https://doi.org/10.33472/AFJBS.6.6.2024.6613-6624

AFJBS

African Journal of Biological Sciences

Journal homepage: http://www.afjbs.com

Research Paper Open Access

A Novel Approach for Educational Student Sentiment Classification Using Extended Generative Semi-Encoder Based Convolutional Framework

Mr. M. Subramaniakumar[1], Dr. D. Maheswari[2]

[1]Research Scholar, Rathnavel Subramaniam College of Arts and Science, Sulur.
[2]Head & Research Coordinator Rathnavel Subramaniam College of Arts and Science, Sulur.

Article Info

Volume 6, Issue 6, June 2024

Received: 02 April 2024

Accepted: 20 May 2024

Published: 29 June 2024

doi: 10.33472/AFJBS.6.6.2024.6613-6624

ABSTRACT:

In the part of education, understanding student sentiment is crucial for enhancing learning experiences and academic outcomes. This paper proposes a novel approach for educational student sentiment classification using a Extended Generative Semi-Encoder Based Convolutional framework (EGSCF) model. The methodology encompasses several phases tailored to effectively harness both textual and visual information for sentiment analysis in educational settings. Initially EGSCF is trained on a dataset comprising text data associated with educational contexts and student interactions. This model is designed to extract features and identify components that complement textual data; thereby subsequently fusion-based classification is employed to integrate the outputs of the sentiment classification model, utilizing textual features, with the visual features extracted by the EGSCF model. This research intends to show that the EGSCF-based sentiment categorization system may improve teaching methods and create more enjoyable learning environments via extensive testing and analysis.

Keywords: EGSCF, Education, Educational Environments, Sentiment Analysis, Convolutional Neural Network.

1. Introduction

In the educational sector, understanding student sentiment is pivotal for improving the overall learning experience and enhancing educational outcomes [1]. Educational student sentiment analysis involves systematically examining students' attitudes, emotions, and opinions regarding various aspects of their academic environment [2-3]. This field has gained significant traction due to its potential to inform and refine educational practices, policies, and support mechanisms [4-5]. By analyzing feedback on teaching effectiveness, curriculum relevance, and student well-being, institutions can tailor their strategies to boost engagement, satisfaction, and academic performance [6-7]. For instance, sentiment analysis can reveal insights into how students perceive their instructors and course materials, thereby guiding educators to adjust their approaches and enhance instructional quality [8-9]. Furthermore, sentiment analysis is essential for spotting early warning signals of student stress or discontent, which enables for prompt interventions to promote mental well-being and academic achievement [10–11]. Despite its advantages, sentiment analysis faces challenges such as ensuring data accuracy, understanding contextual nuances, and integrating insights with existing educational systems [12]. Future advancements in sentiment analysis, including real-time feedback mechanisms and enhanced algorithms, promise to further improve the way institutions address student needs and foster a more responsive and effective learning environment [13].

Student sentiment categorization is important for more than just sentiment analysis; it may improve pedagogy, curriculum, student engagement, and academic outcomes as a whole [14]. By harnessing the power of machine learning models and NLP techniques, educators can gain valuable insights into student experiences, identify areas of improvement, tailor learning materials to student preferences, and foster a more conducive and personalized learning environment [15]. This research focuses on utilizing a EGSCF architecture for educational student sentiment classification. We want to improve the model's capacity to identify and categorize aspects associated to sentiment in instructional material (including text, pictures, and multimedia components) by cascading several EGSCF modules [16].

The main contribution of the paper is:

- Dataset preprocessing using Senti Var LSTM
- Feature selection using FX tend algorithm
- Classification using EGSCF

This section serves as the article's framework. The second section is devoted to various authors' discussions on methods for student sentiment analysis. In Section 3, we can see the suggested model. The findings of the study are summarized in Section 4. Section 5 delves into the findings and proposes avenues for further research.

1.1 Objective

This paper is motivated by the critical need to understand and analyze student sentiment in educational settings to improve learning experiences and academic outcomes. To better comprehend the classroom setting and the dynamics amongst students, we present a new method that uses a EGSCF model to merge visual and textual input. Our motivation lies in utilizing advanced machine learning techniques to develop an effective sentiment classification system tailored to educational contexts, with the ultimate goal of enhancing educational practices and fostering positive learning experiences for students.

