Journal of Engineering and Technology

ISSN 2180-3811 eISSN 2289-814X https://jet.utem.edu.my/jet/index

DOI: https://doi.org/10.54554/jet.2025.16.1.007

CROSS-CULTURAL EMOTION ANALYSIS ON X USING BIDIRECTIONAL ENCODER REPRESENTATIONS FROM TRANSFORMERS (BERT)

A. Nafis*[1] and M. Ghazali[2]

[1] Faculty of Computing, Universiti Teknologi Malaysia, Jalan Iman, 81310 Skudai, Johor, Malaysia

[2] Malaysia-Japan International Institute of Technology, Universiti Teknologi Malaysia, 54100 Kuala Lumpur, Wilayah Persekutuan Kuala Lumpur, Malaysia

corresponding: nafis20@graduate.utm.my

Article history:
Received Date: 21 November 2024
Revised Date: 8 April 2025
Accepted Date: 1 Jun 2025

Keywords: Sentiment Analysis, BERT (Bidirectional

Abstract— It is essential to understand how environmental factors and cultural backgrounds affect emotional responses to build Human-Computer Interaction (HCI) technologies. This study uses the BERT (Bidirectional Encoder Representations from Transformers) model to investigate the emotional landscape of interactions on the X platform across various cultural contexts. The aims of the study are as follows: (1) Gather and prepare X platform data from COVID-19 sources in India, Pakistan, Malaysia, the US, the UK, Australia, and

This is an open-access journal that the content is freely available without charge to the user or corresponding institution licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0).

Encoder Representations from Transformers), Cultural Contexts, Social Media, Emotion Detection	Canada. (2) Use a BERT model to qualitatively analyze tweet sentiment. (3) Assess the accuracy of the model and look for sentiment trends in tweets from nations throughout the pandemic. The BERT model successfully classified sentiment, as seen by its 80.29% accuracy rate on test data. Sentiment research showed that positive sentiment was far more prevalent in the US, Canada, and Australia, indicating that these countries were better able to adjust to the COVID-19 situation. Stability was seen in the balanced sentiment distributions displayed by Pakistan, India, and the United Kingdom. Despite having fewer data points, Pakistan and Malaysia continued to have largely positive attitudes. This study provides the basis for a comparative analysis of emotional responses by taking contextual and cultural aspects into account.

I. Introduction

The study of designing and creating interactive systems is at the core of the field known as human-computer interaction, or HCI [9]. Researchers in human-computer interaction (HCI) are interested in how people engage with technology and in developing user-friendly, effective, and pleasurable systems. Understanding how technology affects human emotions is one of the major issues in HCI [10]. Culture, environment, and personality are just a few of the many variables that can have an impact on emotions, which are intricate and multidimensional. It is critical to comprehend how emotions impact human-computer interaction to develop user-centered solutions. Investigating the emotional dynamics of HCI has become critical in this dynamic field [7]. This chapter introduces the research's multifaceted relationship between contextual factors, cultural diversity, and human emotions. It draws attention to how relevant these elements are today for influencing user experiences, especially in the ever-changing realm of social media site X interactions (previously known as Twitter) [25]. The utilization of the BERT model is proposed as a crucial instrument for exploring the complex subtleties of emotional reactions.

II. Literature Review

A. Emotions and Culture in HCI

Understanding how emotions and culture interact with HCI is crucial, particularly on social media platforms like X [5]. People express a broad range of emotions on social media, reflecting real-world relationships in a virtual setting. Boehner et al. [23] critique the traditional informational model of emotional computing, which treats emotions as measurable internal states, and propose an interactional approach. They emphasize co-constructed and co-interpreted emotional expressions, advocating for a creative approach to assessment and design in HCI. Meanwhile, Cowie et al. [17] highlight the importance of recognizing implicit emotional cues in human communication. They advocate for multidisciplinary efforts to develop systems that combine signal processing, psychology, and language studies to detect and respond to user emotions effectively. Gelderblom and Ford [6] explore how Hofstede's cultural dimensions affect user performance on web-based interfaces. While their study does not definitively prove a direct impact, it suggests that considering cultural traits can improve usability. This literature survey explores the theoretical and historical foundations of emotional dynamics in digital media, highlighting significant problems and gaps. It focuses on how cultural and environmental factors influence emotional reactions in the digital realm.

