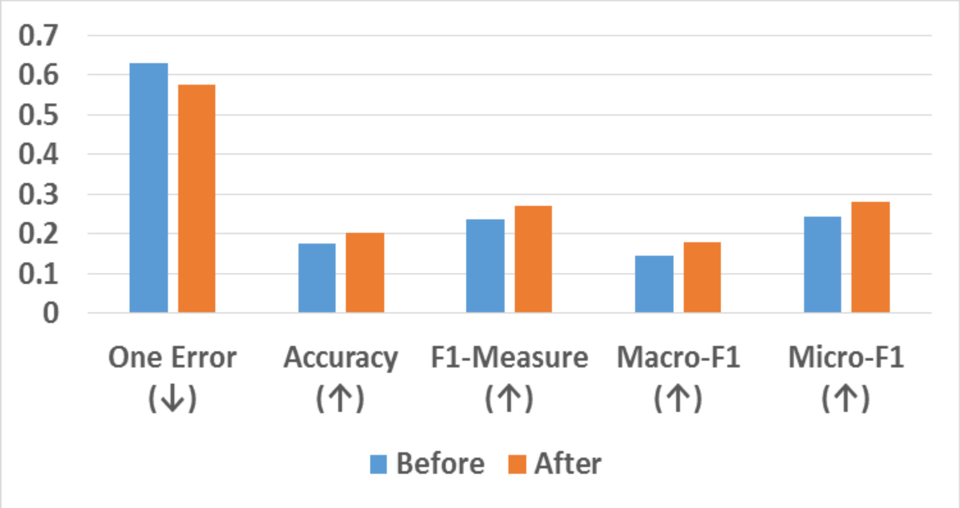


Fig. 10 Comparing MLCET performance on stemmed and lemmatized dataset

It can be noticed from Fig. 10 that applying stemming and lemmatization on MLDB has helped to enhance the performance of MLCET by 9%, 16%, 15%, 23%, and 16%, respectively, for the five metrics used.

All the experiments show that the performance of MLCET that follows the adaptation approach of MLC is enhanced because it uses statistics of neighbors, including prior and likelihood probability, along with feature similarity and label dissimilarity [19].

V. Conclusion

Extracting information from user tweets has become essential because of the increasing usage of social media applications like Twitter. This information is helpful for decision-making to increase business profit and identify and stop the spreading of wrong or sensitive information in critical situations. It is observed that the majority of tweets express multiple emotions, thereby making use of multi-label classification crucial for processing data in them. ML dataset requires specific preprocessing to create the dataset suitable for MLC. When MLCET processed such dataset, it showed improved results compared to RAkEL, BRkNN, and MLkNN. The work also depicted the importance of stemming and lemmatization for enhancing performance on ML text data. It will be interesting to process ML data to perceive multiple emotions by focusing more on label dependencies and correlation.

References

1. Tao and Fang, Toward multi-label sentiment analysis: a transfer learning based approach, Journal of Big Data (2020) 7:1, https://doi.org /10.1186/s40537-019-0278-0
2. Castillo, C. (2016). “Introduction”. In: Big Crisis Data: Social Media in Disasters and Time-Critical Situations. Cambridge University Press, pp. 1–17.
3. Han, B., Cook, P., and Baldwin, T. (2013). “Lexical Normalization for Social Media Text”. In: ACM Trans. Intell. Syst. Technol. 4.1, 5:1–5:27.
4. Nguyen, D. T., Joty, S. R., Imran, M., Sajjad, H., and Mitra, P. (2016). “Applications of Online Deep Learning for Crisis Response Using Social Media Information”. In: CoRR abs/1610.01030. arXiv: 1610.01030.
5. Alan Aipe, Mukuntha N S, Asif Ekbal, Sadao Kurohashi, Deep Learning Approach towards Multi-label Classification of Crisis Related Tweets, WiPe Paper – Social Media Studies
6. Proceedings of the 15th ISCRAM Conference – Rochester, NY, USA May 2018
7. Saif M. Mohammad, Felipe Bravo-Marquez, Mohammad Salameh, and Svetlana Kiritchenko, SemEval-2018 Task 1: Affect in Tweets, In *Proceedings of the International Workshop on Semantic Evaluation (SemEval-2018)*, New Orleans, LA, USA, June 2018.
8. Rohitash Chandra, Aswin Krishna, COVID-19 sentiment analysis via deep learning during the rise of novel cases, 2021
9. Zahir Abbas Khan, Rekha V, Ajit Danti, Naveen J, *Fake News Detection Using Linguistic Inquiry With Syllable Pattern Of Words Based On CNN Approach, Eur. Chem. Bull. 2023, 12(Special Issue 10), 231 - 239*
10. Tidake, Vaishali S., and Shirish S. Sane. "Multi-label classification: a survey." *International Journal of Engineering and Technology* 7.4.19 (2018): 1045-1054.
11. Alec Go et al., Twitter Sentiment Classification using Distant Supervision,

