Emotion Detection using Transformer Model with Deep Learning

Divyang Bharatbhai Joshi
Department of Computer Science and Engineering
Parul University, Vadodara, Gujarat

Asst. Prof. Pankaj Agrawal
Department of Computer Science & Engineering
Parul University, Vadodara, Gujarat

Apoorva Ashokbhai Dhokai
Department of Computer Science and Engineering
Parul University, Vadodara, Gujarat

Amit Kumar
Department of Computer Science and Engineering
Parul University, Vadodara, Gujarat

Nandini Pavankumar Agrawal
Department of Computer Science and Engineering
Parul University, Vadodara, Gujarat

Mr. Avinash Sood
CEO, Binary Qubit Gurugram, Haryana

Abstract — Emotion detection is a crucial aspect of Natural Language Processing (NLP) that helps classify emotions from text. Traditional models, such as Support Vector Machines (SVMs) and Recurrent Neural Networks (RNNs), struggle with contextual understanding. In contrast, Transformer-based models like BERT and RoBERTa significantly enhance performance. This research proposes a Transformer-based deep learning approach to classify emotions into six categories. By implementing effective preprocessing and label encoding, we achieve over 90% accuracy, surpassing conventional deep learning methods (which typically range between 70% and 85%). Our findings emphasize the superiority of Transformer models in capturing semantic nuances, making them valuable for applications in sentiment analysis, mental health monitoring, and human-computer interaction.

Keywords — Emotion Detection, Natural Language Processing (NLP), Sentiment Analysis, BERT, RoBERT, Text Classification, Contextual Embeddings, Self-Attention Mechanism, Mental Health Monitoring, Human-Computer Interaction.

I. INTRODUCTION

Emotion detection in textual data has gained significant attention in recent years due to the growing prevalence of digital communication. Applications such as opinion mining, mental health assessment, customer sentiment analysis, and human-computer interaction benefit from accurate emotion classification. Traditional sentiment analysis techniques, including lexicon-based methods and early machine learning models, have demonstrated some success but often fail to capture deeper linguistic patterns and contextual dependencies. The emergence of Transformer-based models has transformed emotion detection by leveraging self-attention mechanisms and deep contextual embedding, achieving state-of-the-art performance.

A. Background and evolution of sentiment analysis

Early sentiment analysis relied on rule-based and lexicon-based methods, using predefined sentiment dictionaries to determine text polarity. However, these approaches struggled with sarcasm, context variations, and domain-specific language. The introduction of statistical learning techniques, including Naïve Bayes, SVMs, and Decision Trees, improved accuracy but still lacked the ability to fully understand contextual relationships.

With the rise of deep learning, we've seen the emergence of models like Long Short-Term Memory (LSTM) networks and Convolutional Neural Networks (CNNs), which have really enhanced sentiment analysis. But, let's be honest, these models struggled a bit with long-range dependencies. That's where Transformer models, like BERT, came into play, tackling those challenges head-on and taking sentiment classification to the next level with their self-attention mechanisms and bidirectional contextual learning.

B. The rise of transformer-based models

The Transformer architecture, which made its debut with BERT (Bidirectional Encoder Representations from Transformers), tackled many of the challenges that traditional deep learning models faced. Unlike earlier approaches, Transformers use self-attention mechanisms to grasp the context of words in relation to each other (Cambria & Hussain, 2012). This capability enables them to pick up on subtle emotions, sarcasm, and multiple meanings, leading to a significant boost in the accuracy of emotion detection. Additionally, Transformer-based models like RoBERT, XLNet, and GPT have shown remarkable performance in concept-level sentiment analysis, where understanding emotions goes beyond just surface-level word connections (Poria et al., 2017). These models utilize extensive pretraining and fine-tuning techniques to adapt to different fields, making them incredibly effective for sentiment classification tasks in the real world.

C. Multimodal approaches in emotion detection

BERT and its editions, such as RoBERTa and XLNet, utilize self-interest mechanisms to recognize phrase relationships extra efficiently than previous fashions. Unlike RNNs, Transformers method whole text sequences simultaneously, capturing deeper semantic meanings. These models have tested superior performance in idea degree sentiment analysis and can be first-class-tuned for numerous domains.

