2025 European Conference on Computing in Construction

CIB W78 Conference on IT in Construction

Porto, Portugal
July 14-17, 2025

INCORPORATION OF CONTEXT-BASED DATA TO REFINE MACHINE LEARNING MODELS IN INDUSTRIALIZED CONSTRUCTION

Mauricio Melgar[1], Yaser Miranda[1], Mohamed Al-Hussein[2], Beda Barkokebas[3,4]
[1]Faculty of Civil Engineering, Universidad Nacional de Ingeniería, Lima, Peru
[2]Department of Civil and Environmental Engineering, University of Alberta, Edmonton, AB, Canada
[3]Department of Construction Engineering and Management, Pontificia Universidad Católica de Chile, Santiago, Chile
[4]Centro Nacional de Excelencia para la Industria de la Madera (CENAMAD), Pontificia Universidad Católica de Chile, Santiago, Chile

Abstract

Industrialized construction promises to transform the industry by performing activities in shop floors to improve operational efficiency. However, the lack of integration between practical experience and real-time data limits more efficient operations on the shop floor. This study addresses this limitation by combining design parameters from BIM models with real-time production data collected via RFID in a semi-automated shop floor. By including context-based data in machine learning models, cycle time prediction accuracy was improved, compared to models containing only design-based data. highlighting how data granularity optimizes operational planning and production processes in Industrialized Construction.

Introduction

The construction industry faces significant challenges, such as low labor productivity, high volume of rework, widespread supply chain failures, and constantly evolving customer needs (Goulding et al., 2017). However, an even more critical barrier is the industry's inability to learn from its mistakes, as data generated during operations is not effectively leveraged (Bilal et al., 2016). These barriers are intrinsically related to the complex nature of the work performed, and the diversity of contexts found in the construction industry (San & Sothy, 2016).

In this context, industrialized construction (IC) emerges as an innovative approach in which most of the construction components are manufactured in a controlled environment (i.e., shop floor) and assembled on site (Goh & Goh, 2019). This approach enables significant productivity improvements through process simplification, standardization of work, and tighter control of operating fronts. In addition, it fosters continuous improvement in operational efficiency, making it a key strategy to address the structural problems of the industry.

Contemporary literature indicates that moving a significant part of the activities onsite to IC shop floors represents one of the best alternatives, due to the wide margins of productivity improvement it offers (Goh & Goh, 2019). There is also consensus that this transition will be achievable in the foreseeable future (Bowmaster, 2019). Within this context, technologies such as real-time production sensors have been identified as essential to improve competitiveness in this model, according to Dodge Data & Analytics (2020).

Despite its potential, real-time data integration in IC faces significant internal barriers that hinder process improvement in manufacturing plants. A key challenge is the limited set of key indicators for decision-making (Jang et al., 2021). Additionally, the shortage of trained professionals to interpret and utilize this data further restricts its practical application (Goulding et al., 2017). Lundkvist et al. (2010) emphasize that these challenges are deeply interconnected, as the absence of robust methodologies and theoretical frameworks for continuous data integration fosters skepticism among industry professionals. As a result, resistance to change and uncertainty persists, highlighting the need for a structured framework to build trust in data utilization and drive sustainable process improvements.

To address this gap, Qi et al. (2018) proposed a framework that integrates RFID and BIM in the manufacturing process of prefabricated and reinforced concrete components. They found the framework could establish an efficient, low cost, and easily implementable information capturing and sharing system for further planning and management of prefabrication.

Barkokebas et al. (2023) developed an exploratory data analysis (EDA) method to evaluate large volumes of real-time data for the purpose of evaluating improvements in an IC shop floor. This method combined two data regarding the manufacturing process: (1) manufacturing cycle times collected via radio frequency-identifiers (RFID) sensors installed in an IC shop floor and, (2) design-based data of manufactured panels contained in building information model (BIM) models. Despite the large amount of data, machine learning (ML) models did not present relevant predictions of cycle times due to the lack of context-based information about the current manufacturing process (Barkokebas et al., 2023). Moreover, there are still limited studies applying real-time data in IC in order to allow this type of data to be further used for process improvement.

