MUKADIMAH
Jurnal Pendidikan, Sejarah, dan Ilmu-ilmu Sosial
Program Studi Pendidikan Sejarah FKIP - Universitas Islam Sumatera Utara, Medan
e-Journal: http://jurnal.uisu.ac.id/index.php/mkd | e-mail: mukadimah@fkip.uisu.ac.id
ISSN: 2622-1373 (Online)
ISSN: 2614-1159 (Print)
Vol. 9 No. 2
August 2025

Instructor Performance Analysis in Educational Contexts Based on Learner Evaluation Data: Integration of Clustering and Predictive Model

Santi Dwi Desy Lestari*, Master of Human Resource Development, Universitas Airlangga, Indonesia
Hendro Margono, Faculty of Social and Political Science, Universitas Airlangga, Indonesia

ABSTRACT

This study aims to analyze instructor performance in educational contexts by classifying instructors based on learner evaluation data through the K-Means clustering algorithm and developing a predictive model to support effective and targeted instructor development programs. The data were derived from learners' evaluations of instructors, covering aspects such as discipline and professionalism, mastery of subject matter, and pedagogical skills in delivering content. The results indicate that k=3 is the optimal cluster, producing three categories: Superior Instructor, Potential Instructor, and Developing Instructor. Furthermore, the predictive model demonstrates that the Naive Bayes algorithm outperforms XGBoost in performance prediction, achieving higher accuracy, recall, precision, and F1-scores. The integration of clustering and prediction proves effective in enabling faster, objective, and data-driven decisions for instructor development. These findings provide significant implications for educational institutions in establishing adaptive and sustainable systems of instructor evaluation and management.

ARTICLE HISTORY

Received 08/07/2025
Revised 26/07/2025
Accepted 05/08/2025
Published 01/10/2025

KEYWORDS

Data-driven learning; instructor development; instructor performance evaluation; k-means clustering; predictive model.

***CORRESPONDENCE AUTHOR**

santi.dwi.desy-2024@pasca.unair.ac.id

DOI: https://doi.org/10.30743/mkd.v9i2.11772

INTRODUCTION

Instructors play a crucial role in the implementation of learning since the improvement of learning outcomes largely depends on instructor quality. Learner evaluation data on instructors provides a comprehensive overview of their competence in interaction, material delivery, and teaching skills. Instructor performance, therefore, constitutes a critical factor that must be evaluated objectively and continuously. The evaluation results of learners represent a valuable source of data that can be harnessed to assess instructor performance. However, such data are often underutilized. Consequently, clustering and predictive models are needed to classify instructor performance based on learner evaluation results and to accelerate data-driven decision-making processes for designing targeted instructor development programs.

A study by Liu (2022) demonstrates that the K-Means algorithm is effective in identifying patterns and clustering instructors based on learner evaluation data. Instructor segmentation can serve as the basis for validating clusters using metrics such as average within-cluster distance and silhouette scores, which are widely employed in similar studies. These clusters subsequently provide labeled data for building predictive models capable of automatically estimating instructor performance. K-Means, one of the most widely used clustering algorithms in data mining, simplifies large datasets into homogeneous groups by partitioning data into clusters based on centroid distances, thereby uncovering hidden patterns in evaluation data.

In earlier research, Chahuán et al. (2025) successfully classified 314 lecturers into four clusters using K-Means, based on their digital capabilities, which were then used to design training programs tailored to each cluster. This finding confirms the utility of clustering in constructing instructor development strategies according to cluster-specific characteristics. Similarly, Chen and Wang (2023) applied clustering to group instructors based on their behavior in digital learning environments, producing categories of pedagogically active instructors and those more focused on content delivery. The clustering results were further used to provide learning recommendations adapted to each cluster's teaching style. In another study, Argentinia (2021) developed a model for grouping students according to learning styles in project-based learning using the K-Means algorithm. The findings showed that K-Means is effective in profiling learners based on perceptions and behaviors, providing a foundation for strategy development in education. Collectively, these studies affirm the relevance of K-Means in grouping instructors according to learner perceptions of their strengths and weaknesses.

The application of predictive models in instructor performance evaluation is equally valuable for efficiency and objectivity in strategic decision-making. Such models enable early detection of potential instructor performance, allowing institutions to anticipate declines in quality proactively (Aljabri & Khalil, 2021). Predictive approaches further support data-driven policy-making in instructor development (Romero & Ventura, 2020). Hence, predictive models constitute essential tools for modern educational systems that are responsive to ongoing advancements in learning technologies.

