

CLASSIFICATION OF CONSTRUCTION FIRMS BASED ON BIM ROLES AND BIM LEVELS USING MACHINE LEARNING TECHNIQUES

H. Sadeh¹*, C. Mirarchi², A. Pavan³

¹ Dept. of Architecture, Built Environment and Construction Engineering, Politecnico di Milano, Milan, Italy -
hooman.sadeh@polimi.it

² Dept. of Architecture, Built Environment and Construction Engineering, Politecnico di Milano, Milan, Italy -
claudio.mirarchi@polimi.it

³ Dept. of Architecture, Built Environment and Construction Engineering, Politecnico di Milano, Milan, Italy -
alberto.pavan@polimi.it

Commission V, WG V/7

KEYWORDS: BIM Roles, BIM Levels, Machine Learning, BIM Education, Construction Management

ABSTRACT:

Application of Building Information Modelling (BIM) within the AEC industry has been evolving. With new developments and increasing capabilities, BIM is reshaping the design, construction, and operation, and maintenance processes and revolutionizing the entire functions of building life cycles. To maximize BIM benefits and take advantage of its capabilities, it is imperative that project stakeholders define specific roles and responsibilities within projects; to employ professionals with high levels of BIM proficiency, expertise, and knowledge. This study aims to classify the construction firms into different clusters based on their BIM capabilities, implementation, BIM levels, and type of BIM roles they employ for construction projects. It will further predict and classify BIM levels at company level according to its usage. The methodology was based on a survey design which consisted of application an online questionnaire that was distributed to AEC professionals in the industry. 61 suitable responses were analysed, using different supervised and unsupervised machine learning algorithms, including Cluster Analysis, K-Nearest Neighbours algorithm (k-NN), Random Forest, and Gradient Boosting. The findings showed most firms were not applying BIM on their projects and the majority of those that did were not utilizing it in its full potential. Firms were further classified in terms of BIM levels and types of BIM applications they utilize on construction projects. The results showed that Random Forest had the highest performance and the most accuracy, comparing with KNN and Gradient Boosting, even though the performance and predictions results produced by all models were in proximity of one another.

1. INTRODUCTION

Building Information Modelling (BIM) is a current trend within the construction industry. According to NIBS, “BIM is a digital representation of physical and functional characteristics of a facility. As such it serves as a shared knowledge resource for information about a facility forming a reliable basis for decisions during its lifecycle from inception onward” (NIBS, 2021). ISO 16757-1: 2015 defines BIM as “construction of a model that contains the information about a building from all phases of the building life cycle” (ISO, 2015). BIM has been utilized within the construction industry for a long time, and recently, its usage has increased among the BIM-based construction networks, however, collaborations among stakeholders remain challenging (Oraee et al., 2019). BIM has gained significant momentum and evolved during the last decade, however, its implementation and usage are not universally standardized and differ significantly, depending on the context in which it is used for (Hooper, 2015). Its application for designers, architects, contractors, and owner/operators may be different due to different unique intended purposes, however, there is a commonality that all these groups share when utilizing BIM, and that it to share different types of project data and information. Depending on

the size of the firms and their need for a particular purpose of usage, they may employ experts for one or all the following roles such as BIM modeler, BIM specialist, and BIM coordinator. Davies et al. (2017) investigated the definition of different BIM roles and found that there exists so many different names and variations depending on the project and specific country standards and guidelines. They interpret BIM managers as those responsible for overall creation, production, and implementation of BIM plans and protocols on projects, and additionally, describe BIM coordinators as those responsible for exchange of models and information, working directly under the supervision of BIM managers. Further, BIM specialist is a person that does the modelling and has an advance knowledge of information management and the BIM software (“BIM Manager, BIM Specialist and BIM Coordinator roles,” n.d.). Ellis (2020) referred to BIM specialist as ‘BIM Technician’ and stated that they must “be able to understand not only how a building fits together but how a building can be modelled accurately in a BIM environment.” Although these roles are crucial to application and execution of BIM on construction projects, they are not sustainable and are always evolving due to advancements in BIM (Akintola et al., 2017). BIM professionals play the role of change agents within the industry. Although these roles are accepted in the construction community, there is not a universal and standard definition of duties and responsibilities (Bosch-Sijtsema et al., 2019). In