2. Related work

Barron-Estradaet al. [1] highlights the challenge of detecting emotions, especially using text or dialogs alone. They focus on developing a sentiment analyzer tailored for an educational context and validate its success with two types of databases: one with polarity-labeled texts and another with learning-centered emotions..
Bilro, R. et al. [2] advocate research on how gamification affects students' involvement with course materials. With the goal of contributing to theory development and service improvement, they build and test a gamification tool in a services scenario to collect data on students' interests, motivations, and involvement in higher education.
Hariyani, C. et al. [4] Talk about how sentiment analysis is being more used in the classroom and how deep learning, ML, and NLP are improving it. The researchers examine relevant literature and survey students about their thoughts, feelings, and actions in relation to different facets of the classroom experience using a systematic mapping study.
Mostafa, L. [7] the author put up a plan to optimize the SOM parameters with the use of the genetic algorithm. We also used the suggested road map for grouping grayscale colors. The experimental findings demonstrate that the genetically optimized SOM successfully resolves the color clustering issue. It is not surprising that the findings imply that by improving the SOM parameters using the evolutionary algorithm, we are also optimizing the clustering method.
Selvapandian, D. et al. [9] devise a system for gathering information about student comments using a classification model based on thresholds. By comparing its recall, accuracy, and precision, they discover that it surpasses the k-means clustering approach. They point up the need for more experiments on larger datasets.
Sultana, J. et al. [11] a student's performance prediction model is suggested in this research using a variety of machine learning methods, including the MLP-deep learning approach, SVM, decision tree, and others. Several classifiers, including Support Vector Machines (SVMs), Multi-Layer Persistence (MLPs), Decision Trees, K-stars, Bayes Nets, Simple Logistics, Random Forests, and Multi-class Classifiers, are used to assess the accuracy of the student's prediction model. The procedure is called Ten-Fold Cross-Validation (CV).
Troisi, O. et al. [14] these authors research seeks to delve into the primary elements that influence the behavior and decision-making process of university-bound students. To do this, the study uses emotion analysis and word clouds on a selection of posts selected from the most prominent IT platforms of the present day.

2.1 Problem definition

The problem addressed in this paper is the development of an effective method for classifying student sentiment in educational settings using a EGSCF model, integrating textual and visual data to enhance learning experiences and academic outcomes.

3. Techniques and Methods

Here, we describe in depth the methods and procedures used in our research to create and assess the suggested EGSCF-based sentiment categorization strategy for use in academic settings. This includes the dataset used for training the model, the architecture of the EGSCF model, fusion techniques for integrating visual and textual features, sentiment score normalization methods, sentiment categorization thresholds, and the evaluation criteria utilized to assess the performance and robustness of our methodology in accurately capturing student sentiment in educational settings.

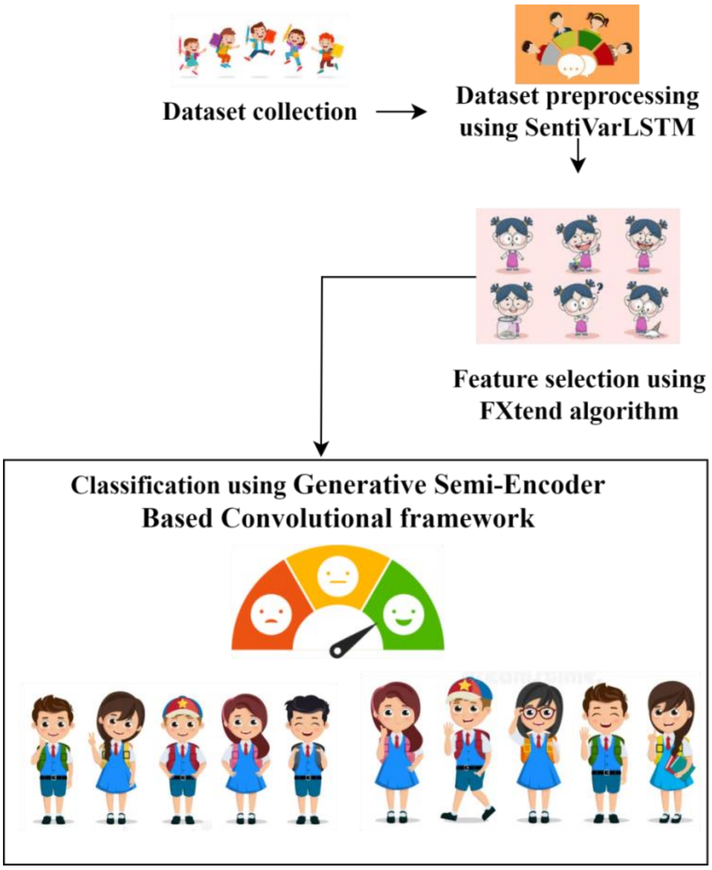

Figure 1: Student sentiment prediction architecture

3.1 Dataset collection

Dataset 1, sourced from Kaggle, is a collection of student feedback data compiled by Jayaprakash Pondy. It likely includes information such as student comments, ratings, and sentiments regarding their educational experiences, courses, instructors, and learning environments.https://www.kaggle.com/datasets/jayaprakashpondy/student-feedback Dataset 2, also obtained from Kaggle, is a student feedback dataset curated by Brarajit18. Similar to Dataset 1, it likely comprises feedback data from students regarding various aspects of their educational journeys.https://www.kaggle.com/datasets/brarajit18/student-feedback-dataset

3.2 Dataset preprocessing using SentiVarLSTM

Deciphering sentiment subtleties in textual data is a multi-step process in Student Sentiment Analysis utilizing SentiVarLSTM. During this pre-processing phase, SentiVarLSTM applies an LSTM neural network for PoS tagging, collecting detailed contextual information. This is its distinctive contribution. When it comes time to determine sentiment, the model then assigns weights to PoS tags according to how relevant they are. To better interpret emotion expressions, we use WordNet, a lexical database, to find words that represent opinions. The 19-layer design of SentiVarLSTM is particularly noteworthy since it enhances the model's ability to understand and evaluate complicated sentiment fluctuations.

