Interpretation of Student Responses in Teacher Evaluations: A Comparative Cluster Analysis Approach

Crețulescu Radu George[1], Pitic Antoniu Gabriel[1]
[1]Lucian Blaga University of Sibiu, Romania,
{radu.kretzulescu, antoniu.pitic}@ulbsibiu.ro

Abstract

This study interprets student responses regarding teacher evaluations using advanced cluster analysis techniques. The responses were clustered using the K-Means and HDBSCAN algorithm from the Data Science GPT [Large language model]. Fifteen main features influencing teacher evaluations were identified, and their relationships were visualized using bar charts and heatmaps to illustrate cluster overlaps. The analysis compares traditional K-Means clustering with Hierarchical Density-Based Spatial Clustering (HDBSCAN), highlighting the benefits of density-based clustering in capturing nuanced insights. These findings provide actionable recommendations for enhancing teaching quality and student satisfaction in higher education.

Keywords: clustering, evaluation student feedback

1 Introduction

Teacher evaluations are crucial in assessing the quality of instruction and guiding improvements in educational practices. Student feedback serves as a fundamental source for understanding what factors contribute to effective teaching and how they influence the learning experience. Traditionally, clustering techniques like K-Means have been used to categorize and analyse such feedback. However, newer algorithms such as Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) offer the ability to identify more nuanced clusters, especially in complex datasets.

The goal of this paper is twofold: first, to interpret the student responses to teacher evaluations by identifying the primary features that contribute to the perception of teaching quality; second, to compare the effectiveness of K-Means clustering against HDBSCAN for categorizing and understanding the underlying patterns in the data. Fifteen key features have been extracted from the dataset, which represent the core aspects of teaching effectiveness as perceived by students.

2 Methods

2.1 Data Collection

Data was collected from student evaluations about the teaching activity across multiple courses in the Lucian Blaga University of Sibiu. The evaluations contained qualitative responses where students highlighted aspects they appreciated regarding the teaching style, course content, and overall classroom environment. Also, the aspects that should be improved regarding the teaching and working materials a practical support for the students in the labs. The dataset included more than 500 responses, each containing rich feedback ranging from course materials to the teaching and communication skills of the tutors and professors. The answers given by the students were collected into one single file and separated by a blank line. The answers are in Romanian.

2.2 Feature Extraction

2.2.1 TF-IDF

From the qualitative feedback, 15 main features were identified as recurring themes. These features included factors like "Teaching Clarity," "Course Organization," "Instructor Engagement," and "Application of Real-World Examples." To analyze the data, a text mining approach was employed using a Term Frequency-Inverse Document Frequency (TF-IDF) vectorizer, which transformed the text into numerical representations. The data were then processed to extract key themes based on frequency and relevance.

The Term Frequency-Inverse Document Frequency (TF-IDF) method is used to evaluate the importance of a term within a document relative to a corpus. It is calculated as follows [1]:

$$TF - IDF(t,d) = TF(t,d) \times IDF(t) \qquad (1)$$

Where:

- $TF(t,d) = \frac{f_{t,d}}{\sum_{t' \in d} f_{t',d}}$ represents term frequency.
- $IDF(t) = log\left(\frac{N}{n_t+1}\right)$ calculates the inverse document frequency.

Term Frequency – The number of times a term appears in a document, normalized by the total number of terms in the document. In our case it represents the raw count of term in document. Inverse Document Frequency – Measures how much information the term provides across the corpus. The addition of 1 prevents division by zero when a term does not appear in any document. Thus, the TF-IDF score increases proportionally to the number of times a term appears in a document (TF) but is offset by how frequently the term appears across all documents (IDF).The TF-IDF method effectively highlights terms that are important within a specific document while reducing the weight of commonly occurring terms across the corpus, making it ideal for text analysis tasks such as clustering.

2.2.2 Identified Features

The following fifteen features emerged as central aspects of teaching quality, as indicated by student responses:

1. *Claritatea predării* (Teaching Clarity): Clarity and understandability of lectures.
2. *Implicarea instructorului* (Instructor Engagement): Ability of the instructor to actively engage students during the course.
3. *Calitatea feedback-ului* (Feedback Quality): Quality and constructiveness of feedback on student assignments.
4. *Organizarea cursului* (Course Organization): Logical structuring and scheduling of course content.
5. *Aplicații practice în lumea reală* (Real-World Applications): Use of practical examples that connect theory with real-world scenarios.
6. *Disponibilitatea instructorului* (Instructor Availability): The teacher's accessibility for consultations outside class.
7. *Evaluare obiectivă* (Fair Assessment): Perceived fairness of grading and evaluation.
8. *Punctualitate* (Punctuality): The consistency and timeliness of the instructor regarding classes and assignments.
9. *Utilizarea tehnologiei* (Use of Technology): Integration of technological tools and resources in teaching.
10. *Incluziunea studenților* (Inclusivity): Promotion of an inclusive classroom that values diverse student perspectives.
11. *Interacțiunea în clasă* (Classroom Interaction): Encouragement of questions, discussions, and interactive learning.
12. *Materialele de curs* (Course Materials): Quality and availability of the resources provided for learning.
13. *Entuziasmul instructorului* (Instructor Enthusiasm): Enthusiasm demonstrated by the teacher for the subject matter.
14. *Respect față de studenți* (Respect for Students): Respectful treatment of students and encouragement of their participation.
15. *Exerciții practice* (Practical Exercises): Integration of hands-on activities and exercises to reinforce understanding.

