PLANNING MALAYSIA:
Journal of the Malaysian Institute of Planners
VOLUME 21 ISSUE 5 (2023), Page 110 – 125

SELECTING A STANDARD SET OF ATTRIBUTES FOR THE DEVELOPMENT OF MACHINE LEARNING MODELS OF BUILDING PROJECT COST ESTIMATION

Hafez Salleh[1], Rui Wang[2], Nur Zahirah Haji Affandi[3], Zulkiflee Abdul-Samad[4]

[1,2,3,4] *Centre of Building Construction and Tropical Architecture (BuCTA), Faculty of Built Environment,*
UNIVERSITI MALAYA

Abstract

Accurate cost estimation is a critical aspect of successful construction projects, and the application of machine learning offers promising advancements in this domain. However, to achieve reliable cost predictions, the selection of a standardized set of attributes that significantly influence model performance is essential. This research addresses the research gap by investigating the systematic clarification of a standard set of attributes for machine learning models in building cost estimation. Firstly, plenty of attributes were summarized by literature review, then by questionnaire surveying and focus group discussion of the Delphi study period, the final 68 ranked attributes were determined and formulated the attribute set of building data. The findings of this research are beneficial to improve the accuracy of estimation by providing the essence of developing a building cost estimation of machine learning because the domain researcher can refer to these listed attributes to determine the lay structure of a new model.

***Keywords**:* Standardized set of attributes, cost estimation model, machine learning

[4] Senior Lecturer at Universiti Malaya. Email: zulkiflee1969@um.edu.my

INTRODUCTION

In the construction industry, accurate building cost estimation is essential for project success, budget planning, and resource allocation (Car-Puši & Mladen, 2020; Elmousalami, 2020). Traditional methods often rely on expert judgment and historical data, leading to time-consuming and biased estimates (Hashemi et al., 2020). The advent of machine learning offers a promising alternative, enabling data-driven approaches to improve accuracy and efficiency (Abed et al., 2022; Hashemi et al., 2020). However, there is a lack of sufficient research focusing on the selection of a standardized attribute set for machine learning models in building cost estimation (Elmousalami, 2020; Pike & Grosse, 2018). This research aims to address the gap by clarifying a standard set of attributes for the development of machine learning models in building cost estimation.

To effectively solve the above problem, this research summarized the long list of attributes of building data by literature review and used the Delphi method including questionnaire surveying and focus group discussion to rank and screen key attributes for cost estimation and further formulate a standard set of attributes for building a cost estimation model. This research holds significant implications for the construction industry and the field of machine learning applications. The establishment of a standardized attribute set will enhance transparency and comparability in cost estimation practices, empowering scholars to make informed references.

LITERATURE REVIEW

The theoretical foundation for developing a cost estimation model by machine learning

With the development of Artificial Intelligence (AI) technology, more and more innovative machine learning models were developed to improve the accuracy of building cost estimation (Elmousalami, 2020). Developing a prediction model is a common process of data mining, which involves using statistical and machine learning techniques to analyze and extract useful patterns and relationships from large datasets (Lu & Zhang, 2022), so it is essential to clarify the main procedures of data mining before developing a machine learning model. In general, the Cross-Industry Standard Process for Data Mining (CRISP-DM) model can provide effective guidance for developing data mining techniques, and it provides a systematic and comprehensive approach to developing machine learning models, making it an ideal choice for building cost estimation models that are reliable, relevant, and aligned with business objectives (Schröer et al., 2021).

As Figure 1, the CRISP-DM model is composed of six steps for data mining: business understanding, data understanding, data preparation, modelling, evaluation and development (Schröer et al., 2021). It is obvious that business understanding is the key step and also the basis of subsequent work. The critical task of business understanding for developing a cost estimation model is to identify the attributes of building data from big data, precise and vital attribute set is beneficial to construct the structure of the cost estimation model and guide the limitation of data collection and data cleaning (Elmousalami, 2020).

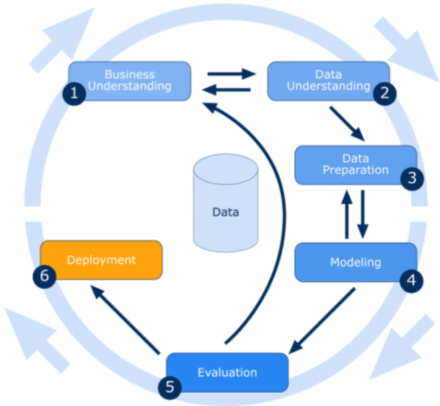

Figure 1: The Cross-Industry Standard Process for Data Mining Model (CRISP-DM)
Source: (Schröer et al., 2021; Wirth & Hipp, 2000)

Attributes of building data for cost estimation
Attributes in building cost estimation refer to the specific characteristics or factors that are considered in the estimation process to determine the cost of constructing a building or structure (Elhag & Boussabaine, 1998; Elmousalami, 2020). These attributes provide a basis for quantifying and evaluating the various elements that contribute to the overall cost.

