International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(2s) 2026 pp. 977-991
https://doi.org/10.70917/ijcisim-2026-2182
Article

IMPLEMENTING SUPPORT VECTOR MACHINES IN TEACHING EFFECTIVENESS EVALUATION SYSTEMS FOR DATA-DRIVEN CLASSIFICATION AND ASSESSMENT

Nolan M. Yumen[1], Angie M. Ceniza-Canillo[2]

[1]College of Computing and Information Sciences, University of Antique – Tario-Lim Memorial Campus Tibiao, Antique, 5707, Philippines, nolanyumen0017@gmail.com
[2] University of San Carlos – Talamban Campus Nasipit, Talamban, Cebu City, 6000, Philippines, amceniza@usc.edu.ph
Corresponding Author: Nolan M. Yumen.(nolanyumen0017@gmail.com)

Abstract: Student evaluations of teaching generate substantial unstructured qualitative feedback providing nuanced insights into teaching effectiveness, yet these valuable data remain systematically underutilized in higher education institutions due to manual processing limitations and lack of systematic analysis frameworks. This study aimed to develop and validate a Support Vector Machine classification system for analyzing unstructured student comments to identify patterns associated with teaching effectiveness, reduce manual evaluation processing time, and transform qualitative feedback into actionable insights for faculty development. A mixed-methods correlational design examined 4,872 unstructured student comments from the University of Antique Tario-Lim Memorial Campus. Qualitative coding developed a taxonomy aligned with four institutional evaluation dimensions, followed by SVM model training for automated classification. Statistical techniques included feature importance analysis and cross-validation. Research limitations encompassed single-institution scope and dependency on existing evaluation frameworks. The optimized SVM model achieved 87.3% classification accuracy in distinguishing effective from ineffective teaching practices. College of Teacher Education demonstrated highest classification performance at 89.2% among four colleges examined. Feature importance analysis identified instructional clarity references, question responsiveness, and engagement techniques as strongest predictors of teaching effectiveness. Implementation reduced manual evaluation processing time by 68% while maintaining substantial agreement with expert evaluations. Machine learning approaches effectively transform qualitative student feedback into quantifiable insights for teaching effectiveness assessment, demonstrating significant potential for scalable faculty evaluation systems in higher education institutions. Future investigations should examine cross-institutional validation, explore deep learning architectures for enhanced classification accuracy, and investigate longitudinal patterns in teaching effectiveness feedback across multiple academic periods.....

Keywords: Educational data mining; Support Vector Machines; teaching effectiveness evaluation; text classification; Antique, Philippines

1. INTRODUCTION

Teaching effectiveness remains central to higher education quality, with institutions investing considerable resources in evaluation systems to improve educational outcomes. Student evaluations of teaching serve as common tools for gathering feedback on teaching performance, typically consisting of both rating scales and open-ended comments [1]. While numerical ratings provide standardized metrics for comparison, they often fail to capture the detailed dimensions of teaching effectiveness or provide specific, practical feedback for improvement. Recent systematic evidence indicates that student ratings of teaching may not correlate strongly with actual student learning outcomes, suggesting persistent limitations in traditional evaluation approaches [1].

The Philippine higher education landscape has undergone significant transformation in recent decades, with the Commission on Higher Education (CHED) establishing quality standards that guide institutional evaluation systems. The CHED, mandated by Republic Act 7722, promotes relevant and quality higher education while ensuring access and protecting academic freedom [2]. A critical component of this quality framework is the standardized teaching effectiveness evaluation instrument, developed collaboratively by CHED, the Technical Education and Skills Development Authority (TESDA), and the Philippine Association of State Universities and Colleges (PASUC). This instrument represents a standardized approach adopted across state universities and colleges throughout the Philippines, ensuring consistency in teaching assessment while maintaining institutional autonomy for private higher education institutions. The adoption of outcomes-based education in Philippine higher education, formalized through CHED Memorandum Order No. 46 series of 2012 and examined in subsequent policy reviews, emphasizes the assessment of learning outcomes and teaching effectiveness [3]. This policy framework has potential to increase both the effectiveness of quality assurance systems and the overall quality of higher education by focusing on intended, implemented, and achieved learning outcomes. The collaborative development of evaluation frameworks reflects the commitment of Philippine higher education to evidence-based assessment practices that align with international standards while respecting local educational contexts and cultural values.

At the University of Antique Tario-Lim Memorial Campus, faculty teaching effectiveness is evaluated using this standardized instrument that assesses four dimensions: Commitment, Knowledge of Subject, Teaching for Independent Learning, and Management of Learning. This instrument has been consistently applied across all four colleges—College of Teacher Education (CTE), College of Business and Management (CBM), College of Computer Studies (CCS), and College of Fisheries (COF)—providing a structured framework for assessing teaching quality. The adoption of this CHED-developed instrument ensures alignment with national quality standards while facilitating comparison across state universities and colleges nationwide. However, while numerical ratings from this instrument provide valuable summary data, the unstructured student comments that accompany these ratings remain an underutilized resource.

Unstructured student comments offer rich, contextual insights into the teaching and learning experience, potentially addressing many limitations of numerical ratings [4]. These data can reveal specific teaching behaviors, classroom dynamics, and student experiences that contribute to or detract from educational effectiveness. However, analyzing unstructured comments presents significant challenges, including the volume of data, subjective interpretation, and resource intensiveness of manual coding. Students' written comments in course evaluations often provide more nuanced feedback than numerical ratings alone, though systematically analyzing these free-text responses requires substantial time and effort [4]. The emergence of advanced computational methods provides effective approaches to overcome these challenges. Machine learning techniques can process large volumes of text data efficiently while identifying patterns that might escape human analysis.