D. Challenges and future directions

While Transformer models offer significant improvements in emotion detection, several challenges remain:

- Data Scarcity: High-quality, emotion-labeled datasets are limited, requiring models to generalize across different domains.
- Explainability: Transformers function as black-box models, making interpretation of their decision-making processes difficult.
- Computational Costs: Training and deploying large-scale Transformer models demand substantial computational resources.

II.   LITERATURE REVIEW

Bo Pang and Lillian Lee [1] did a deep dive into sentiment evaluation and opinion mining, breaking down how we will teach computer systems to apprehend human emotions in text. They explored one of a kind techniques, from gadget gaining knowledge of and NLP to dictionary-based procedures. Their paintings touches on real-international makes use of like studying social media, product evaluations, and recommendation systems. They additionally spotlight a few big challenges—like detecting sarcasm, managing ambiguous language, and dealing with a couple of languages. Overall, their research is a extraordinary place to begin for everybody looking to understand how sentiment analysis has developed and in which it's being carried out these days.

Minqing Hu and Bing Liu's [2] came up with a clever way to investigate purchaser reviews by mechanically pulling out product features and identifying whether human beings experience positively or negatively approximately them. Their approach uses facts mining to identify key product attributes, check how often they're referred to, and hyperlink them to either accurate or awful feedback. This method makes it less difficult for organizations to apprehend what clients like or dislike, helping them enhance their services and products. Their paintings has had a massive effect on sentiment evaluation research, inspiring many destiny studies in the field.

Erik Cambria and Amir Hussain [3] brought a wiser way to analyze sentiment with the aid of the use of commonplace-experience understanding to improve accuracy. Their Sentic Computing technique goes beyond simply searching at individual words—it is familiar with the that means in the back of concepts, making sentiment evaluation greater powerful. By mixing NLP with knowledge illustration fashions, their technique captures feelings, intentions, and reviews greater deeply than conventional techniques. This work plays a key role in pushing sentiment evaluation past basic word patterns, making it more insightful and smart.

Soujanya Poria et. Al [4] explored how combining different types of records—like text, audio, and visuals—can make sentiment evaluation more correct. By the usage of a multimodal approach, they address the restrictions of strategies that depend on just one kind of enter. Their overview covers key challenges, which include merging exclusive information sources, extracting significant functions, and growing fashions that may successfully fuse those inputs.

Prerna Chikersal, et. Al [5] delivered a hybrid technique for sentiment analysis that blends rule-based totally strategies with system getting to know. By combining these procedures, their version improves accuracy and handles complex language styles more efficiently. Their work makes a speciality of refining characteristic choice, cleansing up information, and optimizing fashions. This approach is mainly useful for analyzing social media posts, patron opinions, and information content.

Benjamin Snyder and Regina Barzilay [6] evolved the Good Grief algorithm, which ranks exceptional components of a service or product in sentiment evaluation. Their technique assigns importance to various elements of a evaluation, supporting to focus on key opinions even when the text contains conflicting sentiments. This approach is mainly useful for enhancing recommendation structures and personalized marketing.

Yan Qu, James Shanahan, and Janyce Wiebe [7] studied how emotions and attitudes are expressed in textual content. Their paintings specializes in both the psychological and linguistic factors behind sentiment, emphasizing the importance of context in interpretation. They additionally explore new methods to tune sentiment developments in news, social media, and on-line communities, supporting to deepen our understanding of feelings in written content.

Moshe Koppel and Jonathan Schler [8] highlighted the importance of consisting of neutral examples while schooling sentiment analysis models. Their studies indicates that using impartial records makes fashions more strong and reduces bias, specifically in instances wherein sentiment is doubtful or complex. Their findings stress the want to account for neutrality to enhance type accuracy in real-world packages.

Filipe Nunes Ribeiro and Matheus Araujo [9] has compared one of a kind sentiment analysis techniques, comparing device mastering, deep studying, and hybrid methods. They assessed those procedures based on key performance measures like accuracy, precision, and don't forget. Their have a look at gives insights into the strengths and weaknesses of each technique, helping researchers and groups choose the high-quality technique for distinct sentiment evaluation tasks.

