[1]Myelinda A. Baldelovar

Harmonizing Perceptions: Exploring Alignment of Numerical and Textual Evaluation Through Fine-Grained Sentiment Analysis

Abstract: - The assessment of faculty teaching effectiveness plays a pivotal role in shaping educational practices and ensuring academic excellence. While numerical evaluations provide quantifiable measures of teaching effectiveness, textual comments offer valuable qualitative insights into the faculty's instructional methods and interactions. However, discrepancies between these evaluation formats may hinder the accurate assessment of faculty performance. This study aims to determine the alignment between the numerical and textual evaluations of faculty teaching effectiveness using fine-grained sentiment analysis and determine the relationship between numerical ratings and textual comments to identify patterns of consistency or divergence. A large dataset of anonymized BSCS students' feedback is analyzed through sentiment analysis to extract sentiments expressed in the textual comments. The study results reveal a misalignment between students' subjective perceptions expressed in the textual comments and the numerical ratings. This misalignment arises due to several factors, such as student communication style variations, individual interpretations of rating scales, or subjective biases. This study advances faculty evaluation methodologies and offers useful recommendations for organizations looking to improve assessment procedures as well as educational institutions that want to improve the precision and efficacy of their faculty evaluation systems for better faculty development and decision-making.

Keywords: fine-grained, numerical, sentiment analysis, textual

1. INTRODUCTION

Sentiment Analysis (SA) is an ongoing field of research in the text mining field. It is the computational treatment of opinions, sentiments, and subjectivity of text. It seeks to find opinions, identify the sentiments they express, and then classify their polarity [7]. In the higher education setting, sentiment analysis is vital in considering the sentiments of the students on the performance of faculty members. Faculty performance evaluation is essential in order to ensure high-quality teaching and learning outcomes in higher education institutions. Faculty members are often evaluated based on students' textual comments and numerical evaluations. A comprehensive knowledge of instructor effectiveness is greatly enhanced by blending numerical evaluations with textual comments. The gaps and misalignments between the measurable measurements and the complex sentiments represented in written comments might, however, be shown by this convergence. This study aims to a) determine the alignment between the numerical and textual evaluations of faculty teaching effectiveness using fine-grained sentiment analysis and b) determine the relationship between numerical rating and textual comments.

Sentiment analysis can be classified into coarse-grained and fine-grained analysis [9]. In the study of Xiao et al. [10], they proposed a memory-enhanced collaborative fine-grained interaction transformer (MCFIT) to learn collaborative fine-grained interaction between image and text. In the academic domain, when assessing and analyzing learning management systems, instruction, pedagogical practices, and courses, student feedback is essential [3]. In the study of Lalata et al. [5], they applied an ensemble approach integrating five individual machine algorithms namely Naive Bayes, Logistic Regression, Support Vector Machine, Decision Tree, and Random Forest algorithms to classify the students' comments based on the majority voting principle. The study of Rajput et al. [8] shows a high correlation between sentiment analysis-based metrics and aggregated Likert scale scores. They also use tag clouds, sentiment scores, and frequency-based filters to provide new insights into teacher's performance. An ensemble model that consists of five machine learning algorithms namely Naive Bayes, Logistic Regression, Support Vector Machine, Decision Tree, and Random Forest algorithms was used by Lalata et al. [6] to analyze and identify the polarity of the written comments of the students. The result showed that there is a marginally significant relationship between the numerical rating and the overall sentiment scores. Dake and Gyimah [2] used Naïve Bayes (NB), Support Vector Machine (SVM), J48 Decision Tree (DT), and Random Forest (RF) to classify the sentiments of the qualitative feedback of students.

[1] **DIT** North Eastern Mindanao State University-Tagbina Campus Philippines

The alignment between numerical evaluation scores and the sentiments expressed in accompanying textual remarks has received limited attention in the existing literature. Although sentiment analysis techniques have been extensively applied to extract sentiment from textual comments in various domains, the integration of such textual sentiments with the corresponding numerical ratings using fine-grained sentiment analysis remains an area that has not been thoroughly investigated. This presents an opportunity to explore methods that bridge the gap between numerical evaluations and textual comments, ultimately leading to a more comprehensive understanding of the evaluation process.