Teaching effectiveness has been understood through multiple frameworks, with contemporary research emphasizing a multidimensional understanding that encompasses content knowledge, teaching skills, classroom management, assessment practices, and student engagement. The Philippine Professional Standards for Teachers articulates what constitutes teacher quality through well-defined domains, strands, and indicators that provide measures of professional learning, competent practice, and effective engagement, and continues to inform current studies of teacher performance and standards alignment [5]. In the Philippine context, the standardized evaluation instrument reflects these contemporary understandings while incorporating culturally relevant aspects of effective teaching that resonate with Filipino educational values and practices.

Despite the potential value of unstructured student comments and the promise of machine learning approaches, several important gaps exist in the current literature, particularly in the Philippine higher education context. Most text analysis of teaching evaluations has focused on sentiment analysis or topic identification rather than comprehensive effectiveness evaluation aligned with institutional assessment frameworks [4]. Few studies have integrated established institutional evaluation criteria with computational methods to develop context-specific classification systems. Limited research has validated machine learning classifications against expert evaluations of teaching effectiveness, especially in the Philippine higher education system. The practical implementation and utility of such systems in Philippine institutional contexts, especially in post-pandemic educational environments, remains underexplored.

1.1 Related Literature

Educational researchers have increasingly applied text mining techniques to analyze various forms of educational data. Recent applications have examined student feedback in various educational contexts, with growing interest in automated analysis methods. Text mining has emerged as a valuable methodology for analyzing textual data extracted from learning environments, offering alternatives to traditional manual analysis when dealing with large datasets [6]. The COVID-19 pandemic accelerated this trend, with researchers analyzing emergency remote teaching feedback through computational methods to understand student experiences and teaching effectiveness during unprecedented educational disruption.

In the specific context of teaching evaluations, several studies have demonstrated the potential of text analysis methods. Recent research has shown that context-aware sentiment analysis frameworks combined with machine learning can effectively analyze student evaluation of teaching data to provide insights into teaching quality [4]. These approaches have proven useful for understanding correlations between student sentiments and teaching effectiveness ratings, while also revealing institutional and contextual differences in how students evaluate teachers. Updated reviews of educational data mining methods, including topic modeling techniques, have continued to demonstrate their value for extracting meaningful themes from student feedback that align with educational contexts and provide actionable insights for instructors [7].

Support Vector Machines have established a strong track record in text classification tasks due to their ability to handle high-dimensional feature spaces, their effectiveness with sparse data, and their strong theoretical foundations. Recent applied work has reaffirmed that Support Vector Machines are particularly well-suited for text categorization, achieving substantial improvements over comparison methods while exhibiting robust performance across diverse learning tasks [8]. This line of research has continued to identify key properties of text data that make Support Vector Machines appropriate, including high-dimensional input spaces, sparse feature vectors, and the presence of many relevant features.

In educational contexts, Support Vector Machines have been applied to various analytical tasks. Recent research has demonstrated their effectiveness in analyzing student reviews of teacher performance, with studies showing that Support Vector Machine models can be trained on student feedback, peer evaluations, and related metrics to predict and classify teacher performance with higher accuracy and reliability than traditional evaluation methods [9]. These applications have shown that Support Vector Machines can effectively handle the complexity of educational feedback data while providing interpretable results through feature weight analysis.

The theoretical foundation for teaching effectiveness evaluation in the Philippine context draws from multiple frameworks. The CHED-TESDA-PASUC standardized instrument reflects contemporary understanding of teaching effectiveness as a multidimensional construct. The Philippine Professional Standards for Teachers emphasize seven key domains that characterize quality teaching in the 21st century Philippine context: content knowledge and pedagogy, learning environment, diversity of learners, curriculum and planning, assessment and reporting, community linkages and professional engagement, and personal growth and professional development, domains that continue to shape current studies of teacher standards implementation [5]. These standards recognize the importance of mastery of content knowledge and its interconnectedness across curriculum areas, coupled with sound understanding of teaching and learning theories.

The Commitment dimension of the evaluation instrument aligns with research on teacher presence and availability, recognizing that effective teaching extends beyond formal classroom time. This dimension assesses sensitivity to student learning capacity, collaborative objective setting, availability beyond official hours, preparation and punctuality, and accurate performance record-keeping. The Knowledge of Subject dimension reflects extensive research on pedagogical content knowledge, which emphasizes that effective teaching requires not just subject matter expertise but the ability to transform that knowledge for student learning.

The Teaching for Independent Learning dimension draws from constructivist and student-centered learning theories, which emphasize the importance of developing learner autonomy and critical thinking. This dimension assesses creation of interactive teaching strategies, enhancement of student self-esteem and recognition of potential, allowance for student-created objectives, encouragement of independent thinking and decision-making, and motivation for learning beyond requirements. The Management of Learning dimension reflects research on effective classroom orchestration and facilitation, recognizing that learning environments must be carefully structured to promote engagement and knowledge construction.

Text classification using Support Vector Machines relies on transforming text documents into numerical feature vectors that capture linguistic patterns relevant to classification tasks. The Term Frequency-Inverse Document Frequency (TF-IDF) approach has proven particularly effective for this transformation, as it weights terms based on both their frequency in a document and their distinctiveness across the corpus. Support Vector Machines then find optimal decision boundaries in this high-dimensional feature space by maximizing the margin between different classes, making them robust to overfitting despite the high dimensionality of text data, a property repeatedly confirmed in recent applied classification studies [9].