B. BERT Model

This literature review highlights the importance of Bidirectional Encoder Representations from Transformers (BERT) in analyzing emotional dynamics on social media [14]. BERT's bidirectional approach effectively extracts context from both preceding and following words, making it suitable for in-depth emotional content analysis [26]. Several studies demonstrate BERT's effectiveness in emotion detection, aspect-based sentiment analysis, and sentiment classification [3]. Despite its high accuracy, BERT faces challenges such as computational complexity, fixed input length constraints, and generalization across diverse datasets. The proposed research aims to improve BERT for sentiment analysis by addressing these shortcomings and incorporating cultural and environmental factors. This enhanced BERT model will better represent cultural nuances and support a wide range of emotional responses in the digital space. The review underscores the need for sophisticated models like BERT to understand emotions and culture in HCI. Building on this foundation, the proposed study seeks to extend BERT's capabilities to offer a deeper understanding of emotional reactions on social media platforms.

C. Description of Related Studies

The combination of many approaches and models has led to notable breakthroughs in sentiment analysis, a crucial aspect of natural language processing. Talaat [24] compares deep learning models to conventional machine learning techniques with the goal of improving sentiment analysis accuracy by fusing emojis with DistilBERT. The study shows the success of the model with a noteworthy F1 score of 83.74%. Its limits stem from the peculiarity of the dataset, though, which calls for the investigation of bigger and more varied datasets as well as the improvement of emotion analysis depth for complex sentiments.

Using an ABSA-based Roberta+LSTM technique, Sirisha and Chandana [22] analyze Aspect-Based Sentiment Analysis (ABSA) in tweets about the conflict between Russia and Ukraine. Even though the study's accuracy is high (Neg: 0.96, Pos: 0.92, Neu: 0.93), it admits that its use of Twitter (now X) data has limits, including potential biases and the inability to accurately capture changing moods.

Pipalia et al. [18] show experimental results with a range of accuracies as they study sentiment categorization using pre-trained language models. Although the study's limitations are not stated clearly, they might include issues with generalizability to different domains and possible biases in the IMDB-reviews dataset. When comparing transformer models for emotion recognition, RoBERTa achieves the best F1-score (0.742) [2]. Potential biases in the emotion dataset may also present challenges for generalisation to other datasets and contexts.

The efficiency of BERT, RoBERTa, DistilBERT, and XLNet in extracting emotions from texts is assessed by Adoma et al. [27]. RoBERTa obtained a high F1-score of 0.93. However, obtaining high F1-scores may present difficulties due to computational complexity, fixed input length restrictions, and generalization to other datasets.

Using domain-specific models, Lin and Moh [13] improve sentiment analysis on tweets connected to COVID-19. The BERT model got the best accuracy at 75.08%. Notwithstanding advancements, difficulties persist in utilizing domain-specific models for sentiment analysis in complex situations, underscoring the dynamic character of sentiment.

Faraby et al. [4] compare pre-training models to categorize problems into cognitive science-based groups. RoBERTa achieved the best F1 score of 0.8116. Although the study's accuracy rate is high, it notes that there may be problems with the dataset, which means that further research and development of question classification systems for educational purposes is still necessary.

Research on sentiment analysis has made great strides in a variety of applications. Each study has added insightful information, produced impressive outcomes, and identified possible roadblocks that will hopefully be overcome in the future.

III. Research Methodology

The research begins by identifying the primary issue and reviewing existing studies to understand current knowledge and gaps in the field. This quantitative study collects data from X users in Malaysia, India, Pakistan, USA, UK, Canada, and Australia using pre-existing datasets from Kaggle (Miglani, 2019). Data preprocessing includes standardizing formats, removing null values, and filtering data by country. The BERT model is then selected for sentiment analysis, with its performance evaluated using metrics such as accuracy, precision, recall, and F1 score. The optimized BERT model analyzes social media sentiments, capturing cultural nuances, and cross-cultural comparisons are visualized using graphs and charts. Finally, the findings are analyzed for significance, conclusions are drawn, and recommendations for future research are made. This methodology (Figure 1) provides a structured approach, employing robust data collection and thorough analysis techniques to effectively address the research question.