Considering this, the present research addresses this gap by incorporating context-based data from the actual manufacturing process (actual queue lengths at workstations) in current machine learning models to achieve higher accuracy in predicting cycle times in IC shops. Using the results of Barkokebas et al. (2023) as baseline, the ML models will be updated with the developed context-based data to assess whether the inclusion of context-based data will yield more reliable results. This research contribution is twofold: (1) it presents novel ML methods for process improvement in IC shop floors, and (2) it provides evidence on the importance of context-based data over design-based data for cycle time predictions in IC.

Research methods

Definition of the case study

The case study focuses on a semi-automated industrialized construction operation located in Alberta, Canada, where wall, floor and ceiling panels are manufactured in different areas of the plant. For this analysis, the wall panel area was specifically selected because of its relevance to the overall production.

The manufacturing process starts at the framing station, (W01) where multiple panels are manufactured as per design specifications contained in BIM models. During the observation, it is determined that W01 dictates the production rate throughout the plant, considered a benchmark to evaluate the overall operation.

Barkokebas et al. (2023), performed a study on the same shop floor where a large dataset containing the design features and cycle times of 27,680 panels in W01 was used to predict future performance. The cycle times were collected through RFID antennas, which continuously monitor the locations and timestamps of each panel in production between September 2015 to October 2018. Despite a significant amount of reliable data, the predictive models demonstrated low accuracy, evidencing the need for more context-based data and expert knowledge during the development of machine learning models to achieve meaningful predictions in IC cutting plants. Indeed, the authors argued that more context-related data should allow for more accurate analyses of cycle time variability and optimize predictive models in future studies. Finally, the present study will seek to test the following hypothesis: *"The inclusion of context-based data will render better results over design-related data in the prediction of cycle times in IC shop floors"*.

Point of departure and context-based data preparation

An Exploratory Data Analysis (EDA) approach was adopted in this study due to its successful implementation in past studies handling large volumes of data (Velleman & Hoaglin (2012); Behrens (1997)). To ensure consistency with the comparison between the baseline presented in Barkokebas at el. (2023) and the proposed method, the same dataset is used between both studies resulting in a total of 7,819 units after data cleaning. The features processed in the ML models are presented in Figure 1 whereas the Figure also shows the context-based data generated and included in this present study. In order to create more context-based data, the present study used the timestamps generated with the RFID technology to develop the following production-related data:

- **WIP_1:** Number of panels under production between workstations 1 and 2.
- **WIP_2:** Number of panels under production between workstations 2 and 3.
- **WIP_Total:** Number of panels under production between workstations 1 and 4.

These parameters obtained from the RFID antennas will be combined with the characteristics of the panels (i.e., design-based data), from the BIM models, to subsequently form the production-based dataset (input). Figure 2 shows a breakdown of each of the elements of the production-based dataset. This final dataset is divided into two experiments designed to model the cycle times in W01 (output). For comparison, the R2 values of the following experiments will be compared with the previous results presented in Barkokebas et al. (2023).

E1: Design Parameters, WIP_1 and WIP_2.

E2: Design Parameters, WIP_1 and WIP_Total.

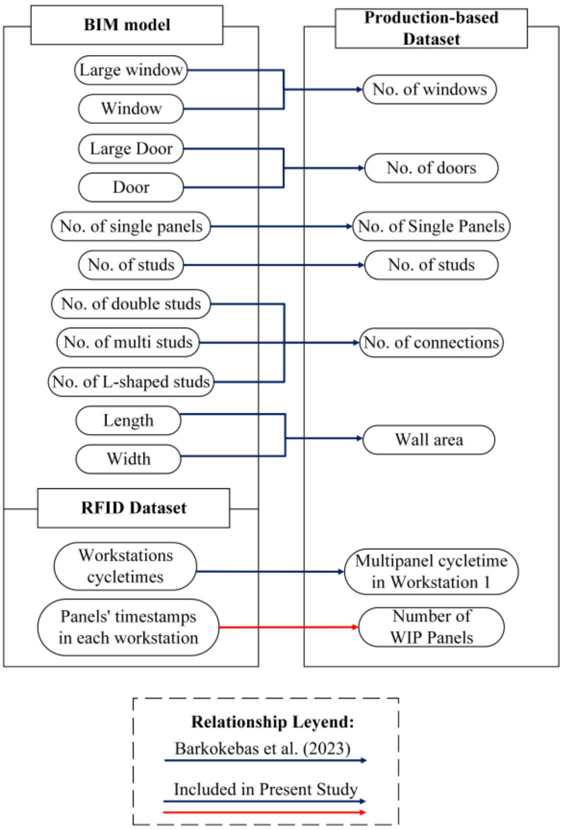

Figure 1: Production-based dataset on multi-panels

Data modelling

This subsection describes the selection of the ML models and other methods to be deployed in addressing the proposed study hypotheses. It is important to note that the parameters used for the models and methods are the ones used in Barkokebas et al. (2023), so a comparison can be made with their results. Additionally, the present study follows the data modeling proposed by Barkokebas et al. (2023) with additions as per indicated in Figure 2.