12. Jishnu Ray Chowdhury, Cornelia Caragea, and Doina Caragea, Cross-Lingual Disaster-related Multi-label Tweet Classification with Manifold Mixup, Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop, pages 292–298, July 5 - July 10, 2020
13. Anuradha Surolia, Shikha Mehta, Ponnurangam Kumaraguru et al. Understanding Emotions: A PoliEMO Datasetand Multi-label Classification in Indian Elections, 21 March 2023, [https://doi.org /10.21203/rs.3.rs-2694319/v1]
14. Xuan Liu, Tianyi Shi, Guohui Zhou, Mingzhe Liu, Zhengtong Yin, Lirong Yin & Wenfeng Zheng, Emotion classification for short texts: an improved multi-label method, HUMANITIES AND SOCIAL SCIENCES COMMUNICATIONS | (2023) 10:306, https://doi.org/10.1057/s41599-023-01816-6
15. Iqra Ameer, Noman Ashraf, Grigori Sidorov, Helena Gomez Adorno, Multi-label Emotion Classification using Content-Based Features in Twitter, ISSN 2007-9737, Computación y Sistemas, Vol. 24, No. 3, 2020, pp. 1159–1164, doi: 10.13053/CyS-24-3-3476
16. Rasha Obeidat, Maram Gharaibeh, Malak Abdullah and Yara Alharahsheh (2022) Multi-label multi-class COVID-19 Arabic Twitter dataset with fine-grained misinformation and situational information annotations, Peer J Computer Science, DOI 10.7717/peerj-cs.1151
17. Mohammed Jabreel, and Antonio Moreno, A Deep Learning-Based Approach for Multi-Label Emotion Classification in Tweets, Appl. Sci. 2019, 9, 1123; doi:10.3390/app9061123
18. Semiu Salawu, Prof. Jo Lumsden, Prof. Yulan He, A Large-Scale English Multi-Label Twitter Dataset for Online Abuse Detection, Proceedings of the Fifth Workshop on Online Abuse and Harms, pages 146–156, August 6, 2021. ©2021 Association for Computational Linguistics
19. Tarle, B., Chintakindi, S. & Jena, S., Integrating multiple methods to enhance medical data classification. Evolving Systems 11, 133–142 (2020). https://doi.org/10.1007/s12530-019-09272-x
20. G. Madjarov, D. Kocev, D. Gjorgjevikj, and S. Dzeroski, "An extensive experimental comparison of methods for multi-label learning", Pattern Recognit., vol. 45, no. 9, pp. 3084-3104, 2012
21. S. Godbole, S. Sarawagi, "Discriminative methods for multi-labeled classification", in: Advances in Knowledge Discovery and Data Mining, Springer, Berlin/ Heidelberg, 2004, pp. 22-30
22. Dai Y, Li Y, Li S-T (2020) "Multi-label learning for concept-oriented labels of product image data", Image and Vision Computing 93, 103821
23. Vaishali S Tidake, Shirish S Sane, A novel algorithm for multi-label classification by exploring feature and label dissimilarities, 2019, Journal Int. J. Comput. Inf. Syst. Indus. Manage. Appl., Volume 11, Pages 161-169
24. Tsoumakas G., Katakis I. and Vlahavas I. (2011). "Random k-Labelsets for Multi-Label Classification". IEEE Transactions on Knowledge Data Engineering, 23. 1079-1089. 10.1109/TKDE.2010.164

D. Spyromitros-Xiou_s, G. Tsoumakas, and I. Vlahavas, "An empirical study of lazy multilabel classification algorithms", in Proc. 5th Hellenic Conf. Artif. Intell., Syros, Greece, 2008, pp. 401{406

25. M. L. Zhang, Z.H. Zhou, "ML-kNN: A lazy learning approach to multilabel learning", Pattern Recognit., vol.40, no.7, pp. 2038-2048, 2007
26. Zhou, Zhi-Hua, and Zhang, Min-Ling. "Multi-instance multi-label learning with application to scene classification", Advances in neural information processing systems. 2006