* Corresponding and main author

addition, each firm uses BIM for a different purpose as mentioned earlier. This could be for design, rendering, information-sharing, clash detection, geometry representation, and other purposes. This has resulted in creation of different BIM levels within the BIM community. These levels start from BIM level 0 and include 2D, 3D, 4D, and 5D (Lorek, 2021). These are also referred to as maturity levels which correspond to level of information exchange in the construction sector. “2D BIM is a digital geometric model that constitutes an X and a Y axis associated with further information. 3D BIM is a digital geometric model that constitutes an X, Y and Z axis associated with further information” (Hamil, 2021). While 4D and 5D correspond to scheduling and cost of a project, respectively. Based on these existing differences across the construction industry as well as existence of different standards and specifications, the need to evaluate and understand the structure of construction firms in terms of BIM usage and utilization is very imperative. This will, in fact, increase the possibility of more collaboration among construction firms of different backgrounds and will assist the industry on building and developing a uniform platform for information sharing and smooth transfer of knowledge across all sectors of the construction sector. The remaining structure of this research is as follows: the purpose of research, research methods and materials, results and discussion, and conclusion.

1.1 Purpose of Research

This research intends to categorize the construction firms into different groups that have similar characteristics, based on different BIM-related attributes such as BIM usage, BIM roles, and BIM levels. Additionally, it will classify and predict the construction firms’ BIM levels in terms BIM roles: BIM specialist, BIM coordinator, BIM manager, and overall, BIM utilization. The purpose is to evaluate the compatibility within each observation (group) along with attributing BIM factors, and to further understand and categorize the different types of BIM levels that are being used in the construction sector.

2. RESEARCH METHODS AND MATERIALS

2.1 Data Collection

The methodology consisted of implementation of a survey design that was developed based on a structured online questionnaire. The survey was part of a larger study, and for the purposes of this paper, some portions of the data were used from the original questionnaire as secondary data. The questionnaire was distributed to the members of the construction industry via e-mail and dispersed in different LinkedIn groups, therefore, making the calculation of response rate not feasible. The questions were closed-ended and multiple-choice questions. A total of 170 responses were received, but only 61 were deemed suitable and complete in the form of following groups: 13 respondents from micro firms with less than 10 employees, 10 respondents from small firms having between 10 and 50 employees, 6 medium firms that had between 50 and 100 employees, and 32 large firms with over 100 employees, as shown in Figure 1.

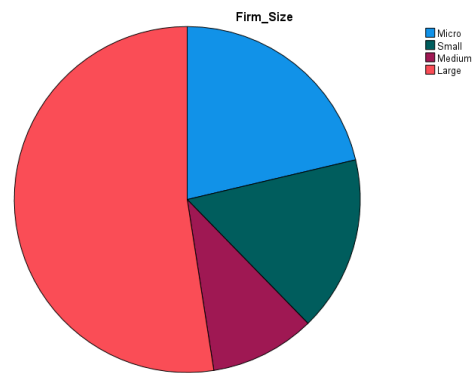


Figure 1. Profile of respondents by firm size.

The profiles of respondents in terms of firms’ specialty and country of origin are shown in Tables 1 and Table 2, respectively. Most of the respondents (over two-thirds) were from the USA and the rest were from other countries across the globe. The ‘Mix’ category, shown in Table 1, represents the firms that engaged in more than one specialty, e.g., performing both commercial and residential works.

		Valid		Cumulative	
		Frequency	Percent	Percent	Percent
Valid	Architectural	5	8.2	8.2	8.2
	Commercial	18	29.5	29.5	37.7
	Consultant	7	11.5	11.5	49.2
	Engineering	1	1.6	1.6	50.8
	Industrial	3	4.9	4.9	55.7
	Mix	17	27.9	27.9	83.6
	Owner	5	8.2	8.2	91.8
	Residential	3	4.9	4.9	96.7
	Subcontractor	2	3.3	3.3	100.0
	Total	61	100.0	100.0	

Table 1. Firm specialty.