The panels were grouped based on the number of clusters generated and the outliers identified using the IQR and Hampel Filter (HF) methods. The clusterization algorithm HDBSCAN was first used for each experiment, as this algorithm provides a number of clusters instead of requiring it as an input. This number of clusters was then used as inputs for the clusterization algorithms K-means and Agglomerative. For the outlier detection, Interquartile range (IQR) and Hampel Filter methods were selected as the former is a classical method for this purpose, and the latter has good performance in detecting outliers in manufacturing cycle times in particular (Nishigaki et al., 2020). The size of each cluster varies depending on the clustering method and the outlier detection algorithm, these values are also used as the sample size used in the regression models (LR, RFR and SVR), while the outliers are discarded.

Finally, to evaluate the present hypothesis, three regression algorithms were implemented: linear regression (LR), random forest regression (RFR) and support vector regression (SVR) due to its implementation in the baseline study. While the databases have similarities, by adding context-based data, we generated a new database, so the hyperparameters required by advanced algorithms such as RFR and SVR were determined using the optimal settings provided by the GridSearchCV library. . A double cross validation was applied with a training and testing ratio of 90% and 10%, respectively, within each set, ensuring that the developed models will not present overfitting problems.

The performance of each model was measured through the coefficient of determination (R^2), evaluating the extent to which the panel production database explains the variability in cycle time forecasts. Finally, a comparison of the obtained performance (R^2) is made with the results of the experiments by Barkokebas et al. (2023), allowing to evaluate how production-related data influence the prediction of cycle times in IC workshops

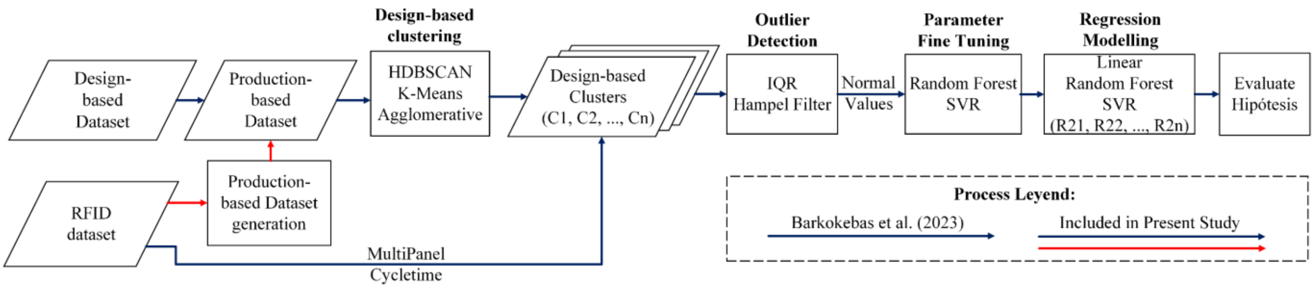

Figure 2: Machine learning modeling to test the hypothesis

Results and discussion

Clusters evaluation

In the baseline study, four clusters were identified through the HDBSCAN algorithm. However, by incorporating production-related parameters, the results varied depending on the experiments performed, as shown in Table 1. It should be noted that the experiments were performed by adding one more cluster to the original number provided by HDBSCAN. Table 2 shows the number of panels in each cluster and its defining features according to each experiment. The size of each cluster varies depending on the clustering method and the outlier detection algorithm, these values are also used as the sample size used in the regression models, while outliers are discarded.

