Ingeniería y
Competitividad
Vol 26 (2) 2024; doi: 10.25100/iyc.v26i2.13759

Vol 26 (2)

ISSN
0123-3033
e- 2027-8284

Original Research

Towards an improved of teaching practice using Sentiment Analysis in Student Evaluation

Hacia una mejora en la práctica docente utilizando el Análisis de Sentimiento en la Evaluación de Estudiantes

Jefferson A. Peña-Torres
Pontificia Universidad Javeriana, Cali, Colombia

Abstract

Student Evaluation of Teaching (SET) serves as an ad hoc way of assessing teaching effectiveness within higher education institutions. This paper introduces an approach to analyzing sentiments expressed in SET comments using a Large Language Model (LLM). By employing natural language processing techniques, the polarity conveyed by students upon course completion is extracted and analyzed, aiming to furnish educators and stakeholders with valuable insights into teaching quality and areas for improvement in teaching practice. This study showcases the effectiveness of LLMs in sentiment analysis of comments, underscoring their potential to enhance the evaluation process. The development of a prototype tool, collaborative labeling of end-of-course evaluations, and a comparison with LLM-based labeling are experimentally explored. Subsequently, the implications for educational institutions are discussed, and future research directions in this domain are proposed.

Resumen

La evaluación del estudiante sobre la enseñanza (SET) es una forma ad hoc de evaluar la efectividad docente en instituciones de educación superior. En este articulo presenta un enfoque para analizar los sentimientos expresados en los comentarios de SET utilizando un modelo de lenguaje grande (LLM). Al emplear técnicas de procesamiento de lenguaje natural, se extrae y analiza la polaridad expresada por los estudiantes al finalizar el curso, con el objetivo de proporcionar a educadores e interesados ideas valiosas sobre la calidad de la enseñanza y elementos a mejorar de la práctica docente. Este estudio demuestra la efectividad de los LLM en el análisis de sentimientos de los comentarios, resaltando su potencial para mejorar el proceso de evaluación. Se experimenta con el desarrollo de una herramienta prototipo, el etiquetado de conjunto de evaluaciones al final del curso de forma colaborativa y se compara con el etiquetado realizado con un LLM. Luego se discuten las implicaciones para las instituciones educativas y se proponen futuras direcciones para la investigación en este ámbito.

Keywords: sentiment analysis, Student feedback, Large Language Models

Palabras clave: análisis de sentimientos, comentarios de los estudiantes, modelos de lenguaje grandes

How to cite?

Peña-torres, J.A. Towards an improved of teaching practice using Sentiment Analysis in Student Evaluation. Ingeniería y Competitividad, 2024, 26(2) e-21013759

https://doi.org/10.25100/iyc.v26i2.13759

Recibido: 19-10-23 Aceptado: 20-04-24

Correspondence:
jeffersonamado.pena@javerianacali.edu.co

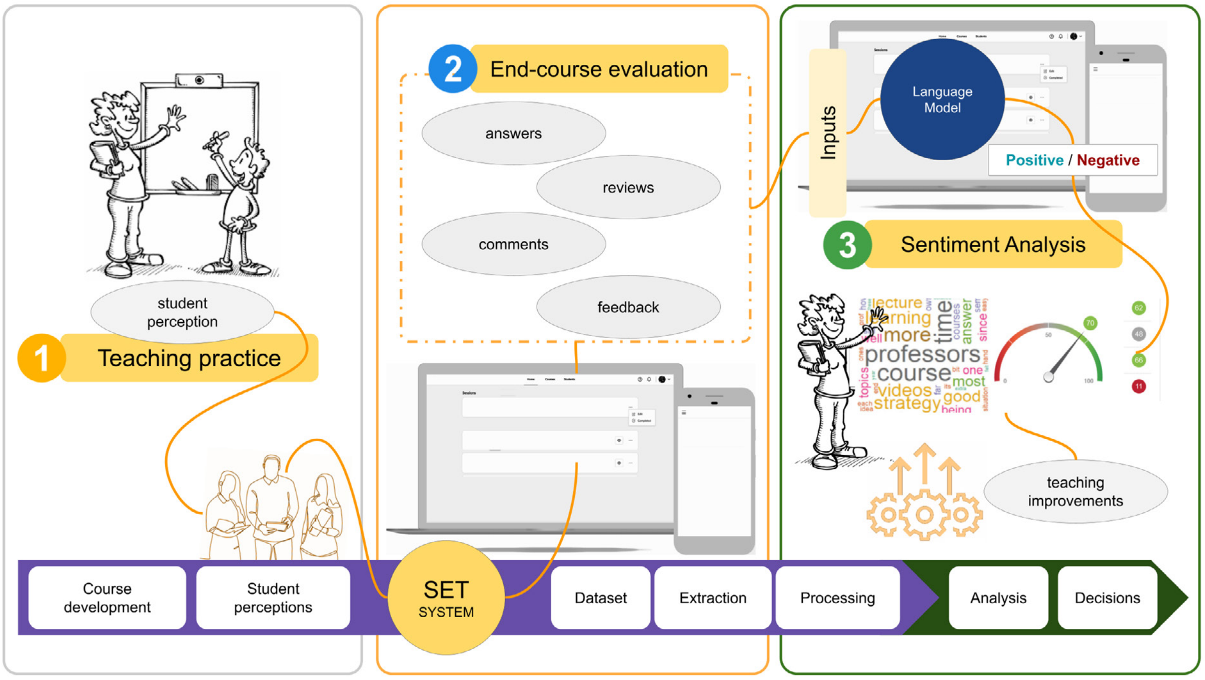
Conflict of interest: none declared

Why was it conducted?:
"Towards an Improved Teaching Practice Using Sentiment Analysis in Student Evaluation" was carried out with the goal of enhancing teaching practices through sentiment analysis of student evaluations at the end of the course. It was identified that the polarity expressed by students at the course's conclusion is an underutilized resource and that a large language model (LLM) can effectively capture the underlying perceptions and emotions that students have regarding teaching practices. In this context, sentiment analysis allows for an automated understanding of student comments, providing valuable insights that can be used to adjust and improve teaching methodologies.