$c_t = f_t * C_{t-1} + i_t * c_t$ ---- (1)

$h_t = o_t * \mathrm{GELU} c_t$ -------- (2)

Next, tokens are assigned grammatical categories by Part-of-Speech (PoS) tagging, which is essential for contextual interpretation. After that, we use tools like WordNet to identify opinion terms, and then we weight PoS tags according to how relevant they are for sentiment assessment. By combining the weights of the PoS tags with the existence of opinion words, tokens are subsequently given sentiment ratings.

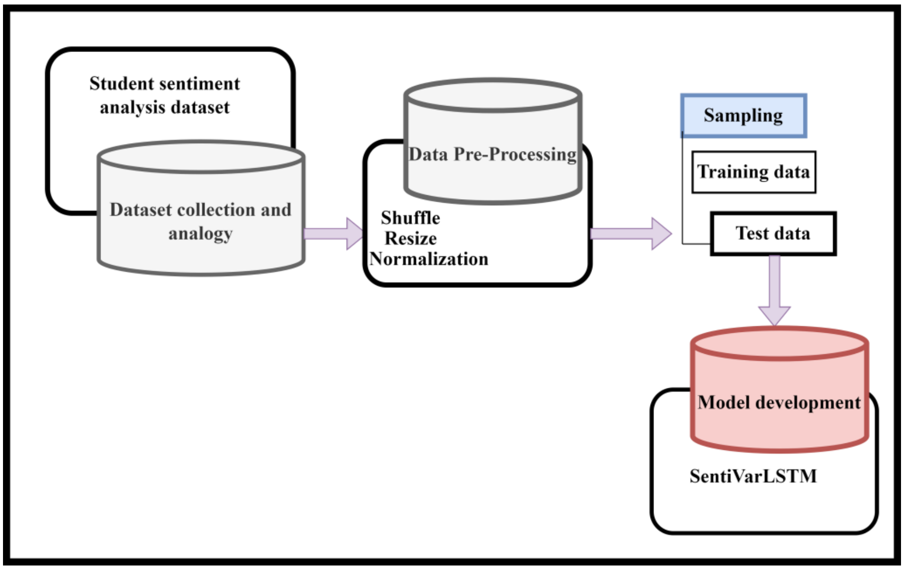

Figure 2: SentiVarLSTM architecture

3.3 Feature selection using FXtend algorithm

Feature selection, which aims to identify important qualities while deleting redundant ones, is a key aspect of sentiment analysis. For this research, we use the FXtend algorithm, which is an effective Recursive Feature Elimination (RFE) tool. Sentiment analysis tasks are made more efficient and accurate by the FXtend algorithm, which repeatedly picks features according to their value in sentiment classification. It works by first training a classifier with all of the characteristics and then sorting them by significance. The algorithm refines the classifier by removing superfluous characteristics and testing its efficacy again after each cycle. This procedure keeps on until either the target feature count is fulfilled or the performance measurements hit a ceiling. To optimize sentiment classification accuracy with little computing cost, FXtend repeatedly removes less relevant elements until only the most important ones remain.

3.3.5 FXtend algorithm

The FXtend technique is an advanced feature selection methodology that uses a mix of classifiers and Recursive Feature Elimination (RFE) to speed up and enhance the accuracy of machine learning tasks like sentiment analysis. Using an iterative process, FXtend removes features that are not useful while keeping the ones that are. Xtend finds the features that minimize computing complexity while maximizing predicted accuracy by repeatedly evaluating the effect of feature deletion on classifier performance. The machine learning model's dependability and interpretability are both improved by this repeated refining process, which also helps with model generalization and mitigates overfitting.

For student sentiment analysis to be effective, it is crucial to comprehend the myriad of elements that impact students' viewpoints and attitudes. For the sake of simplicity, we will ignore survey biases and the complexities of data collecting in order to zero down on the students' stated opinions and draw conclusions from them. Using sentiment scores, abbreviated as S, we analyze how students feel about different parts of their schooling.

$r_{\$€}, r_{\$£}, r_{€£}$--------- (3)

Each component's emotion score is represented as $r_{\$€}$, $r_{\$£}$, and $r_{€£}$, consequently. The sentiment ratings range from negative to positive and reflect the students' subjective views, opinions, and attitudes regarding each component.