Table 1: Number of clusters for each experiment

	Clusters	Clusters + 1
Experiment 1	4	5
Experiment 2	3	4

Table 2: Clusters and outliers created from the design-based dataset according to the proposed models indicating the total number of panels per cluster (top), normal panel durations(bottom left), and number of outliers detected in each cluster (bottom right)

Experiment	Model	Filter	Number of Clusters	Data for each Clúster					Cluster Features
				1	2	3	4	5	
1	HDBSCAN	IQR	4	810 761 | 49	2307 2151 | 156	942 889 | 53	3760 3502 | 258		windows, doors, WIP_E1 y WIP_E2
		HF	4	810 612 | 198	2307 1722 | 585	942 775 | 167	3760 2699 | 1061		
	K-means	IQR	4	1988 1884 | 104	1496 1366 | 130	2686 2499 | 187	1649 1530 | 119		windows, doors, WIP_E1 y WIP_E2
		HF	4	1988 1442 | 546	1496 1158 | 338	2686 1926 | 760	1649 1108 | 541		
		IQR	5	1816 1710 | 106	1397 1274 | 123	1924 1801 | 123	1532 1412 | 120	1150 1068 | 82	
		HF	5	1816 1306 | 510	1397 1083 | 314	1924 1403 | 521	1532 1034 | 498	1150 810 | 340	
	Agg	IQR	4	1997 1871 | 126	3079 2882 | 197	1188 1077 | 111	1555 1459 | 96		windows, doors, WI_E1, WIP_E2 and No. of panels
		HF	4	1997 1471 | 526	3079 2207 | 872	1188 935 | 253	1555 1162 | 393		
		IQR	5	3079 2882 | 197	1555 1459 | 96	1188 1077 | 111	823 754 | 69	1174 1090 | 84	
		HF	5	3079 2207 | 872	1555 1162 | 393	1188 935 | 253	823 599 | 224	1174 676 | 498	
2	HDBSCAN	IQR	3	1011 957 | 54	1242 1177 | 65	5566 5175 | 391			windows, doors, WIP_E1 y WIP_ET
		HF	3	1011 726 | 285	1242 1018 | 224	5566 4079 | 1487			
	K-means	IQR	3	3118 2919 | 199	1598 1454 | 144	3103 2908 | 195			windows, doors, WI_ E1, WIP_ET and No. of panels
		HF	3	3118 2207 | 911	1598 1253 | 345	3103 2231 | 872			
		IQR	4	1906 1800 | 106	1430 1304 | 126	2716 2537 | 179	1767 1643 | 124		
		HF	4	1906 1373 | 533	1430 1114 | 316	2716 1972 | 744	1767 1239 | 528		
	Agg	IQR	3	4311 4040 | 271	2390 2225 | 165	1118 1018 | 100			windows, doors, WI_ E1, WIP_ET and No. of panels
		HF	3	4311 3181 | 1130	2390 1672 | 718	1118 872 | 246			
		IQR	4	2390 2225 | 165	2189 2033 | 156	1118 1018 | 100	2122 1999 | 123		
		HF	4	2390 1672 | 718	2189 1668 | 521	1118 872 | 246	2122 1464 | 658		

The results in Table 2 are consistent with the baseline study (Barkokebas et al. (2023), where it was identified that the number of openings (windows and doors) remains the main defining features factor for multipanels. This finding reinforces previous studies (Chien et al. (2007); Chen et al. (2018)), which concluded that process times at the framing station are mainly influenced by the number of openings. Nevertheless, the newly developed context-based data (e.g., W1 and WIP) was also found significant to determine panels in each cluster. Indeed, these were identified as a key secondary feature in all clusterizations, regardless of the experiment or whether the "clusters + 1" configuration was used.

Performance of regression models

Tables 3 and 4 present the R^2 values obtained through the LR, RFR and SVR regression algorithms applied to predict the cycle times at the framing station (W01). In all experiments, the RFR algorithm presented the best performance, as found by Gyulai et al. (2018) and Lingitz et al. (2018) when predicting cycle times in shop floors. On the contrary, the performance of the SVR algorithm was consistently low, even reaching negative R^2 values in some experiments. This could be explained by its sensitivity to the selection of hyperparameters, which are the same as in the baseline study, and by the high variability in the production plant data, which hinders its ability to generalize

Table 3: R2 values from different regression analyses performed for each cluster for experiment 1.

