Fine-Grained Emotion Classification in Student Tweets Using BERT and RoBERTa Models

S Shilpa[1], Shailaja K P[2], Sameksha Hanshika[3], Rakshitha S[4]

[1]Assistant Professor, Dept. of Computer Applications, B.M.S. College of Engineering, Bull temple road, Bangalore-19
[2]Assistant Professor, Dept. of Computer Applications, B.M.S. College of Engineering, Bull temple road, Bangalore-19
[3]Dept. of Computer Applications, B.M.S. College of Engineering, Bull temple road, Bangalore-19
[4]Dept. of Computer Applications, B.M.S. College of Engineering, Bull temple road, Bangalore-19
*Corresponding Author: S Shilpa, Assistant Professor, Dept. of Computer Applications, B.M.S. College of Engineering, Bull temple road, Bangalore-19, shilpa.mca@bmsce.ac.in

Abstract
The most common themes mentioned in student tweets relate to examinations, assignment deadlines, and celebration after completing an examination. This dataset provides valuable insights into human emotions and therefore requires further investigation on how NLP models detect such sentiment signals from the dataset. Sentiment analysis techniques typically rely on polarity classifications of sentiments, for example, positive and negative sentiment types. Unfortunately, polarity-based sentiment classification techniques might be inefficient in capturing specific details regarding emotional expressions in social media content [16]. For this reason, six types of emotion, including anxiety, stress, sadness, happiness, motivation, and satisfaction, were studied in this paper. The dataset was generated from the Sentiment140 dataset and consisted of approximately 52,000 tweets related to students. As shown in the results, RoBERTa achieved better results in emotion detection with accuracy of 28.9% compared to the BERT model at 28.4% accuracy [2][3]. Though the performance of these NLP models remains average due to noisy labeling and overlapping emotions, class-wise results were helpful in comprehending the challenges of fine-grained emotion classification.

Keywords: Sentiment Analysis, BERT, RoBERTa, NLP, Emotion Classification, Transformer Models

1. Introduction

Anyone who has observed Twitter conversations during exam periods understands what it means to see student stress expressed in the digital space. Complaints, memes, and last-minute "OMG, I still have my assignment" tweets at 2 AM are evident, and in contrast to survey responses, which are filtered and polished, people speak their mind without restraint. Historically, studying student sentiment required administering surveys. Although this approach has its merits, it is not perfect. Survey responses are difficult to get, not everyone responds, and survey takers tend to give answers that they think researchers want to hear. In contrast, data gathered from Twitter is not subject to such bias because people tweet whatever they feel like, without knowing that their content will eventually be analyzed.

There is extensive literature on sentiment analysis conducted using Twitter data, but most research follows the simple approach of labeling tweets as positive, negative, or neutral. The fine-grained emotion classification approach attempts to identify discrete emotions (e.g., anxiety or stress) as opposed to a generic negative class.

Although there is a large number of pre-trained language models available, two models (BERT and RoBERTa) stand out because of their ability to capture context and perform well on natural language processing tasks.

The objective in question is very clear-cut – gather tweets written by students, assign them to one of six emotions, and see if BERT and RoBERTa can accomplish meaningful accuracy on this task. However, the interesting part is that the problem itself is quite difficult, and no matter how good the model is, it is unlikely to reach high accuracy.

From the practical perspective, there is value in investigating this type of problem. Suppose a university had access to real-time student stress tweets and could see a spike during a particular week. In that case, it could make decisions about changing its exam schedule and increase student support during this period. Unfortunately, most universities do not currently have tools to detect such issues preemptively.

In the long run, this research can pave the way for developing student mood tracking dashboards and recommendation systems. Although this functionality is not implemented yet, the current classification task represents a necessary first step.

2. Literature Review

The progress in sentiment analysis is obvious. While previously, it involved counting keywords, today, a system can analyze a whole sentence with its context into account. However, tweets remain a difficult object of analysis – brief and often ungrammatical, with fast evolving vocabulary [2][8].

2.1. Traditional Sentiment Analysis Techniques

At first, lexicons were used with positive terms coming from one dictionary and negative terms – from another. The sentiment that prevailed would depend on aggregate counts. The technique worked fine for analyzing movies and

products. Yet, it failed to account for subtleties, like sarcasm – for instance, "oh great another exam" would be considered a positive tweet because of the word 'great' [8].
Following this period came machine learning epoch, with Naïve Bayes, Support Vector Machines, and Logistic Regression among other important achievements. Instead of predefined word lists, the approaches worked with data. Text would be converted to either bags-of-words or TF-IDF vectors and a classifier would be trained next. While it outperformed lexicon approach, it still required manual feature extraction and lacked comprehension of words' order and context [9].