In the context of building a cost estimation model, attributes refer to the variables or features used as input to the model to predict or estimate costs (Elmousalami, 2020). These attributes capture relevant information about the projects, activities, or resources that influence the cost. The quality and relevance of the attributes significantly impact the accuracy and effectiveness of the cost estimation model (Pike & Grosse, 2018).

The following table shows the collected attributes derived for building Cost estimation from various literature reviews. It also shows the cost-estimating techniques that were adopted for Estimating the cost of the building. There are 11 categories to group the attributes and the categories are Project Strategic, Parties-involved, Site-related, Mechanical and Electrical, Building-component, Design-related, Material-related, Area-related, Ratio-related, Households-related and External Influences.

Table 1: Categorization of attributes of building datasets

No.	Attributes	Citations
Project Strategic Variables (V1)		
1	Bidding environment	(Chan & Park, 2005)
2	Duration	(Jumas et al., 2018), (Ahiaga-Dagbui & Smith, 2014), (Bala et al., 2014), (S. Kim & Shim, 2014), (Elfaki et al., 2014), (Emsley et al., 2002), (G.-H. Kim et al., 2004), (Elhag & Boussabaine, 1998)
3	Estimating Method	(Alshemosi & Alsaad, 2017), (Al-Khaldi, 1990)
4	Importance for the project to be completed within budget	(Chan & Park, 2005)
5	Procurement Strategy	(Emsley et al., 2002)
6	Project Type	(Ahiaga-Dagbui & Smith, 2014), (Elhag & Boussabaine, 1998)
7	Quality of Building	(Emsley et al., 2002)
8	Quality of Project Information	(Riquelme & Serpell, 2013)
9	Tendering Strategy	(Ahiaga-Dagbui & Smith, 2014), (Elfaki et al., 2014), (Emsley et al., 2002)
10	Contract Form/ Type of Contract	(Chan & Park, 2005), (Emsley et al., 2002), (Elhag & Boussabaine, 1998)
11	Purpose/ Type of use	(Jumas et al., 2018), (El-Sawalhi & Shehatto, 2014), (S. Kim & Shim, 2014), (Chan & Park, 2005), (Emsley et al., 2002)
12	Year (Built year)	(S. Kim & Shim, 2014), (G.-H. Kim et al., 2004)
Parties-involved Variables (V2)		
2.1 Consultant		
13	Experience with similar projects	(Chan & Park, 2005)
14	Level of construction sophistication	(Chan & Park, 2005)
15	staffing level to attend to the contractor	(Chan & Park, 2005)
16	No. of DBB/DB projects handled by the consultant in the past	(Chan & Park, 2005)
2.2 Contractor		
17	Financial Management ability	(Chan & Park, 2005)
18	Design capability	(Chan & Park, 2005)
19	Experience with similar size of projects	(Chan & Park, 2005)
20	Experience with similar types of projects	(Chan & Park, 2005)
21	Health and safety management capability	(Chan & Park, 2005)
22	Key personnel's management ability	(Chan & Park, 2005)
23	Prior working relationship with consultants and owner	(Chan & Park, 2005)
24	Quality control and management capability	(Chan & Park, 2005)
25	Staffing level	(Chan & Park, 2005)
26	Technical expertise	(Chan & Park, 2005)
27	Track record for completion on budget, time and acceptable quality	(Chan & Park, 2005)
28	Level of Technologically Advancement	(Chan & Park, 2005)
29	Size of the contractor by paid-up capital (US$)	(Chan & Park, 2005)
30	The magnitude of claims and disputes in contractor's past projects	(Chan & Park, 2005)
31	Adequacy of contractor's plant and equipment	(Chan & Park, 2005)
2.3 Client related		
32	No. of DBB/DB projects handled by owner in the past	(Chan & Park, 2005)
33	Owner's experience with similar projects	(Chan & Park, 2005)