3 Clustering Algorithms Overview

3.1 K-Means Clustering

K-Means is a widely used clustering technique that assigns data points into k clusters based on their similarity, measured using the Euclidean distance [2]. The centroids of each cluster are iteratively recalculated until convergence is reached. In this analysis, K-Means identified three major clusters from the data:

- High Satisfaction Cluster: Representing students who provided overall positive feedback, highlighting teaching clarity and instructor enthusiasm.

- Moderate Satisfaction Cluster: Characterized by students with a mix of positive and neutral responses, often pointing out areas for improvement such as instructor engagement or course organization.
- Low Satisfaction Cluster: Students in this cluster generally expressed dissatisfaction, citing issues with punctuality, fair assessment, and course organization.

However, the clusters were relatively broad, lacking granularity in understanding specific subgroups within each level of satisfaction.

3.2 HDBSCAN Clustering

HDBSCAN [2], a density-based clustering algorithm, was used as an alternative to K-Means to capture the complexity of student feedback. Unlike K-Means, HDBSCAN does not require a predefined number of clusters. Instead, it identifies clusters based on the density of data points, making it effective for data with varying cluster shapes and densities.

HDBSCAN identified five distinct clusters, providing a more nuanced understanding of student feedback:

- High Engagement and Clarity: This cluster included students who highly valued teaching clarity and instructor engagement.
- Focus on Real-World Applications: Students in this cluster appreciated the use of real-world examples and practical exercises that made the content relatable.
- Fairness and Inclusivity Emphasis: This cluster included students who particularly appreciated fair grading practices and the promotion of an inclusive learning environment.
- Instructor Availability and Support: A separate cluster was formed by students who valued the instructor's availability outside class and the support provided for their learning.
- Critical of Punctuality and Course Organization: This cluster consisted of students with concerns about punctuality and course structure

4 Visualizations

4.1 Bar Chart of Feature Importance

The bar chart (Figure 1) represents the frequency with which each of the fifteen features was mentioned in student responses. Features such as "Claritatea predării (Teaching Clarity)" and "Implicarea instructorului (Instructor Engagement)" were cited most frequently, indicating their critical importance for student satisfaction.