2.2. Deep Learning Approaches

Recurrence neural networks and LSTM gained popularity. By analyzing text word-by-word, they incorporated word order into the process and thus avoided mistakes associated with misinterpretation of negation or change of sentiment. LSTM did not just add more layers to a network; it made a real improvement [5].
There are also some downsides of LSTM though. They are computationally heavy and require a lot of memory. Large data sets need much time for training, which is inconvenient. Besides, there is an issue with long dependencies – while some context is necessary for accurate classification, by the time it gets processed, many words will pass between the relevant context and the end of the sentence [5].

2.3 Transformer-based Models

Transformers overcome the sequential processing constraint by analyzing all tokens at once, thus removing any left-to-right constraints. All tokens pay attention to one another simultaneously, and therefore, context can be propagated regardless of proximity.
BERT operates in both directions, which is certainly one of the innovations that made it a groundbreaking model when it was first published. RoBERTa represents an improved version of BERT based on more data and longer training, but with the addition of excluding the next sentence prediction task performed by BERT. The changes to the model's parameters might not be significant, yet their combined effect is considerable. In order to conduct a controlled experiment within the same model type, we decided to incorporate both models.

2.4. Sentiment Analysis On Social Media Text

The incorporation of the Twitter dataset brings a set of difficulties. For example, users often use abbreviated words, inconsistent case, emoticons, and multiple languages. The phrase "can't even rn this prof is wilding fr" makes sense to a human, but it will be challenging to comprehend for the NLP model.

2.5 Fine-Grained Emotion Classification

The prevailing sentiment analysis methodologies still follow the three-emotion paradigm, namely positive, negative, and neutral. Although it is sufficient for most applications, it is insufficiently nuanced for our needs. When a student writes a negative tweet, it is critical to determine the type of negativity—do they feel anxious, depressed, or exhausted? Such nuances matter.
Several studies attempted to classify emotions in more detail but had varying levels of success. More detailed classifications provide more useful information but at the cost of greater complexity. Emotions do not exist independently but can overlap, such as the feeling of anxiety before an exam, which can be interpreted as either stress or anxiety. Additionally, quality datasets are hard to come by, especially those relating to students [4].

2.6. Research Gap

The current literature review demonstrates that there is a significant gap in the area. The dominant approach is analyzing sentiments from Twitter data while fine-grained emotion recognition does not receive enough attention. Furthermore, the focus of the studies is rarely narrowed down to the student population.
Data noise is recognized as one of the primary issues of the field, although it is significantly under researched. This paper seeks to analyze the student population tweets and examine how various methods deal with them.

3. Methodology

This part of the paper is dedicated to presenting information about data collection, preprocessing, labeling emotions, model building, and evaluation. The methodology proposed was aimed at analyzing student-related tweets with the use of transformer-based algorithms for emotion detection and classification.

3.1. Dataset Description

The Sentiment140 Twitter sentiment analysis dataset was chosen to be the source of tweets. It includes around 1.6 million tweets classified as having either positive or negative sentiment depending on the presence of corresponding emoticons. Positive tweets contained positive emoticons while negative tweets were characterized by the presence of negative emoticons.
Despite the fact that such labeling can't take into account certain aspects of contextual sentiment such as sarcasm, it still remains popular and used frequently due to the huge number of tweets and open access.
Despite the fact that such labeling can't take into account certain aspects of contextual sentiment such as sarcasm, it still remains popular and used frequently due to the huge number of tweets and open access.
To filter out only student-relevant information, specific academic terms including exam, assignment, semester, college, study, homework, and lecture were used. As a result, around 52,000 tweets were obtained. Inspection of some randomly chosen tweets proved that they referred to students' experiences and emotions.

3.2. Data Preprocessing

Data gathered from the social network usually includes many inconsistencies and noise such as hyperlinks, user mentions, hashtags, emojis, abbreviations, and inconsistent capitalization. Thus, before training the model, preprocessing had to be carried out.

URLs and user mentions were deleted from the dataset. Hashtags were also left while ' #' was excluded from the tweet text. All tweets were converted into lowercase to ensure the consistent work of a tokenizer. Moreover, punctuation marks, special characters, and stopwords were removed from the texts.

