CCC 2025

Proceedings of the Creative Construction Conference (2025)
Edited by: Miroslaw J. Skibniewski & Miklós Hajdu & Žiga Turk
https://doi.org/10.22260/CCC2025/0085

A CLUSTER ANALYSIS AS THE BASIS FOR A PROFITABILITY ANALYSIS OF BIM PROJECTS

Konrad Neubaur, Hendrik Morgenstern, Katharina Klemt-Albert
RWTH Aachen University, Aachen, Germany

Abstract

Building Information Modeling (BIM) has become increasingly popular in the German construction industry, particularly in transport infrastructure, to achieve project goals more reliably and increase productivity. Due to strong political demand to expand BIM applications, authorities need to demonstrate its added value, especially concerning cost-effectiveness. Existing approaches to examining BIM's economic efficiency often do not account for the specific conditions of German road construction and focus on individual projects, making it difficult to generalize results. As part of a research project commissioned by DEGES GmbH on behalf of the Federal Republic of Germany, an economic feasibility study on the application of BIM methodology in the planning and construction phases of federal trunk road construction is being conducted. With the approach of a project comparison of BIM and non-BIM projects, however, groups of similar projects are required. This article demonstrates how to establish the basis for this profitability analysis of BIM projects. Starting from a database of several hundred projects, the aim is to identify the most homogeneous project groups to serve as a basis for retrospective project analysis. To achieve this goal, a cluster analysis was conducted, focusing on determining which features are suitable for the technical characterization of projects and how these must be prepared. The cluster algorithms K-Means and DBSCAN were implemented and examined regarding optimal parameter settings. The cluster solutions were then interpreted, compared, and an algorithm was selected. The results show that, provided sufficient data is available – both datasets and characteristics – the use of cluster analysis is a useful tool for forming homogeneous groups of construction projects.

Keywords: building information modeling, cluster analysis, feasibility study, transport infrastructure.

Peer-review under responsibility of the scientific committee of the Creative Construction Conference 2025.

1. Introduction

1.1. Initial situation

Construction projects around the world fail to meet their deadlines and cost targets [1]. Germany is no different from other nations. Major public-sector projects are a particular focus here, as they involve the use of public funds [2]. The Building Information Modeling (BIM) method has been praised for many years as an approach to counteract this shortcoming. The then German Federal Ministry of Transport and Digital Infrastructure defines BIM as a "collaborative working method that uses digital models of a building to consistently record and manage the information and data relevant to its life cycle and exchange it between the parties involved in transparent communication or transfer it for further processing" (translated from German) [3]. Reducing costs and completing construction projects on time are only part of the added value attributed to BIM. Other added values experienced through the use of BIM include increased quality in planning and the loss-free transfer of building information between different life cycle phases of a building. [3, 4] BIM implementation is already relatively advanced worldwide, particularly in the Scandinavian countries, but also in the USA, Singapore, the UK and Australia [5]. In Germany, too, there have been increasing efforts to promote the BIM method in the public construction sector in recent years. For this reason, various initiatives have been launched and various strategy papers on this topic have been published. Three particularly important papers should be mentioned here, which are intended to help establish the BIM method across the board in Germany: the phased plan for digital planning and construction, published in 2015 as a basis for the introduction

Corresponding author email address: neubaur@icom.rwth-aachen.de

of IT-supported processes and technologies in the planning, construction and operation of buildings [3], the BIM Masterplan for Federal Buildings published in 2021 [6], in which the objectives and the implementation strategy of the BIM method for federal buildings are described, as well as the Federal Trunk Roads BIM Masterplan, in which an implementation strategy for BIM in federal trunk road construction is described [7]. In the latter, the hypothesis regarding the economic benefits of BIM is that 2-4% acceleration effects and 3-6% cost reductions can be realized [7]. However, this has yet to be proven. Qualitative studies on the added value of using the BIM method in terms of costs and deadlines are not yet available in a form that would be generally applicable or transferable to federal trunk road construction in Germany. Due to a lack of evidence on the cost-effectiveness of BIM in federal trunk road construction in Germany, the Institute of Construction Management, Digital Construction and Robotics in Civil Engineering at RWTH Aachen University was commissioned with a research project aimed at investigating the cost-effectiveness of the BIM method in federal trunk road construction projects.