No.	Attributes	Citations
34	Owner's level of construction sophistication	(Chan & Park, 2005)
35	Owner's staffing level to attend to the contractor	(Chan & Park, 2005)
Site-related variables (V3)		
36	Geology property (Soft, Medium, Hard)	(M.-Y. Cheng & Wu, 2005)
37	Location (Including location index)	(Ahiaga-Dagbui & Smith, 2014), (Alshemosi & Alsaad, 2017), (S. Kim & Shim, 2014), (Al-Khaldi, 1990), (An et al., 2007)
38	Location of the core (e.g., central, peripheral)	(Doğan et al., 2006), (Doğan et al., 2008)
39	Seismic Zone	(J. C. P. Cheng et al., 2010)
40	Site Access	(Ahiaga-Dagbui & Smith, 2014), (Bhokha & Ogunlana, 1999), (Emsley et al., 2002), (Elhag & Boussabaine, 1998)
41	Site Condition (Including ground condition)	(Alshemosi & Alsaad, 2017), (J. C. P. Cheng et al., 2010), (Al-Khaldi, 1990), (Riquelme & Serpell, 2013), (Elhag & Boussabaine, 1998)
42	Soil Type	(Ahiaga-Dagbui & Smith, 2014)
43	Topography	(Emsley et al., 2002)
44	Type of Location	(Emsley et al., 2002)
45	Type of Site	(Emsley et al., 2002)
Mechanical and Electrical related variables (V4)		
46	Air conditioning system	(Emsley et al., 2002)
47	Electrical buried pipe	(Jiang, 2019)
48	Electrical installations	(Emsley et al., 2002)
49	Electro-mechanical infrastructure	(J. C. P. Cheng et al., 2010)
50	Mechanical Installations	(Emsley et al., 2002)
51	**No. of elevators	(Ji et al., 2011), (Ahn et al., 2014), (El-Sawalhi & Shehatto, 2014), (Ji et al., 2019), (Emsley et al., 2002)
52	Protective Installation (fire protection)	(Emsley et al., 2002)
53	Special Installations	(Emsley et al., 2002)
54	Type of Mechanical works	(El-Sawalhi & Shehatto, 2014)
55	Type of electricity works	(El-Sawalhi & Shehatto, 2014)
Building-component variables (V5)		
5.1 Structural		
56	Type of foundation	(Jumas et al., 2018), (Arafa & Alqedra, 2011), (El-Sawalhi & Shehatto, 2014), (Doğan et al., 2006), (Hong et al., 2011), (Doğan et al., 2008), (Latief et al., 2013), (An et al., 2007), (Bhokha & Ogunlana, 1999) , (Feng & Li, 2013), (Ahn et al., 2014)
57	Building envelope	(Emsley et al., 2002)
58	Structural units	(Emsley et al., 2002)
59	Structure form	(Feng & Li, 2013)
60	Structure type	(Ji et al., 2011), (Hong et al., 2011)
61	Substructure	(S. Kim & Shim, 2014) , (Emsley et al., 2002)
62	Superstructure	(S. Kim & Shim, 2014)
63	Retaining Wall	(S. Kim & Shim, 2014)
64	Type of Slab	(El-Sawalhi & Shehatto, 2014)
65	Usage of basement	(An et al., 2007), (G.-H. Kim et al., 2004)
5.2 Architectural		
66	Windows and doors	(Feng & Li, 2013), (Emsley et al., 2002)
67	Wall (Internal and External)	(S. Kim & Shim, 2014), (Emsley et al., 2002)
68	Ceiling	(S. Kim & Shim, 2014)
69	Floor Type	(Doğan et al., 2006), (Doğan et al., 2008)