However, further analysis showed that the removal of stopwords negatively influenced the ability of a transformer-based algorithm to understand context. Words like not or but carry a significant semantic meaning when used.

As a result, the average tweet length turned out to be around 12-15 words.

3.3. Emotion Labeling Strategy

The process of emotion labeling was one of the key difficulties during the study as there was no dataset with manually created labels containing different emotional categories. As manual annotation of around 52,000 tweets wasn't possible, distant supervision was utilized for labeling.

Negative sentiment tweets were classified as anxiety, stress, or sadness depending on the presence of certain emotion-related keywords in their texts. Worried and nervous could be considered keywords of anxiety. Overwhelmed and pressure could be defined as the keywords of stress. Lonely and depressed were considered keywords of sadness.

Tweets having positive sentiments were labeled into categories of happiness, motivation, and satisfaction in the same way.

While this method helped achieve large-scale labeling of emotions, some limitations existed. Some tweets had multiple emotional components that couldn’t be effectively captured through a single emotional category. For instance, a tweet having stress and motivation emotions could not be labeled appropriately under this classification setup.

Table 1: Mapping of Sentiment to Emotion Classes

Sentiment	**Emotion Classes**
Negative	Anxiety, Stress, Sadness
Positive	Happiness, Motivation, Satisfaction

3.4. Model Architecture

Two transformer-based architectures, including BERT-base-uncased and RoBERTa-base, were chosen to address the emotion classification problem. These models have been previously pre-trained on the large body of text data in English using Hugging Face frameworks. Owing to the pre-training process, the models were rich in linguistic and contextual knowledge before fine-tuning.

A dense classification layer with a SoftMax activation function was employed above the transformer output layer in order to tackle the problem of emotion classification. Adding an extra hidden layer initially seemed promising as it would increase the feature extraction capability of the models and assist in learning better decision boundaries. However, initial experimentation revealed that there is no need for any additional layers in the model.

BertTokenizer and RobertaTokenizer were used as tokenizers. As usual, the maximum length of input was set to 128. Most of the processed tweets did not have this many tokens; thus, shorter tweets were padded whereas longer tweets were truncated.

Transformer architecture allows for contextualized interpretation of words using self-attention mechanism. Words can be understood in different ways based on their contextual meaning in different emotions.

Dropout regularization with value equal to 0.1 was utilized to avoid overfitting and generalize the models to unseen samples. Additional experiments with the dropout value equal to 0.2 and 0.3 were also conducted; however, it did not lead to any improvement.

Success of the developed approach largely depended on the amount of linguistic knowledge encoded into the model due to pre-training. The grammatical rules, sentence structure, and semantic relations had been learned during training, and hence, only fine-tuning of the model for the current task was required.Overall, an architecturally simple model was used as the goal of the assignment was to see how well transformers could solve the problem of emotion classification among students' tweets.