		Valid		Cumulative	
		Frequency	Percent	Percent	Percent
Valid	Canada	2	3.3	3.3	3.3
	Egypt	1	1.6	1.6	4.9
	India	3	4.9	4.9	9.8
	Italy	5	8.2	8.2	18.0
	Netherlands	2	3.3	3.3	21.3
	Nigeria	1	1.6	1.6	23.0
	Sweden	1	1.6	1.6	24.6
	UK	2	3.3	3.3	27.9
	USA	44	72.1	72.1	100.0
	Total	61	100.0	100.0	

Table 2. Country of respondents.

2.2 Cluster Analysis

SPSS Modeler was utilized to group up the construction firms, in terms of BIM attributes, into different clusters. TwoStep Cluster Analysis, which is an unsupervised machine learning technique, was deemed appropriate for this computation since the questionnaire contained binary data and was of categorical nature. The data was divided into training and test partitions. 70% of the data was used for the training set and 30% was utilized for testing, respectively.

2.3 Supervised Algorithms for Classification

Classification of BIM levels was conducted by machine different supervised machine learning techniques to model the associations and dependencies between the predicted output target and the input features. The following supervised algorithms were utilized and compared as part of this process. (1) KNN, (2) Random Forest, and (3) Gradient Boosting. 10-fold cross-validation was used to divide the data set into 10 different folds as demonstrated by Vabalas et al. (2019). Each fold was given an opportunity to be used as a subsample for testing. One sample was retained and the remaining of the folds were used for training. The process was repeated 10 times and each of the subsamples were used once as the validation data. The performances of the models were evaluated using class performance and overall performance, comparing area under the curve, F1-score, precision, accuracy, and specificity. The following equations were utilized for this process.

$$\begin{aligned} \text{Accuracy} &= (\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN}) & (1) \\ \text{Precision} &= \text{TP} / (\text{TP} + \text{FP}) & (2) \\ \text{Recall} &= \text{TP} / (\text{TP} + \text{FN}) & (3) \\ \text{Specificity} &= \text{TN} / (\text{TN} + \text{FP}) & (4) \\ \text{F1} &= (2 * \text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall}) & (5) \end{aligned}$$

Where:

TP is the number of positive classes that are correctly predicted by the model.

FP is the number of positive classes that are incorrectly classified by the model.

TN is the number of negative classes that are predicted correctly by the model.

FN represents the number of negative classes that are predicted incorrectly by the model.

3. RESULTS AND DISCUSSION

3.1 Firm Clusters

Five variables of BIM Levels, BIM Usage, BIM Coordinator, BIM Manager, and BIM Specialist were used in the cluster analysis, resulting in 4 clusters as shown in Figure 2. Further, Figure 3 shows the predictor importance for each variable, with BIM Levels having the highest importance of predictability and BIM specialist being the lowest among the 5 variables. The average Silhouette value was 0.6, which means that the results are statistically good. The distributions show the number of variables assigned to each cluster. By default, clusters are sported from left to right based on size. According to Figure 2, cluster 1 has the highest number of variables (24) and cluster 3 has the least number of variables (8), 39.3% and 13.1%, respectively.

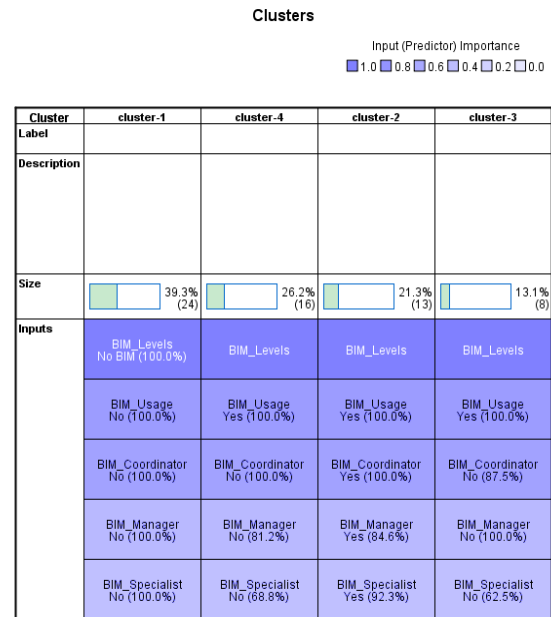


Figure 2. Summary of clusters.