No.	Attributes	Citations
70	Roof (construction, & profile)	(Emsley et al., 2002)
71	Type of roof	(Jumas et al., 2018) , (Ji et al., 2011), (Ahn et al., 2014), (S. Kim & Shim, 2014), (An et al., 2007), (G.-H. Kim et al., 2004)
72	Type of Tiling	(El-Sawalhi & Shehatto, 2014)
5.3 Finishes		
73	Ceiling Finishes	(Emsley et al., 2002)
74	Floor Finishes	(Emsley et al., 2002)
75	Wall Finishes	(Emsley et al., 2002)
76	Roof Finishes	(Emsley et al., 2002)
Design-related variable (V6)		
77	Building height	(Jumas et al., 2018) , (Alshemosi & Alsaad, 2017) , (Bala et al., 2014), (Jin et al., 2012), (Bhokha & Ogunlana, 1999), (Emsley et al., 2002)
78	Level of Design Complexity	(Chan & Park, 2005)
79	No. of buildings	(Hong et al., 2011)
80	No. of floors	(Ji et al., 2011), (Ahn et al., 2014), (Arafa & Alqedra, 2011), (Bala et al., 2014), (S. Kim & Shim, 2014), (Jin et al., 2012), (Ji et al., 2019), (J. C. P. Cheng et al., 2010), (Doğan et al., 2006), (Doğan et al., 2008), (Feng & Li, 2013), (Sonmez, 2004)
81	No. of units	(Jiang, 2019), (An et al., 2007), (G.-H. Kim et al., 2004)
82	No. of similarly constructed buildings	(Hong et al., 2011)
83	Shape Complexity	(Emsley et al., 2002)
84	Type of Ground Plan (e.g., open space/ compartmentalised)	(Hong et al., 2011)
Material-related variables (V7)		
85	Concrete	(Jiang, 2019)
86	Masonry	(Jiang, 2019)
87	Steel bar	(Jiang, 2019)
Area-related variables (V8)		
88	External Wall area	(Jumas et al., 2018), (Bala et al., 2014)
89	Area per unit	(Latief et al., 2013), (Sonmez, 2004), (Ji et al., 2019), (An et al., 2007)
90	Building Area	(Amin, 2017), (Sonmez, 2004), (Shin, 2015)
91	Compactness (external wall area/ gross external floor area)	(Jumas et al., 2018), (Bala et al., 2014)
92	Gross External Floor Area	(Bala et al., 2014)
93	Gross Floor Area	(Jumas et al., 2018), (Ji et al., 2011), (Ahn et al., 2014), (Latief et al., 2013), (An et al., 2007), (Hong et al., 2011), (Shin, 2015), (G.-H. Kim et al., 2004), (Elhag & Boussabaine, 1998)
94	Functional Area	(Bhokha & Ogunlana, 1999)
95	The gross floor area of the subsidiary facilities	(Hong et al., 2011)
96	Ground Area	(Jin et al., 2012)
97	Ground Floor Area	(Arafa & Alqedra, 2011)
98	Land Area	(Amin, 2017)
99	Landscape Area	(Jin et al., 2012), (Hong et al., 2011)
100	Site area	(Jin et al., 2012), (J. C. P. Cheng et al., 2010), (Hong et al., 2011)
101	Structural Parking Area	(Arafa & Alqedra, 2011), (Sonmez, 2004)
102	Total area	(Alshemosi & Alsaad, 2017), (S. Kim & Shim, 2014), (Doğan et al., 2006), (Doğan et al., 2008)
103	Typical Floor Area	(Arafa & Alqedra, 2011), (El-Sawalhi & Shehatto, 2014)

No.	Attributes	Citations
104	*Underground area	(Jin et al., 2012), (Hong et al., 2011)
105	Lot area	(Ahn et al., 2014)
Ratio-related variables (V9)		
106	*Floor Area ratio	(S. Kim & Shim, 2014), (Jin et al., 2012)
107	*Building Coverage ratio	(S. Kim & Shim, 2014), (Jin et al., 2012)
108	Building ratio	(Hong et al., 2011)
109	Building to-plan ratio	(Hong et al., 2011)
110	Number of Units per Number of Storeys Ratio	(Latief et al., 2013)
111	The ratio of floor area to total area	(Doğan et al., 2006), (Doğan et al., 2008)
112	The ratio of the footprint area to the total area	(Doğan et al., 2006), (Doğan et al., 2008)
113	The ratio of typical floor area to GFA	(Jumas et al., 2018)
114	Wall-to-floor ratio	(Emsley et al., 2002)
Households-related variables (V10)		
115	No. of households	(Ji et al., 2011), (Hong et al., 2011), (Ahn et al., 2014), (Arafa & Alqedra, 2011), (J. C. P. Cheng et al., 2010)
116	No. of households per piloti	(Ji et al., 2011), (Ahn et al., 2014)
117	No. of households per unit floor	(Ji et al., 2011), (Ahn et al., 2014)
118	No. of households per building	(Ji et al., 2011), (Ahn et al., 2014)
119	Type of household	(Hong et al., 2011)
External Influences variables (V11)		
120	Earthquake impact (Low, High)	(M.-Y. Cheng & Wu, 2005)
121	Economic Instability	(Alshemosi & Alsaad, 2017), (Al-Khaldi, 1990)
122	Weather Conditions	(Riquelme & Serpell, 2013)
123	Market Status	(Alshemosi & Alsaad, 2017), (Elhag & Boussabaine, 1998)

RESEARCH METHODOLOGY

The application of the Delphi study in this research is to poll a group of experts to reach a group consensus regarding the attributes of Big Data Analytics in building cost estimation. The Delphi study was conducted in June 2023 in the Faculty of Built Environment, University of Malaya and mainly includes two rounds: ranking the different attributes from the literature review by questionnaire; validating the result of the attribute set of the building by focus group discussion. The 14 experts of the Delphi study are shown in Table 2.