1.2. Aim of the study and connection with the cluster analysis

The aim of the research work presented here is to establish a basis in the form of groups of similar transportation infrastructure projects in federal trunk road construction in Germany for a subsequent comparison to investigate the effectiveness of the BIM method. In this context, the term effectiveness is defined as a noticeable positive difference in the cost and schedule development of a BIM project compared to a comparable conventional project that was not carried out using BIM (henceforth referred to as a non-BIM project). This paper shows how cluster analysis can be used to form groups of similar projects that can be used as a basis for investigating the effectiveness of BIM.

2. Cluster analyses

A cluster analysis serves as an instrument for dividing a heterogeneous population into homogeneous groups [8]. In contrast to classification methods, the labels of the cluster groups are not specified in advance but are identified by the algorithm used in each case from the information in the input data. This means that cluster analysis methods belong to the field of unsupervised machine learning. [9] In the following subsections, a brief overview of different clustering methods is given, the two cluster algorithms used in the analysis are presented and, finally, possible methods for selecting the optimal cluster solution are discussed.

2.1. Cluster analysis methods

Procedures for carrying out a cluster analysis can be divided into three large groups. *Incomplete cluster analysis methods*, also known as geometric methods [10], representation or projection methods [11, 12] are used to calculate the spatial representation of classification objects in a low-dimensional space. When using such methods to form clusters, this is done by interpreting the spatial representation. Graphical clustering therefore requires a maximum of three dimensions to be available.

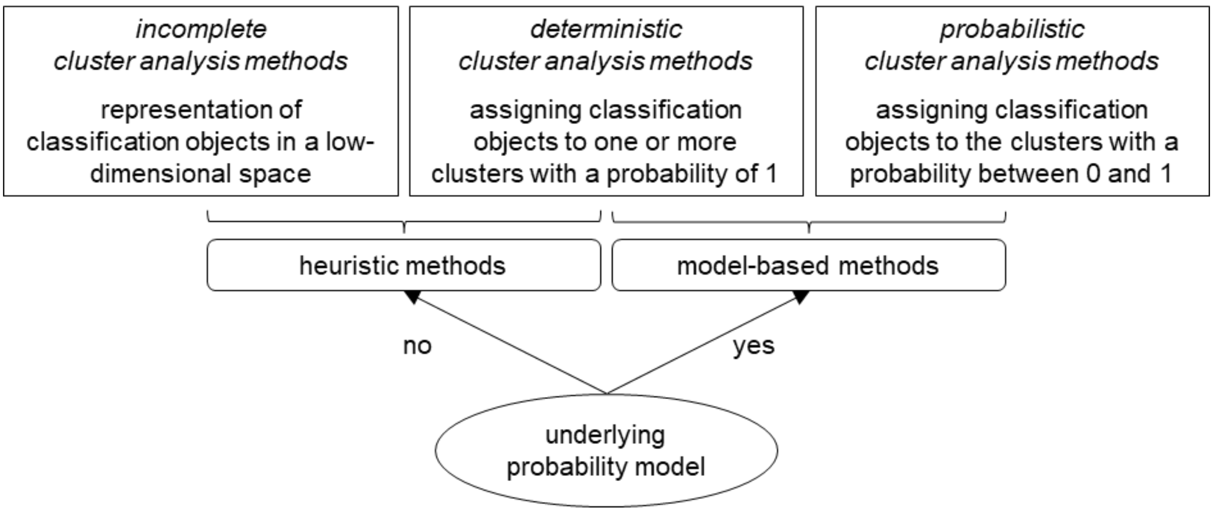

Fig. 1. Assignment of objects to the clusters [13]

Further possibilities for using this group of methods lie either in their use as a method for factor analysis or as a supporting method for calculating new data matrices for *deterministic cluster analysis methods*, the second major group of methods. In deterministic cluster analysis methods, classification objects are assigned to one or more clusters with a probability of 1. In contrast, in *probabilistic cluster analysis methods*, the third major group of methods, the classification objects are assigned to the clusters with a probability between 0 and 1. This classification is aimed at the type of assignment of the classification objects to the clusters. [13] This can also be divided into model-based or heuristic methods [14, 15]. Model-based methods are characterized by the fact that their use is based on the assumption of a probability model generated from the input data. This is not the case with heuristic methods. [13] The relationship between these two typologies is shown in figure 1. For the present research question, the decision was made to use deterministic cluster analysis methods due to the clear assignment of the classification objects to individual clusters and the resulting simplified interpretation of the cluster solution.