This study has a significant impact on the field of contemporary education. The results of employing fine-grained sentiment analysis to examine the consistency between numerical ratings and textual remark evaluations, institutions can make educated decisions regarding faculty development and establish successful strategies for enhancing teaching effectiveness. This will also have practical implications for higher education institutions' efforts to improve teaching effectiveness and support faculty development. By understanding the consistency between numerical ratings and textual comments evaluations using fine-grained sentiment analysis, institutions can make informed decisions about faculty development and create effective strategies for improving teaching effectiveness.

2. METHODS

This section presents the process for employing sentiment analysis to examine how numerical and textual assessments of faculty teaching effectiveness align. The process identifies the polarity of sentiments as moderately negative, negative, neutral, moderately positive, and positive. Fig. 1 below shows the sentiment analysis of the textual and numerical evaluation.

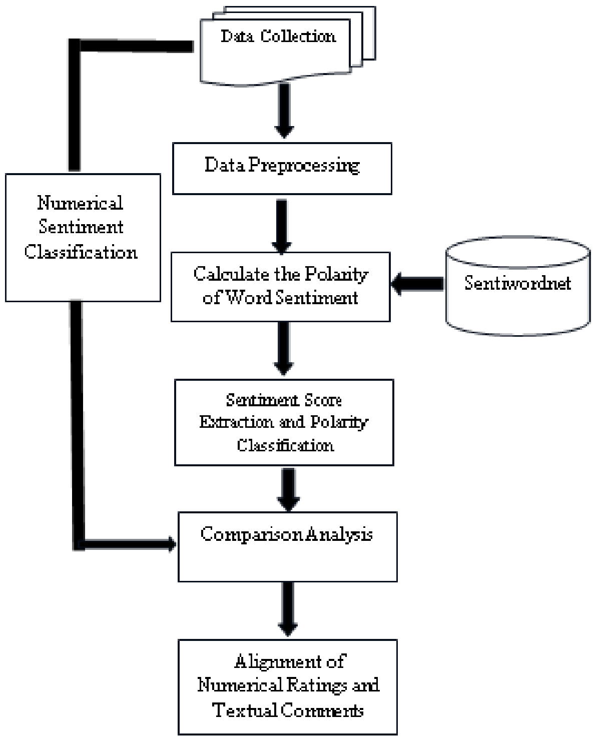

Fig. 1 - Sentiment analysis of textual and numerical data evaluation process

2.1 Data Collection

In sentiment analysis, relevant text data that communicate opinions, emotions, and attitudes are systematically gathered as part of the data collection process. The results of the BSCS students' evaluation of BSCS faculty performance for the second semester of the academic year 2022-2023 served as the basis for the data used in this study. The four essential components of the faculty performance evaluation are commitment, subject-matter expertise, teaching for autonomous learning, and management of learning. Students rank their faculty performance in every component on a scale of 1 to 5, with 5 being the highest and 1 being the lowest, and each area contains critical indicators. Students can also discuss their strengths and weaknesses in the narrative comments section.

2.2 Data Preprocessing

This process aims to transform unstructured raw text input into a refined and structured format suited for precise sentiment classification. This involves the following steps:

2.2.1 Tokenization: This includes the process of splitting text into words or subword tokens.

2.2.2 Stopword Removal: This includes the process of eliminating words that are widely used but may not convey much emotion.

2.2.3 Lemmatization: This involves breaking down each word to its "lemma," or equivalent base form such as, “building”, and “builder” into “build”.

2.2.4 Punctuation Removal: This method standardizes the text and removes punctuation like commas, periods, and question marks, making the content cleaner and simpler.

2.2.5 Stop Word Removal: Stop words are commonly used words such as articles, pronouns, and prepositions. Removal of these stop words helps to increase the accuracy of classification as there are fewer and only meaningful tokens left.

2.3 Calculate the Polarity of Word Sentiment

To calculate the polarity of word sentiment, SentiWordNet was used. It is a lexical tool for sentiment analysis that gives words ratings based on their emotional resonance [4], [1]. It is an expansion of the WordNet lexical database, which classifies words into synsets (sets of synonyms) and details their semantic relationships. By adding sentiment scores to these synsets, SentiWordNet expands WordNet and creates a vocabulary that classifies words as either positive, negative, or neutral based on their semantic meaning. SentiWordNet uses a word sense disambiguation method to assign emotion polarity. This implies that it takes into account the various meanings or senses a word may have and assesses the sentiment of each sense independently. The total sentiment score of a word is then calculated by adding the sentiment scores of the word's multiple meanings and weighting them based on how frequently each sense occurs in a sizable corpus of literature.