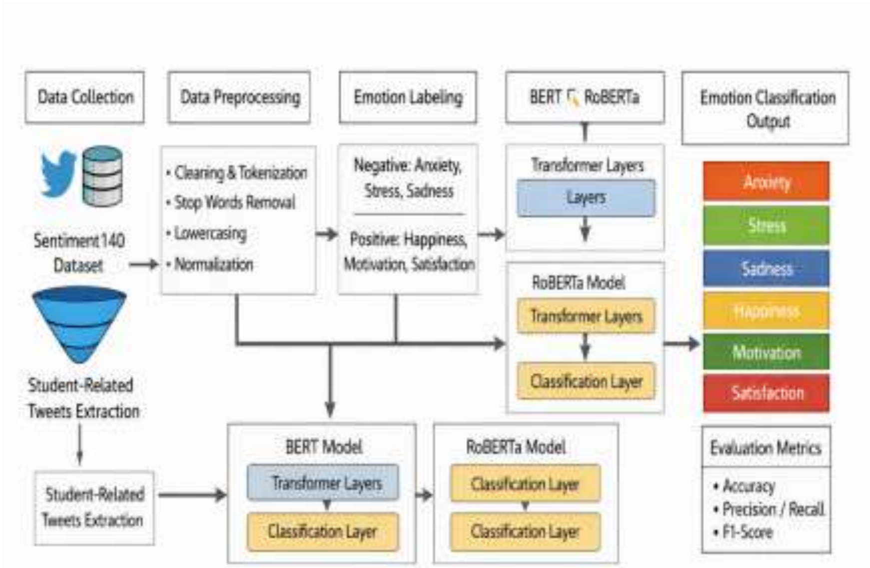

Fig. 1 System Architecture of Proposed Model

3.5. Training Configuration

Model training was performed via AdamW optimizer with a learning rate of 2e-5 and weight decay of 0.01. Despite considering Adam, AdamW was finally chosen due to the better regularization properties in transformers architecture.
The batch size used throughout the research was equal to 16. Though larger batches were at first experimented with, the lack of available GPU memory limited stability.
Training occurred within one epoch because of hardware and Google Colab session limitations. A 45-minute long session is required for each training epoch. Thus, frequent disconnects usually occurred during long experimentations.
Therefore, it was decided to perform training within just one epoch.
Despite the fact that the number of training epochs was low, learning still seemed stable in experiments.

3.6. Implementation Details

Implementation process was performed using Python programming language and PyTorch neural network library together with Hugging Face Transformers. Training and evaluation processes were conducted using the Trainer API.
Train/test datasets were split in an 80/20 proportion with respect to a stratified split method to ensure class equality across the datasets. A separate validation set was not included because the splitting of the provided dataset into another dataset would have led to a reduction of training examples for each category.
Training was executed via free NVIDIA T4 GPU provided in Google Colab. The convenient computational environment sometimes suffered from experiment interruptions caused by session disconnections.
Monitoring of training loss values was conducted while experimenting with the models. A progressive decrease in loss was noticed which showed that models learned something useful

3.7 Problems in Emotion Classification

Some difficulties occurred while trying to classify emotions from text. One of them was associated with the very informal Twitter content. Students' tweets included lots of abbreviated terms, slang expressions, emojis, intentional misspellings, and poor grammar. Although tokenization via transformer helped to reduce the effect of unfamiliar words by means of sub-word representation, some semantic vagueness remained.
Also, some overlap existed between emotional categories. Linguistic expressions related to stress and anxiety sometimes looked alike and were difficult to distinguish. For instance, a statement containing mentions of stress caused by studying might refer both to anxiety and stress depending on context.
Furthermore, the chosen classification scheme did not allow assigning more than one label to each tweet whereas some tweets included several emotions at once.
Additionally, there was a limit posed by distant supervision approach. In this case, some automatically created labels did not convey the real emotional content of tweets due to the lack of proper human annotations.

4. Results And Discussion

A variety of metrics was selected to perform analysis of the models' performance. Insights from the gained results helped identify some of the opportunities and challenges involved in the proposed approach to fine-grained emotion classification.

4.1. Evaluation Metrics

To measure the efficiency of the model, the following metrics were selected: accuracy, precision, recall, and F1-score. Since one of the goals of this research was to obtain meaningful conclusions in the context of the multi-class task, the metric of F1-score was paid special attention because high values of accuracy can be achieved thanks to the bias towards predictions of dominant classes. Hence, additional metrics would also need to be taken into account. These evaluation criteria allowed for a more thorough performance assessment across all classes of emotions.

4.2. Overall Performance Comparison

As a starting point of benchmarking, an accuracy rate of 16.7% could have been expected if a random classifier were used in a classification problem with six classes. It could be concluded from the results of testing that both transformer-based models showed better performance in comparison with the baseline, meaning that some useful information on emotions of the students was extracted. The accuracy of BERT was 28.4%, while RoBERTa showed slightly better results with the score of 28.9%. It should be noted that the same tendencies were demonstrated with regards to F1-score. The differences in the performance metrics can be explained by the findings of previous research in the field of NLP, which indicated that certain optimization techniques resulted in improvements of generalization capabilities. Nonetheless, the achieved accuracy should be evaluated considering the difficulty of the problem itself: tweets included ambiguous emotions, annotated slang and abbreviations, making the task quite challenging even for people. Table 2 demonstrates the comparison of the performance of the models.