2.2. Algorithms

For this research question, it was decided to use two algorithms and compare the results with each other. The two selected algorithms are briefly presented in the following subsection.

2.2.1. K-Means

Classification objects are assigned in the K-Means algorithm by minimizing the sum of squares of the scatter in the clusters. This target criterion can be expressed as the sum of the squared euclidean distance between each object g and the respective assigned cluster center k. The number of clusters must be specified. The initial point of the clustering procedure can be selected either manually or randomly for each cluster. The procedure can be described in four steps. [13]

1. Manual or random assignment of classification objects to K clusters.
2. Calculation of the cluster centers.
3. Reassignment of the classification objects so that the scatter sum of squares in the clusters is minimized.
4. Check whether step 3 results in a changed assignment of the classification objects to the clusters. Steps 2 & 3 are repeated until no further change occurs.

2.2.2. DBSCAN

The DBSCAN algorithm (abbreviation for Density Based Spatial Clustering of Applications with Noise) was introduced by Ester, et al. [16] and is a density-based clustering method. The basic idea of this algorithm is that there must be a minimum number of neighbors (*minPts*) for each point of a cluster within a specified radius (*eps*). A distinction is made between core points and border points. It is a core point if the condition applies that the number of all points within the eps neighborhood is greater than or equal to minPts. The relationship between the points can be described as directly density-reachable, density-reachable and density-connected points. The group of points that cannot be assigned to a cluster due to these conditions is referred to as *noise* and contains outliers. To perform the cluster analysis with this algorithm, only the two parameters eps and minPts need to be specified. [16]

2.3. Determining the parameters of the algorithms

When using both K-Means and DBSCAN, parameters must be determined so that the cluster analysis can take place. The selection of parameters should be assessed with regard to the quality of the cluster solution. This section looks at selected methods that can be used to make this assessment.

When using the K-Means algorithm, the number of clusters must be determined in advance. The elbow method is suitable for this, in which the sum of squared errors (SSE) of the euclidean distance between each classification object and the mean value of each cluster is displayed in a graph for an increasing number of clusters. The optimal number of clusters is at the point where the rate of change of the SSE decreases sharply. This can be recognized in a graph by a kink, the so-called elbow. The elbow method is not always suitable for determining the optimum number of clusters if, for example, several kinks or no kink at all can be detected in the graph. [17] Figure 2 shows two elbow diagrams as examples. At this point, the silhouette score method is very well suited to support the decision and also to confirm the formal validity of the cluster solution [13]. The average distance between points within a cluster and the average distance between different clusters is calculated. The silhouette coefficient (SC) can be calculated for the entire cluster solution and can lie between -1 and 1. The closer the coefficient is to 1, the better the cluster solution is in the sense that the clusters are homogeneous in themselves but different from each other.[18]

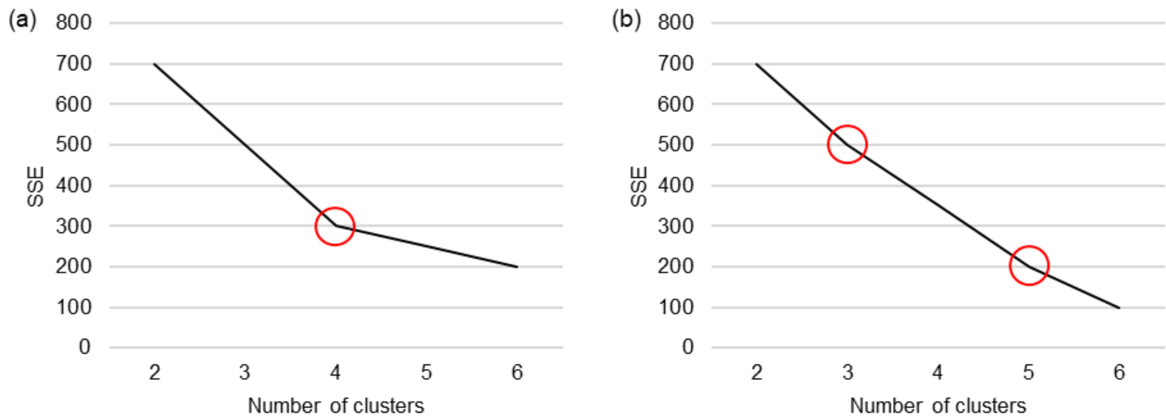

Fig. 2. (a) Unambiguous determination of the number of clusters with the elbow method possible; (b) Unambiguous determination of the number of clusters with the elbow method not possible