Table 2: Expert panel list of the Delphi method

Experts	Age	Gender	Position	Working Experience
A1	33	Male	Construction data technical staff	7 years (Enterprise)
A2	41	Female	Construction data technical staff	16 years (Enterprise)
A3	32	Male	Construction data technical staff	5 years (Enterprise)
A4	38	Male	Academia in quantity surveying	8 years (Institute)
A5	31	Female	Academia in quantity surveying	7 years (Institute)
A6	42	Male	Academia in quantity surveying	8 years (Institute)
A7	36	Female	Academia in quantity surveying	5 years (Institute)
A8	35	Female	Academia in quantity surveying	4 years (Institute)
A9	44	Male	Manager of Building Cost Services	8 years (Enterprise)

Experts	Age	Gender	Position	Working Experience
A10	51	Male	Manager of Building Cost Services	12 years (Enterprise)
A11	38	Female	Manager of Building Cost Services	7 years (Enterprise)
A12	36	Female	Cost engineer	5 years (Enterprise)
A13	47	Male	Cost engineer	12 years (Enterprise)
A14	44	Male	Cost engineer	8 years (Enterprise)

Round 1: Ranking the different attributes

The first round of the Delphi Study is conducted using Questionnaire Survey, selected panel experts will be asked to evaluate the attributes compiled from the literature review according to the suitability of the attributes to be used for building cost estimation. The Likert scale in the first round of the Delphi Study ranges from Unsuitable to Highly Suitable as shown in Table 3. Subsequently, the data from the questionnaire will be regularised and averages calculated to determine the suitability of the attributes for use in construction cost estimation.

Table 3: Likert Scale used in the First Round of the Delphi Study

Score	1	2	3	4	5
Measure	Not Suitable	Less Suitable	Moderately Suitable	Fairly Suitable	Highly Suitable

Round 2: Validating the result of the attribute set of the building

In round 2 of the Delphi Study Method, the validation of the attributes is made through focus group sessions. A focus group is also known as a group interview which is moderated and the outcome of this interview will be studied. Participants commented on the results of the attribute set of building obtained in the previous round based on their own research and work experience and ultimately voted to approve or disapprove of the output finding after deliberation.

RESULT AND DISCUSSION

Ranked attributes according to the suitability

From the data collection and data analysis, the attributes have been ranked according to the panel experts' votes using the Likert scale. The score for each attribute is obtained by averaging the scores of the 14 experts. According to the ranked version of the attributes, this research concluded that the highest-ranking attributes are in the categories of Project Strategic, Design Strategic, Area Related and Ratio related. Whereas some of the lowest ranked attributes are in the categories of Household related and Parties Involved. Table 4 lists the attributes with mean points of more than 4.500 and subsequently voted as the most suitable sets of attributes.

Table 4: Attributes with the Highest Ranking

No.	Attributes	Categories	Mean
1	Duration	V1	**4.833**
2	Quality of Project Information	V1	
3	Design Complexity	V6	
4	Type of Ground Plan (e.g., open space/ compartmentalised)	V6	
5	Concrete	V7	
6	Gross Floor Area	V8	
7	Wall to Floor Ratio	V9	
8	Building to Plan Ratio	V9	
9	Total Area	V8	**4.667**
10	Typical Floor Area to GFA Area ratio	V9	
11	Footprint Area to Total Area ratio	V9	
12	Floor Area to Total Area ratio	V9	
13	Floor Area Ratio	V9	
14	Functional Area	V8	
15	Area Per Unit	V8	
16	Number of Similar Constructed Buildings	V6	
17	Estimating Method	V1	
18	Tendering Strategy	V1	**4.500**
19	Consultant Level of Construction Sophistication	V2	
20	Contractor Financial Management	V2	
21	Contractor Experience with similar project	V2	
22	Contractor Key's Personnel Management Ability	V2	
23	Site Condition (including ground condition)	V3	
24	Location (including location index)	V3	
25	Gross External Floor Area	V8	
26	Lot Area	V8	
27	Soil Type	V3	
28	Steel Bar	V7	
29	Market Status	V11	
30	Economic Instability	V11	
31	Weather condition	V11	
32	Location of the core (e.g., central, peripheral)	V3	
33	Ground Floor Area	V8	
34	Ground Area	V8	
35	Number of Units	V6	
36	Number of floors	V6	
37	Height	V6	
38	Roof Finishes	V5	
39	Wall Finishes	V5	
40	Floor Finishes	V5	
41	Ceiling Finishes	V5	

No.	Attributes	Categories	Mean
42	Type of Tiling	V5	
43	Floor Type	V5	
44	Ceiling Finishes	V5	
45	Walls Finishes	V5	

Therefore, the attributes listed in Table 4, should be taken into account when developing a building cost estimation model. However, the specific rank of each attribute may depend on the purpose and objective of the constructed building. Each building project has unique elements and characteristics to it. Meanwhile, the procedure of selecting the right attributes, whereby the scope of the project must be determined first.

Validated attribute set of building

Round 2 of the Delphi study first invited 14 experts to comment on the result of the ranking evaluation, together voting on whether the attribute list is sufficient and precise. According to the Pie Chart in Figure 2, 10 out of 14 agreed on the listed attributes and categorized attributes. However, 4 out of 10 suggest an improvisation to the listed attributes.