Model	**R2 value**																	
	C1			**C2**			**C3**			**C4**			**C5**			**Average**		
	LR	**RFR**	**SVR**	**LR**	**RFR**	**SVR**	**LR**	**RFR**	**SVR**	**LR**	**RFR**	**SVR**	**LR**	**RFR**	**SVR**	**LR**	**RFR**	**SVR**
HDBSCAN + IQR	0.20	0.24	-0.19	0.23	0.24	0.04	0.09	0.14	-0.10	0.18	0.23	0.18	-	-	-	0.17	0.21	-0.02
HDBSCAN + HF	0.36	0.30	-0.28	0.25	0.27	0.11	0.10	0.06	0.00	0.20	0.28	0.20	-	-	-	0.23	0.23	0.01
K-means (4) + IQR	0.17	0.16	-0.05	0.14	0.17	-0.13	0.11	0.12	-0.22	0.30	0.32	0.16	-	-	-	0.18	0.19	-0.06
K-means (4) + HF	0.19	0.22	-0.15	0.06	0.06	-0.43	0.16	0.15	-0.17	0.30	0.33	-0.01	-	-	-	0.18	0.19	-0.19
K-means (5) + IQR	0.19	0.19	-0.04	0.11	0.11	-0.06	0.07	0.08	-0.31	0.30	0.32	-0.02	0.21	0.18	-0.07	0.18	0.17	-0.10
K-means (5) + HF	0.04	0.10	-0.54	0.07	0.04	-0.55	0.14	0.18	-0.51	0.25	0.22	-0.18	0.24	0.24	-0.11	0.15	0.16	-0.38
Agg. (4) + IQR	0.26	0.31	0.14	0.19	0.20	-0.06	0.06	0.05	-0.56	0.26	0.26	-0.01	-	-	-	0.19	0.21	-0.12
Agg. (4) + HF	0.40	0.37	0.13	0.10	0.15	-0.31	0.12	0.09	-0.62	0.23	0.25	-0.03	-	-	-	0.21	0.22	-0.21
Agg. (5) + IQR	0.14	0.16	-0.03	0.19	0.21	0.13	0.12	0.12	-0.39	0.19	0.26	-0.22	0.12	0.13	-0.23	0.15	0.18	-0.15
Agg. (5) + HF	0.10	0.14	0.27	0.15	0.18	0.33	0.03	0.01	-0.44	0.24	0.20	-0.11	0.00	0.04	-0.99	0.10	0.11	-0.19
																0.17	0.19	-0.14

Note: HF = Hampel filter, Agg. = Agglomerative, LR = Linear regression, RFR = Random forest regression, SVR = Support vector regressor.

Table 4: R2 values from different regression analyses performed for each cluster for experiment 2.

Model	**R2 value**														
	C1			**C2**			**C3**			**C4**			**Average**		
	LR	**RFR**	**SVR**	**LR**	**RFR**	**SVR**	**LR**	**RFR**	**SVR**	**LR**	**RFR**	**SVR**	**LR**	**RFR**	**SVR**
HDBSCAN + IQR	0.11	0.15	-0.24	0.01	0.12	-0.18	0.12	0.22	0.09	-	-	-	0.08	0.16	-0.11
HDBSCAN + HF	0.10	0.13	-0.50	0.11	0.15	-0.11	0.21	0.28	0.12	-	-	-	0.14	0.19	-0.16
K-means (3) + IQR	0.22	0.27	0.13	0.14	0.13	-0.13	0.13	0.18	-0.11	-	-	-	0.16	0.19	-0.04
K-means (3) + HF	0.21	0.29	0.07	0.07	0.09	-0.50	0.20	0.24	-0.08	-	-	-	0.16	0.20	-0.17
K-means (4) + IQR	0.25	0.24	-0.02	0.11	0.09	-0.21	0.14	0.16	-0.04	0.27	0.33	0.07	0.19	0.20	-0.05
K-means (4) + HF	0.16	0.19	-0.19	0.10	0.15	-0.23	0.14	0.18	-0.05	0.34	0.39	0.11	0.19	0.23	-0.09
Agg. (3) + IQR	0.19	0.22	0.05	0.16	0.13	-0.21	0.02	0.08	-0.10	-	-	-	0.12	0.14	-0.08
Agg. (3) + HF	0.23	0.28	0.06	0.13	0.13	-0.19	0.09	0.06	-0.29	-	-	-	0.15	0.16	-0.14
Agg. (4) + IQR	0.16	0.18	-0.05	0.23	0.30	0.12	0.04	0.10	-0.27	0.11	0.14	-0.04	0.13	0.18	-0.06
Agg. (4) + HF	0.10	0.12	0.22	0.27	0.32	0.11	0.11	0.11	-0.42	0.13	0.18	-0.23	0.15	0.18	-0.08
													0.15	0.18	-0.10

Note: HF = Hampel filter, Agg. = Agglomerative, LR = Linear regression, RFR = Random forest regression, SVR = Support vector regressor.