BIM Levels consisted of four different categories. Firms that only used 3D geometrical models, firms that only used 3D geometrical representations, firms that employed information geometrical models linked to other software, and firms that did not utilize BIM at all. Firms in cluster 1, which is the largest cluster, do not utilize BIM and could be classified as non-digitalized firms. Firms in cluster 2 are advanced in terms of BIM usage and digitalization, with all having a BIM coordinator role, 92.3% having a BIM specialist, and 84.6% having a BIM manager. The BIM level categories, i.e., the purpose of usage is equally distributed among these firms, utilizing BIM for all the categories. Cluster 3 could be considered firms that are at a low-intermediate level in terms of BIM usage, based on the percentage of BIM roles that they employ within their companies. None of the firms in this group have a BIM manager, 87.5% do not have a BIM coordinator, only 37.5% have a BIM specialist, and they only utilize BIM (66.7%) as an information geometrical model linked to other software. Lastly, cluster 4 firms could be deemed as intermediate level firms regarding BIM usage. These firms only employ BIM managers and specialists, 21.4% and 25%, respectively. Moreover, these firms apply BIM in the context of 3D geometrical models (60%) as well as 3D geometrical

representations (66.7%).

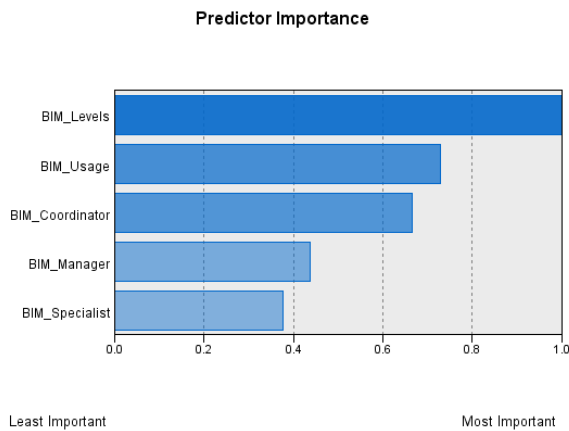


Figure 3. Variable predictor importance.

Tables 3, 4, 5, 6, and 7 show the cluster profiles for the 5 variables that were used in the cluster analysis.

		Cluster				
		1	2	3	4	Combined
Only 3D	Frequency	0	4	0	6	10
Geometrical Models	Percent	0.0%	40.0%	0.0%	60.0%	100.0%
Only 3D	Frequency	0	5	0	10	15
Geometrical Representation	Percent	0.0%	33.3%	0.0%	66.7%	100.0%
Information	Frequency	0	4	8	0	12
Geometrical Model Linked to other Tools	Percent	0.0%	33.3%	66.7%	0.0%	100.0%
No BIM	Frequency	24	0	0	0	24
	Percent	100.0%	0.0%	0.0%	0.0%	100.0%

Table 3. BIM levels.

		Cluster				
		1	2	3	4	Combined
Yes	Frequency	0	13	1	0	14
	Percent	0.0%	92.9%	7.1%	0.0%	100.0%
No	Frequency	24	0	7	16	47
	Percent	51.1%	0.0%	14.9%	34.0%	100.0%

Table 4. BIM coordinator.

		Cluster				
		1	2	3	4	Combined
Yes	Frequency	0	11	0	3	14
	Percent	0.0%	78.6%	0.0%	21.4%	100.0%
No	Frequency	24	2	8	13	47
	Percent	51.1%	4.3%	17.0%	27.7%	100.0%

Table 5. BIM manager.

		Cluster				
		1	2	3	4	Combined
Yes	Frequency	0	12	3	5	20
	Percent	0.0%	60.0%	15.0%	25.0%	100.0%

	No	Frequency	24	1	5	11	41
	Percent	58.5%	2.4%	12.2%	26.8%	100.0%	

Table 6. BIM specialist.

		Cluster				
		1	2	3	4	Combined
Yes	Frequency	0	13	8	16	37
	Percent	0.0%	35.1%	21.6%	43.2%	100.0%
No	Frequency	24	0	0	0	24
	Percent	100.0%	0.0%	0.0%	0.0%	100.0%

Table 7. BIM usage.