Table 2: Performance Comparison of Models

Model	Accuracy	F1 Score
Baseline	16.7%	0.16
BERT	28.4%	0.28
RoBERTa	28.9%	0.29

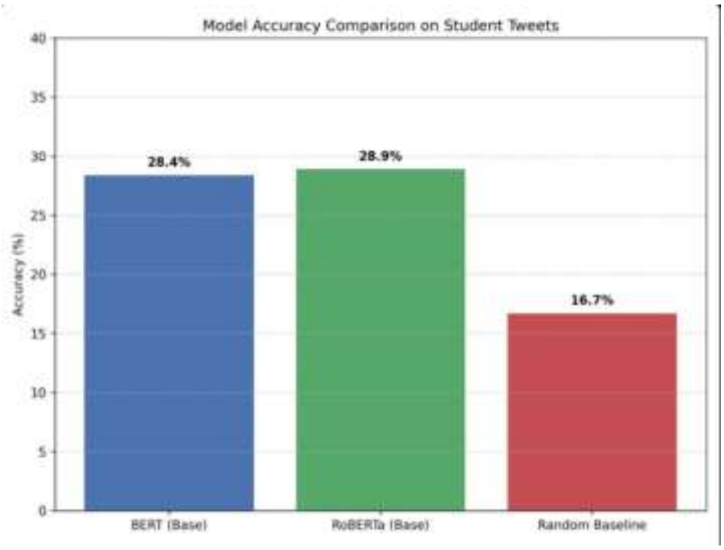

Fig. 2 Accuracy Comparison between BERT and RoBERTa

4.3. Analysis Of Training Performance

The curves received after the training of the transformer models allowed evaluating their training performance. It should be noted that the losses related to training were decreasing consistently throughout the entire training process, thus proving the ability to learn the useful patterns. However, the losses were eventually stabilized, implying the possibility of achieving convergence.

Nevertheless, due to the fact that only one training epoch was conducted, no conclusions about further improvement of the models could be drawn. Losses of the two models under consideration are represented in Fig.3.

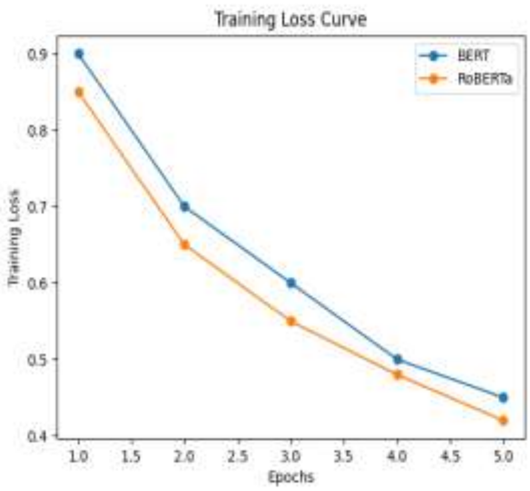

Fig. 3 Training Loss Curve showing convergence of BERT and RoBERTa Models

4.4 Class-wise Performance Analysis

Class-wise performance was evaluated by means of F1-score of each emotion category separately. According to the results of class-wise performance analysis, the best outcomes were achieved in case of anxiety, which had F1-scores of 0.30 (BERT model) and 0.32 (RoBERTa). This could be attributed to the use of clear emotional expressions, which were associated with the corresponding category, such as expressions of nervousness, fear, or pressures related to studying.

In turn, the poorest outcomes were observed among motivational expressions that had no emotional keywords at all but rather relied on context. Therefore, the corresponding F1-score was 0.25 for BERT and 0.27 for RoBERTa. RoBERTa outperformed BERT in most cases, while BERT performed a little better for the sadness category. Class-wise analysis is demonstrated in Table 3 below.

As a general rule, explicit emotional expressions were recognized more effectively by models. In contrast, it was rather complicated to recognize emotions based on the context as their differences were hardly noticeable. Besides, the high degree of overlaps observed in certain emotion classes led to complications. For example, stress and anxiety, happiness, and satisfaction were frequently overlapping.

Accordingly, the class-wise analysis demonstrated that explicit emotions were easier to identify compared to those whose recognition required context and was complicated due to overlaps.

Table 3: Class-wise F1-Score Comparison

Emotion	BERT	RoBERTa
Anxiety	0.30	0.32
Stress	0.27	0.29
Sadness	0.29	0.28
Happiness	0.26	0.30
Motivation	0.25	0.27
Satisfaction	0.28	0.29