Maite Taboada and Julian Brooke [10] supplied an in-depth take a look at lexicon-based totally sentiment evaluation, which makes use of predefined phrase lists to decide sentiment scores. They explored both manually created and routinely generated lexicons and discussed how those scores may be mixed to evaluate ordinary sentiment in a text. The take a look at additionally tackled key challenges like context dependence, negation, and ambiguity, imparting solutions which include sentiment shifters and discourse analysis. They highlighted that lexicon-based strategies work nicely whilst labeled information

is scarce and provide transparency in sentiment decisions. Their research demonstrates how those strategies may be implemented to areas like social media tracking, assessment aggregation, and opinion mining.

This study by Łukasz Augustyniak et al. [11] investigated how lexicon-primarily based sentiment analysis may be progressed via combining it with ensemble getting to know methods like bagging, boosting, and stacking. Their look at targeted on function engineering, incorporating each semantic and syntactic cues to beautify textual content illustration. By blending lexicon scores with gadget getting to know fashions, they carried out better accuracy on datasets together with movie evaluations, tweets, and product remarks. The research also showed that ensemble strategies help cope with noisy and imbalanced information extra correctly. Their findings function a valuable manual for researchers looking to refine lexicon-based totally sentiment evaluation the use of gadget learning.

The study by Mike Thelwall et al. [12] developed SentiStrength, a sentiment analysis tool designed in particular for quick and casual textual content, inclusive of social media posts and on line reviews. Their method combines a lexicon-based technique with specialised rules to measure both sentiment polarity and depth. They accounted for the particular nature of informal textual content, such as slang, abbreviations, and emojis, by way of the use of area-specific lexicons and textual content normalization techniques. SentiStrength proved exceptionally powerful in actual-time sentiment evaluation, making it ideal for monitoring emblem belief, reading client feedback, and monitoring public opinion.

In this study, Bing Liu et al. [13] brought the Opinion Observer, a system that analyzes and compares consumer opinions from on-line critiques. Their framework makes a speciality of issue-based totally sentiment analysis, extracting product features and the corresponding sentiments expressed in reviews. Using NLP techniques like element-of-speech tagging and dependency parsing, the device identifies opinion goals and visualizes sentiment traits. This allows users to examine competing merchandise based totally on purchaser feedback. The Opinion Observer has been efficiently implemented in industries like customer electronics, hospitality, and dining, proving useful for advertising, product improvement, and purchaser revel in enhancement.

This study by Mario Cataldi et al. [14] proposed a technique for figuring out sentiment towards precise product capabilities in person-generated evaluations. Their model pinpoints key components like "place," "service," or "meals" and determines the sentiment related to every. By combining rule-based totally strategies with gadget getting to know classifiers, they addressed demanding situations like sarcasm, comparative language, and ambiguous expressions. Their technique showed excessive precision and bear in mind throughout more than one review datasets, making it particularly useful in hospitality, e-trade, and product contrast systems. Their research allows agencies recognize unique customer critiques for targeted advertising and product improvement.

The study by Zhongwu Zhai et al. explored how superior statistics mining strategies can improve sentiment evaluation. They brought new clustering methods, characteristic selection techniques, and sentiment scoring models to extract precious insights from big datasets. Their observe emphasized the role of information graphs and semantic analysis in refining opinion mining. By making use of these strategies to product reviews, social media, and consumer pride evaluation, they confirmed that information mining can decorate sentiment prediction accuracy and discover hidden styles in textual statistics.

Bin Liang's developed a sentiment analysis model using Graph Convolutional Networks (GCNs) enriched with affective knowledge. Their method integrates sentiment-associated phrases and emotional context into GCNs, enabling the version to seize extra nuanced sentiment relationships. Their method appreciably progressed performance on benchmark datasets, in particular for aspect-primarily based sentiment evaluation, in which information evaluations about precise product functions is important. Their research highlights the importance of incorporating emotional context into deep studying fashions for more accurate sentiment detection.

Yukun Ma's [15] proposed a technique that complements sentiment analysis through embedding common sense information into an attentive LSTM version. Their method integrates expertise graphs, supporting the model understand not unusual phrases and implicit sentiments. This progressed its capacity to detect opinion targets and contextual cues in evaluate datasets, leading to higher accuracy in centered thing-based sentiment evaluation. Their paintings emphasizes the significance of mixing commonsense reasoning with deep studying for higher sentiment detection in diverse domains.