The number of clusters does not have to be specified in advance for the DBSCAN algorithm. However, the parameters eps and minPts must be determined here. This can be done empirically using the silhouette score. By iterating over a specified range of values for the eps and minPts parameters and calculating the silhouette score for the respective cluster solution, the parameter values for which the best solution is output can then be selected. [19]

3. Presentation of the data basis

Characteristics are required for clustering into groups of similar projects. This section presents the composition of the data basis for the cluster analysis. The data basis provided by DEGES (Deutsche Einheit Fernstraßenplanungs- und -bau GmbH (DEGES): Project management company for the implementation of transportation infrastructure projects in Germany) for the cluster analysis is compiled from three different data sources: a database containing master data on the projects, a table containing information on the engineering structures of the projects and a document type containing information in the form of a project description for projects. The data basis was created from these three data sources in consultation with the client and is presented below.

One feature that characterizes a project is the type of measure. A distinction is made between three different types of measure: New construction measures, reconstruction measures and maintenance measures. Due to the differences between these three types of measure and the resulting poor comparability of projects with different types of measure, it was decided not to use this characteristic as a variable in the cluster analysis, but to divide the data basis into three groups in advance on the basis of this characteristic. Accordingly, the cluster analysis is carried out individually for each group of measure types.

Since federal trunk road construction projects are being examined, the length of a construction project is a suitable project characteristic. The route length is specified in kilometers and used as a metric variable in the cluster analysis. Furthermore, the road type and the cross-section of the road can be specified to characterize federal trunk roads. For the road type, a distinction is made between a freeway and federal road types [20]. This variable is used as a binary variable, whereby the value 1 stands for a freeway and the value 0 for a federal road. The variable cross-section describes on an ordinal scale how many lanes (sum of both directional lanes) a measure has. A road can have either 2, 3, 4, 6 or 8 lanes. The binary variable rest area indicates whether the corresponding construction measure contains rest areas or not. The value 1 indicates that there are rest areas. The value 0 indicates that there are no rest areas. Two further binary variables provide information on which phases of the project DEGES has been entrusted with. The variable LPH_P indicates whether services from planning phases have been commissioned, while the variable LPH_A contains information on services from execution phases. A value of 1 means that services have been commissioned, while a value of 0 indicates the opposite. The last variable is "Kompl_Ing" and provides information on whether the corresponding project includes engineering structures and, if so, what complexity is to be expected here in terms of project execution. This is an interval-scaled variable with a value range from 0 to 1: Kompl_Ing = 0 means that there are no engineering structures. As the value increases, so does the complexity of the existing engineering structures. Table 1 shows an example of a data set from the data basis.

Table 1. Exemplary data set of the data basis for the cluster analysis

Type of measure	Length (km)	Road type	Cross section	Rest area	LPH_P	LPH_A	Kompl_Ing
New construction	16,80	1	4	0	0	1	0,260342

The data basis comprises a total of 301 projects, consisting of 168 new construction measures, 103 reconstruction measures and 30 maintenance measures.

4. Conduct the cluster analysis

Performing the cluster analysis includes the steps of determining the optimal parameters of the algorithm used so that the best formal solution is achieved and the selection of the cluster solution with regard to the algorithm used. The cluster analysis is implemented in Python in the Spyder development environment. The algorithms K-Means and DBSCAN from the free software library Scikit-learn for machine learning in version 1.5.1 are used.

4.1. Determining the optimum parameters of K-Means and DBSCAN

4.1.1. K-Means

First and foremost, the optimal number of clusters must be determined. For this purpose, the Ellbow criterion was used in conjunction with the silhouette score for cluster solutions of two to ten clusters across all possible combinations (a total of eight variants) of the different parameters of the algorithm. Further information on the parameters of the algorithm can be found in the documentation of the corresponding library [21].