Despite the growing body of research on clustering and predictive modeling in education, there remains a significant gap in integrating these two approaches specifically for instructor evaluation in educational contexts. Many institutions continue to rely on traditional evaluation systems that are descriptive rather than prescriptive, often resulting in delayed interventions and less effective development strategies. Addressing this issue is critical because the quality of instructors directly shapes student learning outcomes and institutional competitiveness in an increasingly data-driven educational landscape. By combining clustering with predictive modeling, this study contributes not only to methodological innovation but also to practical solutions for educational institutions seeking to enhance adaptive, objective, and sustainable instructor management systems.

The Naive Bayes algorithm is effective and efficient when handling unbalanced data and has the ability to generate stable predictions (Zhang, 2004). In contrast, XGBoost is powerful for classification tasks because it combines boosting with regularization, thereby minimizing overfitting and improving accuracy (Chen & Guestrin, 2016). These two models are employed and evaluated to measure performance in predicting instructor performance classifications based on learner evaluation data. The results of these models are expected to provide a foundation for more objective and data-driven decisions in instructor development.

However, a key problem that persists in educational practice is the underutilization of learner evaluation data. Although institutions regularly collect evaluations from learners, the data are often processed in a descriptive manner only, with limited application for strategic instructor development. This leads to delays in identifying performance issues, difficulties in differentiating between instructors with varying strengths and weaknesses, and the absence of proactive measures to prevent a decline in instructional quality. Furthermore, existing evaluation methods are frequently subjective and lack mechanisms for early detection of performance trends. These challenges highlight the pressing need for integrating machine learning approaches into educational evaluation systems to ensure faster, more reliable, and evidence-based decision-making.

By referring to these previous research results, this article aims to cluster instructors based on learner evaluation data using the K-Means clustering algorithm, and to construct a prediction model capable of objectively detecting improvements or declines in instructor performance. The integration of clustering and predictive models is expected to map instructor performance more accurately and to accelerate data-driven decision-making in the design of development programs, thereby making instructor training and coaching more effective and efficient.

METHOD

This study adopts a quantitative research design using the Cross-Industry Standard Process for Data Mining (CRISP-DM) approach. The quantitative orientation is emphasized because the research relies on numerical evaluation scores collected from structured questionnaires. These scores were processed statistically and analyzed using clustering and predictive modeling techniques to generate measurable, replicable, and objective outcomes. The stages of the CRISP-DM methodology applied in this study are illustrated in Figure 1.

Figure 1. Research Methodology Flow Based on the CRISP-DM Approach

This flowchart illustrates the stages of the research methodology, beginning with *Business Understanding*, followed by *Data Understanding* and *Data Preparation*. The process then continues to the *Modeling* stage, where K-Means clustering is applied for classification, and predictive models are built using Naïve Bayes and XGBoost. The *Evaluation* stage employs the Elbow method, Silhouette coefficient, and performance metrics to validate the models. Finally, the *Deployment* stage provides recommendations for instructor development programs tailored to the clusters formed.

1. Business Understanding

   The primary objective of this research is to develop a data-driven system that can classify and predict instructor performance based on learner evaluation results. Through clustering and predictive models, the system supports objective decision-making in formulating instructor development strategies.

2. Data Understanding

   The population in this study consists of instructors actively teaching at a corporate education institution. Samples were selected using purposive sampling, with the requirement that instructors had received evaluations from all their learners. The evaluation scores were obtained from questionnaires that measured learners' perceptions of instructors across indicators such as professionalism/code of ethics, mastery of subject matter, and pedagogical abilities. The evaluation scores, expressed on a 1–10 scale, were processed numerically during cluster analysis.

Table 1. Sample Data from Survey Results on Participant Evaluation of Instructors

NO	INSTRUCTOR NAME	IN QUALITY STRUCTURE					
		Instructor is disciplined with time	Instructor looks neat and professional	Instructor explains the purpose and objectives of the material presented	Instructor builds an interesting and fun learning atmosphere	Instructors master the material presented	Instructor delivers the material clearly
1	VERY FERNANDO	9.4	9.4	9.2	9	9.4	9.2
2	SONI ASMAUL FUADI	9.4	9	9	9	9	9
3	KUSWANTONO	9.54	9.62	9.62	9.54	9.54	9.54
4	RONALD LAPASAU	9.5	9.5	9.5	9.38	9.5	9.5
5	FIRMAN RAHARJA	9.5	9.5	9.5	9.13	9.25	9.25
6	BARZALIUS AKBAR	9.25	9.5	9.25	8.5	9.5	9.5
7	SEPTIA WINIATI	9.75	9.63	9.63	9.38	9.63	9.63
8	PRAHARA LUKITO EFFENDI	9.5	9.67	9.67	9	9.5	9.5
9	MUDJIANTO	9	8.89	8.89	8.67	9	8.89
10	MARYANTO	9.57	9.57	9.43	9.14	9.43	9.29
11	SONI ASMAUL FUADI	8.8	9.6	9.8	9.2	9.4	9.4
12	R. ARI WARSONO	9.68	9.64	9.68	9.59	9.68	9.64