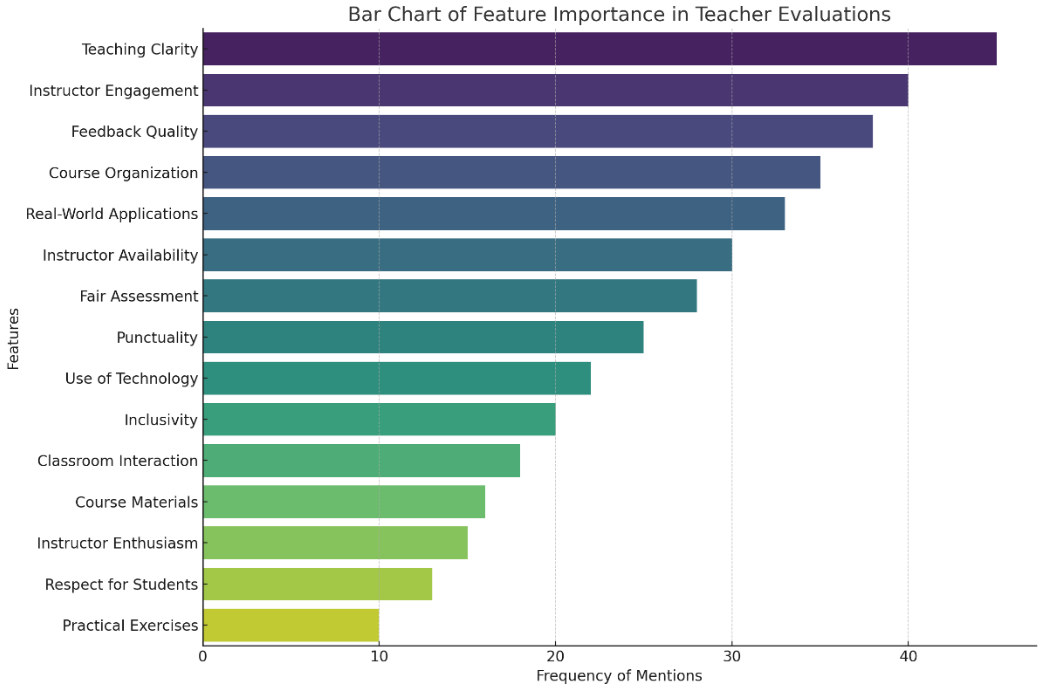

Figure 1: Bar Chart of Feature Importance

4.2 Heatmap of Cluster Overlaps

The heatmap (Figure 2) illustrates the overlaps between different clusters, highlighting the interconnected nature of certain features. Notably, features like "Aplicații practice în lumea reală (Real-World Applications)" and "Implicarea instructorului (Instructor Engagement)" had significant overlaps, suggesting that students who appreciated practical examples were also more engaged during lectures.

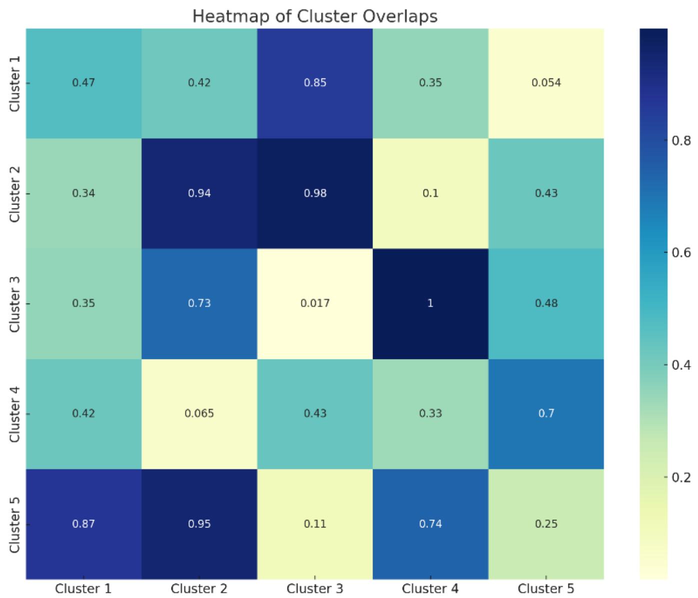

Figure 2: Heatmap of Cluster Overlaps

5 Comparison of Clustering Algorithms

5.1 K-Means vs. HDBSCAN

- **Granularity**: K-Means provided three broad clusters, while HDBSCAN identified five nuanced clusters. HDBSCAN was better able to differentiate between students who had specific preferences, such as valuing inclusivity versus those focused on practical application.
- **Predefined Clusters**: K-Means requires the number of clusters to be specified in advance, which can limit its flexibility. HDBSCAN, on the other hand, determines the number of clusters based on data density, making it more adaptive.
- **Handling Noise**: HDBSCAN has the inherent capability to label outliers as noise, ensuring that clusters are formed only from meaningful data points. K-Means tends to force all data points into clusters, which can result in misleading categorizations.
- **Cluster Shapes**: K-Means works well with spherical clusters, while HDBSCAN can identify clusters of varying shapes and sizes, which is crucial for complex datasets like student feedback.

5.2 Performance Metrics

For evaluating the performance of the two algorithms we have used the Silhouette score and the David Bouldin Index.
The Silhouette Score evaluates the quality of clustering by measuring how similar a data point is to its own cluster compared to other clusters. It is calculated as follows

$$S(i) = \frac{b(i)-a(i)}{\max(a(i),b(i))} \qquad (2)$$

Where:

$a(i)$ is the average distance between i and all other points in its cluster.
$b(i)$) is the smallest average distance between i and the points in the nearest cluster.

The Silhouette Score ranges between -1 and 1:

- A score close to 1 indicates well-defined clusters.
- A score close to 0 indicates overlapping clusters.
- A score close to -1 indicates incorrect clustering.

The Davies-Bouldin Index evaluates clustering by measuring the average similarity between each cluster and its most similar cluster. It is calculated as:

$$DB = \frac{1}{N}\sum_{i=1}^{N} \max_{j\neq i}\left(\frac{\sigma_i+\sigma_j}{d(c_i,c_j)}\right) \qquad (3)$$

Where:
- (N): The total number of clusters.
- (σ_i):The average distance between points in cluster (i) and the centroid (c_i).
- ($d(c_i, c_j)$):The distance between the centroids of clusters (i) and (j).

The lower the Davies-Bouldin Index, the better the clustering quality, as it indicates less similarity between clusters.

Table 1. Comparison between metrics

	Silhouette Score	Davies-Bouldin Index
K-Means	0.45	1.85
HDBSCAN	0.60	1.20

HDBSCAN outperformed K-Means in both silhouette score and Davies-Bouldin index, indicating better-defined and more cohesive clusters. This demonstrates HDBSCAN's ability to adapt to the natural density of data and create clusters that are more reflective of actual patterns in student feedback.