A. How BERT works

Since the purpose of BERT is to produce a language model, only the encoder mechanism is employed. Tokens are supplied into the Transformer encoder in a specific order. Prior to being processed by the neural network, these tokens are initially integrated into vectors. Contextualized representations are produced as a series of vectors in the output, each of which corresponds to an input token. It can be difficult to define a prediction aim while training language models. Many models take a directed approach

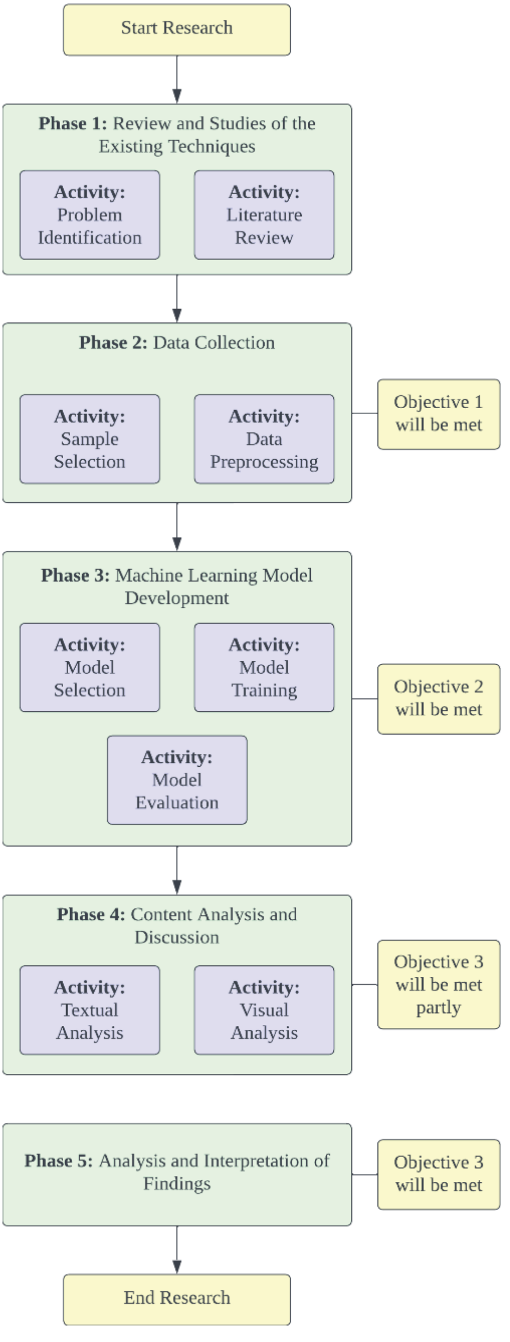

Figure 1: Flow of the Research Methodology

and could restrict context learning by predicting the word that will appear next in a sequence. BERT uses two cutting-edge training techniques to meet this challenge: Masked Language Model (MLM) and Next Sentence Prediction (NSP).

B. Masked Language Modeling (MLM)

Before feeding word sequences into BERT, 15% of the words are replaced with a [MASK] token. BERT predicts the original value of these masked words using the context from the unmasked words. This involves:

- Adding a classification layer on top of the encoder output (Figure 2).
- Multiplying the output vectors by the embedding matrix to transform them into the vocabulary dimension (Figure 2).
- Calculating the probability of each word in the vocabulary using SoftMax.

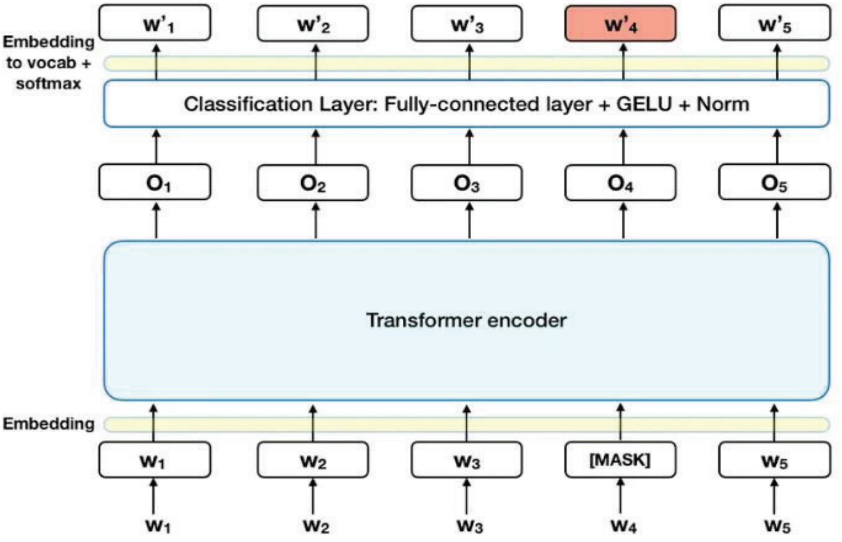

Figure 2: Masked Language Modeling (MLM)