2.4 Sentiment Score Extraction and Polarity Classification.

Textual data were converted into numerical values, or sentiment scores, in order to be used as textual data input for sentiment analysis. Sentiwordnet, a lexical resource that assigns sentiment ratings to words based on their positivity, negativity, and neutrality, was used to determine the sentiment scores of the textual comments. It is an expansion of WordNet, a sizable lexical database of links between English words. The resulting sentiment ratings are then categorized based on the polarity of the sentiment reflected in a textual comment—positive, negative, or neutral. Table I below displays the emotion score and its polarity classification.

Table I. Polarity Classification of Sentiment Scores

Sentiment Score	Polarity Classification
0.51 to 1.0	Positive
0.01 to 0.50	Moderately Positive
0	Neutral
-0.01 to -0.50	Moderately Negative
-0.51 to -1.0	Negative

2.5 Comparison Analysis

Analysis of variance is used to determine the alignment of numerical ratings and textual comments of the faculty performance evaluation. Using Analysis of Variance in Python, the alignment between the numerical ratings of the faculty members and the equivalent sentiment scores of the textual comments from the faculty performance evaluation is determined.

2.6 Numerical Sentiment Classification

Likert Scale Conversion. Likert scales are numeric data and do not directly provide a measure of sentiment polarity. To easily analyze the sentiments of the Likert scale values provided by students in the faculty performance evaluation, each scale is given an equivalent polarity classification as shown in Table II below.

Table II. Equivalent Polarity Classification for the Likert Scale

Likert Scale Range	Description	Polarity Classification
4.21 – 5.00	Strongly agree	Positive
3.41 – 4.20	Agree	Moderately Positive
2.61 – 3.40	Neutral	Neutral
1.81 – 2.60	Disagree	Moderately Negative
1.0 – 1.80	Strongly Disagree	Negative

3 RESULTS AND DISCUSSIONS

This section presents the results of the study.

Based on the study's findings, it can be seen in Table III below that there is a slight misalignment between some of the BSCS faculty members' numerical ratings and textual remarks given by students. This misalignment suggests that a more thorough evaluation procedure is required, one that incorporates both textual and numerical rating metrics to appropriately assess faculty performance. Textual analyses are frequently subjective and susceptible to the effect of individual perceptions, feelings, and experiences. On the other side, numerical evaluations make an effort to offer a more objective measurement. Numerical ratings may not adequately represent the context and subtle nuances, but textual evaluations can. People can go into detail about particular features they liked or didn't like, which may not be fully captured by a single number. A person's communication style can vary from another person's, even if someone's text is written in a more upbeat tone, they might nonetheless give it a slightly lower score.

Table III. Alignment of Numerical Rating and Sentiment Score

Faculty	Numerical Rating	Polarity Description	Sentiment Score	Description	Alignment (Yes/No)
1	4.55	Positive	0.48	Moderately Positive	No
2	4.80	Positive	0.64	Positive	Yes
3	4.50	Positive	0.64	Positive	Yes
4	4.60	Positive	0.70	Positive	Yes
5	4.85	Positive	0.67	Positive	Yes
6	3.52	Positive	0.66	Moderately Positive	No
7	3.45	Positive	0.49	Moderately Positive	No

According to Table IV below, there is a significant difference between numerical ratings and textual comments in faculty evaluations since the F-value is higher than the F-critical value and the p-value is lower than the significance level at 0.05. This signifies that students' perceptions, as represented in their textual comments, are at odds with the numerical ratings they provide. This could imply that the straightforward numerical grading system is not adequately capturing the more complex input that students are offering in the comments. The numerical ratings might not accurately represent the students' true feelings or experiences. If textual comments reveal concerns, issues, or positive aspects that are not reflected in the numerical scores, the institution might need to reevaluate the effectiveness of its evaluation system.