The analysis showed that ten clusters should be selected for the "new construction" group of measures, seven clusters for the "reconstruction" group of measures and two clusters for the "maintenance" group of measures. As only a cluster number of two to ten was examined and the optimum cluster number in the group of new construction measures would have been ten, the analysis was carried out again for a cluster number of two to twenty in order to rule out the possibility that the optimum solution was not, for example, eleven. Repeating the analysis and adjusting the maximum number of clusters from ten to twenty produced different results for the group of new construction and maintenance measures. For new construction measures, the optimum number of clusters should be 19, for maintenance measures 14. Although both cluster solutions show a slightly improved silhouette score of approx. 0.05, it was decided not to use these solutions in order to avoid an excessively large number of clusters making it more difficult to interpret the clusters. Table 2 contains the parameters for the selected cluster solutions with K-Means.

Table 2. Parameters and silhouette score (sc) of the best cluster solution with K-Means

Group of measures	Number of clusters	init	n_init	algorithm	sc
new construction	10	k-means++	10	lloyd	0.493622
reconstruction	7	k-means++	1	elkan	0.607542
maintenance	2	Random	1	elkan	0.495379

4.1.2. DBSCAN

The parameters eps and minPts must be specified for the execution of the DBSCAN algorithm. In addition, a decision can be made regarding the distance measure used to calculate the distance between points. The euclidean distance dimension is used here, which is why all other parameters of the DBSCAN algorithm can remain at the default setting. [22]

As described in subsection 2.3, the parameters eps and minPts are determined by iterating over a predefined range of values and then selecting the best solution using the silhouette score. For the parameter eps, a value range of 0.01-5.00, divided into 500 intervals, was identified as practicable. A value range of 7 to 14 was specified for the minPts parameter. Table 3 contains the parameters for the selected cluster solutions with DBSCAN. Since DBSCAN can recognize outliers, the number of projects identified as outliers is also indicated.

Table 3. Parameters and silhouette score (sc) of the best cluster solution with DBSCAN

Group of measures	**Number of clusters**	**eps**	**minPts**	**Outliers**	**sc**
new construction	5	2.10	7	8.33 % (14)	0.433327
reconstruction	3	2.00	8	14.56 % (15)	0.53567
maintenance	2	1.47	14	33.33 % (10)	0.240945

4.2. Comparison and selection of the cluster solution

For all three groups of measures, a decision must then be made as to whether the cluster solution found with K-Means or with DBSCAN should be selected. This is done for each type of measure and is explained briefly below.

Tables 4 to 6 contain cross-tabulations of the assignments of how projects clustered with K-Means were clustered by DBSCAN. In the group of new construction measures, it can be seen that the DBSCAN solution is made up of one or two clusters of the K-Means solution. However, this does not include all classification objects, as outliers are detected by DBSCAN. This can also be observed in the group of reconstruction measures. Only in the group of maintenance measures the two cluster solutions do not match.

Table 4. Cross-tabulation of the assignment of projects by DBSCAN and K-Means for new construction measures

		DBSCAN						
	Cluster	*1*	*2*	*3*	*4*	*5*	*Outliers*	*Total*
K-Means	*1*	-	11	-	-	-	-	11
	2	15	-	-	-	-	-	15
	3	-	-	-	-	8	4	12
	4	16	-	-	-	-	1	17
	5	-	-	24	-	-	-	24
	6	-	-	-	15	-	4	19
	7	-	26	-	-	-	-	26
	8	-	-	9	-	-	3	12
	9	-	-	30	-	-	-	30
	10	-	-	-	-	-	2	2
	Total	31	37	63	15	8	14	168

Table 5. Cross-tabulation of the assignment of projects by DBSCAN and K-Means for reconstruction measures

		DBSCAN				
	Cluster	*1*	*2*	*3*	*Outliers*	*Total*
K-Means	*1*	67	-	-	-	67
	2	-	11	-	-	11
	3	-	-	-	3	3
	4	-	-	-	5	5
	5	-	-	-	6	6
	6	-	-	10	-	10
	7	-	-	-	1	1
	Total	67	11	10	15	103

Table 6. Cross-tabulation of the assignment of projects by DBSCAN and K-Means for maintenance measures

		DBSCAN			
K-Means	*Cluster*	*1*	*2*	*Outliers*	*Total*
	1	14	6	7	27
	2	-	-	3	3
	Total	14	6	10	30

In addition to formal criteria such as the silhouette score, content-related aspects are also decisive for the selection of the cluster solution (whether K-Means or DBSCAN). Since the cluster analysis presented here was carried out in order to divide a large number of projects into homogeneous groups and then compare BIM and non-BIM projects within these groups, the presence of BIM projects in the clusters formed is of great importance. Figure 3 contains bar charts for all measure groups, showing the breakdown into BIM and non-BIM projects. For the new construction and reconstruction measures, the K-Means solution was chosen due to the better silhouette score and more potential BIM projects to be considered in the subsequent comparison. For the group of maintenance measures, the DBSCAN solution was chosen despite the better silhouette score of the K-Means solution. The reason for this is that the clusters identified by K-Means are very heterogeneous in terms of their size and not particularly homogeneous in terms of their project characteristics.

