13th International Conference on Applied Informatics
Eszterházy Károly Catholic University
Eger, Hungary, February 19–21, 2026

Student Opinion Mining: Automated Topic Extraction from Student Feedback

Olivér Czimbalmos[a], Zsolt Szántó[a], Gábor Kőrösi[a], Péter Becsei[a], Beáta Udvardi[b], Richárd Farkas[a]

[a]Institue of Informatics, University of Szeged,
czimbalmos@inf.u-szeged.hu,

[b]Directorate of Academic Affairs, University of Szeged

Introduction

Student Evaluation of Teaching Work questionnaires are a cornerstone of quality assurance in Hungarian higher education. While quantitative ratings are straightforward to analyze, students also provide large volumes of unstructured textual feedback. The manual processing of these textual reviews is labour-intensive, which often results in their underutilization. The goal of this study is the automatic analysis of student reviews. In this paper, we introduce a pipeline that analyses thousands of student reviews via high-level topic classification followed by fine-grained topic clustering and sentiment analysis. The output of these analyses' steps are visualized in a user-friendly way.

The faculties of the University of Szeged employ various questionnaires, but recently, the Directorate of Academic Affairs developed a new, standardized category system. The first step of our proposed pipeline (re)classifies student reviews originating from diverse questionnaires into the unified set of six standardized classes by employing Large Language Models (LLMs). To obtain a more detailed picture of the topics reported by students, we broke these classes down into smaller latent topics, thereby revealing critical institutional strengths and issues that are often overlooked by manual analysis. The fine-grained latent topics are automatically labeled. Lastly, sentiment scores are assigned to each of the student reviews.

In this paper, we compare different methods and machine learning models to classify and clusterize Hungarian student reviews. Our results demonstrate that text mining can transform raw student feedback into actionable policy insights for university management.

Methodology and Results

For the automatic analysis of student reviews, we used data from two faculties, comprising a corpus of about 34,000 unique reviews. We developed a processing pipeline that integrates four main stages: (1) classify the old questionnaires into a standardized category system, (2) automatic generation of fine-grained topic clusters, (3) labeling of the clusters, and (4) sentiment analysis of the reviews.

Recent studies have increasingly utilized Natural Language Processing (NLP) to analyze student feedback [1]. To the best of our knowledge, no automated NLP framework currently exists for Hungarian Student Evaluation of Teaching Work analysis, and universities typically publish only aggregated quantitative statistics. Our methodology aligns most closely with prior work on Hungarian patient feedback [7]. We advance this approach by replacing smaller models with high-parameter LLMs to leverage their extensive semantic knowledge for nuanced sentiment detection and by introducing a fully automated topic labeling mechanism.

To map the reviews to the new standardized category system, we employed Large Language Models using a few-shot prompting strategy. For evaluating the classification accuracy, we constructed a benchmark set of 200 manually annotated examples and compared the performance of three LLMs. Llama-70B-Instruct [3] achieved the highest weighted macro F1 score (0.72 for classification), outperforming the Qwen2-72B-Instruct [8] and DeepSeek-llm-67b-chat [2] models. Based on these results, we selected the Llama model for all subsequent experiments.

After the mapping, we further subdivided each class into fine-grained latent topics. LLMs are not feasible to use for topic modeling a corpus of ten thousand of reviews. Instead of we apply textual embeddings to convert reviews to vectors and classterized them with HDBSCAN. The adoption of Qwen3-embedding-8B [9] provided a substantial performance leap over previous BERT-based models (e.g., Jina-V2 [4]). However, the pipeline presents specific challenges regarding stability. We observed that HDBSCAN [5] is highly sensitive to hyperparameter selection; minor adjustments to the *minimum cluster size* can lead to either excessive fragmentation or over-generalization of topics. Furthermore, distinct clusters occasionally exhibit semantic overlap, suggesting that a future hierarchical merging step is necessary to optimize the distinctness of the identified themes. Figure 1 show the subdivision of a category after 2D dimension reduction of a reviews by UMAP [6].