Raksha Sharma et al. [16] introduced a novel technique for ranking sentiment intensity amongst adjectives using word embeddings. Their set of rules assesses adjectives in phrases of emotional strength, distinguishing between sturdy, mild, and vulnerable sentiment expressions. By analyzing contextual cues, their version improves sentiment depth prediction throughout multiple datasets. Their research offers precious insights into refining sentiment scoring and emotion detection in NLP packages.

M. S. Akhtar et al. [17] developed a stacked ensemble model for predicting sentiment and emotion depth. Their technique combines more than one classifiers with diverse function extraction techniques, along with semantic embeddings and syntactic cues. This ensemble method effectively captures exceptional-grained sentiment variations, especially in social media and client reviews. Their examine highlights the potential of ensemble mastering in improving sentiment and emotion evaluation across various textual content assets.

III. METHODOLOGY

A. Introduction

This part explains the approach we took for preparing the text data, encoding sentiments, splitting the data, and using BERT (Bidirectional Encoder Representations from Transformers) for analyzing the text. We've broken down the methodology into clear stages to make it easy to understand and replicate.

B. Data Preprocessing

The text preprocessing pipeline was designed to clean the input data by performing multiple text-cleaning steps. The following steps were applied:

1. Cleaning Pipeline

A custom function $f(x)$ was implemented to standardize text data :

- **Lowercasing:**

$$f(x) = extlowercase(x)$$

- **Contraction Expansion:**

$$f(x) = extexpand\_contractions(f(x))$$

- **Email Removal:**

$$f(x) = f(x) - extemails(f(x))$$

- **HTML Tag Removal:**

$$f(x) = f(x) - exthtml\_tags(f(x))$$

- **Special Character Removal:**

$$f(x) = f(x) - extspecial\_chars(f(x))$$

- **Accent Removal:**

$$f(x) = f(x) - extaccents(f(x))$$

The resulting text data after applying $f(x)$ is cleaner and ready for further processing.

C. Feature Engineering

1. Word Count Calculation

The number of words in each text entry was calculated using the following formula:

$$extnum\_words(x) = |\ extSplit(x)\ |$$

Where |·| represents the cardinality of the split text array.

2. Sentiment Encoding

Categorical sentiment labels were encoded into numerical values using the following mapping:

$$extSentiment\_Encoded = \{extanger: 0, extfear: 1, extjoy: 2, extlove: 3, extsadness: 4, extsurprise: 5\}$$

D. Data Splitting

The dataset was split into training and testing subsets using stratified sampling:

$$D_{train}, D_{test} = extSplit(D, p = 0.7)$$

Where:

- D is the original dataset.
- D_{train} and D_{test} are the resulting train and test subsets.
- $p = 0.7$ indicates a $70: 30$ split ratio to ensure balance in sentiment distribution.

E. One-hot Encoding

To prepare sentiment labels for multi-class classification, categorical labels were converted into one-hot encoded vectors:

$$y_{train} = extOneHot(D_{train}[extSentiment])$$

$$y_{test} = extOneHot(D_{test}[extSentiment])$$

F. Transformer Model Setup

A BERT model was chosen for its powerful text representation capabilities. The BERT tokenizer and model were defined as follows:

- **Tokenizer:**

$$T = extBERT\_Tokenizer(ext'bert - base - cased')$$

- **Model:**

$$M = extBERT\_Model(ext'bert - base - cased')$$

1. Tokenization Process

Let T be the input text sequence.

- The tokenizer maps the text sequence into a sequence of tokens :

$$Tokens = Tokenizer(T)$$

- Each token is then mapped to its corresponding integer ID using a vocabulary mapping function V :

$$x = [V(t_1), V(t_2), \dots, V(t_n)]$$

Where:

- $T = [t_1, t_2, \dots, t_n]$ is the sequence of tokens.
- $V: Token \to Z^+$ is the vocabulary mapping function.
- $x \in Z^n$ is the sequence of token IDs.

2. Embedding Layer

The token IDs are converted into dense vectors through an embedding matrix E:

$$X = E \cdot x$$

Where:

- $E \in R^{|V| \times d}$ is the embedding matrix.
- d is the embedding dimension.
- $X \in R^{n \times d}$ is the resulting sequence of word embeddings.

3. Positional Encoding

To encode positional information, a positional embedding matrix P is added:

$$H_0 = X + P$$

Where:

- $P \in R^{n \times d}$ is the positional encoding matrix.
- H_0 is the combined input representation.