What were the most relevant results?
The sentiment analysis models used achieved polarity classification similarly to human assessments. Using a dataset labeled through crowdsourcing, the best model demonstrated 93% accuracy.
Negative comments tended to be longer compared to positive or neutral ones. Positive comments were brief and direct, like "good course," whereas negative comments contained more words to express dissatisfaction or displeasure.
The analysis of the polarity distribution of comments showed that neutral comments were not recognized, suggesting a possible bias or imbalance in data collection.
The ASET system prototype proved to be robust and adaptable, allowing the creation, updating, and deployment of new versions and machine learning models for sentiment analysis. The microservices-based architecture improved the system's resilience and reliability.

What do these results contribute?
The implementation of large language models for sentiment analysis in student feedback offers an effective tool for educators, enabling automatic preliminary analysis that can help improve teaching practices more quickly and efficiently. The analysis results can provide teachers with valuable insights into specific aspects of their teaching that need adjustments, based on the polarity of student comments at the end of the course.

Graphical Abstract



doi: 10.25100/iyc.v26i1.13759

Introduction

Student evaluation of teaching (SET) is an ad-hoc way of assessing teaching effectiveness in higher education institutions (1). Feedback from students can assist teachers in comprehending the students behaviors and improve teaching practice (2,3). Receiving feedback can highlight different issues that arose during the course, related to the material, readings, course tools and even teaching practice. For this reason, educational institutions, centers, and faculty staff rely on tools to gather information and aid instructors in supporting student learning (4–6).

Despite end-course systems showing surveys with dichotomous, closed, or likert questions being quickly processed, the same cannot be said for open-ended questions or free-text responses, which often demand more time and effort for comprehensive analysis. This poses a challenge for teachers who must decipher student feedback, opinions, and comments from evaluation software, resulting in time lost if the lecturer should to understand the text. In addition, it leaves it up to you to assign the polarity of the comments; It is the teacher who decides if they were positive or negative (5,7).

The reason behind using student feedback for the improvement of teaching is to give a comprehensive view of teaching from the students' perspective, which might result in valuable information or data for teachers (8). Interest in student perception is increasingly becoming a prominent method for evaluating multiple elements of the academic context, teaching practices, student engagement, and even the achievement of learning objectives (9–12). Collecting short feedback messages at the end of a course offers numerous benefits for both the lecturer and future students, including adjustments to teaching practices, slides, readings, activities, student behavior and also provides to lecturer an summarized overview of student opinions.

Recently, artificial intelligence (AI) has made remarkable advancements across multiple domains, stretching the limits of what was once deemed impossible and opening up a new era of exploration and innovation. The large linguistic models (LLMs) can be considered a relevant development for the Natural Processing Language (NLP) because these can perform a wide variety of tasks, such as summarizing, synthesizing, translating content, and analyzing the sentiment of sentences, comments and reviews. LLMs use a transformative architecture and have been incorporated into several popular tools like Google's Bidirectional Encoder Representations from Transformers (BERT) (13), OpenAI's Generative Pre-trained Transformer (GPT)(14) and among others not so well known.

Addressing the situation of teachers reading, understanding and evaluating students feedback comments nowadays can be automated using sentiment analysis, opinion mining and other approaches (15–18). In particular, the analysis of the comments at the end of the course, which is when the teacher should receive information about his or her practice in a way that allows him or her to modify or adjust it for a next iteration.

The paper aims to classify the polarity expressed in teaching feedback at the end of the course using a Large Linguistic Model (LLM), present word clouds, statistics, trends and other analyzes from the comments through a prototype tool. The sentiment analysis of Student evaluation of teaching (SET) can be considered as a strategy to improve teaching practice. The analysis of student comments allows us to identify elements of teaching practice that caused difficulties for students. The rest of the paper has been organized as follows. Related research is presented in section 2. The research method is presented in Section 3. Followed by results and discussion in Section 4. To finish, conclusions and future work are outlined in Section 5.

Related Work

Sentiment Analysis (SA) for education can vary depending on the context and multiple tasks that can be addressed. Commonly, sentiments surrounding education can range from positive to negative. In this context, the emotions affect the motivation and the outcome of the learning process (19–21). The role of Artificial Intelligence (AI) in sentiment analysis is crucial for aiding

doi: 10.25100/iyc.v26i1.13759

specific processes and tasks related to student comments (22). Modern AI-powered tools excel at discovering polarity, classifying, and predicting emotions within unlabeled sets of comments (23). Numerous studies have explored sentiment analysis (SA) in education, with a substantial focus on e-learning, classroom learning, and daily sessions and real-time interventions (25, 24).

Similarly, there is an interest in feedback and student comments due to their potential to offer valuable insights into student learning behaviors and to enhance teaching practices. For example, the student perception of teacher feedback (26), the relationship between feedback and learning motivation(27), the feedback as part of student satisfaction (28), sentiment analysis of student feedback (29–32) and this last using techniques as Support Vector Machines(33), Naive Bayes (34), dictionaries, lexicons (29), and more recently, Deep Learning-based models(35) and Language Learning Models (LLMs)(16,17,36,37). Making important advances in data analysis within the educational context aimed at understanding, automating and improving the learning experience.