C. Next Sentence Prediction (NSP)

BERT also learns to predict if the second sentence in a pair is the subsequent sentence in the original document. During training, 50% of the sentence pairs are sequential from the document, while the other 50% are random pairs. The process involves:

- Adding a [CLS] token at the beginning and a [SEP] token at the end of each sentence (Figure 3).
- Adding sentence embeddings to distinguish between Sentence A and Sentence B (Figure 3).
- Adding positional embeddings to indicate token positions (Figure 3).

D. BERT Model for Sentiment Analysis

The BERT (Bidirectional Encoder Representations from Transformers) model is a state-of-the-art natural language processing model developed by Google. It uses a transformer architecture to understand the context of words in a sentence by looking at the words that come before and after it. This bidirectional approach allows BERT to grasp the nuanced meaning of words based on their context, making it highly effective for sentiment analysis.

BERT identifies sentiments by transforming the input text into embeddings, which are then passed through multiple layers of transformers. These transformers apply self-attention

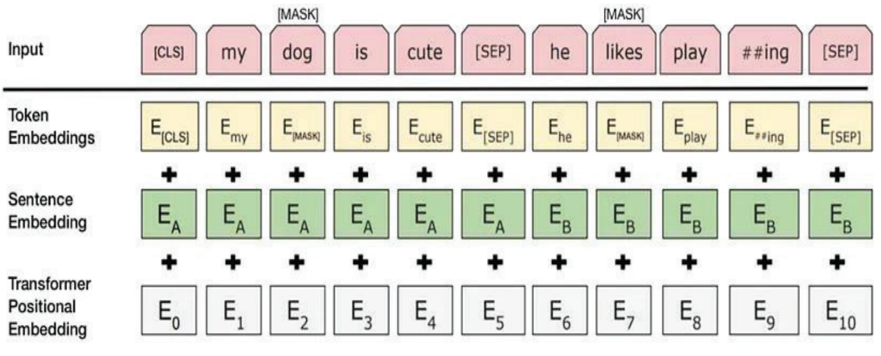

Figure 3: Masked Language Modeling (MLM)

mechanisms to weigh the importance of each word in relation to others in the sentence. The final layer outputs a vector representation that can be classified into sentiment categories (Figure 4).

For example, in sentiment analysis, a tweet like "I love this new phone!" would be tokenized and transformed into embeddings. BERT's transformers would then analyze the context, understanding that "love" is a strong positive word, leading to a positive sentiment classification.