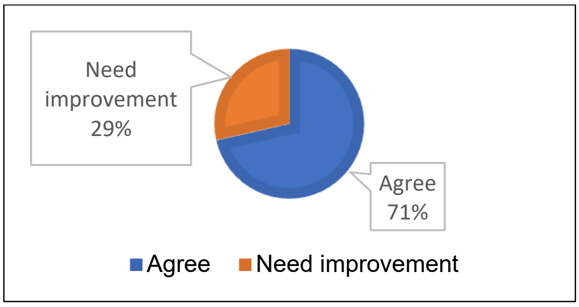

Figure 2: Pie Chart illustrates the Acceptability of the Attributes

Most experts keep the approval views on the listed attributes of the highest ranking (Table 4). On the other hand, experts A3, A6, A10 and A13 proposed that the listed attributes should be improved including revision and supplements.

Supplements

- ***Expert A3:*** *suggested including Data Volume as one of the attributes to be considered in this research where the data includes all the V5 of Big Data, which is the Volume, Variety, Velocity, Veracity and Value. Other than that,* *Expert A3 also suggested including the type of financing of the project such as government initiative or private funding. Besides that, Expert A6 also suggested Implications of Innovation and Technology such as BIM, Digital fabrication and automation, robotics, etc.*

- ***Expert A6:*** *suggested Contingency cost as one of the attributes to be included in the sets of attributes.*

Revision

- ***Expert A10:*** *suggested simplifying the attributes that are quite similar to each other. For Example, attributes like Level of Design Complexity and Shape Complexity can be integrated into a single attribute as a 'Design Complexity'. Another suggestion that the Panel Expert made is the integration between the Building Area to be Gross Floor Area, as the meaning of the two attributes is quite similar. Therefore, with this suggestion, we need to take into consideration any attributes that might have a similar meaning and can be integrated together as 1 attribute.*

- ***Expert A13:*** *suggested ranking the categories instead of each of the attributes. For example, if the Design-related attributes are mostly ranked at the top, then the categories of the attributes as a whole are put at the very top and thus accordingly. However, the attributes are not equally distributed, and this method may need another round of the Delphi method, which may take a longer time to reach out to each of the Panel Experts, therefore, this might be done in further research.*

Therefore, taking into account the second round of the Delphi study in coming up with the standard sets of attributes, whereby focus group discussion is conducted to validate the findings. Figure 3 finalises sets of attributes that can be utilised in developing a cost estimation model of machine learning.

To summarize the above findings, 68 finalised attributes have been formed as the standard sets of attributes for developing a building cost estimation model by machine learning algorithm. However, the listed attributes could also be revised based on the specific project's condition, relevant cost estimation researchers can refer to this attribute set to complete the step of the Business Understanding regarding the CRISP-DM model when establishing a machine learning model.

CONCLUSION

Current cost estimation techniques (e.g., traditional and probabilistic methods) can not satisfy the requirement of the construction industry due to the need for a more accurate result, more and more scholars gradually focus on the usage of machine learning techniques to develop innovative cost estimation models. Importantly the attribute set of building data is the basis of subsequent research regarding the CRISP-DM model, so this research aims to clarify the attribute set of building data by using Delphi methods with 2 rounds. By questionnaire surveying and focus group discussion of the Delphi study period, the final 68 ranked attributes were determined and formulated to the attribute set of building data. The findings of this research are beneficial to improve the accuracy of estimation by providing the essence of developing a building cost estimation of machine learning because the domain researcher can refer to these listed attributes to determine the layer structure of a new model.

Funding: Funded by the Ministry of Higher Education, Fundamental Research Grant Scheme (FRGS), Project No. FP087-2022.