The data in Table 1 demonstrate the quantitative nature of the study, in which each evaluation aspect is numerically represented, enabling statistical grouping through clustering and predictive modeling.

3. Data Preparation

   Data preparation included cleansing activities such as removing duplicates, handling missing values, and addressing outliers. Additionally, K-Means cluster labeling was performed, which later served as the target label in building the predictive model.

4. Modeling

   At this stage, K-Means clustering was applied as an unsupervised learning algorithm capable of detecting hidden patterns without requiring prior labels. The clustering output generated groups of instructors with similar evaluation patterns, which then became the foundation for tailored coaching and development initiatives. Subsequently, predictive models were developed using the Naïve Bayes and XGBoost algorithms to forecast instructor performance trends.

5. Evaluation

   The analysis was conducted using RapidMiner software. For clustering validation, the Elbow method and Silhouette Coefficient were applied. The cluster results were labeled and interpreted to provide concrete recommendations for instructor development programs tailored to the characteristics of each cluster.

RESULT AND DISCUSSION

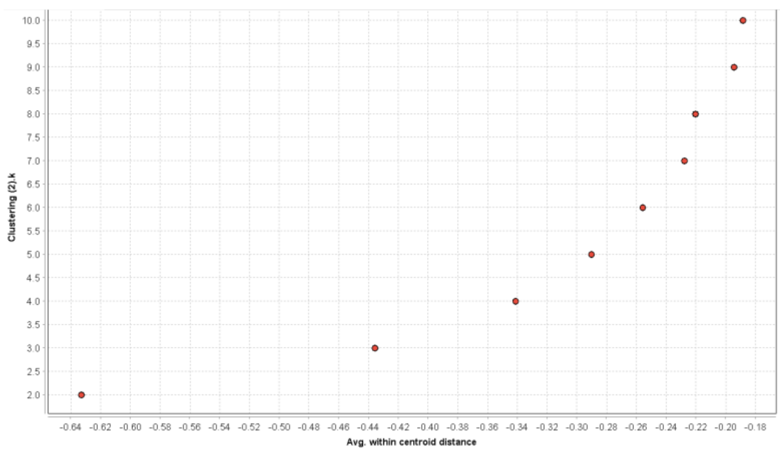

Figure 2. Elbow Method Results

Table 2. Sillhoute Score Results

k	Avg_Distance
3	0.44
2	0.63
4	0.34
8	0.22
10	0.19
6	0.26
5	0.29
7	0.23
9	0.19

Based on the determination of the optimal number of clusters, the difference between Elbow Method and Silhouette Score is shown. The elbow graph indicates an elbow at k=3 which is considered the optimal number of clusters in terms of computation and variability. Whereas from the Silhouette Score results, the highest value is at k= 2, which means that the best quality of cluster separation in terms of distance is at k=2. Therefore, the selection of the number of clusters is adjusted to the purpose of the analysis, namely the focus on interpreting strategies and variations in Instructor performance, so k = 3 is more relevant (Rousseeuw, 1987; Ketchen & Shook, 1996; Tan et al., 2019). This is supported by the distribution of clusters generated with k = 3, namely cluster 0 = 71 items, cluster 1 = 70 items, cluster 2 33 items. The work design in Rapidminer is described as follows:

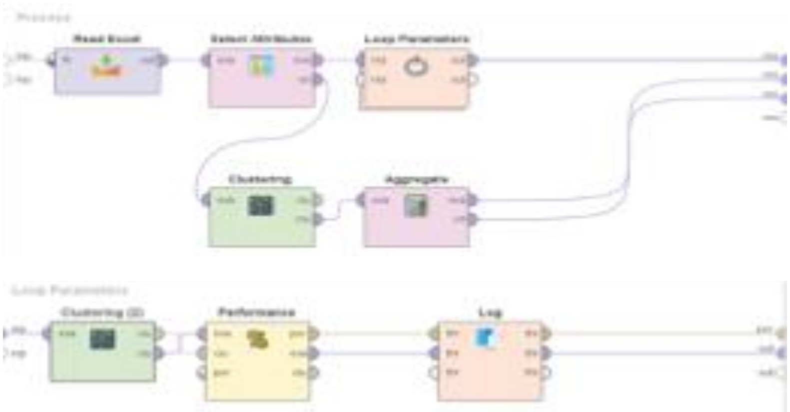

Figure 3. Parameter Determination and Clustering Process

The relationship between variables in the cluster is described as follows:

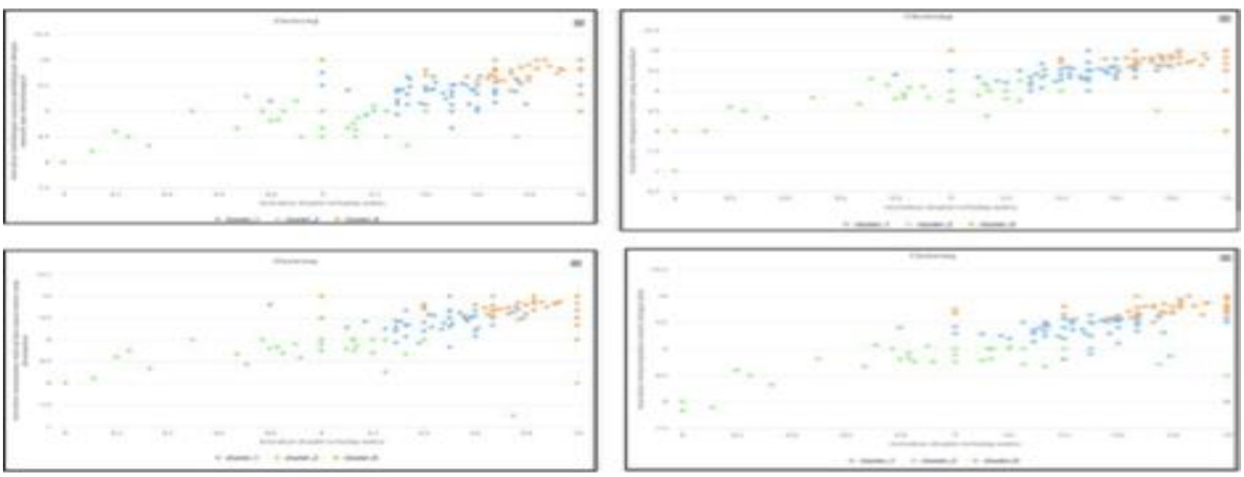

Figure 4: Cluster Distribution and Relationship Between Variables

Visualization of clustering results in each indicator shows the same pattern, where cluster 0 occupies the highest position in all aspects of participant assessment. Cluster 1 (blue) shows Instructors with medium and stable performance and Cluster 2 (green) has the lowest scores in almost all indicators. This indicates that K-means can segment by mapping significant differences in instructor quality as a basis for an appropriate and incremental Instructor development strategy. The labels and definitions of the three clusters are given below:

Table 3. Cluster Results and Definitions

No	Label	Number	Cluster Name	Description
1	Cluster 0	70	Potential Instructors	Instructors with good and stable performance but still need strengthening in some pedagogical aspects (delivery of material).
2	Cluster 1	33	Developing Instructors	Instructors with low performance, need basic training and intensive mentoring for competency improvement
3	Cluster 2	71	Superior Instructors	Instructors with very high performance in all, all aspects, ready to become role models and trainers for other instructors.

The development program for Instructors in cluster 2 (Superior Instructors) is focused on strengthening Instructors as leaders in the learning value chain and also as mentors for other Instructors with a development program themed "Leader Instructor Acceleration". This development program includes training Train the Trainer, engage actively in the curriculum and materials development as well as, and initiators in the company's strategic expert forums. This is consistent with the study of Chahuán-Jiménez et al. (2025) which states that highly competent instructors have the ability to act as learning leaders for an organization. This opinion is in line with the idea of distributed expertise, where superior instructors help strengthen the overall capacity and capability of other Instructors (Smith & MacGregor, 2020).

The development program for Instructors in cluster 0 (potential Instructors) with stable performance but still have some things to improve, especially in terms of teaching in delivering material and building classroom interaction, is more emphasized on the learning soft skills improvement program through the "Competency Elevation Path" program design. This program consists of activities such as instructional training and certification, microteaching sessions and feedback and peer review among instructors. This is in line with Liu's (2022) study that evaluation and practice-based training approaches can significantly improve instructional skills. Active learning training can increase Instructor motivation and professionalism in the learning process (Carless & Boud, 2018).