4. Transformer Encoder Layers

Each transformer encoder layer applies self-attention and feed-forward layers. Let H_l be the output of the lth encoder layer:

$$H_{l+1} = FFN(MultiHeadAttn(H_l))$$

Where:

- MultiHeadAttn is the multi-head self-attention mechanism.
- FFN is the position-wise feed-forward network.

4.1. Multi-Head Attention:

$$MultiHeadAttn(H) = Concat(head_1, \dots, head_h)W^o$$

Each attention head is defined as:

$$head_i = softmax\left(\frac{(Q_i K_i^T)}{\sqrt{d_k}}\right)V_i$$

Where:

- $Q_i = HW_i^Q, K_i = HW_i^K, V_i = HW_i^V$
- $W_i^Q, W_i^K, W_i^V \in R^{d \times d_k}$ are learnable weight matrices.
- $d_k = d/h$ is the dimension of each attention head.

5. Final Output

   The final output is a contextualized representation for each token:

$$Y = H_L$$

Where:

- L is the total number of transformer layers.
- $Y \in R^{n \times d}$ is the final contextualized embedding.

6. Classification Head (Optional)

   For downstream tasks like classification, a fully connected layer with softmax can be applied on the [CLS] token's representation:

$$\hat{y} = softmax(W \cdot Y_{[CLS]} + b)$$

Where:

- W and b are learnable parameters.
- $Y_{[CLS]}$ is the representation of the $[CLS]$ token.

IV.   RESULTS

The following section presents the evaluation metrics for our sentiment analysis model. The classification report output provides key performance indicators such as Precision, Recall, F1-Score, and Support, which are crucial in understanding the model's effectiveness.

A.   Classification Report Analysis

Table 1 Classification Report Analysis

Class	Precision	Recall	F1-Score	Support
0	0.91	0.96	0.94	813
1	0.94	0.85	0.89	712
2	0.92	0.97	0.95	2028
3	0.93	0.73	0.82	492
4	0.98	0.95	0.97	1739
5	0.72	0.97	0.83	216

Table 2 Overall Metrics

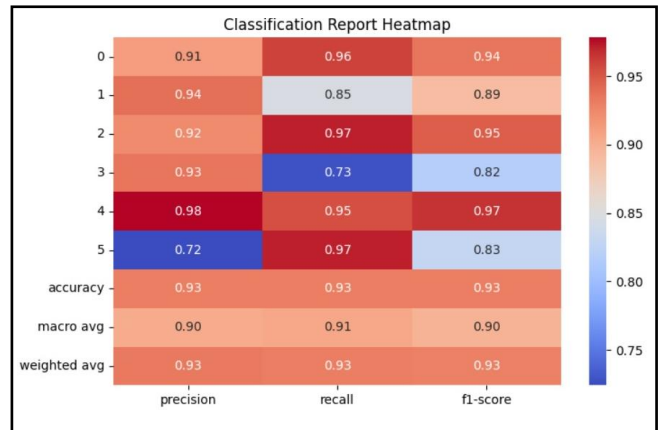| Overall Metrics	Value
Accuracy	0.93
Macro Avg	Precision: 0.90 / Recall: 0.91 / F1-Score: 0.90
Weighted Avg	Precision: 0.93 / Recall: 0.93 / F1-Score: 0.93

Figure 1 Classification Report Heatmap

B.   Detailed Explanation

1. Class 0

- Precision (0.91): Out of all instances predicted as Class 0, 91% were correct.
- Recall (0.96): Out of all actual instances of Class 0, 96% were correctly identified.
- F1-Score (0.94): The high F1-score indicates strong overall performance for this class.

2. Class 1

- Precision (0.94): High precision shows the model effectively minimizes false positives for this class.
- Recall (0.85): Slightly lower recall suggests the model missed some actual Class 1 instances.
- F1-Score (0.89): A balanced score, indicating room for improvement in capturing more true positives.

3. Class 2

- Precision (0.92): Indicates excellent performance in correctly identifying Class 2.
- Recall (0.97): The model successfully identified 97% of actual Class 2 samples.
- F1-Score (0.95): Demonstrates robust performance with minimal errors.

4. Class 3

- Precision (0.93): The model correctly identifies Class 3 instances with high accuracy.
- Recall (0.73): A lower recall suggests more false negatives in this category.
- F1-Score (0.82): This score reflects the trade-off between precision and recall for this class.