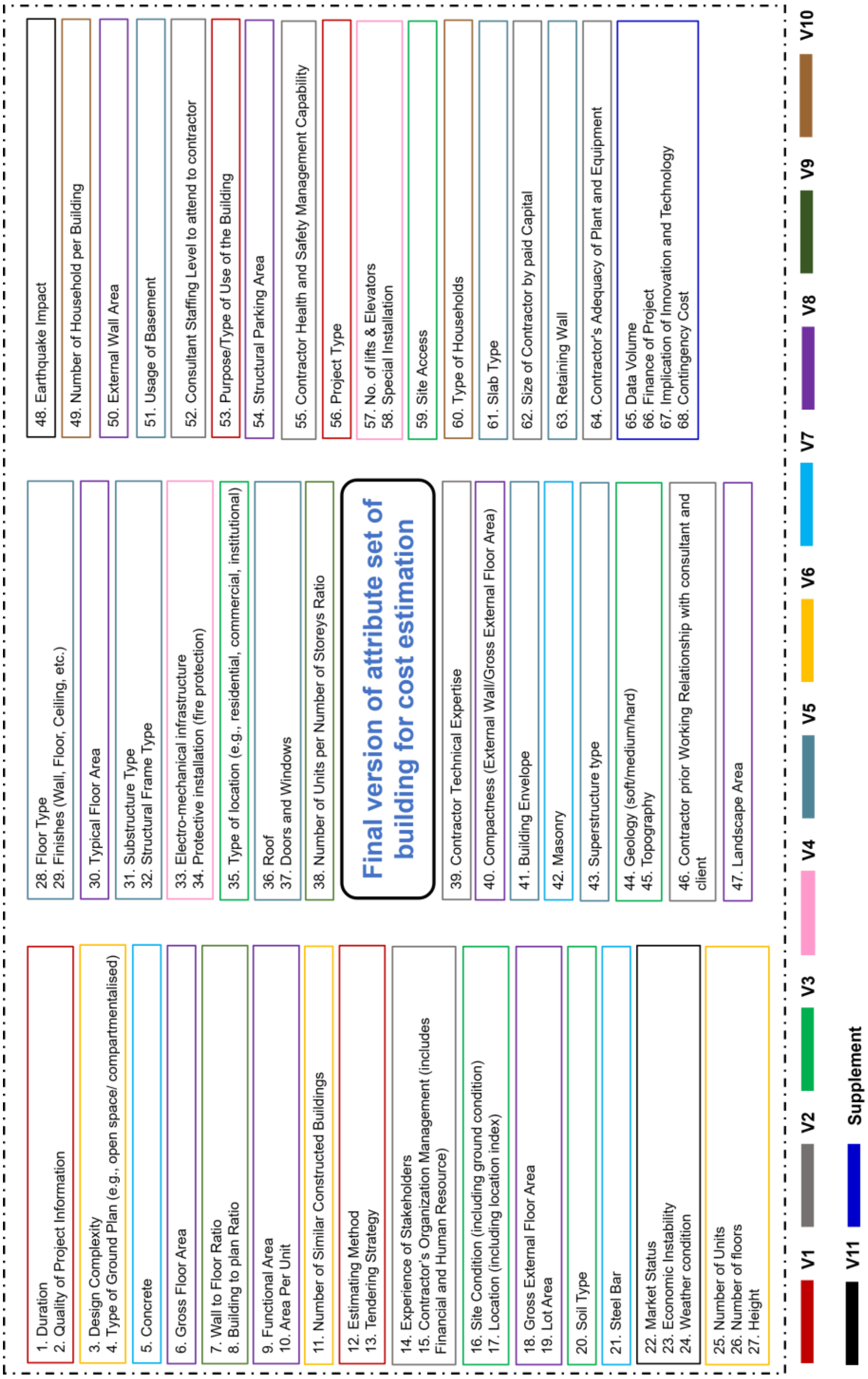

Figure 3: Final version of attribute set of building for cost estimation

REFERENCES

Abed, Y. G., Hasan, T. M., & Zehawi, R. N. (2022). Machine learning algorithms for constructions cost prediction: A systematic review. *INTERNATIONAL JOURNAL OF NONLINEAR ANALYSIS AND APPLICATIONS*, *13*(2), 2205–2218. https://doi.org/10.22075/ijnaa.2022.27673.3684

Ahiaga-Dagbui, D. D., & Smith, S. D. (2014). Rethinking construction cost overruns: Cognition, learning and estimation. *Journal of Financial Management of Property and Construction*.

Ahn, J., Ji, S.-H., Park, M., Lee, H.-S., Kim, S., & Suh, S.-W. (2014). The attribute impact concept: Applications in case-based reasoning and parametric cost estimation. *Automation in Construction*, *43*, 195–203.

Al-Khaldi, Z. S. (1990). *Factors affecting the accuracy of construction costs estimating in Saudi Arabia*. King Fahd University of Petroleum and Minerals (Saudi Arabia).

Alshemosi, A. M. B., & Alsaad, H. S. H. (2017). Cost estimation process for construction residential projects by using multifactor linear regression technique. *Criterion*, *71*, 7.

Amin, M. (2017). Development of cost estimation model for residential building. *Int J Civ Struct Eng Res*, *5*(1), 1–4.

An, S.-H., Kim, G.-H., & Kang, K.-I. (2007). A case-based reasoning cost estimating model using experience by analytic hierarchy process. *Building and Environment*, *42*(7), 2573–2579.

Arafa, M., & Alqedra, M. (2011). Early stage cost estimation of buildings construction projects using artificial neural networks. *Journal of Artificial Intelligence*, *4*(1), 63–75.