Further, what can be obtained based on these 4 clusters is that the contractors do not fully engage BIM in all their operations. There are only limited number of firms that use BIM equally based on its full potential across all design and construction activities. Moreover, the results are relatively interesting, given that the sample size consisted mostly of larger firms, and large firms are the ones that should be advanced in BIM utilization.

3.2 BIM Levels Classification

Four variables of BIM Usage, BIM Specialist, BIM Manager, and BIM Coordinator were utilized as input variables and BIM Levels was chosen as the target variable for the machine learning algorithms. The purpose was to classify the type of BIM usage among construction firms based on BIM roles within those firms.

Figure 4 below is the ROC (Area Under the Receiver Operating Characteristics) curve for the three algorithms, which shows the true positive rate against the false positive rate. According to this graph, the area is high and close to 1, which is an indication of good performance by all three models.

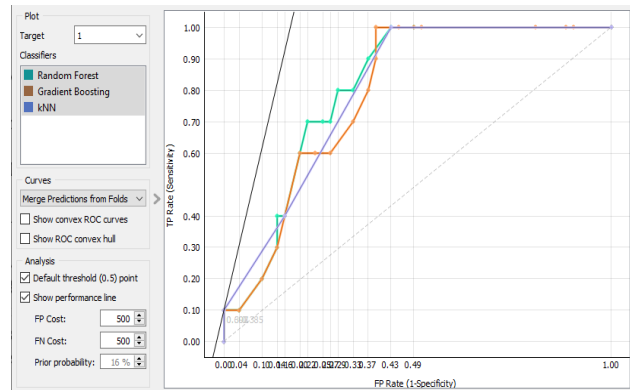


Figure 4. ROC curve.

Table 8 below is a summary of the machine learning performance statistics for the three models, among which Random Forest has the highest performance (AUC), comparing to the other two models: Gradient Boosting and KNN. However, all three models are very close, with AUC values of 0.861, 0.843, and 0.845.

Model	AUC	CA	F1	Precision	Recall
Random Forest	0.861	0.656	0.653	0.652	0.656
Gradient Boosting	0.843	0.639	0.616	0.615	0.639

KNN	0.846	0.623	0.621	0.622	0.623
-----	-------	-------	-------	-------	-------

Table 8. Model accuracy.

Although all models were close in terms of accuracy, Random Forest was the better model comparing with the other two models. With a 65% accuracy, it implies that 4.5 out of every 10 BIM levels were predicted correctly. Although the accuracy is somewhat considered low, it is contributed to the small sample of the data and is not a representation of the model performance. Based on the Recall values of around 65%, every 4.5 out of 10 BIM levels are misclassified.

Figures 5, 6, and 7 represent the confusion matrix for our machine learning models. Confusion matrices are utilized for performance assessment of classification algorithms by comparing the target values and with the predicted values. It also provides and shows the different types of errors made by the model.

		Predicted				Σ
		1	2	3	4	
Actual	1	5	2	3	0	10
	2	3	8	4	0	15
	3	4	5	3	0	12
	4	0	0	0	24	24
Σ		12	15	10	24	61

Figure 5. Random Forest confusion matrix.

		Predicted				Σ
		1	2	3	4	
Actual	1	2	4	4	0	10
	2	2	11	2	0	15
	3	2	8	2	0	12
	4	0	0	0	24	24
Σ		6	23	8	24	61

Figure 6. Gradient Boosting confusion matrix.

		Predicted				Σ
		1	2	3	4	
Actual	1	4	3	3	0	10
	2	4	7	4	0	15
	3	4	5	3	0	12
	4	0	0	0	24	24
Σ		12	15	10	24	61

Figure 7. KNN confusion matrix.

4. CONCLUSION

Digitalization of the construction sector has been an on-going debate and discussion among researchers and the construction