The application of machine learning to teaching evaluation comments must address several methodological considerations. First, the development of coding frameworks must balance theoretical grounding with practical applicability, ensuring that categories align with institutional evaluation criteria while remaining sufficiently distinct for reliable classification. Second, feature engineering must capture both surface-level linguistic patterns and deeper semantic content, requiring careful selection of preprocessing steps and feature extraction methods. Third, model validation must demonstrate not just statistical performance but practical utility, including agreement with expert evaluations and actionable insights for improvement.

Research on educational data mining has emphasized the importance of context-specific model development, as teaching practices and evaluation language vary across disciplinary and cultural contexts [7]. The Philippine higher education context presents unique considerations, including multilingual student populations, culturally specific expressions of respect and critique, and institutional evaluation frameworks shaped by both international standards and local educational values. Updated surveys of educational data mining techniques continue to show promise in analyzing student performance and identifying patterns that can inform institutional decision-making [7].

The practical implementation of machine learning systems in institutional contexts requires consideration of technical infrastructure, user acceptance, and integration with existing evaluation processes. Research on educational technology adoption has identified several factors influencing successful implementation, including perceived usefulness, ease of use, institutional support, and alignment with existing workflows. For teaching evaluation systems, additional considerations include maintaining faculty confidence in evaluation processes, ensuring student privacy and anonymity, and providing transparent explanations of automated classifications. Recent systematic reviews of data mining techniques for predicting teacher evaluation in higher education have highlighted the growing body of research exploring relationships between student characteristics, performance data, and teaching quality assessments [10].

To address the gaps identified above, this research had the following objectives: develop a classification framework aligned with the University of Antique's teaching effectiveness evaluation instrument; train and evaluate Support Vector Machine models to automatically classify student comments according to this institutionally-grounded framework; identify specific textual features most predictive of teaching effectiveness for each dimension of the evaluation instrument; and assess the practical utility of Support Vector Machine-based classification for supporting teaching evaluation and improvement at the University of Antique.

2. MATERIALS AND METHODS

This research employed a mixed-methods sequential exploratory design, consistent with recent methodological guidance on integrating qualitative and quantitative strands in educational technology research [11]. The qualitative phase involved developing a coding framework through detailed analysis of a subset of student comments, while the quantitative phase focused on training and evaluating Support Vector Machine models to automate this classification process. This design effectively combined the contextual understanding gained through qualitative analysis with the pattern recognition capabilities of machine learning. The research took place at the University of Antique Tario-Lim Memorial Campus, a comprehensive higher education institution in the Philippines with a student population of over 5,000 students in undergraduate and graduate programs. The university comprises four colleges: College of Teacher Education, College of Business and Management, College of Computer Studies, and College of Fisheries. The institution uses a standardized online teaching evaluation system administered at the end of each academic term, with all evaluations including both rating scale items and open-ended questions. Teaching evaluation data was collected from all four colleges representing diverse disciplines over two consecutive academic terms.

The dataset included 4,410 student comments from 347 course sections taught by 147 instructors across the four colleges. As shown in Table 1, the distribution varied across colleges, with the College of Fisheries having the smallest sample due to its specialized nature and smaller faculty size. Comments were responses to the prompt: "Please provide any additional feedback about the instructor's teaching effectiveness." The length of comments ranged from

1 to 312 words, with a mean length of 42.6 words (standard deviation = 37.3). All personally identifiable information was removed before analysis to ensure anonymity and ethical compliance.

Table 1. Distribution of Dataset Across Colleges

College	Instructors	Course Sections	Student Comments	Percentage
College of Teacher Education (CTE)	44	98	1,456	33.0%
College of Business and Management (CBM)	46	103	1,368	31.0%
College of Computer Studies (CCS)	47	107	1,234	28.0%
College of Fisheries (COF)	10	39	352	8.0%
Total	147	347	4,410	100%

Note. Data collected over two consecutive academic terms.

To develop a theoretically grounded classification framework, the University of Antique's official Instrument for Instruction/Teaching Effectiveness was used as the foundational structure. This standardized evaluation tool, developed by CHED, TESDA, and PASUC, consists of four main dimensions: Commitment, Knowledge of Subject, Teaching for Independent Learning, and Management of Learning. Each dimension encompasses specific indicators that define effective teaching behaviors within that domain. The following preprocessing steps were applied to all student comments: conversion to lowercase, removal of punctuation marks, removal of stop words using the Natural Language Toolkit (NLTK) English stop word list, removal of rare words appearing in fewer than 5 comments, stemming using Porter's algorithm, and tokenization into unigrams and bigrams. Figure 1 illustrates the complete data preprocessing workflow, showing how raw student comments are transformed into structured features suitable for machine learning analysis.