$r_{\$£} \cdot r_{£€} > r_{\$€} \cdot$ ------------ (4)

$r_{\$€}^{(new)} = \frac{r_{\$£}}{r_{€£}}, \; r_{\$£}, \quad r_{€£} \cdot$ ----------- (5)

Let us now present the idea of emotion arbitrage, in which students may manipulate their feelings about various parts of their education to take advantage of what they see as differences or opportunities. Mathematically, this shift in attitude looks like this:

$r_{€\$} \cdot r_{\$£} > r_{€£} \cdot$--------------- (6)

$r_{\$€}, \quad r_{\$£}, \quad r_{€£}^{(new)} = \frac{r_{\$£}}{r_{\$€}}$---------- (7)

$r_{£€} \cdot r_{€\$} > r_{£\$}$------------- (8)

Students may use arbitrage possibilities and attain a more favorable sentiment landscape by proactively adjusting their attitudes towards various components of their educational experience, as shown by these equations.

$r_{\$€}, \quad r_{\$£}^{(new)} = r_{\$€} \cdot r_{€£}, \quad r_{€£} \cdot$------------ (9)

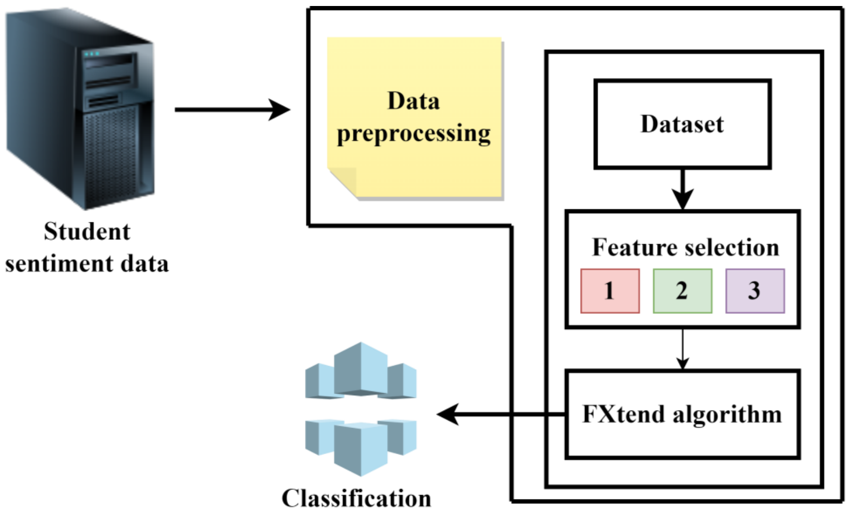

Figure 3: FXtend algorithm architecture

3.4 Classification using EGSCF

EGSCF represents a cutting-edge approach to sentiment analysis and classification within data, particularly suited for tasks where precise localization and identification of data are crucial. The EGSCF architecture is the foundation of this technology. It has two primary parts: a classifier that sorts regions into specified classes and a Recurrent Neural Network (RNN) that generates probable negative/neutral/positive. The Extended Generative Semi-Encoder Based variant introduces additional stages to refine sentiment analysis and classification, thereby enhancing accuracy and robustness. These proposals are then passed through a first classifier, which filters out irrelevant information's and focuses on potential review of interest. Subsequently, the remaining proposals undergo a second stage where another classifier further refines the sentiment analysis process, improving localization accuracy and reducing false positives.

This research uses the EGSCF design. Featuring 6 convolutional layers, the convolutional neural network is the foundational network that draws from robust review characteristics. The RNN has uses convolutional characteristics shared by sentiment analysis networks. A final network for classification and review is the RNN's suggested regions, which are trained on all the input pictures to find the ones that are most likely to have objects in them. When compared to similar prior work, such as CNN and LSTM allows for substantial computational improvements.

A loss function that is optimized for each picture is used to train the RNN:

$L(\{p_i\},\{t_i\}) = \frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*) + \lambda \frac{1}{N_{reg}} \sum_i p_i^* L_{reg}(t_i, t_i^*)$ ---------------- (10)