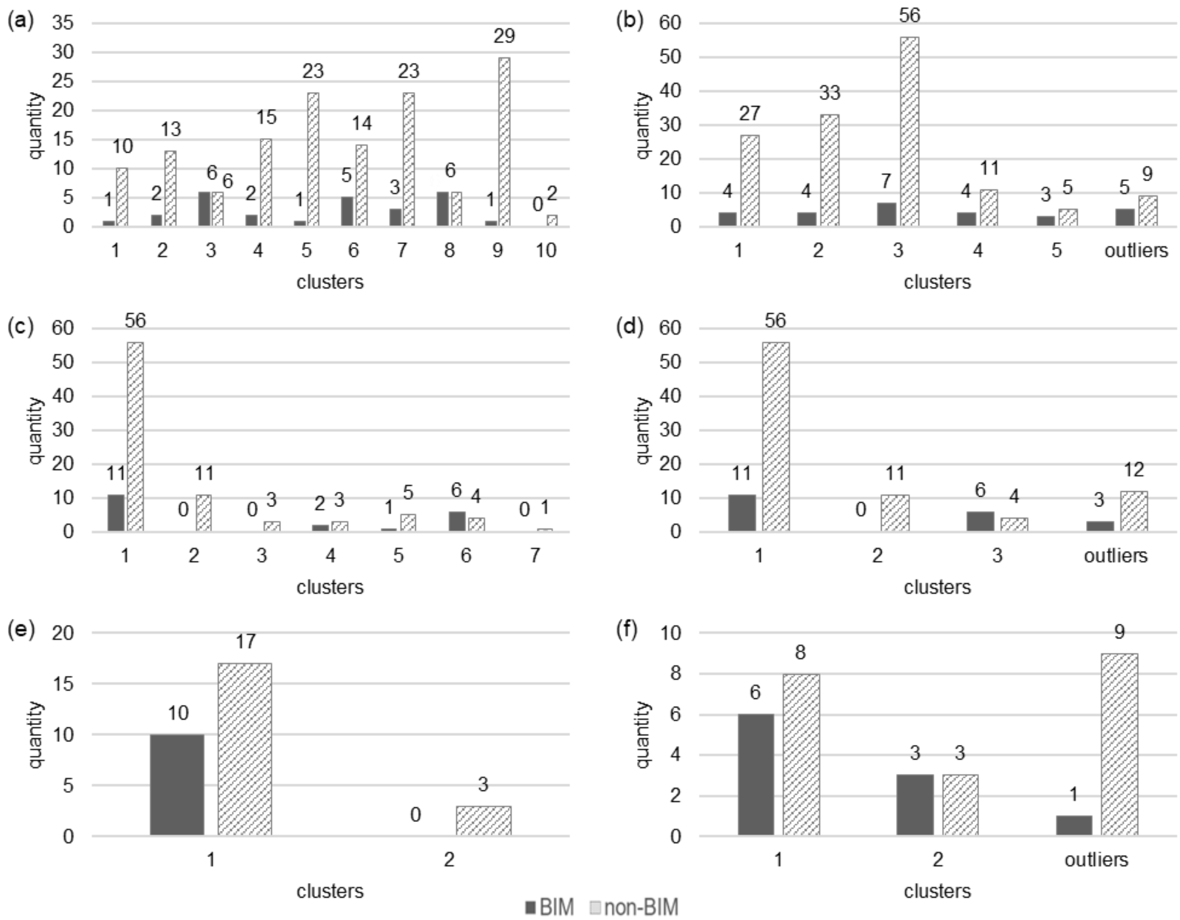

Fig. 3. Distribution of projects by BIM and non-BIM projects per measure group and cluster: (a) K-Means for new construction measures; (b) DBSCAN for new construction measures; (c) K-Means for reconstruction measures; (d) DBSCAN for reconstruction measures; (e) K-Means for maintenance measures; (f) DBSCAN for maintenance measures

5. Interpretation of the cluster solutions

When interpreting the individual cluster solutions for each type of measure, it can be seen that the various clusters differ in particular on the basis of the characteristics of the categorical variables road type, cross-section, LPH_P, LPH_A and rest area. The metric characteristics of length and Kompl_Ing are not the main determinants of the composition of the different clusters, but contribute to a finer subdivision, particularly in the case of measure groups with many clusters. Figure 4 shows an example of the characteristics of the individual features of the projects for two clusters of the reconstruction measure group.