Comparison with previous studies

Table 5 compares the average R^2 values obtained in this study with those reported by Barkokebas et al. (2023). The results show a marked improvement in the current experiments, which is attributed to the inclusion of additional production-related parameters that provide a deeper understanding of plant performance. These results are consistent with the ones found in (Mohsen, 2021).

Table 5: Comparison between R2 mean values for each regression analysis compared baseline

		LR	**RFR**	**SVR**
Barkokebas et al. (2023)		0.07	0.10	0.00
Experiment 1	**R2**	0.17	0.19	-0.14
	% Improvement	258%	189%	0%
Experiment 2	**R2**	0.15	0.18	-0.10
	% Improvement	217%	187%	0%

Despite the improvement in the results, it is important to clarify that these are still not enough to make the model predictively reliable, as R2 values are well below 0.70 in each cluster. This could be attributed to:

1. The high variability inherent in the production plant (Wen et al., 2017).
2. The inadequacy of design parameters as predictors of cycle times (Mohsen, 2021).
3. More context-based data is required to predict cycle time in the shop floor

Conclusions

The present study implemented an innovative approach based on digitization to optimize processes in industrialized construction systems (IC). This approach combined real-time data collected using RFID antennas, design features extracted from BIM models, and advanced analysis techniques such as Exploratory Data Analysis (EDA) and machine learning algorithms. These elements enabled rigorous analysis of large volumes of data, providing a more robust and accurate approach than traditional simulation-based methods.

The inclusion of production parameters in the clustering tools had a significant impact on the identification of key patterns and relationships within the data. This resulted in changes in the number of clusters identified in the experiments. Nevertheless, the number of panels at the first workstation (WIP01) was also consistently identified as a significant secondary feature in all clusterization algorithms, supporting the hypothesis that context-based data are important in predicting manufacturing cycle times. That study concluded that production rate in W01 can generate imbalances in later seasons, affecting overall system efficiency. Both analyses underscore the need for balanced strategies that integrate semi-automated and manual stations to improve production flow.

Moreover, the present hypothesis was confirmed by demonstrating that the average linear correlation coefficients (R^2) in the four experiments significantly exceeded the values reported in the baseline study, with relative improvements ranging from 187% to 258%. However, despite these relative improvements, the absolute R^2 values remain low (ranging from 0.15 to 0.19), indicating that the models still lack sufficient predictive reliability for practical implementation. These improvements were evident in the regressions performed with the Linear Regression (LR) and Random Forests (RFR) models, while SVR was excluded from the analysis due to its poor results in both the baseline and present study. The present results are consistent with the baseline study, which indicates that SVR is not a suitable algorithm for predicting cycle times compared to LR and RFR.

This study presents a structured and replicable framework for evaluating improvement measures in CI systems. However, it is considered essential to extend its application to additional contexts, such as plants with lower levels of automation or smaller data sets. Since the analysis was limited to the use of the base study algorithms for a fair comparison, future research should explore hyperparameter optimization in clustering tools such as HDBSCAN, K-means and Agglomerative, as well as in regression algorithms such as LR, RFR and SVR. These elements are crucial, as they significantly influence the definition of the clusters and the R^2 coefficients obtained. It is also recommended as future research to include more context-based data to develop more accurate prediction models while performing specific studies to evaluate significant features to determine the similarity of panels thought clusterization.

This study focuses on the case study of an IC shop floor of manufactured panels. Then, carrying out the experiment in other IC environments is proposed as future research. Moreover, the development of a practical framework of guidelines for practitioners on how to collect, process, and leverage similar data from this study in their operations is proposed as future research as well, Finally, the applied method, and the results obtained not only strengthen the planning and adaptability of the IC shop floors in the face of changes in demand but also serve as a reference and motivation for other factories to adopt similar systems. This would allow analysis of their current production situation, generating new databases that in turn will feed future research and contribute to the development of new recommendations.

Acknowledgments

The authors appreciate AQBUILT for their collaboration and support of this paper. Also, the authors acknowledge the support of ANID BASAL FB210015 CENAMAD and ANID Fondecyt Iniciación 11241027.