Experience that is then evaluated by students at the end of the course. For a century, numerous institutions around the world have implemented Student Evaluations of Teaching (SET) systems for storing, retrieving, and processing student evaluations of courses and teaching (38). Course and teaching evaluations tools designed to present surveys, evaluate academic performance and that have been maturing until today as a study object to improve the teaching (39–41).

Modern course evaluation SET tools incorporate Likert scale, numeric and open-ended questions, allowing students to provide textual feedback, creating a notable gap in the comprehensive analysis of sentiments and opinions after the course is completed (42–44). In this paper, we propose the sentiment analysis of textual end-of-course feedback using a LLM-based approach to enhance the teaching and learning processes.

Methodology

In Fig. 1 the schematic process carried out in our sentiment analysis for end-course students feedback. While drawing inspiration from existing frameworks like (45), the approach suits our own requirements as the customizations in preprocessing techniques, dataset annotation procedures, model evaluation methodologies, and preprocessing new teacher inputs. Student comments are sourced from official SET systems, then they are preprocessed before being fed into the Large Language Model (LLM) operating within the sentiment analysis system. The analysis facilitates the generation of visualizations such as word clouds and other graphical representations, providing to the teacher a comprehensive understanding of sentiment patterns and trends within the student feedback.

In this paper, we consider a quantitative and qualitative approach. Quantitatively we employ performance measures to assess three pre-trained LLM-based models, this provides valuable insights into potential of Machine Learning models. Qualitatively, we describe a case study and the characteristics of our prototype tool.

doi: 10.25100/iyc.v26i1.13759

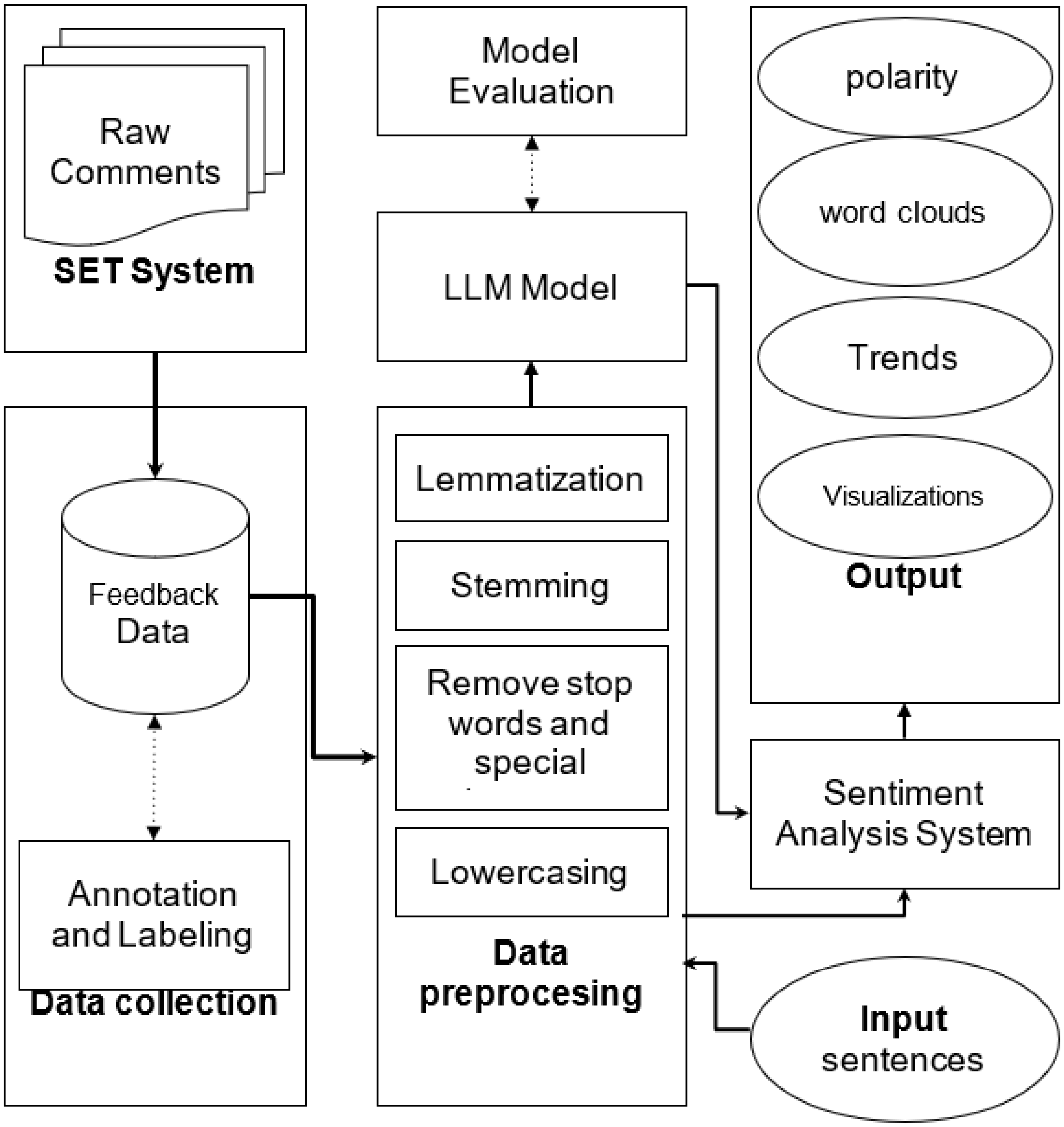


Figure 1. Sentiment Analysis System for End-course student evaluation. Source:Authors

Dataset description: Our dataset comprises 365 Spanish comments collected between 2018 and 2023 encompassing feedback messages from end-of-course evaluation at Universidad del Valle and report from Pontificia Universidad Javeriana, both at Cali, Colombia. Due to scarcity of spanish labeled dataset, the dataset was annotated with positive, neutral and negative using crowdsourcing approach. Participants chose a class for students' comments and the more frequent orientation was established as a label. Table I. shows a few annotated examples. Dataset distribution was 60%, 32% and 8% as Positive, Negative and Neutral respectively.