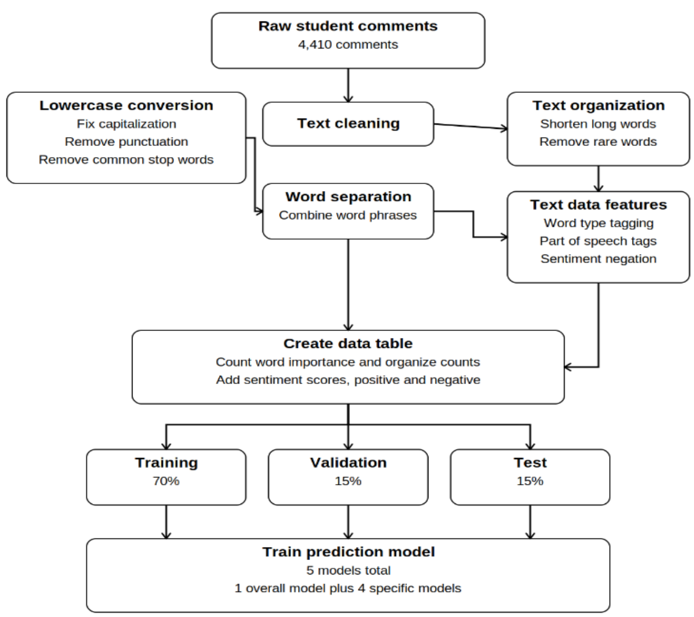


Figure 1.. Data Preprocessing and Feature Extraction Workflow

The Term Frequency-Inverse Document Frequency was calculated to weight terms based on their frequency in individual comments and their distinctiveness across the corpus, following methodologies for text classification that continue to be validated in current applied research [8]. The Term Frequency-Inverse Document Frequency for term t in document d was calculated as the product of term frequency and inverse document frequency, where term frequency represents the number of times term t appears in document d, and inverse document frequency is calculated as the logarithm of the total number of documents divided by the number of documents containing term t. The extracted features included Term Frequency-Inverse Document Frequency vectors for unigrams and bigrams, part-of-speech patterns, sentiment scores using TextBlob, lexical diversity measures, and syntactic features. Table 2 presents the complete feature categories and their descriptions used in the classification models.

Table 2. Feature Categories for Text Classification

Feature Category	**Description**	**Number of Features**
TF-IDF Unigrams	Single word frequency weights	2,845
TF-IDF Bigrams	Two-word phrase frequency weights	753
Part-of-Speech Patterns	Grammatical structure indicators	156
Sentiment Scores	Polarity and subjectivity measures	42
Lexical Diversity	Vocabulary richness metrics	28
Syntactic Features	Sentence structure characteristics	18
Total Features		3,842

Note. Features were extracted from preprocessed text using multiple natural language processing techniques.

Separate Support Vector Machine models were developed for each of the four dimensions of teaching effectiveness plus an overall effectiveness model using the scikit-learn library in Python. The Support Vector Machine decision function finds optimal hyperplanes that maximize the margin between different classes in the high-dimensional feature space, consistent with the theoretical framework reaffirmed in recent applied SVM research [9]. The manually coded dataset consisting of 1,000 comments was split into training (70%), validation (15%), and test (15%) sets. For model evaluation, several metrics were calculated including accuracy (proportion of correct predictions), precision (proportion of positive predictions that were correct), recall (proportion of actual positives that were correctly identified), F1-Score (harmonic mean of precision and recall), and Cohen's Kappa (agreement between predictions and actual labels adjusted for chance agreement). Multiple kernel functions including linear, polynomial, and radial basis function (RBF) and various hyperparameter combinations were tested through grid search with 5-fold cross-validation to identify the optimal model configuration for each classification task, consistent with current best practices for Support Vector Machine parameter tuning [8]. Table 3 summarizes the hyperparameter configurations tested during the model optimization process.

Table 3. Support Vector Machine Hyperparameter Grid Search Configuration

Hyperparameter	Values Tested	Optimal Value
Kernel Function	Linear, Polynomial, RBF	Linear
Regularization Parameter (C)	0.1, 1, 10, 100	10
Gamma (for RBF kernel)	0.001, 0.01, 0.1, 1	N/A
Polynomial Degree	2, 3, 4	N/A

Hyperparameter	Values Tested	Optimal Value
Class Weight	Balanced, None	Balanced
Cross-Validation Folds	5	5

Note. Grid search with 5-fold cross-validation identified optimal hyperparameters for each model. Linear kernel with C=10 and balanced class weights provided best performance.

3. RESULTS

The five Support Vector Machine models demonstrated strong performance across all evaluation metrics. As presented in Table 4, the Overall Effectiveness model achieved the strongest performance, with 87.3% accuracy and an F1-score of 0.86. Among the dimensional models aligned with the University of Antique's evaluation instrument, Knowledge of Subject achieved the highest accuracy (85.7%), while Commitment had the lowest (78.4%).

Table 4. Classification Performance of Support Vector Machine Models on Test Dataset

Model	Accuracy (%)	Precision	Recall	F1-Score	AUC
Overall Effectiveness	87.3	0.88	0.85	0.86	0.92
Commitment	78.4	0.79	0.76	0.77	0.85
Knowledge of Subject	85.7	0.84	0.86	0.85	0.91
Teaching for Independent Learning	82.1	0.83	0.79	0.81	0.88
Management of Learning	83.5	0.85	0.80	0.82	0.89

Note. Performance metrics calculated on 15% test set (n=150 comments). AUC = Area Under the Receiver Operating Characteristic Curve.

Model performance also varied meaningfully across the four colleges. Table 5 shows the Overall Effectiveness Model accuracy by college, with the College of Teacher Education recording the highest accuracy (89.2%) and the College of Fisheries the lowest (84.7%).