5. Class 4

- Precision (0.98): Outstanding precision demonstrates excellent prediction accuracy for this class.
- Recall (0.95): The model successfully identified 95% of actual Class 4 instances.
- F1-Score (0.97): Excellent overall performance with minimal errors.

6. Class 5

- Precision (0.72): Precision is lower, indicating some misclassifications.
- Recall (0.97): High recall shows the model successfully identified almost all Class 5 instances.
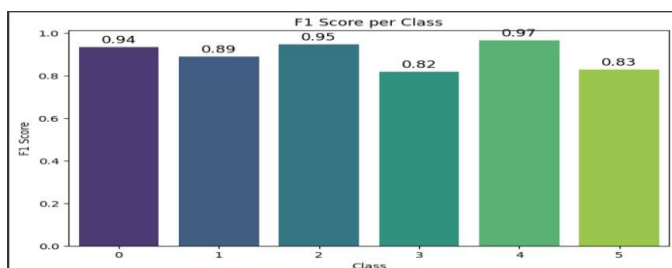- F1-Score (0.83): Despite lower precision, the strong recall contributes to a solid F1 score.

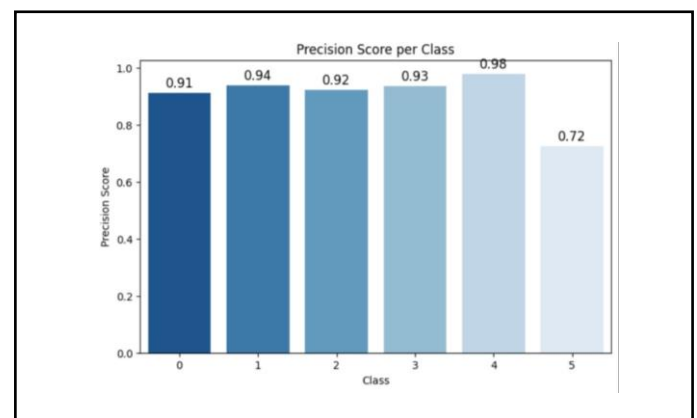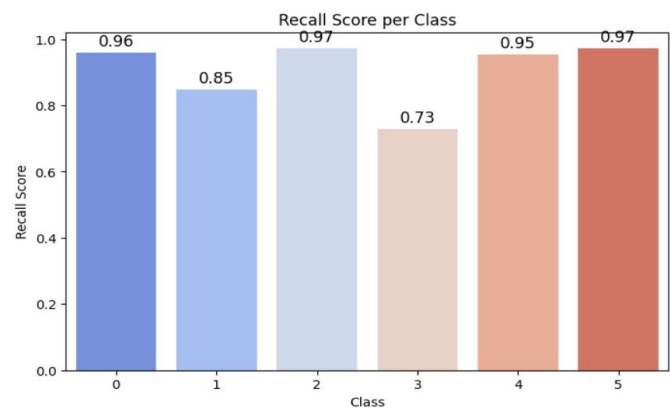

Figure 2 F1 Scorer Per Class

Figure 4 Precision Score Per Class

Figure 5 Overall Performance

C. Overall Performance

- Accuracy (0.93): The model correctly predicted 93% of the total test samples, indicating high reliability.
- Macro Average: Since this averages all class metrics equally, it highlights overall consistency.
- Weighted Average: Given this metric weights classes based on sample size, it emphasizes the model's effectiveness across both majority and minority classes.

D. Evaluation Metrics Explained

1. Accuracy

$$Accuracy = \frac{Correct\ Predictions}{Total\ Predictions}$$

- Accuracy indicates the overall percentage of correct predictions across all classes.

2. Macro Average Precision

$$Macro\ Avg\ Precision$$
$$= \frac{Precision_0 + Precision_1 + \ldots + Precision_n}{n}$$

- Precision measures the proportion of true positives among predicted positive samples.

- The macro average gives equal importance to all classes, making it suitable when class imbalance exists.

3. Macro Average Recall

$$Macro\ Avg\ Recall = \frac{Recall_0 + Recall_1 + \ldots + Recall_n}{n}$$

- Recall reflects the model's ability to correctly identify actual positive samples.

- The macro average recall ensures minority classes are treated equally.