industry. Many research projects are being devoted for implementation of new technologies within the construction industry. These efforts are being made in response to low productivity and lack of technological advancement. The construction sector is lagging, compared with other sectors such as manufacturing, in terms of innovation. BIM is one of the most important technological advances that is being utilized by construction contractors and subcontractors due to many benefits it provides. However, due to complexity of usage, financial cost of implementation, training, and compatibility issues with other partners in the industry, its adaptation has been faced with some challenges by some contractors. BIM levels and BIM roles are critical elements to successful operation of BIM during construction projects. The purpose of this paper was to classify the construction contractors according to their BIM usage, BIM roles and levels. A cluster analysis was conducted, and the results showed that most of the firms within the sample did not fully take advantages of what BIM has to offer. They mostly did not use BIM or used it in a limited capacity. This proves that lack of knowledge and proficiency is evident, although most of the firms within the sample were large U.S. firms. Secondly, supervised machine learning techniques were utilized to classify the firms' BIM levels based on their BIM roles. Random Forest, Gradient Boosting, and KNN were used for classification and prediction as part of this process. Random Forrest was found to be the most accurate model out of the three algorithms. The results showed that the models were almost equal in terms of performance, though the accuracy was low, consistency of results was evident comparing all the three models. The limitations of this study included the small sample size that affected the precision and accuracy of the algorithms for this study. Moreover, BIM roles have different definitions, name variations, and responsibilities in different projects, guidelines, and countries, therefore, this may have impacted the responses, producing a less accurate results by machine learning algorithms. Additionally, most of the firms were larger construction firms from the USA, therefore the results can not be generalized for all sectors and all countries. It is suggested that future studies use larger samples that also include more smaller and medium-sized contractors. Moreover, it is suggested that the firms' BIM experience levels be taken into consideration.

REFERENCES

- Akintola, A., Venkatachalam, S., & Root, D. (2017). New BIM Roles' Legitimacy and Changing Power Dynamics on BIM-Enabled Projects. *Journal of Construction Engineering and Management*, 143(9), 04017066. [https://doi.org/10.1061/\(asce\)co.1943-7862.0001366](https://doi.org/10.1061/(asce)co.1943-7862.0001366)
- BIM Manager, BIM Specialist and BIM Coordinator roles and responsibilities. (n.d.). BibLus. Retrieved January 5, 2022, from <https://biblus.accasoft.com/en/bim-manager-bim-specialist-and-bim-coordinator-roles-and-responsibilities/>
- Bosch-Sijtsema, P. M., Gluch, P., & Sezer, A. A. (2019). Professional development of the BIM actor role. *Automation in Construction*, 97, 44–51. <https://doi.org/10.1016/j.autcon.2018.10.024>
- Building Information Management (BIM) Council | National Institute of Building Sciences. (2021, November 1). National Institute of Building Science. <https://www.nibs.org/bimc> ISO

16757–1:2015. (2015). ISO.
<https://www.iso.org/standard/57613.html>

Ellis, M. (2020). BIM Roles and Responsibilities in the AEC Industry. REBIM®. Retrieved January 5, 2022, from <https://rebim.io/bim-specialists-roles-and-responsibilities-in-the-aec-industry/>

Hamil, D. S. (2021, September 9). Bim dimensions - 3D, 4d, 5D, 6d BIM explained. NBS. Retrieved January 6, 2022, from <https://www.thenbs.com/knowledge/bim-dimensions-3d-4d-5d-6d-bim-explained>

Hooper M. (2015). BIM standardisation efforts - the case of Sweden, ITcon Vol. 20, pg. 332-346, <https://www.itcon.org/2015/21>
ISO 16757–1:2015. (2015). ISO.
<https://www.iso.org/standard/57613.html>

Kathryn Davies, Suzanne Wilkinson, Dermott McMeel (2017). A review of specialist role definitions in BIM guides and standards. Journal of Information Technology in Construction (ITcon), Vol. 22, pg. 185-203, <http://www.itcon.org/2017/10>

Lorek, S. (2021). *What is BIM? | Building Information Modeling | Levels of BIM*. Constructible. Retrieved January 5, 2022, from <https://constructible.trimble.com/construction-industry/what-is-bim-building-information-modeling>

Oraee, M., Hosseini, M. R., Edwards, D. J., Li, H., Papadonikolaki, E., & Cao, D. (2019). Collaboration barriers in BIM-based construction networks: A conceptual model. International Journal of Project Management, 37(6), 839–854. <https://doi.org/10.1016/j.ijproman.2019.05.004>

Vabalas, A., Gowen, E., Poliakoff, E., & Casson, A. J. (2019). Machine learning algorithm validation with a limited sample size. PLOS ONE, 14(11), e0224365. <https://doi.org/10.1371/journal.pone.0224365>