doi: 10.25100/iyc.v26i1.13759

Table. 1. Sample Feedback Comments in dataset. Source: Authors

No.	Sentence	Label
1	El profesor manejó una buena metodología en el curso	Positive
2	Agregar más uso de terminal de forma práctica en clase.	Neutral
3	Sería mejor si el curso se dictará de manera presencial.	Positive
4	Contestar el correo y tener otros medios para comunicarse	Negative
5	Nirguna, been curso y professor.	Positive

Preprocessing: Student feedback is unstructured texts and to analyze them a preprocessing stage is needed. At this stage, the text was divided into sentences and junk elements like stop words, numerical values, and certain special characters were removed to reduce noise in the data set. Using NLTK (46) and spaCy (47) for these tasks, all words were transformers to lower case and stemming and lemmatization were optional in experimentation and model evaluation stage. Additionally, person names and feedback messages such as "ninguno", "sin comentarios", "sin observaciones" were manually removed because they had no relevant content.

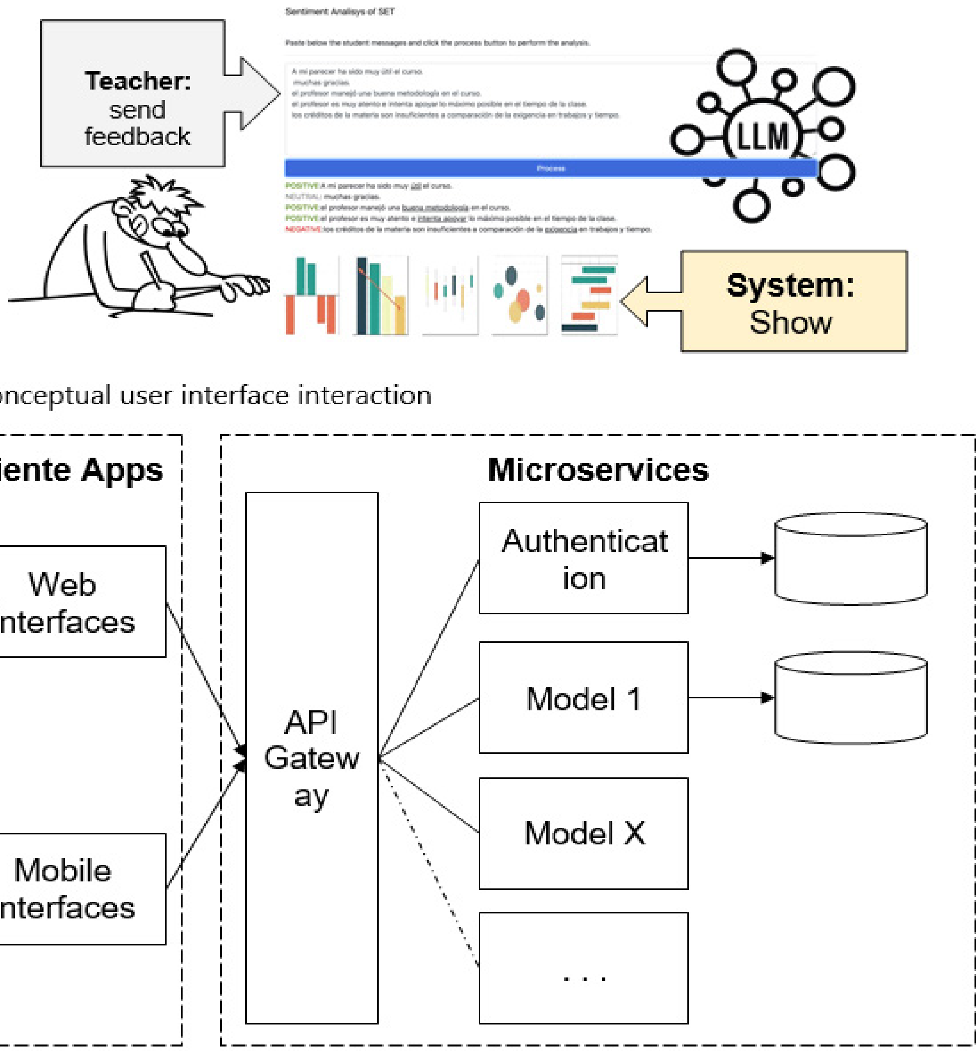
Pre Trained LLM: After the preprocessing the dataset three deep pretrained models for spanish language were used: i) Py-sentimiento, a Robertuito and RoBERTa-based model trained in spanish tweets (48–50), ii) a customized version of VADER (Valence Aware Dictionary and sEntiment Reasoner) wich translate comments and classify the text polarity (51) and iii) Distilbert-based model, a small, fast and light Transformer by distilling BERT base (52,53).

Fine-tuning plays a critical role in the training of models with initial parameters. However, there are instances where these parameters need modification to incorporate or adapt to specific tools or contexts. The Hugging Face Transformers library offers pretrained models, and within it, the TrainingArguments function, which furnishes an intuitive and user-friendly interface for managing key aspects of the training process (54,55).

Tool design and development: The implementation of a tool to support improving teaching and learning using sentiment analysis entails a comprehensive approach focused on leveraging advanced technologies and software development methodologies.As part of study approach Rapid Application Development (RAD) methodology was adopted for designing and development of ASETool, focusing on the rapid build of prototypes focused on the end user. With a containerized-microservices architecture, ASETool separate services for data preprocessing, sentiment analysis, and user authentication allowing scalability, portability and other advantages.