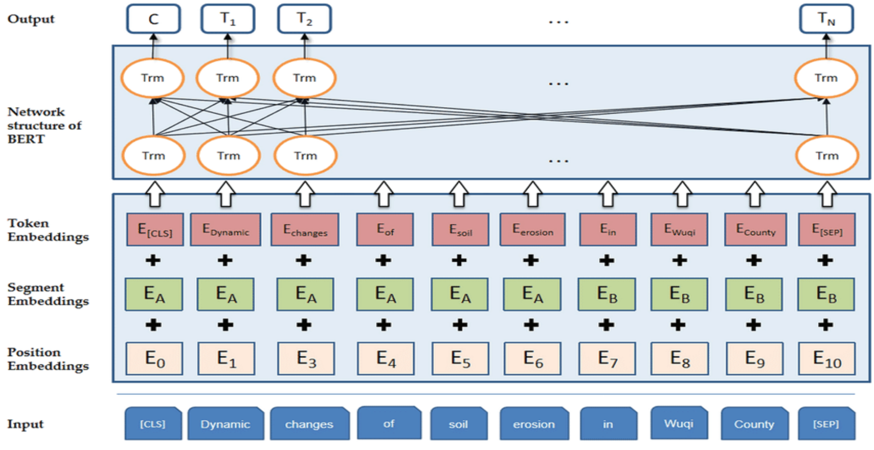

Figure 4: Overall structure of the BERT Model

IV. Results and Discussions

The analysis of sentiment counts within the training and test datasets offers valuable insights into the distribution of sentiment labels. The training dataset contains 5,165 unique tweets, while the test dataset has 1,294 unique tweets [14]. Training the BERT model for two epochs (Figure 5) allowed for the monitoring of training loss and validation accuracy (Figure 6) to identify potential overfitting or underfitting issues. Graphical representations of these metrics depict the training progress. The model achieved an accuracy of 80.29% on the test data (Figure 7), indicating strong generalization and effective sentiment classification.

```
Epoch 1/2
  Training Loss: 0.9218
  Validation Accuracy: 0.78
Epoch 2/2
  Training Loss: 0.5818
  Validation Accuracy: 0.97
```

Figure 5: Training Loss and Validation Accuracy

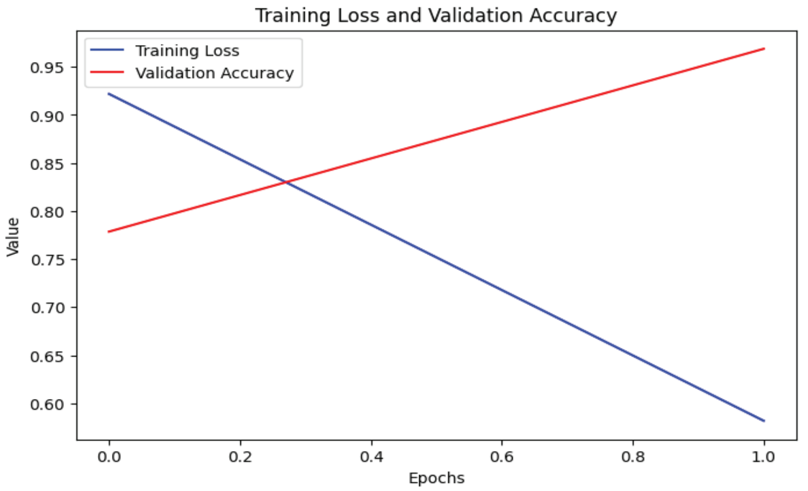

Figure 6: Training Loss and Validation Accuracy Graph

```
Test Metrics:
  Accuracy: 0.8029366306027821
  Precision: 0.8034241386714372
  Recall: 0.8029366306027821
  F1 Score: 0.8017472021720762
```

Figure 7: Test metrics

The achieved accuracy of 80.29% in sentiment analysis is demonstrating robust performance when compared to other state-of-the-art approaches. For example, Talaat [24] combined DistilBERT with emoji-based sentiment analysis, resulting in an F1 score of 83.74%. Although this is slightly higher, the model's effectiveness is constrained by its dataset specificity, whereas the BERT model offers broader applicability. Similarly, Lin and Moh [13] applied a domain-specific BERT model to COVID-19 related tweets, achieving an accuracy of 75.08%, which is lower than the accuracy reported in this study. This highlights the robustness of the BERT model without requiring domain-specific adjustments. Lastly, Adoma et al. [27] obtained a high F1-score of 0.93 using RoBERTa for emotion extraction, underscoring its strength in this specific task. However, the BERT model's accuracy of 80.29% remains competitive, particularly in sentiment analysis tasks. In summary, the BERT model's accuracy is commendable, especially when compared to these advanced models, showcasing its effectiveness and versatility across a range of sentiment analysis scenarios.

In addition, the sentiment distribution across different countries was also examined (Figure 8 and Figure 9). The datasets contained tweets from various locations, each labeled with sentiments ranging from 'Extremely Negative' to 'Extremely Positive'. The sentiment distribution revealed significant variations across countries.