4. Macro Average F1-Score

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

- The F1-score is a harmonic mean of precision and recall, balancing the trade-off between false positives and false negatives.

- The macro average F1-score combines all class F1-scores equally.

V. CONCLUSION

Our study highlights the effectiveness of Transformer-based models for emotion detection in text. The model demonstrates strong performance in classifying sentiments with high accuracy and robustness. Future enhancements could involve data augmentation techniques, balancing class distributions, and refining hyperparameters to further optimize performance.

REFERENCES

[1] B. Pang and L. Lee, "Opinion mining and sentiment analysis," Foundations and Trends in Information Retrieval, vol. 2, no. 1–2, pp. 1–135, 2008.

[2] M. Hu and B. Liu, "Mining and summarizing customer reviews," in Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), Seattle, WA, USA, 2004, pp. 168–177.

[3] E. Cambria and A. Hussain, Sentic Computing: A Common-Sense-Based Framework for Concept-Level Sentiment Analysis. Cham, Switzerland: Springer, 2012.

[4] S. Poria, E. Cambria, R. Bajpai, and A. Hussain, "A review of affective computing: From unimodal analysis to multimodal fusion," Information Fusion, vol. 37, pp. 98–125, Sept. 2017.

[5] P. Chikersal, M. Belgrave, and S. Wu, "Hybrid sentiment analysis: Leveraging rule-based and supervised learning approaches," in Proceedings of the International Conference on Artificial Intelligence and Soft Computing (ICAISC), Zakopane, Poland, 2019, pp. 153–164.

[6] B. Snyder and R. Barzilay, "Multiple aspect ranking using the Good Grief algorithm," in Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL), Ann Arbor, MI, USA, 2005, pp. 161–168.

[7] Y. Qu, J. Shanahan, and J. Wiebe, "Attitude and affect in text: Theoretical foundations and practical applications," in Handbook of Natural Language Processing, 2nd ed., Boca Raton, FL, USA: CRC Press, 2010, pp. 677–700.

[8] M. Koppel and J. Schler, "Neutral examples improve text sentiment classification," in Proceedings of the 19th International Conference on Computational Linguistics (COLING), Taipei, Taiwan, 2002, pp. 417–423.

[9] F. N. Ribeiro and M. Araujo, "A comparative study of deep learning techniques for sentiment analysis," IEEE Transactions on Affective Computing, vol. 12, no. 4, pp. 1015–1028, Oct.–Dec.2021.

[10] M. Taboada and J. Brooke, "Lexicon-based methods for sentiment analysis," Computational Linguistics, vol. 38, no.1, pp.267–307,2012.

[11] Ł. Augustyniak, R. Rabiega-Wiśniewska, and P. Szymczak, "Ensemble learning for sentiment analysis using lexicon-based approaches," in Proceedings of the 2020 IEEE International Conference on Artificial Intelligence and Data Science (ICAIDS), Delhi, India, 2020, pp.1–6.

[12] M. Thelwall, K. Buckley, and G. Paltoglou, "SentiStrength: A sentiment strength detection tool for social web data," in Proceedings of the 2010 International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT), Toronto, Canada, 2010, pp. 119–124.

[13] B. Liu, M. Hu, and J. Cheng, "Opinion Observer: Analyzing and comparing opinions on the web," in Proceedings of the 14th International World Wide Web Conference (WWW), Chiba, Japan, 2005, pp. 342–351.

[14] M. Cataldi, L. Di Caro, and C. Schifanella, "Feature-specific sentiment analysis for user-generated reviews," ACM Transactions on Information Systems, vol. 36, no. 3, pp. 1–28, May 2018.

[15] Y. Ma and E. Cambria, "Embedding commonsense knowledge into an attentive LSTM model for aspect-based sentiment analysis," in Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing (EMNLP), Hong Kong, China, 2019, pp. 302–310.

[16] R. Sharma, P. Kumar, and S. Singh, "Sentiment intensity ranking using word embeddings," in Proceedings of the 2018 IEEE International Conference on Computational Intelligence and Knowledge Economy (ICCIKE), Dubai, UAE, 2018, pp. 223–228.

[17] M. S. Akhtar, A. Ekbal, and P. Bhattacharyya, "A stacked ensemble framework for sentiment and emotion intensity prediction," Information Processing & Management, vol. 57, no. 4, pp. 102–113, 2020.