The foundational user interface concept of the proposed system is delineated in Fig. 2a, while the architectural depiction, featuring pivotal elements and model integration, is presented in Fig. 2b.

doi: 10.25100/iyc.v26i1.13759



Sentiment Analisys of SET

Paste below the student messages and click the process button to perform the analysis.

A mi parecer ha sido muy útil el curso.
muchas gracias.
el profesor manejó una buena metodología en el curso.
el profesor es muy atento e intenta apoyar lo máximo posible en el tiempo de la clase.
los créditos de la materia son insuficientes a comparación de la exigencia en trabajos y tiempo.

POSITIVE A mi parecer ha sido muy útil el curso.
NEUTRAL muchas gracias.
POSITIVE el profesor manejó una buena metodología en el curso.
POSITIVE el profesor es muy atento e intenta apoyar lo máximo posible en el tiempo de la clase.
NEGATIVE los créditos de la materia son insuficientes a comparación de la exigencia en trabajos y tiempo.

(a) Conceptual user interface interaction

(b) Conceptual architecture design

Figure 2. Overview of ASET tool design. Source:Authors

ASET tool presents a user-friendly and web-based design for ease of use by both technical and non-technical educators. With the interface interaction educators possess the capability to submit multiple comments via the interface and subsequently conduct textual analysis. The architectural design allows developers and stakeholders the opportunity to seamlessly integrate novel models and features into the existing architecture, thereby ensuring a harmonious and adaptable system.

Results and discussion

In the absence of a standardized and labeled end-of-course assessment data set, crowdsourcing makes the labeling possible rapidly and cost-effectively (56,57). As part of the approach, the completely labeled dataset was used as a baseline to compare the classification performance from each model following the key idea that crowdsourced sentiment is more accurate (57). The accuracy metric was employed to evaluate the correspondence between the predictions from each model and the crowdsourced labels. In this context, accuracy indicates how well the model can discern the polarity of the texts and whether the predictions are equal to human judgment.

doi: 10.25100/iyc.v26i1.13759

Table 2. Performance Analysis of Analysis of textual feedback. Source:Authors

Model	**Accuracy**
Pysentimiento	.85
Distilbert	.93
Vader	.79

Table No. 2. show the performance of models. For the real-world dataset used in this study the models were consistent for crowdsourcing labels validating the viability for sentiment analysis in end-of-course evaluations. In fact, two of three models considered present an accuracy major than 80%. Coinciding with other studies where popular and non-free LLMs such as GPT (14) have demonstrated the potential for student feedback sentiment analysis achieving relevant performance (16).

Large language models (LLMs) have the potential to automate analysis and improve information richness to improve teaching. LLMs can influence teaching practice based on student's messages, which can be shaped by our conceptual approach.

Accuracy scores indicate that the LLM predictions are similar to those made by a human. The results show that the multilingual-distilbert model adapted for sentiment analysis can classify the polarity of a text in a way that is similar to a person, making these tools can be used for automatic analysis. In particular, for this study, the positive polarity from end-course evaluation raises can reveal a course satisfaction and the teacher could continue with its methodology and practice. However, It is expected that students do not feel satisfied, write negative or neutral comments that can be automatically analyzed. Fig 3. shows the amount of the feedback comments labeled from crowdsourcing (Crowd) and automatically using Vader (M1), Distilbert (M2) and Pysentimiento (M3).

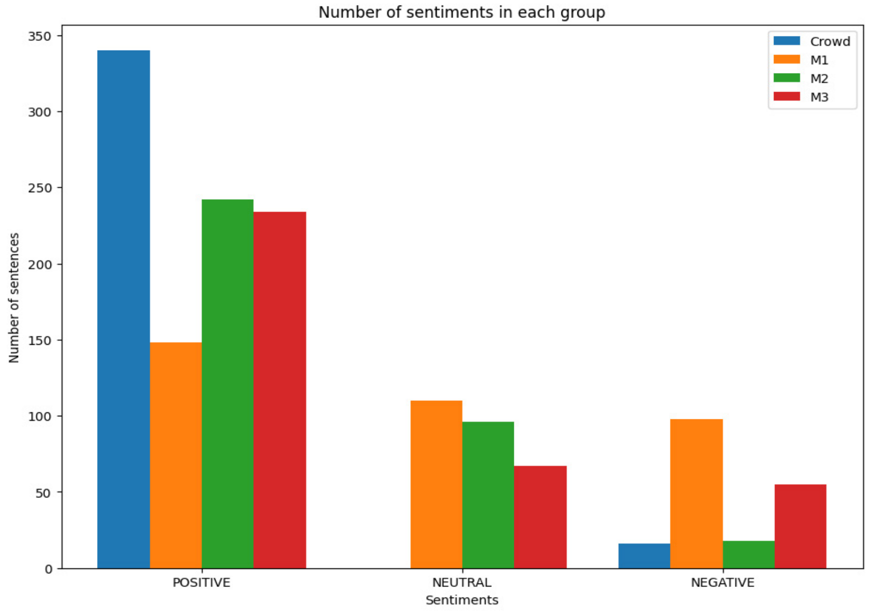
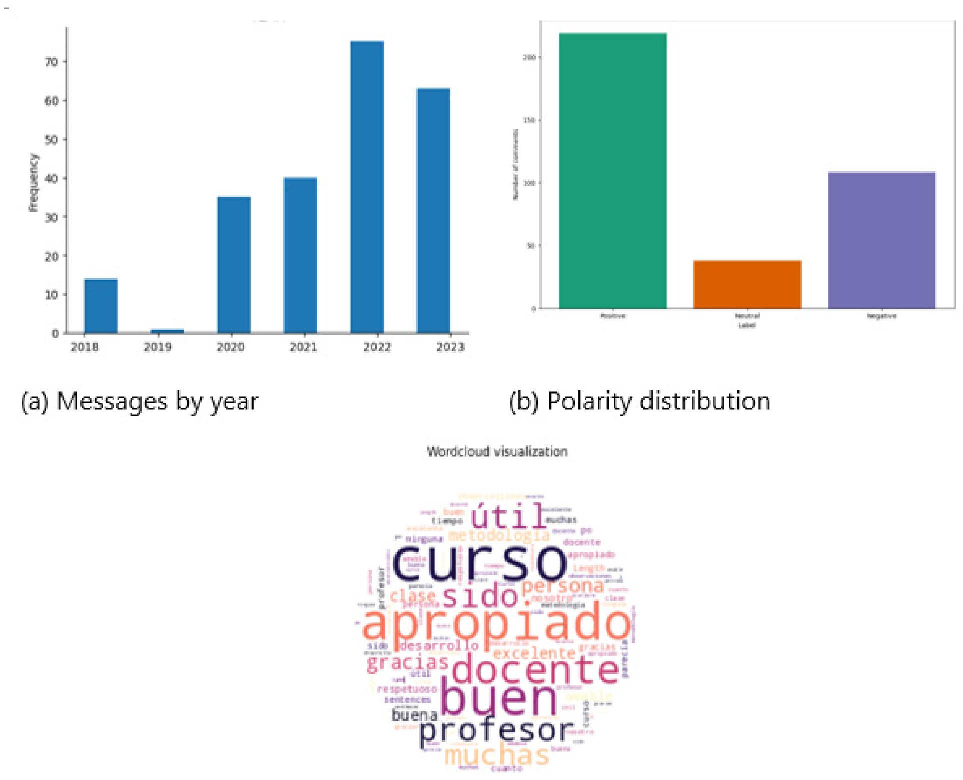