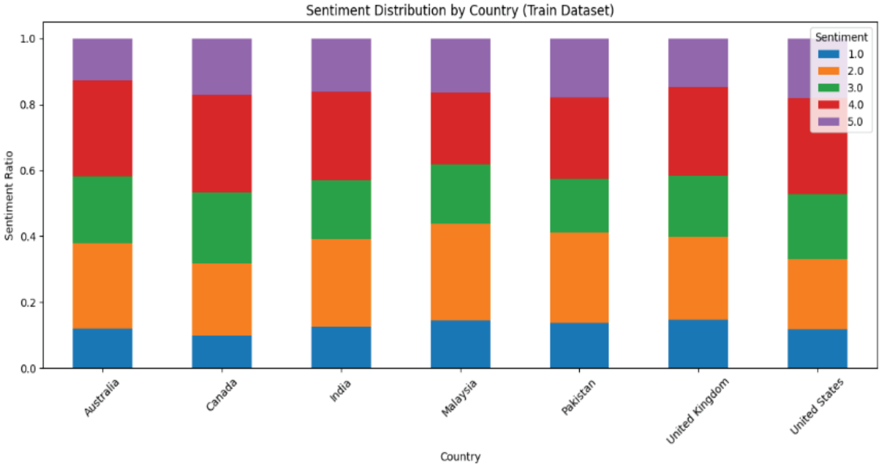

Figure 8: Sentiment Distribution by Country (Train Dataset)

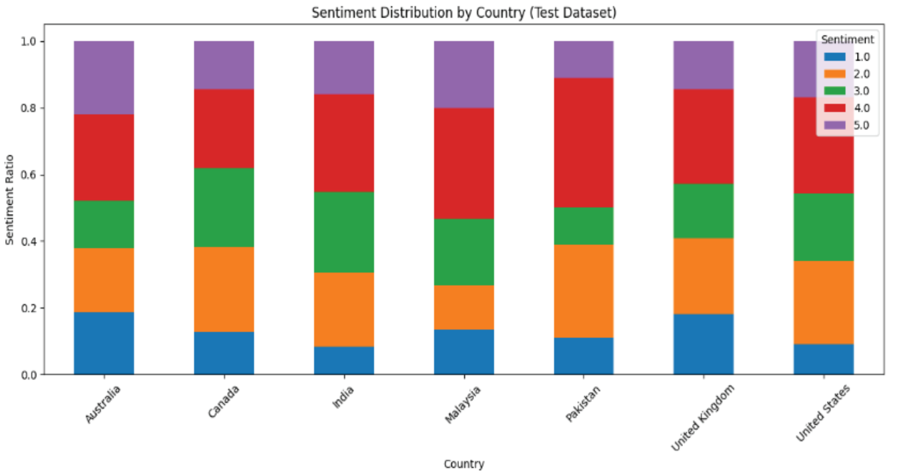

Figure 9: Sentiment Distribution by Country (Test Dataset)

V. Conclusion

This study delves into the emotional responses of individuals across different countries during the COVID-19 pandemic, utilizing sentiment analysis on the social media platform X (formerly Twitter) through the BERT model. The findings indicate distinct emotional landscapes across countries, highlighting how cultural and environmental contexts shape emotional reactions. Australia, Canada, and the United States exhibited predominantly positive sentiments, reflecting strong adaptability and a positive mindset. India and the United Kingdom, while displaying a mix of sentiments, leaned towards positivity, indicating resilience. Malaysia and Pakistan, despite limited data, also showed generally positive sentiments. These results underscore the importance of understanding cultural and contextual factors in designing user-centered HCI systems. The proposed methodology, involving comprehensive data collection, preprocessing, and sentiment analysis using the BERT model, has proven effective in capturing the emotional nuances in tweets during the pandemic. The study not only advances the understanding of cross-cultural emotional dynamics but also sets the stage for future research to further explore the interplay between technology, culture, and emotions. This foundational work offers valuable insights for developing more empathetic and culturally aware technological solutions in the field of HCI.

Table 1: Sentiment Analysis in Different Countries

Country	Train Dataset	Test Dataset	Interpretation
Australia	Sentiment values show a fairly even distribution, with a slightly higher number of positive sentiments (4.0 and 5.0) compared to very negative sentiments (1.0).	The distribution remains consistent, with the highest number of sentiments being positive (4.0 and 5.0).	Australians seem to have a relatively positive outlook, suggesting good adaptability and a positive mindset towards the new pandemic situation.
Canada	Sentiment values are fairly balanced, with a slight inclination towards positive sentiments (4.0 and 5.0).	Positive sentiments (4.0 and 5.0) continue to be prevalent.	Canadians exhibit a generally positive sentiment, indicating resilience and a positive mindset during the pandemic.
India	Sentiment values are fairly distributed, with a notable amount of positive sentiment (4.0 and 5.0).	Positive sentiments (4.0) are more prevalent, though there's a balanced distribution.	India shows a mix of sentiments but leans towards positivity, suggesting a relatively good adaptability to the new situation.
Malaysia	Sentiment values are lower in number but show a fairly even distribution.	There are very few sentiments, but the majority are positive (4.0 and 5.0).	Although data is limited, Malaysians appear to maintain a positive outlook, indicating resilience.
Pakistan	Similar to Malaysia, sentiment values are low but show a positive trend.	Positive sentiments (4.0) are slightly more prevalent.	Pakistan shows a tendency towards positive sentiments, indicating an ability to adapt to the pandemic situation.