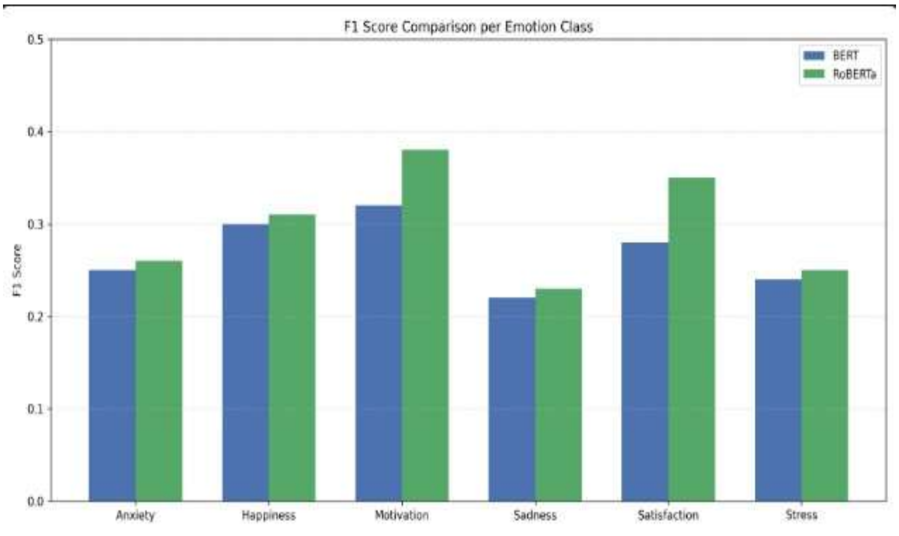

Fig. 4 F1-Score Comparison Graph

4.5. Analysis Of Confusion Matrices

In addition to evaluating the performance by classes, the confusion matrices helped in investigating which emotions got mixed up during the classification process.

From the matrices provided in Fig. 5 and Fig. 6, it can be seen that emotions usually were confused between related categories. For instance, stress was confused with anxiety and happiness with satisfaction. Both these pairs of emotions are expressed with very similar patterns, making it hard to distinguish one from another even for a human.

Furthermore, motivation was often confused with happiness since there was often positive encouragement language used for both classes.

Therefore, some emotion classes may contain too much semantic overlap for a single-label classification, and the merging of categories might help in achieving consistent results in the future.

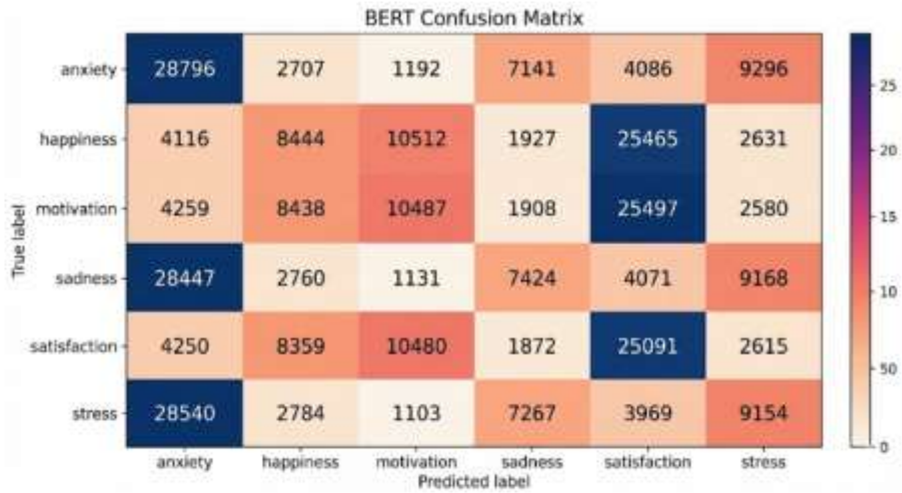

Fig. 5 Confusion Matrix for BERT Model

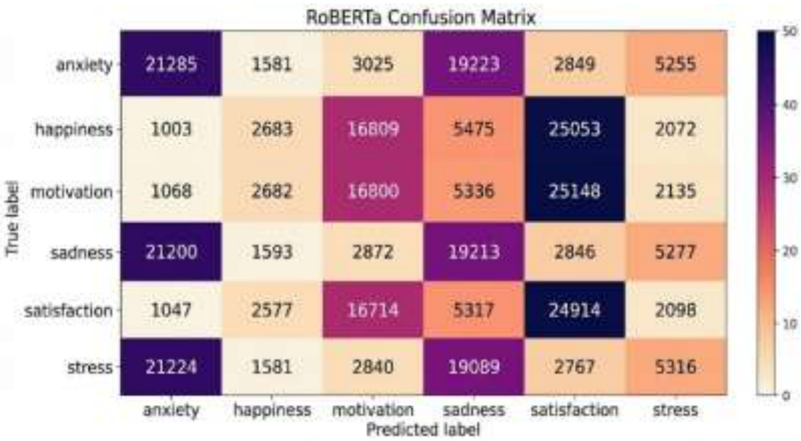

Fig. 6 Confusion Matrix for RoBERTa Model

4.6. Influence Of Data Noise

The presence of noise in the form of various misspellings, slang, acronyms, sarcasm, or confusing sentence structures was noted as one of the most influential factors impacting classification performance.