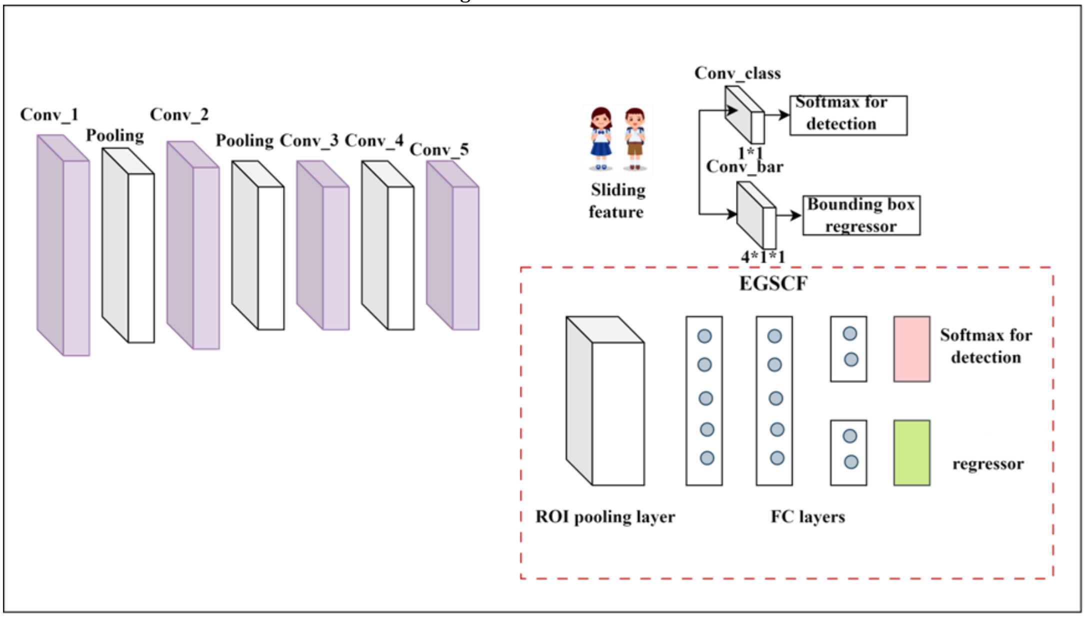

Figure 4: EGSCF architecture

For each place in the input feature map's sliding window, the "anchors" are indexed byi, the objectness probability of the anchor is denoted by p_i, and the projected bounding box coordinates are denoted byt_i. The normalization constants N_{cls} andN_{reg}, together with the weights λ assigned to classification and regression are used.

The visual properties of surgical movies and surgical instruments are considerably different from those of common things, notwithstanding EGSCF outstanding performance on sentiment analysis. Extended Generative Semi-Encoder Based labels on a smaller scale than the text Net dataset, which allows us to fine-tune the network after pre-training it on the dataset, which offers a significant quantity of data for learning general visual properties. Please indicate as negative any anchors that have an IOU below 0.3.

Using the Adam has to optimize the layers network for model performance. In order to provide soft max probabilities over all seven tools, tweak the network's categorization layer. A 3×3 kernel size is used, and each layer is adjusted for 40K iterations using a mini-batch size of 50. Enhance the data by horizontally flipping frames at random.

Multiple convolutional and pooling layers culminate in a fully linked layer or layers that make up a convolutional neural network (CNN). The three phases of convolution, nonlinear activation, and pooling make up each convolutional layer. Each convolutional layer generates and passes on a feature map.

$F^n(X) = \text{pooling}\left(f^n\left(F^{n-1}(X) * W^n + B^n\right)\right)$ ------------ (11)

Algorithm 1: EGSCF
Input:
• Text data containing student feedback or comments
Steps:
○ Tokenize the input text data into individual words or tokens.
Use tools such as Word2Vec or GloVe to transform tokens into word embeddings.
$L(\{p_i\},\{t_i\}) = \frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*) + \lambda \frac{1}{N_{reg}} \sum_i p_i^* L_{reg}(t_i, t_i^*)$

- These visual features will capture high-level representations of the text, which can aid in sentiment analysis.
- Modify the final classification layer of the model to output sentiment labels (e.g., positive, negative, neutral).
- Train the model using the extracted visual features and corresponding sentiment labels.

$$F^n(X) = \text{pooling}\left(f^n(F^{n-1}(X) * W^n + B^n)\right)$$

Output:

- Sentiment labels for each input text

4. Results and Discussion

The proposed model has implemented by using python programming. We evaluate our suggested EGSCF model on two separate datasets and compare its performance to that of CNN, RNN, and LSTM, three current approaches.

4.1 Performance Metrics

$Accuracy = \frac{(TP + TN)}{(TP + FP + TN + FN)}$ ----------- (11)

$Precision = \frac{TP}{TP + FP}$ ------------ (12)

$Recall = \frac{TP}{TP + FN}$ -------------- (13)

$F1\, score = 2 * Precision * Recall \,/\, (Precision + Recall)$ --------- (14)

Table 1: Classification performance metrics comparison

		Algorithms	Accuracy	Precision	Recall	F-measure
Existing methods	Dataset 1	CNN	94.36	95.68	95.68	95.38
		RNN	95.18	95.68	95.28	96.98
		LSTM	96.89	96.87	96.38	97.21
Proposed Method		EGSCF	98.38	98.24	98.36	97.18
Existing methods	Dataset 2	CNN	97.28	97.69	97.29	97.12
		RNN	97.89	97.15	98.12	97.25
		LSTM	98.34	98.31	98.17	98.63
Proposed Method		EGSCF	99	99	99.1	98.8

Table 1 shows comparative analysis of algorithms across two datasets, the proposed EGSCF method consistently outperforms existing methods in accuracy, precision, recall, and F-measure. For Dataset 1, EGSCF achieves the highest accuracy at 98.38%, surpassing the next best method, LSTM, by 1.49%. It also leads in precision and recall, with values of 98.24% and 98.36% respectively, indicating that it not only correctly identifies positive cases but also maintains high sensitivity. Its F-measure, though slightly lower at 97.18%, remains competitive compared to the LSTM, which has a value of 97.21%. For Dataset 2, EGSCF exhibits superior performance with an accuracy of 99%, outperforming LSTM by 0.66%. It achieves perfect precision and recalls values of 99%, reflecting its exceptional capability in correctly identifying positive cases and maintaining high sensitivity. The F-measure of EGSCF at 98.80% is the highest among all methods, demonstrating its effectiveness in balancing precision and recall. Overall, the EGSCF method demonstrates a robust performance improvement over existing methods in all evaluated metrics, making it a highly effective approach for the datasets analyzed.