Country	Train Dataset	Test Dataset	Interpretation
United Kingdom	A substantial number of sentiments, with a balanced amount being positive (4.0 and 5.0) and negative (1.0 and 2.0). There are a notable number of neutral sentiments as well.	Positive sentiments (4.0) are more prevalent, though a balanced distribution exists.	The UK shows a considerable amount of positive and negative sentiment, suggesting mixed adaptability and resilience.
United States	The highest number of sentiments, with a notable amount being positive (4.0 and 5.0).	A significant number of positive sentiments (4.0 and 5.0).	The US has a substantial number of positive sentiments, indicating a strong positive mindset and adaptability during the pandemic.

VI. References

[1] I. Ameer, N. Bölücü, M. H. F. Siddiqui, B. Can, G. Sidorov, and A. Gelbukh, "Multi-label emotion classification in texts using transfer learning," *Expert Systems with Applications*, vol. 213, p. 118534, 2023. doi: 10.1016/j.eswa.2022.118534.

[2] D. Cortiz, "Exploring Transformers models for Emotion Recognition: a comparison of BERT, DistilBERT, RoBERTa, XLNET, and ELECTRA," in ACM International Conference Proceeding Series, 2022, pp. 230–234. doi: 10.1145/3562007.3562051.

[3] M. D. Deepa and A. Tamilarasi, "Bidirectional Encoder Representations from Transformers (BERT) language model for sentiment analysis task: Review," in Turkish Journal of Computer and Mathematics Education, vol. 12, no. 7, 2021.

[4] S. A. Faraby, Adiwijaya, and A. Romadhony, "Educational question classification with pre-trained language models," in 2022 Seventh International Conference on Informatics and Computing (ICIC), 2022, pp. 1–6. doi: 10.1109/ICIC56845.2022.10006957.

[5] Z. Gao and J. Huang, "Human-computer interaction emotional design and innovative cultural and creative product design," *Frontiers in Psychology*, vol. 13, 2022. doi: 10.3389/fpsyg.2022.982303.

[6] J. H. Gelderblom and G. Ford, "The effects of culture on performance achieved through the use of human-computer interaction," 2003.

[7] F. Gurcan, N. E. Cagiltay, and K. Cagiltay, "Mapping human–computer interaction research themes and trends from its existence to today: A topic modeling-based review of past 60 years," *International Journal of Human–Computer Interaction*, vol. 37, no. 3, pp. 267–280, 2021. doi: 10.1080/10447318.2020.1819668.

[8] A. S. Imran, S. M. Daudpota, Z. Kastrati, and R. Batra, "Cross-cultural polarity and emotion detection using sentiment analysis and deep learning on COVID-19 related tweets," *IEEE Access*, vol. 8, pp. 181074–181090, 2020. doi: 10.1109/ACCESS.2020.3027350.

[9] T. Issa and P. Isaias, "Usability and human–computer interaction (HCI)," in Sustainable Design: HCI, Usability and Environmental Concerns, T. Issa and P. Isaias, Eds. Springer London, 2022, pp. 23–40. doi: 10.1007/978-1-4471-7513-1_2.

[10] M. Jeon, "Emotions and affect in human factors and human–computer interaction: Taxonomy, theories, approaches, and methods," in Emotions and Affect in Human Factors and Human-Computer Interaction, M. Jeon, Ed. Academic Press, 2017, pp. 3–26. doi: 10.1016/B978-0-12-801851-4.00001-X.

[11] A. Koufakou, "Deep learning for opinion mining and topic classification of course reviews," Education and Information Technologies, 2023. doi: 10.1007/s10639-023-11736-2.

[12] I. Lazrig and S. L. Humpherys, "Using machine learning sentiment analysis to evaluate learning impact," *Information Systems Education Journal (ISEDJ)*, vol. 1, 2022.