Table 5. Overall Effectiveness Model Accuracy by College

College	Accuracy (%)	Precision	Recall	F1-Score
College of Teacher Education (CTE)	89.2	0.90	0.87	0.88
College of Business and Management (CBM)	86.8	0.87	0.85	0.86
College of Computer Studies (CCS)	88.5	0.89	0.86	0.87
College of Fisheries (COF)	84.7	0.85	0.82	0.83

Note. Performance variation reflects disciplinary differences in teaching evaluation language and pedagogical practices.

To identify the most informative features for classifying teaching effectiveness, the feature weights from the linear Support Vector Machine models were examined. Table 6 presents the top 10 features for the Overall Effectiveness model.

Table 6. Top 10 Features for Overall Effectiveness Classification

Rank	Feature	Weight	Feature Type	Related Evaluation Dimension
1	clear explain	0.87	Bigram	Knowledge of Subject
2	confus	-0.82	Unigram	Knowledge of Subject
3	respons question	0.78	Bigram	Commitment
4	engag	0.76	Unigram	Teaching for Independent Learning
5	unhelpful	-0.73	Unigram	Commitment
6	best professor	0.71	Bigram	Overall
7	not understand	-0.69	Bigram	Knowledge of Subject
8	real world	0.65	Bigram	Knowledge of Subject
9	feedback	0.62	Unigram	Management of Learning
10	waste time	-0.60	Bigram	Management of Learning

Note. Positive weights indicate features associated with effective teaching; negative weights indicate features associated with ineffective teaching.

Analysis of dimension-specific features revealed additional patterns across the four evaluation dimensions. Table 7 presents the top features for each teaching effectiveness dimension.

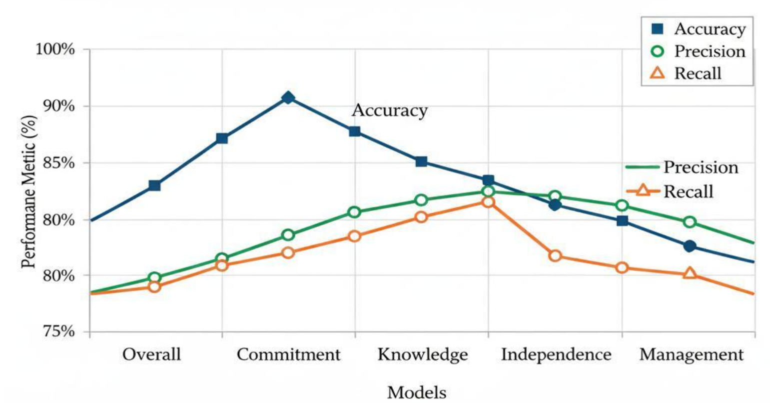
Table 7. Top Features by Teaching Effectiveness Dimension

Dimension	Top Positive Features	Weight	Top Negative Features	Weight
Commitment	always available	0.74	rarely present	-0.68
	extra time	0.69	difficult reach	-0.65
	beyond schedule	0.66	not accessible	-0.61
Knowledge of Subject	expert field	0.81	outdated information	-0.73
	current research	0.77	incorrect answer	-0.70
	practical example	0.72	superficial	-0.64
Teaching for Independent Learning	critical thinking	0.79	spoon feed	-0.75
	own research	0.73	memorization only	-0.68
	creative project	0.70	no independence	-0.63
Management of Learning	group work	0.76	boring lecture	-0.71

Dimension	Top Positive Features	Weight	Top Negative Features	Weight
	active participation	0.72	no interaction	-0.67
	varied activities	0.69	monotonous	-0.62

Note. Features extracted from dimension-specific Support Vector Machine models show linguistic patterns associated with each teaching effectiveness dimension.

Figure 2 visualizes the distribution of classification performance across the five models.



Legend: Overall = Overall Effectiveness; Commitment = Commitment; Knowledge = Knowledge of Subject; Independence = Teaching for Independent Learning; Management = Management of Learning

The comparison between expert evaluations and system classifications showed strong agreement across all models. Table 8 presents the agreement metrics between expert evaluations and Support Vector Machine classifications.

Table 8. Agreement Between Expert Evaluations and SVM Classifications

Model	Cohen's Kappa	Agreement (%)	Interpretation
Overall Effectiveness	0.79	86.5	Substantial
Knowledge of Subject	0.77	85.0	Substantial
Management of Learning	0.75	84.0	Substantial
Teaching for Independent Learning	0.71	82.5	Substantial
Commitment	0.68	79.5	Substantial

Note. Cohen's Kappa interpretation follows current guidance on selecting and interpreting kappa statistics in inter-rater reliability research: 0.61–0.80 = Substantial agreement; 0.81–1.00 = Almost perfect agreement [12].

The implementation study demonstrated substantial time savings through use of the Support Vector Machine-based system. Table 9 presents a comparison of manual versus automated comment analysis. Manual coding of 100 comments required an average of 4.2 hours, while the automated system processed the same volume in under 2 minutes; including time for human review of system outputs, total analysis time was reduced by 68% compared to fully manual methods.

Table 9. Efficiency Comparison: Manual vs. Automated Comment Analysis

Analysis Method	Time for 100 Comments	Time for 1,000 Comments	Human Review Time	Total Time Reduction
Manual Coding	4.2 hours	42 hours	N/A	Baseline
Automated SVM	2 minutes	20 minutes	1.5 hours	68%
Hybrid Approach	1.5 hours	15 hours	1.5 hours	64%

Note. Hybrid approach includes automated classification with selective human verification of uncertain predictions. Time reduction calculated against fully manual baseline.