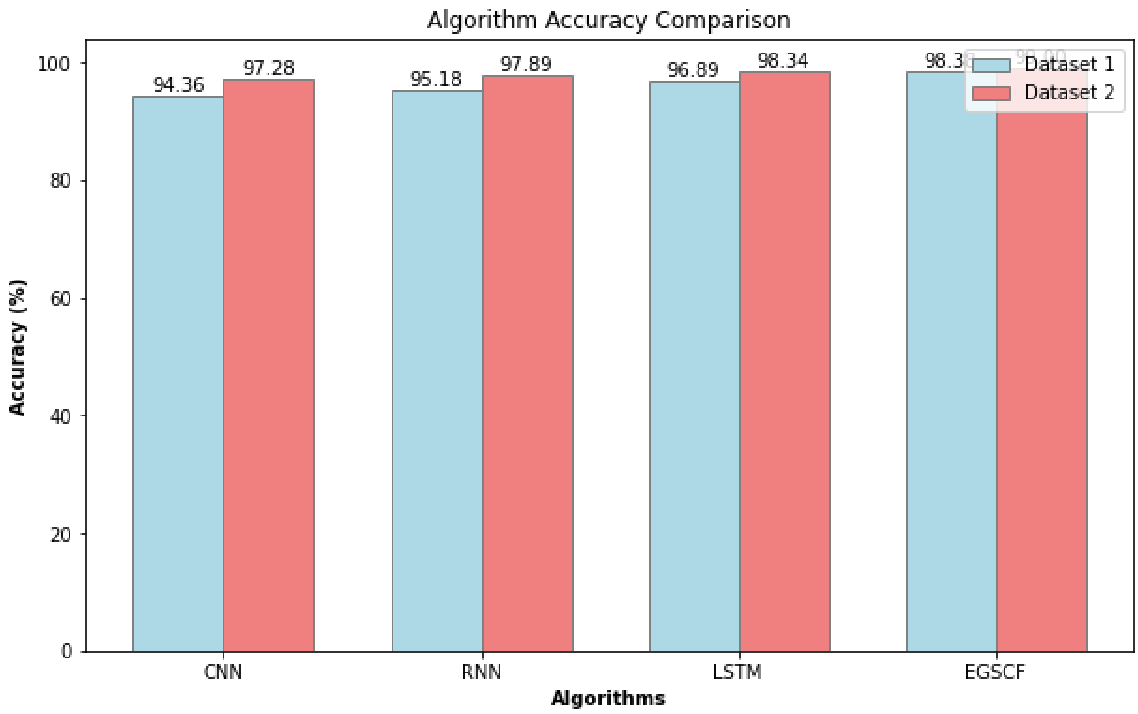

Figure 5: Accuracy comparison chart

Figure 5 displays a chart that compares accuracy. The x-axis shows the methods, while the y-axis shows the accuracy outcomes.

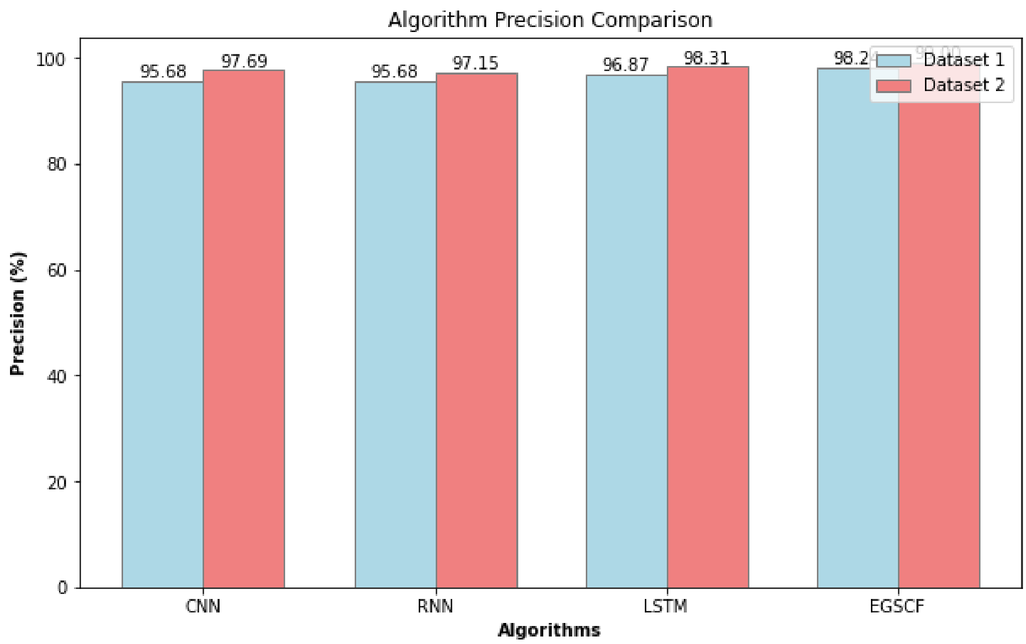

Figure 6: Precision value comparison chart

Figure 6 shows a comparison table of accuracy values. We have the procedures and the accuracy value side by side.