Figure 3. Proportion of comments labeled. Source: Authors

From the study dataset, we observed that negative comments tend to have a greater length compared to positive or neutral ones. Sentences such as “buen curso”, “me gustó”, “estuvo chevere!” were used by students to express positive feedback while in negative cases more words to express disgust, displeasure, dissatisfaction which leads to negative polarity.

Based on crowdsourced labeling, we obtained positive or negative messages for the feedback messages in the data set, with no neutral sentences being recognized. While crowdsourcing has been widely employed for labeling tasks, in our case, the absence of neutral labels poses an intriguing aspect to explore, especially if our objective were to train machine learning models. The lack of neutral labels may suggest a bias or imbalance in the data collection.

The ASET system prototype has great potential, taking raw comments and performing an analysis of sentiments and emotions in feedback when the course has ended to improve teaching and consequently the learning using end-course feedback. The used microservices architecture required real-time sentiment classifier model which was situated in an isolated container. As evaluated the ASET architecture shows various advantages such as the creation, update and deployment of new versions and ML models for sentiment analysis. Furthermore, isolated environments allow fix issues without impacting other parts of the application, such as user interfaces, models and other services. Thus improving the overall resilience and reliability of the application.



(a) Messages by year

(b) Polarity distribution

(c) Wordcloud

Figure 4. Analysis results. Source: Authors

In the analysis of end-of-course comments collected over recent years, the first graph shows the volume of comments for each course is an underutilized resource that can be effectively processed with machine learning techniques. For this reason, we perform the analysis of end-of-course comments collected over the last six years using ASET tool, Fig 4a shows the distribution of comment counts per year, providing a clear visualization of trends over time. Using the dataset, Fig 4b illustrates the distribution of comment polarity, highlighting the proportion of positive, neutral, and negative comments labeled using Distilbert model. To show popular used words within the comments in Fig 4c a word cloud offers a visual representation, where the size of words reflects their frequency in the comments.

doi: 10.25100/iyc.v26i1.13759

Conclusions

Recent artificial intelligence tools will have an impact on sentiment analysis and emotion research. In this paper the possibilities of LLM use for sentiment analysis was evaluated. In particular, the textual analysis of feedback from students at the end-course. This paper reveals that LLMs are not only competent in sentiment analysis but that can be used to support key tasks such as textual analysis of student feedback and provide to teachers a preliminary analysis.

Future work includes conducting aspect-level analysis, refining the software prototype, integrating a generative LLM model capable of providing directives to teachers based on student feedback and, to integrating LLM models to the Student Evaluation of Teaching (SET) system.

Code availability

The dataset, code, plots, and results are publicly available on https://github.com/japeto/llm-set-analisys/ (Only when the manuscript camera-ready stage). For further information or clarification regarding the results, please contact the corresponding author via email.

Acknowledges

We express our gratitude to the Universidad del Valle and its Faculty of Engineering, as well as to the Computer Science Department at Pontificia Universidad Javeriana, for providing access to their evaluation systems and end-course evaluation results.

References

1. Omer K, Jacobs S, Bettger B, Dawson J, Graether S, Murrant C, et al. Evaluating and Improving the Formative Use of Student Evaluations of Teaching. Can J Scholarsh Teach Learn. 2023;14(1):n1.

2. Mancenido Z. Impact evaluations of teacher preparation practices: Challenges and opportunities for more rigorous research. Rev Educ Res. 2023;00346543231174413.

3. Cunningham S, Laundon M, Cathcart A, Bashar MA, Nayak R. First, do no harm: automated detection of abusive comments in student evaluation of teaching surveys. Assess Eval High Educ. 2023;48(3):377–89.

4. Gravestock P, Gregor-Greenleaf E. Student course evaluations: Research, models and trends. Higher Education Quality Council of Ontario Toronto; 2008.

5. Uttl B. Lessons learned from research on student evaluation of teaching in higher education. Stud Feedback Teach Sch Using Stud Percept Dev Teach Teach. 2021;237–56.