Table IV. Analysis of Variance

F-Statistic	F-Crit	p-Value
7.5625	5.987378	0.02505

4 CONCLUSION

In order to shape the landscape of educational practices and sustain the norms of academic quality, the evaluation of faculty teaching efficiency is of utmost importance. It is clear that both quantitative and qualitative metrics are essential instruments for thoroughly assessing teacher performance as educational institutions strive for academic excellence. In this study, the fact that there is an insignificant misalignment between the textual and numerical evaluation outcomes in faculty evaluations highlights how complicated and varied the teaching experience is. Textual comments dig into the intricate world of student attitudes, classroom dynamics, and pedagogical nuances while numerical ratings provide a quantifiable window into the perceived success of faculty members. This discrepancy is an important reminder that education is not only quantitative but also flourishes on the basis of unique experiences, interactions, and learning journeys. Institutions may promote a culture of continual improvement by acknowledging the relevance of this discrepancy and having deliberate conversations about both numerical and textual evaluations. This path of improvement not only improves the caliber of instruction but also demonstrates an unwavering dedication to developing each student's potential, driving education's progress toward a more thorough and fulfilling experience for all.

5 RECOMMENDATIONS

It is suggested that sentiment analysis's application be broadened to take into account different languages and cultures as a direction for future research. The alignment between textual remarks and numerical evaluations in a particular context has been explored by this study, although the dynamics of sentiment expression can differ dramatically between languages. Researchers can gain a greater grasp of how cultural differences and linguistic variances affect students' views and assessments of faculty teaching performance by employing sentiment analysis approaches that are language-inclusive.

REFERENCES

[1] Baccianella S, Esuli A, Sebastiani F. Sentiwordnet3.0: an enhanced lexical resource for sentiment analysis and opinion mining. In: Proceedings of the 7th International Conference on Language Resources and Evaluation. 2010, 2200–2204

[2] Dake, D.K., Gyimah, E. Using sentiment analysis to evaluate qualitative students’ responses. Educ Inf Technol 28, 4629–4647 (2022). https://doi.org/10.1007/s10639-022-11349-1

[3] Elfeky, A. I. M., Masadeh, T. S. Y., & Elbyaly, M. Y. H. (2020). Advance organizers in flipped classroom via e-learning management system and the promotion of integrated science process skills. *Thinking Skills and Creativity*, *35*, 100622. https://doi.org/10.1016/j.tsc.2019.100622

[4] Esuli A, Sebastiani F, Moruzzi V G. SENTIWORDNET: a publicly available lexical resource for opinion mining. In: Proceedings of the 5th International Conference on Language Resources and Evaluation. 2006, 417–422

[5] Lalata, Jay-Ar and Gerardo, Bobby and Medina, Ruji. 2019. A correlation analysis of the sentiment analysis scores and numerical ratings of the students in the faculty evaluation. AIPR '19: Proceedings of the 2nd International Conference on Artificial Intelligence and Pattern Recognition. 140-144. 10.1145/3357254.3357287.

[6] Lalata, Jay-Ar and Gerardo, Bobby and Medina, Ruji. 2019. A Sentiment Analysis Model for Faculty Comment Evaluation Using Ensemble Machine Learning Algorithms. In Proceedings of the 2019 International Conference on Big Data Engineering (BDE 2019). Association for Computing Machinery, New York, NY, USA, 68–73. https://doi.org/10.1145/3341620.3341638

[7] Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, *5*(4), 1093-1113. https://doi.org/10.1016/j.asej.2014.04.011

[8] Rajput, Quratulain and Haider, Sajjad and Ghani, Sayeed. 2016. Lexicon-Based Sentiment Analysis of Teachers’ Evaluation. Applied Computational Intelligence and Soft Computing. doi: 10.1155/2016/2385429

[9] Xiao, Y., Li, C., Thürer, M., Liu, Y., & Qu, T. (2022). User preference mining based on fine-grained sentiment analysis. Journal of Retailing and Consumer Services, 68, 103013. https://doi.org/10.1016/j.jretconser.2022.103013

[10] Xiao, X., Pu, Y., Zhou, D., Cao, J., Gu, J., Zhao, Z., & Xu, D. (2023). Collaborative fine-grained interaction learning for image–text sentiment analysis. *Knowledge-Based Systems*, *279*, 110951. https://doi.org/10.1016/j.knosys.2023.110951