Program heads and faculty members reported generally positive experiences with the system. Figure 3 illustrates the user satisfaction ratings across different aspects of the Support Vector Machine-based evaluation system. On a 5-point scale where 1 represented not useful and 5 represented extremely useful, the average usefulness rating was 4.1 (standard deviation = 0.7).

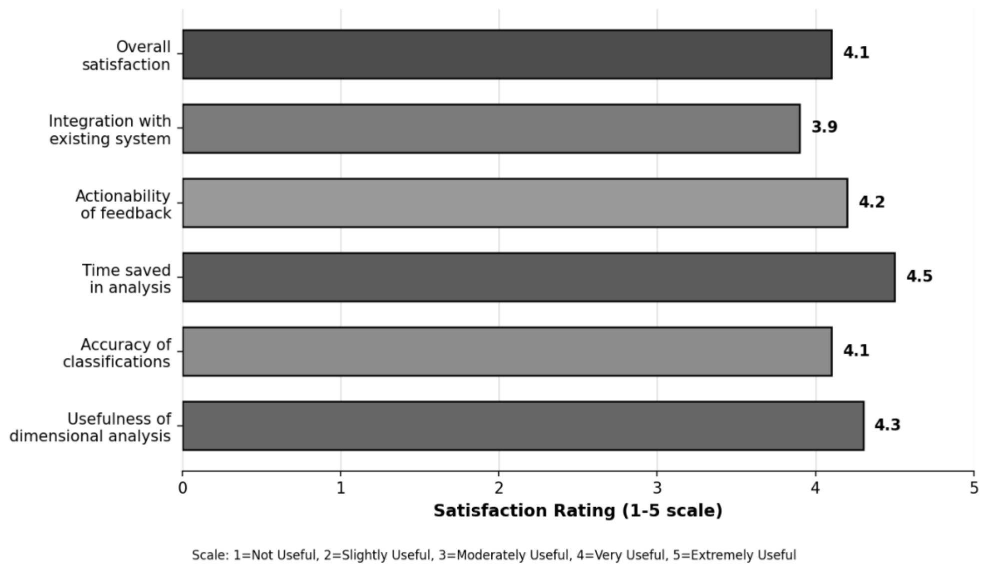


Figure 3. User Satisfaction with SVM-Based Evaluation System

Note. Ratings from program heads/chairs (n=4) and faculty members (n=20) after one semester of system use. Overall satisfaction mean = 4.1 (SD = 0.7).

4. DISCUSSION

The performance metrics achieved in this research demonstrate competitive results when compared to similar applications of machine learning in educational evaluation contexts. The classification accuracy achieved aligns with findings from recent applications of Support Vector Machines for analyzing student feedback and evaluating teacher performance [9]. The results validate the effectiveness of Support Vector Machines for text categorization tasks in educational settings, consistent with applied research confirming their suitability for learning with text data [8]. The variation observed between dimensional models reflects the linguistic expression of different teaching aspects, with subject matter expertise being more explicitly expressed in student feedback compared to dispositional characteristics such as commitment.

The College of Teacher Education's superior performance (89.2% accuracy) may reflect the field's emphasis on explicit pedagogical terminology and assessment language. This finding aligns with observations that education-focused disciplines tend to employ more standardized vocabulary when discussing teaching practices. The College of Business and Management achieved 86.8% accuracy, reflecting the discipline's balance between theoretical knowledge and practical application in student feedback. The College of Computer Studies demonstrated 88.5% accuracy, with students frequently commenting on technical competence and problem-solving approaches. The College of Fisheries' relatively lower accuracy (84.7%) may reflect challenges in analyzing discourse from specialized technical fields, where domain-specific terminology may not align with general teaching effectiveness vocabulary.

The identification of specific linguistic features in Table 6 and Table 7 supports theoretical frameworks of teaching effectiveness. The prominence of "clear explain" (weight = 0.87) as the strongest positive predictor aligns with pedagogical research identifying clarity of explanation as a fundamental component of teaching quality. The high negative weight of "confus" (-0.82) and "not understand" (-0.69) reflects patterns where confusion-related terms predict poor teaching evaluations. The positive weight of "respons question" (0.78) supports research emphasizing teacher responsiveness and availability as critical components of teaching effectiveness, consistent with the Commitment dimension of the evaluation framework. The feature "engag" (weight = 0.76) reflects the growing emphasis on active learning and student engagement in contemporary pedagogical research. The bigram "real world" (weight = 0.65) indicates students value connections between course content and practical applications, supporting experiential learning theories. The unigram "feedback" (weight = 0.62) aligns with research on the importance of timely and constructive feedback in the learning process. Negative features like "waste time" (weight = -0.60) and "unhelpful" (weight = -0.73) provide specific indicators of teaching practices that detract from effectiveness, offering actionable insights for faculty development.

The agreement levels reported in Table 8 demonstrate substantial concordance between automated classifications and expert human judgments across all five models, ranging from Cohen's Kappa of 0.68 (Commitment) to 0.79 (Overall Effectiveness). These results validate the approach while suggesting that the automated system can reliably support, though not replace, human judgment in teaching evaluation processes.

The 68% time reduction reported in Table 9 demonstrates significant efficiency gains through machine learning implementation. These efficiency gains are particularly valuable in institutional contexts where evaluation data must be processed for large numbers of courses and instructors each academic term. Participants in the satisfaction survey particularly valued the dimensional analysis and the extraction of representative comments, features that provide actionable insights for faculty development initiatives. Faculty members reported that the categorized feedback helped them identify specific areas for improvement more efficiently than reading through all comments individually, and department chairs noted that the system facilitated more focused conversations about teaching improvement during faculty evaluation meetings. The time saved in analysis (rated 4.5 out of 5) was identified as the most valuable aspect, followed by the usefulness of dimensional analysis (4.3 out of 5) and the actionability of feedback (4.2 out of 5).