References

Barkokebas, B., Martinez, P., Bouferguene, A., Hamzeh, F., & Al-Hussein, M. (2023). Digitalization-based process improvement and decision-making in offsite construction. *Automation in Construction*, *155*, 105052. https://doi.org/10.1016/j.autcon.2023.105052

Behrens, J. T. (1997). Principles and procedures of exploratory data analysis. *Psychological Methods*, *2*(2), 131–160. https://doi.org/10.1037/1082-989X.2.2.131

Bilal, M., Oyedele, L. O., Qadir, J., Munir, K., Ajayi, S. O., Akinade, O. O., Owolabi, H. A., Alaka, H. A., & Pasha, M. (2016). Big Data in the construction industry: A review of present status, opportunities, and future trends. *Advanced Engineering Informatics*, *30*(3), 500–521. https://doi.org/10.1016/j.aei.2016.07.001

Chen, Q., García de Soto, B., & Adey, B. T. (2018). Construction automation: Research areas, industry concerns and suggestions for advancement. *Automation in Construction*, *94*, 22–38. https://doi.org/10.1016/j.autcon.2018.05.028

Chien, C.-F., Wang, W.-C., & Cheng, J.-C. (2007). Data mining for yield enhancement in semiconductor manufacturing and an empirical study. *Expert Systems with Applications*, *33*(1), 192–198. https://doi.org/10.1016/j.eswa.2006.04.014

D. Gyulai, A. Pfeiffer, G. Nick, V. Gallina, W. Sihn, L. Monostori, Lead time prediction in a flow-shop environment with analytical and machine learning approaches, IFACPapersOnLine. 51 (2018) pp. 1029–1034. https://doi.org/10.1016/j.ifacol.2018.08.472.

Goh, M., & Goh, Y. M. (2019). Lean production theory-based simulation of modular construction processes. *Automation in Construction*, *101*, 227–244. https://doi.org/10.1016/j.autcon.2018.12.017

Jang, J., Ahn, S., Cha, S. H., Cho, K., Koo, C., & Kim, T. W. (2021). Toward productivity in future construction: mapping knowledge and finding insights for achieving successful offsite construction projects. *Journal of Computational Design and Engineering*, *8*(1), 1–14. https://doi.org/10.1093/jcde/qwaa071

J. Bowmaster, J. Rankin, A Research Roadmap for Off-Site Construction:Automation and Robotics, Modular and Offsite Construction Summit Proceedings, 2019, pp. 173–180, https://doi.org/10.29173/mocs91

J.S. Goulding, F.P. Rahimian, M. Arif, A. Akintoye, I. Faraj, Offsite Manufacturing in Architecture Engineering and Construction: An Unfolding Story, in: Proceedings of the International Research Conference 2017, University of Salford, Manchester, 2017, pp. 862–871, https://nrl.northumbria.ac.uk/id/eprint/34608/

Lundkvist, R., Meiling, J., & Vennström, A. (2010). Digitalization of Inspection Data; A Means for Enhancing Learning and Continuous Improvements? *Association of Researchers in Construction Management*, *2*, 6–8.

L. Lingitz, V. Gallina, F. Ansari, D. Gyulai, A. Pfeiffer, W. Sihn, L. Monostori, Lead time prediction using machine learning algorithms: A case study by a semiconductor manufacturer, Procedia CIRP. 72 (2018) pp. 1051–1056. https://doi.org/10.1016/j.procir.2018.03.148

Qi, Bing & Chen, Ke & Costin, Aaron. (2018). RFID and BIM-Enabled Prefabricated Component Management System in Prefabricated Housing Production. 591-601. https://ascelibrary.org/doi/10.1061/9780784481264.058

San, S. D., & Sothy, S. (2016). Analyzing Delays of Road Construction Projects in Cambodia: Causes and Effects. *Journal of Management in Engineering*, *32*(6), 05016020. https://doi.org/10.1061/(ASCE)ME.1943-5479.0000467

S. Nishigaki, K. Saibara, T. Ootsuki, H. Morikawa, Scheduling simulator by ensemble forecasting of construction duration, in: 2020 Proceedings of the 37th ISARC, 2020, https://doi.org/10.22260/ISARC2020/0062.

Velleman, P. F., & Hoaglin, D. C. (2012). Exploratory data analysis. In *APA handbook of research methods in psychology, Vol 3: Data analysis and research publication.* (pp. 51–70). American Psychological Association. https://doi.org/10.1037/13621-003