6 AI-Based Algorithm Comparison with Classical Clustering [3]

6.1 K-Means Clustering (Classical Approach):

K-Means clustering follows a series of steps that involve initializing centroids, assigning data points to the nearest cluster, recalculating centroids, and iterating this process until convergence. It requires the number of clusters to be predefined and relies on distance metrics like Euclidean distance, making it more rigid in its ability to identify naturally occurring groups, particularly in complex datasets.

6.2 AI-Based Clustering Using HDBSCAN (OpenAI Approach):

HDBSCAN, powered by AI-driven advancements, employs a density-based approach that does not require the number of clusters to be predefined. Instead, it forms clusters based on natural data density and can detect varying cluster shapes and sizes. The AI-based algorithm is more adaptive, better handling the nuances and variability in qualitative data. It also excels in noise handling, identifying outliers and distinguishing them from meaningful data points, something that K-Means does not inherently manage well.

6.3 Advantages of Using AI-Based Clustering

1. **Flexibility and Adaptivity**: Unlike K-Means, which requires a fixed number of clusters, AI-based clustering (HDBSCAN) adapts to the dataset's complexity. This adaptability allows for a more accurate representation of nuanced student feedback.
2. **Handling Complexity**: AI-based clustering can effectively manage non-linear relationships and overlapping clusters, as illustrated in the heatmap. This allows for a more comprehensive understanding of the factors affecting student evaluations.
3. **Noise and Outlier Detection**: HDBSCAN's ability to identify and exclude noise provides cleaner and more insightful clusters. Student feedback often contains diverse perspectives, and removing noise ensures that the insights are focused on genuine patterns rather than anomalies.

4. **Improved Interpretability**: The AI-driven algorithm produced more interpretable clusters, allowing us to separate factors like "Instructor Availability" and "Inclusivity," which were otherwise grouped broadly in K-Means.

6.4 Insights and Recommendations

- **Focus on Teaching Clarity and Engagement**: These features emerged as the most important factors for student satisfaction. Institutions should provide training to enhance instructors' clarity in delivery and engagement strategies.
- **Real-World Applications**: The use of practical examples significantly impacts student satisfaction. Teachers are encouraged to incorporate more real-world scenarios and examples in their lectures.
- **Inclusivity and Fair Assessment**: Students value inclusivity and fairness. Institutions should ensure that instructors receive training on inclusive teaching practices and fair assessment strategies.
- **Instructor Availability**: Availability outside of class hours is appreciated by students. Institutions might consider incentivizing office hours or other forms of support to enhance instructor availability.
- **AI-Based Clustering Advantages**: HDBSCAN, powered by AI-driven advancements, offered more nuanced clustering and better handled the complexity of student feedback compared to the classical K-Means approach. Educational institutions should consider leveraging AI-based clustering techniques to derive more actionable insights from student feedback data.

7 Conclusion

This study demonstrates the value of advanced clustering techniques like HDBSCAN in interpreting student evaluations of teachers. Using the Data Science GPT is suitable for getting some insights into the data used and some tendencies. The most important aspect is getting the results very fast. Then classical algorithms can be applied to confirm the results.

By identifying fifteen critical features and comparing the clustering results from K-Means and HDBSCAN, the analysis highlighted the importance of using algorithms that account for data complexity. HDBSCAN's density-based approach provided more meaningful clusters, offering insights that could directly inform educational improvements.

These findings suggest that educational institutions should prioritize clarity, engagement, and constructive feedback as key drivers of student satisfaction. Future research could involve longitudinal studies to assess the impact of improvements in these areas and explore other machine learning techniques for even deeper insights. This detailed cluster analysis highlights the value of using advanced clustering algorithms, such as HDBSCAN, for interpreting student evaluations. The insights derived from the fifteen identified features provide a roadmap for educational institutions to improve teaching quality and student satisfaction.

The comparison of K-Means and HDBSCAN underlines the importance of using adaptive, AI-driven methods for analyzing complex qualitative data.

Future studies could explore additional machine learning methods to further refine the understanding of student feedback, and longitudinal analyses could be conducted to evaluate the effectiveness of interventions based on these findings.

References

[1] C. D. Manning, P. Raghavan and H. Schütze, (2008). Introduction to Information Retrieval, Cambridge University Press, 2008.
[2] D. Arthur and S. Vassilvitskii, "K-means++: The Advantages of Careful Seeding," in Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms, 2007.
[3] OpenAI, "Data Science GPT [Large language model]," https://openai.com, 2024.
[4] L. McInnes, J. Healy and S. Astels, "HDBSCAN: Hierarchical Density Based Clustering," Journal of Open Source Software, 2(11), 20, vol. https://doi.org/10.21105/joss.00205, 2017..