4.1 Implications for Institutional Practice

These findings suggest that the University of Antique should consider full-scale implementation of the Support Vector Machine-based classification system as a supplement to existing teaching evaluation processes. Table 10 presents a phased implementation roadmap for integrating the automated classification system into institutional evaluation workflows.

Table 10. Phased Implementation Roadmap for SVM-Based Evaluation System

Phase	Timeline	Key Activities	Deliverables	Responsible Unit
Phase 1: Pilot Testing	Months 1-3	System integration testing, user training, sample data processing	Pilot test report, user manual	IT Services, Assessment Office
Phase 2: Limited Deployment	Months 4-6	Deploy in 2 colleges, gather user feedback, refine classifications	Deployment report, feedback analysis	College Deans, Faculty
Phase 3: Full Implementation	Months 7-9	Deploy across all colleges, train all users, establish support system	Full deployment, training materials	All Colleges, HR

Phase	Timeline	Key Activities	Deliverables	Responsible Unit
Phase 4: Evaluation	Months 10-12	Assess system impact, compare with manual methods, plan improvements	Impact assessment, improvement plan	Assessment Office, Research Unit

Note. Timeline assumes one-year implementation period with quarterly evaluation checkpoints.

The system should be integrated into the institutional evaluation workflow to provide automatic categorization of student comments according to the four effectiveness dimensions. Faculty members should receive training on interpreting the automated classifications and using the dimensional feedback for continuous improvement. Department chairs and academic administrators should utilize the aggregated results to identify college-wide or program-specific patterns in teaching effectiveness, informing targeted faculty development initiatives.

Implementation should maintain human oversight through a validation process where a sample of automated classifications is reviewed by evaluation experts each semester. This quality assurance mechanism ensures continued accuracy while allowing the system to adapt to evolving evaluation language. An effective faculty feedback report would be organized around a small set of recurring elements that the institution can render in whatever reporting format it adopts. At minimum, such a report should present the faculty member's overall effectiveness rating alongside the corresponding college benchmark; a breakdown of scores across the four dimensions (Commitment, Knowledge of Subject, Teaching for Independent Learning, and Management of Learning), each likewise compared to its benchmark; a short list of recurring strengths drawn from the positive-feature analysis, together with the number of student comments mentioning each one; a parallel list of recurring areas for improvement drawn from the negative-feature analysis; and a small set of representative, anonymized student comments illustrating both the positive feedback and the developmental feedback for that faculty member. Presenting the report in this structured but format-agnostic way keeps the emphasis on the underlying data points that should be communicated to faculty, rather than on any single visual layout, and allows each college to adapt the presentation to its own reporting tools.

The institution should also establish feedback mechanisms to continuously improve the classification system. Faculty members and administrators should be encouraged to report misclassifications or ambiguous categorizations, which can inform periodic model retraining. The system should be updated with new training data each academic year to capture changes in student evaluation language and teaching practices. Regular assessment of inter-rater reliability between the automated system and human evaluators should be conducted to monitor system performance over time.

4.2 Implications for Faculty Development

The feature importance findings should inform faculty development programs by highlighting specific teaching practices that students associate with effectiveness. Table 11 outlines recommended faculty development initiatives based on the linguistic features identified in the analysis.

Table 11. Recommended Faculty Development Initiatives Based on Feature Analysis

Teaching Effectiveness Dimension	Key Features Identified	Recommended Workshop Topics	Target Audience
Commitment	Availability, responsiveness, preparation	"Beyond the Classroom: Effective Student Support"	All faculty
Knowledge of Subject	Clear explanations, current knowledge, practical examples	"Clarity in Teaching: Communication Strategies"	New faculty
Teaching for Independent Learning	Engagement, critical thinking, creativity	"Fostering Student Independence and Creativity"	All faculty

Teaching Effectiveness Dimension	Key Features Identified	Recommended Workshop Topics	Target Audience
Management of Learning	Active learning, varied activities, feedback	"Interactive Teaching Methods and Assessment"	Mid-career faculty

Note. Workshops should be offered semesterly with optional attendance for experienced faculty and mandatory participation for new faculty.

Workshops and training sessions should address the key positive features identified in the analysis, such as explanation clarity, question responsiveness, and engagement techniques. The negative features should be used to help faculty identify and avoid practices that students perceive as ineffective, such as confusing explanations, limited availability, or lack of real-world connections.

4.3 Limitations and Future Research Directions

This study is limited by its single-institution scope and its dependence on the existing University of Antique evaluation framework, which may constrain the generalizability of the classification framework to other institutional contexts. Future research should expand the application of Support Vector Machine classification to other Philippine state universities and colleges using the CHED-TESDA-PASUC standardized evaluation instrument. Cross-institutional validation would demonstrate the generalizability of the classification approach and potentially enable benchmarking of teaching effectiveness across institutions. Researchers should investigate whether a unified model trained on multi-institutional data performs better than institution-specific models, or whether local contextual factors require customized approaches.