[13] H. Y. Lin and T. S. Moh, "Sentiment analysis on COVID tweets using COVID-Twitter-BERT with auxiliary sentence approach," in Proceedings of the 2021 ACMSE Conference, 2021, pp. 234–238. doi: 10.1145/3409334.3452074.

[14] A. Miglani, "Coronavirus tweets NLP - text classification," Kaggle, 2019.

[15] M. Mujahid, K. Kanwal, F. Rustam, W. Aljadani, and I. Ashraf, "Arabic ChatGPT tweets classification using RoBERTa and BERT ensemble model," *ACM Transactions on Asian and Low-Resource Language Information Processing*, vol. 22, no. 8, 2023. doi: 10.1145/3605889.

[16] K. Nimmi, B. Janet, A. K. Selvan, and N. Sivakumaran, "Pre-trained ensemble model for identification of emotion during COVID-19 based on emergency response support system dataset," *Applied Soft Computing*, vol. 122, 2022. doi: 10.1016/j.asoc.2022.108842.

[17] O. E. Ojo, H. T. Ta, A. Gelbukh, H. Calvo, O. O. Adebanji, and G. Sidorov, “Transformer-based approaches to sentiment detection,” in Recent Developments and the New Directions of Research, Foundations, and Applications, S. N. Shahbazova, A. M. Abbasov, V. Kreinovich, J. Kacprzyk, and I. Z. Batyrshin, Eds. Springer Nature Switzerland, 2023, pp. 101–110. doi: 10.1007/978-3-031-23476-7_10.

[18] K. Pipalia, R. Bhadja, and M. Shukla, “Comparative analysis of different transformer based architectures used in sentiment analysis,” in 2020 9th International Conference System Modeling and Advancement in Research Trends (SMART), 2020, pp. 411–415. doi: 10.1109/SMART50582.2020.9337081.

[19] R. Qasim, W. H. Bangyal, M. A. Alqarni, and A. A. Almazroi, “A fine-tuned BERT-based transfer learning approach for text classification,” *Journal of Healthcare Engineering*, vol. 2022, p. 3498123, 2022. doi: 10.1155/2022/3498123.

[20] L. Rundo, R. Pirrone, S. Vitabile, E. Sala, and O. Gambino, “Recent advances of HCI in decision-making tasks for optimized clinical workflows and precision medicine,” *Journal of Biomedical Informatics*, vol. 108, p. 103479, 2020. doi: 10.1016/J.JBI.2020.103479.

[21] A. Singh and G. Jain, “Sentiment analysis of news headlines using simple transformers,” in 2021 Asian Conference on Innovation in Technology (ASIANCON), 2021, pp. 1–6. doi: 10.1109/ASIANCON51346.2021.9544806.

[22] U. Sirisha and B. S. Chandana, “Aspect based sentiment and emotion analysis with RoBERTa, LSTM,” in IJACSA) International Journal of Advanced Computer Science and Applications, vol. 13, no. 11, 2022.

[23] A. Sweidan, N. El-Bendary, and H. Al-Feel, “Sentence-level aspect-based sentiment analysis for classifying adverse drug reactions (ADRs) using hybrid ontology-XLNet transfer learning,” *IEEE Access*, pp. 1–7, 2021. doi: 10.1109/ACCESS.2021.3091394.

[24] A. S. Talaat, “Sentiment analysis classification system using hybrid BERT models,” *Journal of Big Data*, vol. 10, no. 1, 2023. doi: 10.1186/s40537-023-00781-w.

[25] J. Taylor, “X, formerly Twitter, strips headlines from news story links to improve their look,” The Guardian, 2023.

[26] B. Wan, P. Wu, C. K. Yeo, and G. Li, “Emotion-cognitive reasoning integrated BERT for sentiment analysis of online public opinions on emergencies,” *Information Processing & Management*, vol. 61, no. 2, p. 103609, 2024. doi:

10.1016/J.IPM.2023.103609.

[27] A. F. Adoma, N.-M. Henry, and W. Chen, "Comparative analyses of BERT, RoBERTa, DistilBERT, and XLNet for text-based emotion recognition," in 2020 17th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP), 2020, pp. 117-121. doi: 10.1109/ICCWAMTIP51612.2020.9317379.