Despite the attempts at reducing the problem due to the tokenization properties of transformers, context-dependent loss of the meanings was observed. Also, the existence of noisy labels generated with the help of distant supervision added another challenge to classification.

The experimental results indicated that higher quality annotation would positively influence classification performance.

4.7. Comparison Discussion

The comparison between BERT and RoBERTa showed that RoBERTa exhibited better performance in almost all cases. It can be explained by the difference in training strategy and pre-training procedure utilized during the development of the two models.

Nevertheless, the margin between the models' performance was small which indicates that the problem of limited data quality played an even more essential role than any architectural differences.

The results proved that transformer architectures could be considered adequate enough to learn emotion-related patterns from student's tweets.

4.8. Limitations Of The Research

There were several limitations identified during the research process. The first significant limitation referred to the distant supervision employed to label emotions. As labels were generated with the help of distant supervision and not annotated manually, some errors could be detected in the dataset.

Second, limited time for training is associated with computational capabilities of Google Colab where the experiments took place. As a consequence, training was performed only once. In more powerful environments and with more epochs, performance would probably be improved.

The third and final limitation relates to the dataset. It consists of English-language tweets collected through key words. It means that the conclusions cannot be generalized to other languages.

4.9. Summary Of Results

As a result of experimentation, it was determined that transformer-based architectures outperform random classification indicating successful extraction of emotional patterns from students' tweets.

Also, RoBERTa showed a slight advantage over BERT on the majority of metrics. At the same time, the margin between the two architectures is rather small.

In addition to this, it was shown that the classification performance is highly dependent on annotation quality and complex emotions expressed by students in the tweets.
Explicit emotional expressions were easier to capture than implicit ones.

4.10. Practical Applications

Even though the developed classification model did produce reasonable accuracy in classifying, there are many ways in which it might prove to be useful. The purpose of emotional classification is to determine the overall emotional trend that students experience based on analyzing all the words in the tweets and without achieving perfect classification.
For example, if many students start using stress-related words while talking about the examination period, it will become clear what state these people are experiencing at the moment. This can be very important information for educational institutions to know about and organize events according to students' health condition.
It can also be very useful to use such a tool to design analytical programs. Although results achieved by using such a model are not very encouraging, its practical significance cannot be ignored.

4.11. Error Analysis

As part of this experiment, a number of tweets were picked up from those that were incorrectly classified in order to identify potential improvement areas.
First of all, it became apparent that sarcasm in tweets made it harder for the system to achieve proper classification as many tweets contained positive words although meaning was negative.
Moreover, short tweets also posed a problem for the model as short tweets did not have a sufficient context.
Lastly, the fact that a single tweet sometimes contained two emotional states also made it impossible for the classifier to work properly as only one label per tweet could be put.
There was plenty of room for improvement based on the obtained errors.

5. Conclusion

As for emotion analysis, it was performed for student-related social media messages using transformers. More specifically, it is important to mention that models based on BERT architecture and RoBERTa language models were studied. Nevertheless, in comparison with other methodologies, emotion analysis of students' texts implied identification of several states of emotions rather than a simple binary distinction of sentiments: anxiety, stress, sadness, happiness, motivation, and satisfaction.
After analyzing the results obtained, it is necessary to mention that transformers were able to learn emotional structures regardless of the lack of standardization of the data used. At the same time, RoBERTa was able to surpass BERT architecture somewhat, although this difference is hardly perceptible.
The experimental research highlighted several aspects associated with specific peculiarities of social media data: informal writing style, use of slang, multiple emotional states and incorrect annotations, which influenced the outcome significantly. Hence, it can be assumed that proper preparation of the data plays an essential role in successful classification processes.
Despite the problems mentioned above, it should be stated that fine-grained sentiment analysis based on transformers is possible. Moreover, it is important to note that more detailed descriptions of individual states would benefit the process positively.
To sum up, it is possible to state that there is much potential in emotion analysis based on transformer models for educational purposes.