When comparing the cluster solutions determined using K-Means and DBSCAN, it can be seen that the solutions for the new construction and reconstruction measure groups match very well and only differ in that fewer clusters are identified in the solution determined using DBSCAN, as either several clusters identified by K-Means are merged or a classification object is classified as an outlier. In the group of maintenance measures, there is no agreement between the two cluster solutions. In fact, one of the two clusters identified by K-Means is completely classified as an outlier in DBSCAN and the other is divided into two separate clusters. This can also be attributed to the fact that the maintenance measures group contains the fewest projects and these are rather dissimilar.

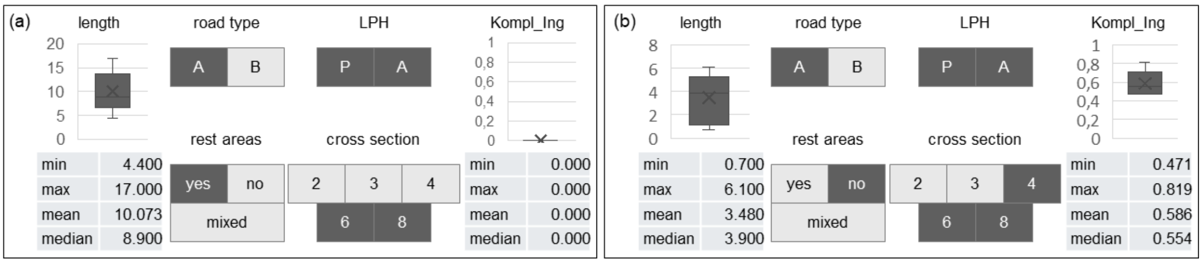

Fig. 4. (a) Characteristics of cluster 2 of the reconstruction measure group;
(b) Characteristics of cluster 6 of the reconstruction measure group

6. Conclusion

This study underscores the importance of cluster analysis as a preparatory step for evaluating the profitability of BIM projects in transportation infrastructure. By utilizing multiple clustering algorithms, such as K-Means and DBSCAN, the robustness and reliability of the cluster solutions were ensured. The findings revealed that the resulting clusters are distinctly characterized by their inherent attributes, which facilitates meaningful comparisons between similar projects. However, the effectiveness of the cluster analysis is highly dependent on the quality and comprehensiveness of the available data. Variations in data sources or the inclusion of additional project features could lead to different clustering outcomes, highlighting the need for thorough and high-quality datasets in future analyses. In summary, cluster analysis serves as a foundational step that enables the comparison of homogeneous project groups, paving the way for subsequent profitability assessments of BIM implementations.

Acknowledgements

The preparation of the content of this report was financed and supported by DEGES GmbH on behalf of the Federal Republic of Germany.

References

[1] F. Barbosa *et al.*, "Reinventing construction through a productivity revolution," McKinsey Global Institute, 2017.

[2] Federal Ministry of Transport and Digital Infrastructure (BMVI), "Reformkommission Bau von Großprojekten: Endbericht: Komplexität beherrschen - kostengerecht, termintreu und effizient," ed. Berlin: Bundesministerium für Verkehr und digitale Infrastruktur, 2015.

[3] Federal Ministry of Transport and Digital Infrastructure (BMVI), "Stufenplan Digitales Planen und Bauen: Einführung moderner, IT-gestützter Prozesse und Technologien bei Planung, Bau und Betrieb von Bauwerken," ed. Berlin: Bundesministerium für Verkehr und digitale Infrastruktur, 2015.

[4] K. Eschenbruch, A. Malkwitz, J. Grüner, A. Poloczek, and C. K. Karl, "Maßnahmenkatalog zur Nutzung von BIM in der öffentlichen Bauverwaltung unter Berücksichtigung der rechtlichen und ordnungspolitischen Rahmenbedingungen,"

Gutachten zur BIM Umsetzung im Forschungsprogramm Zukunft Bau, Bundesinstitut für Bau-, Stadt-und Raumforschung (BBSR), 2014,

[5] A. Borrmann, M. König, C. Koch, and J. Beetz, *Building Information Modeling: Technologische Grundlagen und industrielle Praxis*, 2. ed. Wiesbaden; Heidelberg: Springer Vieweg, 2021. https://doi.org/10.1007/978-3-658-33361-4.