The development program for Instructors in cluster 1 (developing Instructors) requires intensive coaching through the "Fundamental Teaching Reinforcement" program. The main focus of development in this cluster is to build confidence, teaching skills and material delivery. Activities carried out in the development program include basic pedagogy training, thorough classroom observation (mirroring) and mentoring by senior mentors or Superior Instructors, this is because coaching and learning approaches are effective in improving performance (Darling-Hammond et al., 2017).

Furthermore, the cluster results are used as labels in prediction modeling. By dividing the training data and testing data by 70:30, the process description on rapidminer is as follows:

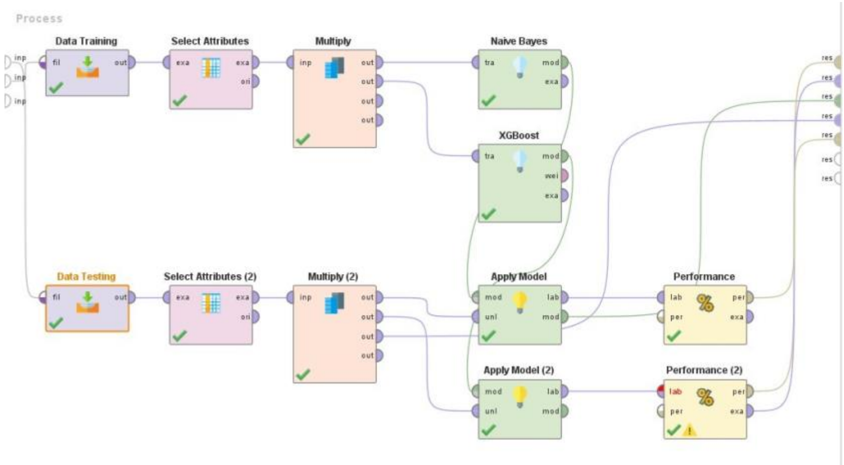

Figure 5: Predictive Modeling Process

Modeling using Naive Bayes and XGBoost algorithms with the following results:

Table 4. Confusion Matrix Results with Naive Bayes

accuracy: 71.15%

	true Instruktur Potensial	true Instruktur Unggul	true Instrutkur Berkembang	class precision
pred. Instruktur Potensial	12	3	2	70.59%
pred. Instruktur Unggul	3	16	2	76.19%
pred. Instrutkur Berkembang	3	2	9	64.29%
class recall	66.67%	76.19%	69.23%	

Table 5. Confusion Matrix Results with XGBoosT

Table View Plot View

accuracy: 67.31%

	true Instruktur Potensial	true Instruktur Unggul	true Instrutkur Berkembang	class precision
pred. Instruktur Potensial	12	4	3	63.16%
pred. Instruktur Unggul	3	15	2	75.00%
pred. Instrutkur Berkembang	3	2	8	61.54%
class recall	66.67%	71.43%	61.54%	

From the confusion matrix results, it can be seen that the Naive Bayes algorithm is better than XGBoost as an Instructor performance prediction model. The results of the evaluation of accuracy, recall, precision and F1 Score of the two models are as follows:

Table 6. Algorithm Comparison Results

	XGBoost	Naïve Bayes
Accuracy	67.31%	71.15%
Recall	66.54%	70.70%
Precision	66.57%	70.35%
F1-Score	66.55%	70.52%

From the evaluation results of the algorithm comparison, it can be seen that Naive Bayes is superior to XGBoost in predicting Instructor performance based on trainee evaluations, as seen from the accuracy, recall, precision and F1 Score. These results indicate that Naïve Bayes is better able to classify instructor performance classes consistently, especially in data distributions that tend to be unbalanced, which is a common characteristic in learning evaluation (Zhang, 2004). The superiority of Naïve Bayes in this case is in line with previous findings that it works effectively on high-dimensional data with independent features and simple data distribution. On the other hand, although XGBoost is known for its ability to handle complex data through boosting techniques, its performance lags slightly in this scenario. This could be due to the dataset size not being large enough or the parameter tuning performed on the XGBoost model not being optimized (Chen & Guestrin, 2016). Therefore, for this context of trainee evaluation data-based prediction, Naïve Bayes is a more practical, efficient, and accurate choice of algorithm.