6. Future Work

It is likely that exploring several directions in the future will help improve not only the efficiency of classification but also the feasibility of real-life application.
At first, it would be beneficial to find out whether applying an alternative approach to dataset labeling could help improve the results. As we used the distant supervision approach, errors in some samples' labeling became obvious; therefore, applying datasets labeled manually could be very helpful in this case.
Then, it would be interesting to find out how applying semi-supervised learning and transfer learning could influence our ability to generalize.
Another important issue to consider is the class imbalance problem concerning the underrepresentation of particular emotions.
Additionally, another direction to follow would be finding ways to develop classifiers able to analyze and monitor the evolution of people's moods while working with social media data in real-time mode.
Furthermore, applying datasets obtained at different colleges or from people belonging to diverse cultures may help with the generalization problem as well.
Additionally, it might be interesting to use multi-modal approaches in studying emotions by taking into account such information as emojis, images, and posting frequency.
Moreover, we should also pay attention to developing an algorithm based on explainable AI to get more interpretable results.
Lastly, deployment of the proposed framework may give us valuable insights regarding its applicability.

7. Acknowledgements

Funding: This research received no external funding.

8. References

[1] Akdogan, Y. E., & Anbar, A. (2024). More than just sentiment: Using social, cognitive, and behavioural information of social media to predict stock markets with artificial intelligence and big data. *Borsa Istanbul Review, 24*(Suppl. 2), 61–82.

[2] Bello, A., Ng, S.-C., & Leung, M.-F. (2023). A BERT framework to sentiment analysis of tweets. *Sensors, 23*(1), 506.

[3] Deng, J., & Liu, Y. (2025). Research on sentiment analysis of online public opinion based on RoBERTa-BiLSTM-Attention model. *Applied Sciences, 15*(4), Article 2148.

[4] Frontiers in Artificial Intelligence. (2021). Sentiment analysis of students' feedback in MOOCs: A systematic literature review. *Frontiers in Artificial Intelligence, 4*, Article 728708.

[5] Kastrati, Z., et al. (2021). A deep learning sentiment analyser for social media comments in low-resource languages. *Electronics, 10*(10), Article 1133.

[6] Kohanová, D., et al. (2025). Social media behaviour and patterns of use among nursing students: A systematised review. *Nurse Education in Practice, 83*, Article 104277.

[7] Mahalakshmi, V., et al. (2024). Twitter sentiment analysis using conditional generative adversarial network. *International Journal of Cognitive Computing in Engineering, 5*, 161–169.

[8] Maurya, C. G., & Jha, S. K. (2024). Sentiment analysis: A hybrid approach on Twitter data. *Procedia Computer Science, 235*, 990–999.

[9] McGarry, K. (2023). Analyzing social media data using sentiment mining and bigram analysis for the recommendation of YouTube videos. *Information, 14*(7), Article 368.

[10] Nijhawan, T., Attigeri, G., & Ananthakrishna, T. (2022). Stress detection using natural language processing and machine learning over social interactions. *Journal of Big Data, 9*(1), Article 33.

[11] Rodríguez-Ibáñez, M., et al. (2023). A review on sentiment analysis from social media platforms. *Expert Systems with Applications, 223*, Article 119862.

[12] Smith-Mutegi, D., et al. (2025). Perceptions of STEM education and artificial intelligence: A Twitter (X) sentiment analysis. *International Journal of STEM Education, 12*(1), 9.

[13] Tafesse, W. (2022). Social networking sites use and college students' academic performance: Testing for an inverted U-shaped relationship using automated mobile app usage data. *International Journal of Educational Technology in Higher Education, 19*(1), 16.

[14] Valarmathi, B., et al. (2024). Sentiment analysis of Covid-19 Twitter data using deep learning algorithm. *Procedia Computer Science, 235*, 3397–3407.

[15] Vidyashree, K. P., et al. (2024). A tweet sentiment classification approach using an ensemble classifier. *International Journal of Cognitive Computing in Engineering, 5*, 170–177.

[16] Xu, Q. A., Chang, V., & Jayne, C. (2022). A systematic review of social media-based sentiment analysis: Emerging trends and challenges. *Decision Analytics Journal, 3*, Article 100073.

[17] Zhang, T., Yang, K., Ji, S., & Ananiadou, S. (2023). Emotion fusion for mental illness detection from social media: A survey. *Information Fusion, 92*, 231–246.