[6] Federal Ministry of the Interior, Building and Community (BMI), Federal Ministry of Defense (BMVg), "Masterplan BIM für Bundesbauten," ed. Berlin & Bonn: Bundesministerium des Innern, für Bau und Heimat, Bundesministerium der Verteidigung, 2021.

[7] A. Meister, F. Scholz, and S. Banemann, "Masterplan BIM Bundesfernstraßen: Digitalisierung des Planens, Bauens, Erhaltens und Betreibens im Bundesfernstraßenbau mit der Methode Building Information Modeling (BIM)," ed. Berlin: Bundesministerium für Verkehr und digitale Infrastruktur, 2021.

[8] K. Backhaus, B. Erichson, S. Gensler, R. Weiber, and T. Weiber, *Multivariate Analysemethoden: Eine anwendungsorientierte Einführung*, 16th ed. Wiesbaden: Springer Fachmedien Wiesbaden, 2021. https://doi.org/10.1007/978-3-658-32425-4.

[9] M. Plaue, *Data Science: Grundlagen, Statistik und maschinelles Lernen*, 1st ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2021. https://doi.org/10.1007/978-3-662-63489-9.

[10] A. D. Gordon, *Classification: methods for the exploratory analysis of multivariate data* (Monographs on applied probability and statistics). London [u.a.]: Chapman and Hall (in Englisch), 1981.

[11] A. K. Jain and R. C. Dubes, *Algorithms for clustering data*. Prentice-Hall, Inc., 1988.

[12] O. Opitz, *Numerische Taxonomie*. Stuttgart: Fischer, 1980.

[13] J. Bacher, A. Pöge, and K. Wenzig, *Clusteranalyse: Anwendungsorientierte Einführung in Klassifikationsverfahren*, 3rd ed. Berlin; Boston: Oldenbourg Wissenschaftsverlag, 2010. https://doi.org/10.1524/9783486710236.

[14] C. Fraley and A. E. Raftery, "MCLUST: Software for Model-Based Cluster Analysis," *Journal of Classification,* vol. 16, no. 2, pp. 297-306, 1999/07/01 1999, https://doi.org/10.1007/s003579900058.

[15] S. Frühwirth-Schnatter, *Finite mixture and Markov switching models*. Springer, 2006. https://doi.org/10.1007/978-0-387-35768-3.

[16] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," presented at the Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, Portland, Oregon, 1996. [Online]. Available: https://www.dbs.ifi.lmu.de/Publikationen/Papers/KDD-96.final.frame.pdf.

[17] D. J. KETCHEN and C. L. SHOOK, "THE APPLICATION OF CLUSTER ANALYSIS IN STRATEGIC MANAGEMENT RESEARCH: AN ANALYSIS AND CRITIQUE," *Strategic Management Journal,* vol. 17, no. 6, pp. 441-458, 1996, https://doi.org/10.1002/(SICI)1097-0266(199606)17:6<441::AID-SMJ819>3.0.CO;2-G.

[18] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *Journal of Computational and Applied Mathematics,* vol. 20, pp. 53-65, 1987/11/01/ 1987, https://doi.org/10.1016/0377-0427(87)90125-7.

[19] F. Hashmi. "How to create clusters using DBSCAN in Python." https://thinkingneuron.com/how-to-create-clusters-using-dbscan-in-python/ (accessed 08.02., 2025).

[20] *Bundesfernstraßengesetz in der Fassung der Bekanntmachung vom 28. Juni 2007 (BGBl. I S. 1206), das zuletzt durch Artikel 1 des Gesetzes vom 22. Dezember 2023 (BGBl. 2023 I Nr. 409) geändert worden ist*, 2023.

[21] scikit-learn-developers. "KMeans." https://scikit-learn.org/1.5/modules/generated/sklearn.cluster.KMeans.html (accessed 09.02., 2025).

[22] scikit-learn-developers. "DBSCAN." https://scikit-learn.org/1.5/modules/generated/sklearn.cluster.DBSCAN.html (accessed 09.02., 2025).