We assigned labels to newly generated clusters, where comparative experiments on different prompting strategies showed that including a small number (10) of randomly selected examples from each cluster within a single prompt yielded better results than more sophisticated approaches, such as using cluster centroids or issuing separate prompts for each cluster.

To further enhance the interpretability of the clusters, we assigned sentiment values to them. Using a few-shot prompt, each review was analyzed separately, and the resulting sentiment scores were then aggregated at the cluster level.

Our framework classifies student feedback into expert-defined categories and is further capable of automatically decomposing these categories into latent clus-

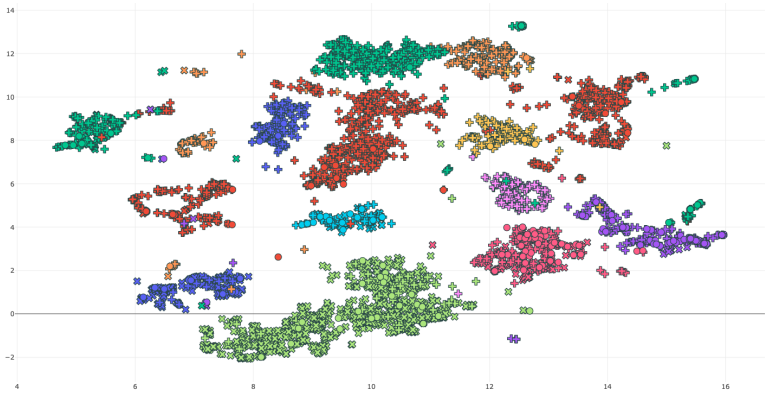

Figure 1. 2D projection of latent topics in one of the categories.

ters. The interpretation of these clusters is supported through automatic labeling and sentiment detection. Overall, we demonstrate an automated framework that transforms raw student feedback into actionable insights, supporting institutional decision-making and educational policy development.

References

[1] A. Abdi, G. Sedrakyan, B. Veldkamp, J. van Hillegersberg, S. M. van den Berg: *Students feedback analysis model using deep learning-based method and linguistic knowledge for intelligent educational systems*, Soft Computing 27.19 (2023), pp. 14073–14094.

[2] X. Bi, D. Chen, G. Chen, S. Chen, D. Dai, C. Deng, H. Ding, K. Dong, Q. Du, Z. Fu, et al.: *Deepseek llm: Scaling open-source language models with longtermism*, arXiv preprint arXiv:2401.02954 (2024).

[3] A. Grattafiori, A. Dubey, A. Jauhri, A. Pandey, A. Kadian, A. Al-Dahle, A. Letman, A. Mathur, A. Schelten, A. Vaughan, et al.: *The llama 3 herd of models*, arXiv preprint arXiv:2407.21783 (2024).

[4] M. Günther, J. Ong, I. Mohr, A. Abdessalem, T. Abel, M. K. Akram, S. Guzman, G. Mastrapas, S. Sturua, B. Wang, M. Werk, N. Wang, H. Xiao: *Jina Embeddings 2: 8192-Token General-Purpose Text Embeddings for Long Documents*, 2023, arXiv: 2310.19923 [cs.CL].

[5] L. McInnes, J. Healy, S. Astels, et al.: *hdbscan: Hierarchical density based clustering.* J. Open Source Softw. 2.11 (2017), p. 205.

[6] L. McInnes, J. Healy, J. Melville: *Umap: Uniform manifold approximation and projection for dimension reduction*, arXiv preprint arXiv:1802.03426 (2018).

[7] M. Osváth, G. Yang Zijian, K. Kósa: *Magyar páciensek narratív tapasztalatainak elemzése BERT témamodellezéssel és szentimentelemzéssel*, XVIII. Magyar Számítógépes Nyelvészeti Konferencia, Szeged (2022), pp. 311–324.

[8] *Qwen2 Technical Report* (2024).

[9] Y. Zhang, M. Li, D. Long, X. Zhang, H. Lin, B. Yang, P. Xie, A. Yang, D. Liu, J. Lin, F. Huang, J. Zhou: *Qwen3 Embedding: Advancing Text Embedding and Reranking Through Foundation Models*, arXiv preprint arXiv:2506.05176 (2025).