6. Gaoa G, Pangb M, Pengc J, Lud Y. A Hierarchical Probe Evaluation Method for Teaching Effect of University Engineering Courses Based on the Keypoints of Knowledge. In: 3rd International Conference on Education, Language and Art (ICELA 2023). Atlantis Press; 2024. p. 376–84.

7. Alshammari E. Student evaluation of teaching. Is it valid? J Adv Pharm Educ Res Apr-Jun. 2020;10(2):97.

8. Röhl S, Bijlsma H, Rollett W. The process model of student feedback on teaching (SFT): A theoretical framework and introductory remarks. Stud Feedback Teach Sch Using Stud Percept Dev Teach Teach. 2021;1–11.

9. Jensen E, Dale M, Donnelly PJ, Stone C, Kelly S, Godley A, et al. Toward automated feedback on teacher discourse to enhance teacher learning. In: Proceedings of the 2020 chi conference on human factors in computing systems. 2020. p. 1–13.

10. Mandouit L, Hattie J. Revisiting "The Power of Feedback" from the perspective of the learner. Learn Instr. 2023;84:101718.

11. Lim LA, Dawson S, Gašević D, Joksimović S, Pardo A, Fudge A, et al. Students' perceptions of, and emotional responses to, personalised learning analytics-based feedback: an exploratory study of four courses. Assess Eval High Educ. 2021;46(3):339–59.

12. Chen S, Nieminen JH. Towards an ecological understanding of student emotions in feedback: a scoping review. Assess Eval High Educ. 2024;1–18.

13. Devlin J, Chang MW, Lee K, Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers) [Internet]. Minneapolis, Minnesota: Association for Computational Linguistics; 2019. p. 4171–86. Available from: https://aclanthology.org/N19-1423

14. Radford A, Wu J, Child R, Luan D, Amodei D, Sutskever I. Language Models are Unsupervised Multitask Learners. 2019;

15. Altrabsheh N, Cocea M, Fallahkhair S. Learning sentiment from students' feedback for real-time interventions in classrooms. In: International conference on adaptive and intelligent systems. Springer; 2014. p. 40–9.

16. Shaikh S, Daudpota SM, Yayilgan SY, Sindhu S. Exploring the potential of large-language models (LLMs) for student feedback sentiment analysis. In: 2023 International Conference on Frontiers of Information Technology (FIT). IEEE; 2023. p. 214–9.

17. Häkkinen J, Ramadan Z. A Study on the Perception of Feedback with Varying Sentiment Generated Using a Large Language Model. 2023.

18. Mouronte-López ML, Ceres JS, Columbrans AM. Analysing the sentiments about the education system trough Twitter. Educ Inf Technol. 2023;28(9):10965–94.

19. Shen L, Wang M, Shen R. Affective e-learning: Using "emotional" data to improve learning in pervasive learning environment. J Educ Technol Soc. 2009;12(2):176–89.

20. Zhou J, Ye J min. Sentiment analysis in education research: a review of journal publications. Interact Learn Environ. 2023;31(3):1252–64.

21. Cutroni L, Paladino A. Peer-ing in: A systematic review and framework of peer review of teaching in higher education. Teach Teach Educ. 2023;133:104302.

22. Zhu JJ, Chang YC, Ku CH, Li SY, Chen CJ. Online critical review classification in response strategy and service provider rating: Algorithms from heuristic processing, sentiment analysis to deep learning. J Bus Res. 2021;129:860–77.

23. Cox A. Exploring the impact of Artificial Intelligence and robots on higher education through literature-based design fictions. Int J Educ Technol High Educ. 2021;18(1):1–19.

24. Tian F, Zheng Q, Zhao R, Chen T, Jia X. Can e-Learner's emotion be recognized from interactive Chinese texts? In: 2009 13th International Conference on Computer Supported Cooperative Work in Design. 2009. p. 546–51.

25. Ortigosa A, Martín JM, Carro RM. Sentiment analysis in Facebook and its application to e-learning. Comput Hum Behav. 2014;31:527–41.

26. Poulos A, Mahony MJ. Effectiveness of feedback: The students' perspective. Assess Eval High Educ. 2008;33(2):143–54.

27. Gan Z, He J, Zhang LJ, Schumacker R. Examining the relationships between feedback practices and learning motivation. Meas Interdiscip Res Perspect. 2023;21(1):38–50.

doi: 10.25100/iyc.v26i1.13759

28. Kanwar A, Sanjeeva M. Student satisfaction survey: A key for quality improvement in the higher education institution. J Innov Entrep. 2022;11(1):27.

29. Giang NTP, Dien TT, Khoa TTM. Sentiment analysis for university students' feedback. In: Advances in Information and Communication: Proceedings of the 2020 Future of Information and Communication Conference (FICC), Volume 2. Springer; 2020. p. 55–66.

30. Dalipi F, Zdravkova K, Ahlgren F. Sentiment analysis of students' feedback in MOOCs: A systematic literature review. Front Artif Intell. 2021;4:728708.

31. Neumann M, Linzmayer R. Capturing student feedback and emotions in large computing courses: A sentiment analysis approach. In: Proceedings of the 52nd ACM Technical Symposium on Computer Science Education. 2021. p. 541–7.

32. Ren P, Yang L, Luo F. Automatic scoring of student feedback for teaching evaluation based on aspect-level sentiment analysis. Educ Inf Technol. 2023;28(1):797–814.

33. Katragadda S, Ravi V, Kumar P, Lakshmi GJ. Performance analysis on student feedback using machine learning algorithms. In: 2020 6th international conference on advanced computing and communication systems (ICACCS). IEEE; 2020. p. 1161–3.