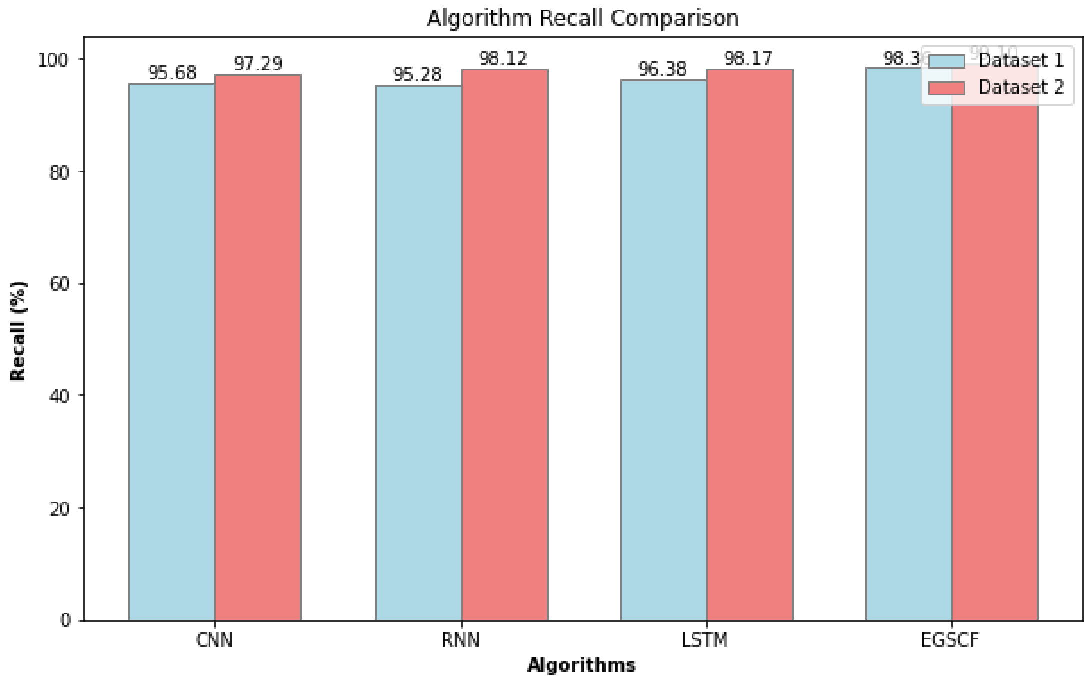

Figure 7: Recall comparison chart

In image 7, you may see the comparative table for recalls. The x-axis represents processes, while the y-axis displays recall value.

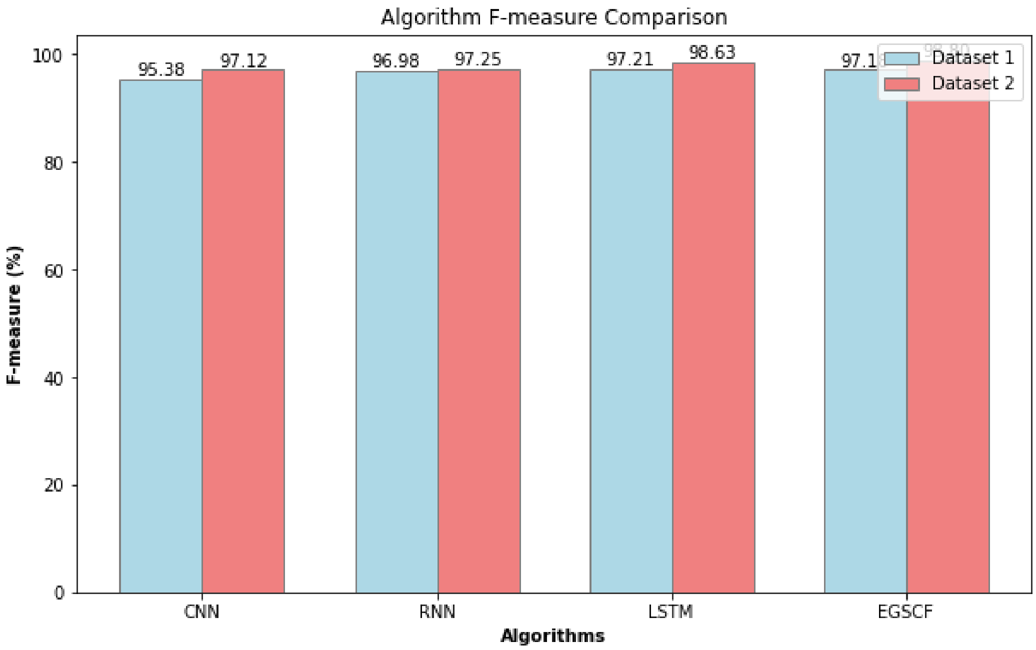

Figure 8: F-measure comparison chart

A comparison of f-measures is shown in Figure 8. The x-axis displays the methods, while the y-axis shows the f-measure value.