Additional research directions include exploring more advanced natural language processing techniques, such as transformer-based models, to determine whether they provide significant improvements in classification accuracy or interpretability compared to Support Vector Machines. Longitudinal studies should examine how teaching effectiveness feedback patterns change over multiple semesters, potentially identifying developmental trajectories in teaching practice. Future investigations should also incorporate additional data sources, such as classroom observation protocols and student performance data, to develop more comprehensive models of teaching effectiveness that integrate multiple evidence sources.

5. CONCLUSIONS

This research successfully demonstrated that Support Vector Machines can effectively classify unstructured student comments according to dimensions of teaching effectiveness at the University of Antique Tario-Lim Memorial Campus. The models achieved high accuracy across the four dimensions of the University's teaching effectiveness evaluation instrument, with performance levels that demonstrate the viability of machine learning approaches for educational evaluation. The strong agreement with expert evaluations validates the approach, while the feature importance analysis supports the institution's evaluation framework. The integration of machine learning approaches with traditional evaluation methods offers a promising pathway for developing more comprehensive, efficient, and effective teaching assessment systems that can support both summative evaluation and formative faculty development processes in Philippine higher education institutions. While limitations related to context-sensitivity, linguistic complexity, and generalizability remain, this research establishes a foundation for more sophisticated applications of machine learning in teaching evaluation within the Philippine higher education context. The successful implementation at the University of Antique demonstrates that computational approaches can transform qualitative feedback into actionable insights while maintaining the nuanced understanding that effective teaching requires multiple dimensions of practice.

List of Abbreviations

AUC: Area Under the Receiver Operating Characteristic Curve; CBM: College of Business and Management; CCS: College of Computer Studies; CHED: Commission on Higher Education; COF: College of Fisheries; CTE: College of Teacher Education; NLTK: Natural Language Toolkit; PASUC: Philippine Association of State Universities and Colleges; RBF: Radial Basis Function; SVM: Support Vector Machine; TESDA: Technical Education and Skills Development Authority; TF-IDF: Term Frequency-Inverse Document Frequency.

Acknowledgments

The authors gratefully acknowledge the administration, faculty, staff, and students of the University of Antique, Tario Lim Memorial Campus, who served as evaluators of the system during this study, and the University of San Carlos for institutional and academic support throughout this research.

Funding Information

This research did not receive any external funding.

Author Contribution

N.M.Y.: Conceptualization, methodology, data collection, software development, formal analysis, investigation, writing original draft. A.M.C.: Supervision, validation, methodology review, writing review and editing. All authors have read and agreed to the published version of the manuscript.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The anonymized data supporting the findings of this study are available from the corresponding author upon reasonable request. Raw student comments are not publicly available in order to protect student and faculty privacy, consistent with the anonymization procedures described in the Materials and Methods section...

References:

1. F. Quansah, A. Cobbinah, K. Asamoah-Gyimah, J. E. Hagan Jr. "Validity of Student Evaluation of Teaching in Higher Education: A Systematic Review", Frontiers in Education, 9, Article 1329734, 2024.
2. Commission on Higher Education. "A.C.H.I.E.V.E. 2030: CHED's Roadmap for Philippine Higher Education", Available at: https://unifast.gov.ph/assets/pdf/infgrfx/ched-achieve.pdf (Accessed on: 22 June 2026), 2024.
3. G. Largoza, C. G. Fernandez. "Review of CHED Policies, Standards, and Guidelines (PSGs) Pre- and Post-K to 12 Reforms", PIDS Discussion Paper Series No. 2025-13, Philippine Institute for Development Studies, 2025.
4. A. Pineda-Briseño, J. Oblitas Cruz, L. Cleofas Sánchez, W. Sanchez, R. Baltazar. "A Context-Aware Framework for Sentiment Analysis of Student Feedback to Inform Educational Strategies in Latin America", Education Sciences, 16(3), Article 399, 2026.
5. J. Ormilla. "Bridging the Gap: A Qualitative Analysis of the Philippine Professional Standards for Teachers and Teaching Internship Experiences of Practice Teachers", International Journal of Learning, Teaching and Educational Research, 24, 2025.
6. M. B. Garcia. "Text Mining as an Educational Evaluation Methodology: Analyzing Textual Data Extracted from Online Learning Environments", Learning Environments Research, 2025.
7. C. Romero, S. Ventura. "Educational Data Mining and Learning Analytics: An Updated Survey", arXiv:2402.07956, 2024.
8. Z. Quan, L. Pu. "An Improved Accurate Classification Method for Online Education Resources Based on Support Vector Machine (SVM): Algorithm and Experiment", Education and Information Technologies, 28(7), pp. 8097–8111, 2023.
9. Y. Li, J. Liu, W. Ji, Y. Li. "Application of Support Vector Machine (SVM) in Teacher Performance Evaluation", in Intelligent Communication Technologies and Applications (ICAICT 2025), Smart Innovation, Systems and Technologies, Vol. 452, Springer, 2025.
10. A. Verma, K. Lamsal, P. Verma. "Data Mining Techniques for Predicting Teacher Evaluation in Higher Education: A Systematic Literature Review", Heliyon, 9(3), Article e14227, 2023.
11. M. Peters, S. Fàbregues. "Missed Opportunities in Mixed Methods EdTech Research? Visual Joint Display Development as an Analytical Strategy for Achieving Integration in Mixed Methods Studies", Educational Technology Research and Development, 2023.
12. M. Li, Q. Gao, T. Yu. "Kappa Statistic Considerations in Evaluating Inter-Rater Reliability between Two Raters: Which, When and Context Matters", BMC Cancer, 23, Article 803, 2023.