34. Mabunda JGK, Jadhav A, Ajoodha R. Sentiment analysis of student textual feedback to improve teaching. In: Interdisciplinary Research in Technology and Management. CRC Press; 2021. p. 643–51.

35. Reddy SS, Gadiraju M, Maheswara Rao V. Analyzing student reviews on teacher performance using long short-term memory. In: Innovative Data Communication Technologies and Application: Proceedings of ICIDCA 2021. Springer; 2022. p. 539–53.

36. Agostini D, Picasso F. Large Language Models for Sustainable Assessment and Feedback in Higher Education: Towards a Pedagogical and Technological Framework. In: Proceedings of the First International Workshop on High-Performance Artificial Intelligence Systems in Education Co-Located with 22nd International Conference of the Italian Association for Artificial Intelligence (AIxIA 2023). 2023.

37. Parker MJ, Anderson C, Stone C, Oh Y. A large language model approach to educational survey feedback analysis. ArXiv Prepr ArXiv230917447. 2023;

38. Freyd M. A graphic rating scale for teachers. J Educ Res. 1923;8(5):433–9.

39. Boring A, Ottoboni K. Student evaluations of teaching (mostly) do not measure teaching effectiveness. Sci Res. 2016;

40. Hoel A, Dahl TI. Why bother? Student motivation to participate in student evaluations of teaching. Assess Eval High Educ. 2019;44(3):361–78.

41. Chen Y. Does students' evaluation of teaching improve teaching quality? Improvement versus the reversal effect. Assess Eval High Educ. 2023;48(8):1195–207.

42. Newman H, Joyner D. Sentiment Analysis of Student Evaluations of Teaching. In: Penstein Rosé C, Martínez-Maldonado R, Hoppe HU, Luckin R, Mavrikis M, Porayska-Pomsta K, et al., editors. Artificial Intelligence in Education. Cham: Springer International Publishing; 2018. p. 246–50.

43. Dake DK, Gyimah E. Using sentiment analysis to evaluate qualitative students' responses. Educ Inf Technol. 2023;28(4):4629–47.

44. Falcon S, Leon J. How do teachers engaging messages affect students? A sentiment analysis. Educ Technol Res Dev. 2023;71(4):1503–23.

45. Rani S, Kumar P. A Sentiment Analysis System to Improve Teaching and Learning. Computer. 2017 May;50(5):36–43.

doi: 10.25100/iyc.v26i1.13759

46. Bird S, Loper E. NLTK: The Natural Language Toolkit. In: Proceedings of the ACL Interactive Poster and Demonstration Sessions [Internet]. Barcelona, Spain: Association for Computational Linguistics; 2004 [cited 2021 Jul 3]. p. 214–7. Available from: https://aclanthology.org/P04-3031

47. Honnibal M, Montani I. spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing. Appear. 2018;

48. Pérez JM, Giudici JC, Luque F. pysentimiento: A Python Toolkit for Sentiment Analysis and SocialNLP tasks. 2021.

49. Pérez JM, Furman DA, Alonso Alemany L, Luque FM. RoBERTuito: a pre-trained language model for social media text in Spanish. In: Proceedings of the Thirteenth Language Resources and Evaluation Conference [Internet]. Marseille, France: European Language Resources Association; 2022. p. 7235–43. Available from: https://aclanthology.org/2022.lrec-1.785

50. García-Vega M, Díaz-Galiano M, García-Cumbreras M, Del Arco F, Montejo-Ráez A, Jiménez-Zafra S, et al. Overview of TASS 2020: Introducing emotion detection. In: Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2020) Co-Located with 36th Conference of the Spanish Society for Natural Language Processing (SEPLN 2020), Málaga, Spain. 2020. p. 163–70.

51. Hutto C, Gilbert E. VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. Proc Int AAAI Conf Web Soc Media. 2014 May;8(1):216–25.

52. Sanh V, Debut L, Chaumond J, Wolf T. DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. ArXiv Prepr ArXiv191001108. 2019;

53. Adoma AF, Henry NM, Chen W. Comparative analyses of bert, roberta, distilbert, and xlnet for text-based emotion recognition. In: 2020 17th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP). IEEE; 2020. p. 117–21.

54. Jain SM. Fine-Tuning Pretrained Models. In: Jain SM, editor. Introduction to Transformers for NLP: With the Hugging Face Library and Models to Solve Problems [Internet]. Berkeley, CA: Apress; 2022 [cited 2024 Jun 12]. p. 137–51. Available from: https://doi.org/10.1007/978-1-4842-8844-3_6

55. Jiang W, Synovic N, Hyatt M, Schorlemmer TR, Sethi R, Lu YH, et al. An Empirical Study of Pre-Trained Model Reuse in the Hugging Face Deep Learning Model Registry. In: 2023 IEEE/ACM 45th International Conference on Software Engineering (ICSE) [Internet]. 2023 [cited 2024 Jun 12]. p. 2463–75. Available from: https://ieeexplore.ieee.org/abstract/document/10172757?casa_token=RXsNoo9rLCQAAAAA:D5FiHpAk-1M-2V0O-7OqXzVVFd8Rb5KyxH8L9eitj9i0prdgOIeGi2ZtMhS46c3rfdpxy2CLzHn0

56. Hsueh PY, Melville P, Sindhwani V. Data quality from crowdsourcing: a study of annotation selection criteria. In: Proceedings of the NAACL HLT 2009 workshop on active learning for natural language processing. 2009. p. 27–35.

57. Borromeo RM, Toyama M. Automatic vs. Crowdsourced Sentiment Analysis. In: Proceedings of the 19th International Database Engineering & Applications Symposium [Internet]. New York, NY, USA: Association for Computing Machinery; 2015. p. 90–5. (IDEAS '15). Available from: https://doi.org/10.1145/2790755.2790761