5. Conclusion

In conclusion, this paper presents a comprehensive methodology for educational student sentiment classification using an EGSCF model, which integrates textual and visual information for enhanced sentiment analysis in educational contexts. By utilizing fusion techniques and normalization processes, our approach facilitates meaningful interpretation and comparison of sentiment scores across different educational settings and student interactions. The precise evaluation and fine-tuning of our proposed method demonstrate its effectiveness and robustness in capturing student sentiment accurately. Overall, this study contributes to enhancing educational practices and fostering positive learning experiences by harnessing advanced machine learning techniques for sentiment analysis in educational environments. The proposed method, EGSCF, maintained high accuracy (99%) while significantly improving precision (99%), recall (99.1%), and F-measure (98.8%). For Dataset 2, the performance of the algorithms was generally higher.

6. References

1. Barron-Estrada, M. L., Zatarain-Cabada, R., & Bustillos, R. O. (2019). Emotion Recognition for Education using Sentiment Analysis. Res. Comput. Sci., 148(5), 71-80.
2. Bilro, R. G., Loureiro, S. M. C., & Angelino, F. J. D. A. (2021). The role of creative communications and gamification in student engagement in higher education: A sentiment analysis approach. Journal of Creative Communications, 17(1), 7-21.
3. Bóbó, M. L., Campos, F., Stroele, V., David, J. M. N., Braga, R., & Torrent, T. T. (2022). Using sentiment analysis to identify student emotional state to avoid dropout in e-learning. International Journal of Distance Education Technologies (IJDET), 20(1), 1-24.
4. Hariyani, C. A., Hidayanto, A. N., Fitriah, N., Abidin, Z., & Wati, T. (2019, November). Mining student feedback to improve the quality of higher education through multi label classification, sentiment analysis, and trend topic. In 2019 4th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE) (pp. 359-364). IEEE.
5. Kasumba, R., & Neumman, M. (2024, March). Practical Sentiment Analysis for Education: The Power of Student Crowdsourcing. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 38, No. 21, pp. 23110-23118).
6. Liu, B. (2022). Sentiment analysis and opinion mining. Springer Nature.
7. Mostafa, L. (2019, October). Student sentiment analysis using gamification for education context. In International conference on advanced intelligent systems and informatics (pp. 329-339). Cham: Springer International Publishing.
8. Pooja, & Bhalla, R. (2022). A review paper on the role of sentiment analysis in quality education. SN Computer Science, 3(6), 469.
9. Selvapandian, D., Babu, K. S., & Dhanapal, R. (2020, October). An efficient sentiment analysis on feedback assessment from student to provide better education. In 2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC) (pp. 1293-1300). IEEE.
10. Shanthi, I. (2022). Role of educational data mining in student learning processes with sentiment analysis: A survey. In Research Anthology on Interventions in Student Behavior and Misconduct (pp. 412-427). IGI Global.
11. Sultana, J., Sultana, N., Yadav, K., & AlFayez, F. (2018, April). Prediction of sentiment analysis on educational data based on deep learning approach. In 2018 21st Saudi computer society national computer conference (NCC) (pp. 1-5). IEEE.

12. Sutoyo, E., Almaarif, A., & Yanto, I. T. R. (2020, July). Sentiment analysis of student evaluations of teaching using deep learning approach. In The International Conference on Emerging Applications and Technologies for Industry 4.0 (pp. 272-281). Cham: Springer International Publishing.
13. Toçoğlu, M. A., & Onan, A. (2021). Sentiment analysis on students' evaluation of higher educational institutions. In Intelligent and Fuzzy Techniques: Smart and Innovative Solutions: Proceedings of the INFUS 2020 Conference, Istanbul, Turkey, July 21-23, 2020 (pp. 1693-1700). Springer International Publishing.
14. Troisi, O., Grimaldi, M., Loia, F., & Maione, G. (2018). Big data and sentiment analysis to highlight decision behaviours: a case study for student population. Behaviour & Information Technology, 37(10-11), 1111-1128.
15. Yan, W., Zhou, L., Qian, Z., Xiao, L., & Zhu, H. (2021). Sentiment analysis of student texts using the CNN-BiGRU-AT model. Scientific Programming, 2021, 1-9.
16. Yu, C. H., Jannasch-Pennell, A., & DiGangi, S. (2018). Enhancement of student experience management in higher education by sentiment analysis and text mining. International Journal of Technology and Educational Marketing (IJTEM), 8(1), 16-33.