CONSLUSION

Segmenting instructors with the K-Means algorithm resulted in three distinct clusters: Excellent Instructors, Potential Instructors, and Developing Instructors. The visualization of these clusters confirmed that one group consisted of high-performing instructors, another reflected instructor with stable yet improvable performance, and the last represented those requiring more intensive coaching. This segmentation not only provides a clear map of instructor quality but also offers a practical foundation for designing targeted development strategies that respond directly to real conditions in the field.

The predictive modeling further demonstrated that the Naïve Bayes algorithm outperformed XGBoost in predicting instructor performance classification. Its consistency in handling imbalanced data and ability to represent probability-based distributions effectively enabled accurate classification of instructor groups derived from clustering. This confirms the potential of predictive models to support automated evaluation processes and to enhance decision-making in instructor management.

The integration of clustering and predictive models in this study illustrates a feasible approach to transforming instructor evaluation systems from being descriptive and subjective into becoming adaptive, objective, and data driven. By relying on learner evaluation data, institutions can anticipate instructor performance trends earlier and design more effective development programs.

Future research can expand this framework by incorporating additional variables, such as teaching experience, educational background, or direct classroom observations, to further enrich the predictive capacity of the models. For educational institutions, adopting a data-driven evaluation system will not only improve the effectiveness of instructor development but also contribute to building more sustainable, efficient, and adaptive learning governance.

REFERENCES

Aljabri, H. M., & Khalil, M. I. (2021). A predictive analytics model for improving faculty performance evaluation. *Education and Information Technologies, 26*(4), 4175–4192. https://doi.org/10.1007/s10639-021-10531-z

Argentinia, R. (2021). Using K-means to determine learner typologies for project-based learning. *International Journal of Computer Applications, 178*(43), 18–24. https://doi.org/10.5120/ijca2021920320

Borko, H. (2019). Professional development and teacher learning: Mapping the terrain. *Educational Researcher, 33*(8), 3–15. https://doi.org/10.3102/0013189X033008003

Carless, D., & Boud, D. (2018). The development of student feedback literacy: Enabling uptake of feedback. *Assessment & Evaluation in Higher Education, 43*(8), 1315–1325. https://doi.org/10.1080/02602938.2018.1463354

Chahuán, J, K., Soto Silva, W. E., & Maturana, S. (2025). Cluster analysis of digital competencies among professors in higher education. *Frontiers in Education, 10,* 45. https://doi.org/10.3389/feduc.2025.00045

Chen, B., & Wang, M. (2023). Online instructor clusters: Implementation frequency of instructional activities. *Education and Information Technologies, 28,* 3421–3440. https://doi.org/10.1007/s10639-023-11721-9

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). ACM. https://doi.org/10.1145/2939672.2939785

Darling-Hammond, L., Hyler, M. E., & Gardner, M. (2017). *Effective teacher professional development.* Learning Policy Institute. https://learningpolicyinstitute.org/product/effective-teacher-professional-development-report

Ketchen, D. J., & Shook, C. L. (1996). The application of cluster analysis in strategic management research: An analysis and critique. *Strategic Management Journal, 17*(6), 441–458. https://doi.org/10.1002/(SICI)1097-0266(199606)17:6<441::AID-SMJ819>3.0.CO;2-G

Liu, R. (2022). Data analysis of educational evaluation using K-means clustering. *Computational Intelligence and Neuroscience, 2022,* 3762431. https://doi.org/10.1155/2022/3762431

Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 10*(3), e1355. https://doi.org/10.1002/widm.1355

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics, 20,* 53–65. https://doi.org/10.1016/0377-0427(87)90125-7

Smith, B. L., & MacGregor, J. T. (2020). Learning communities and the quest for quality. *Change: The Magazine of Higher Learning, 52*(3), 36–43. https://doi.org/10.1080/00091383.2020.1758334

Tan, P.-N., Steinbach, V., & Kumar, V. (2019). *Introduction to data mining* (2nd ed.). Pearson.

Zhang, H. (2004). The optimality of Naive Bayes. In *Proceedings of the 17th International Florida Artificial Intelligence Research Society Conference (FLAIRS)* (pp. 1–6). AAAI Press. https://www.aaai.org/Papers/FLAIRS/2004/Flairs04-019.pdf

Zhang, J., Chen, Y., & Luo, X. (2023). Online instructor clusters: Implementation frequency of instructional activities. *Education and Information Technologies, 28,* 347–362. https://doi.org/10.1007/s10639-023-11345-4