Bala, K., Ahmad Bustani, S., & Shehu Waziri, B. (2014). A computer-based cost prediction model for institutional building projects in Nigeria: An artificial neural network approach. *Journal of Engineering, Design and Technology*, *12*(4), 519–530. https://doi.org/10.1108/JEDT-06-2012-0026

Bhokha, S., & Ogunlana, S. O. (1999). Application of Artificial Neural Network for the forecast of building cost at the pre-design stage. *Seventh East-Asia Pacific Conference on Structural Engineering and Construction, Kochi, Japan*.

Car-Puši, D., & Mladen, M. (2020). *Early Stage Construction Cost Prediction in Function of Project Sustainability*. 631–638. Scopus. https://doi.org/10.23967/dbmc.2020.048

Chan, S. L., & Park, M. (2005). Project cost estimation using principal component regression. *Construction Management and Economics*, *23*(3), 295–304.

Cheng, J. C. P., Law, K. H., Zhang, Y., & Han, C. S. (2010). WEB-ENABLED MODEL-BASED CAD FOR THE ARCHITECTURE, ENGINEERING AND CONSTRUCTION INDUSTRY. In J. Teng (Ed.), *PROCEEDINGS OF THE FIRST INTERNATIONAL CONFERENCE ON SUSTAINABLE URBANIZATION (ICSU 2010)* (pp. 1199–1208). HONG KONG POLYTECHNIC UNIV, FAC CONSTRUCTION & ENVIRONMENT.

Cheng, M.-Y., & Wu, Y.-W. (2005). Construction conceptual cost estimates using support vector machine. *22nd International Symposium on Automation and Robotics in Construction ISARC*, 1–5.

Doğan, S. Z., Arditi, D., & Günaydın, H. M. (2006). Determining attribute weights in a CBR model for early cost prediction of structural systems. *Journal of Construction Engineering and Management*, *132*(10), 1092–1098.

Doğan, S. Z., Arditi, D., & Murat Günaydin, H. (2008). Using decision trees for determining attribute weights in a case-based model of early cost prediction. *Journal of Construction Engineering and Management*, *134*(2), 146–152.

Elfaki, A. O., Alatawi, S., & Abushandi, E. (2014). Using intelligent techniques in construction project cost estimation: 10-year survey. *Advances in Civil Engineering*, *2014*.

Elhag, T. M. S., & Boussabaine, A. H. (1998). An artificial neural system for cost estimation of construction projects. *14th Annual ARCOM Conference*, *1*, 219–226.

Elmousalami, H. H. (2020). Artificial intelligence and parametric construction cost estimate modeling: State-of-the-art review. *Journal of Construction Engineering and Management*, *146*(1), 03119008. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001678

El-Sawalhi, N. I., & Shehatto, O. (2014). A Neural Network Model for Building Construction Projects Cost Estimating. *Journal of Construction Engineering and Project Management*, *4*(4), 9–16. https://doi.org/10.6106/JCEPM.2014.4.4.009

Emsley, M. W., Lowe, D. J., Duff, A. R., Harding, A., & Hickson, A. (2002). Data modelling and the application of a neural network approach to the prediction of total construction costs. *Construction Management & Economics*, *20*(6), 465–472.

Feng, G. L., & Li, L. (2013). Application of Genetic Algorithm and Neural Network in Construction Cost Estimate. *Advanced Materials Research*, *756–759*, 3194–3198. https://doi.org/10.4028/www.scientific.net/AMR.756-759.3194

Hashemi, S., Ebadati E., O. M., & Kaur, H. (2020). Cost estimation and prediction in construction projects: A systematic review on machine learning techniques. *SN Applied Sciences*, *2*, 1–27. https://doi.org/10.1007/s42452-020-03497-1

Hong, T., Hyun, C., & Moon, H. (2011). CBR-based cost prediction model-II of the design phase for multi-family housing projects. *Expert Systems with Applications*, *38*(3), 2797–2808.

Ji, S.-H., Ahn, J., Lee, H.-S., & Han, K. (2019). Cost estimation model using modified parameters for construction projects. *Advances in Civil Engineering*, *2019*.

Ji, S.-H., Park, M., & Lee, H.-S. (2011). Cost estimation model for building projects using case-based reasoning. *Canadian Journal of Civil Engineering*, *38*(5), 570–581.

Jiang, Q. (2019). Estimation of construction project building cost by back-propagation neural network. *Journal of Engineering, Design and Technology*, *18*(3), 601–609. https://doi.org/10.1108/JEDT-08-2019-0195

Jin, R., Cho, K., Hyun, C., & Son, M. (2012). MRA-based revised CBR model for cost prediction in the early stage of construction projects. *Expert Systems with Applications*, *39*(5), 5214–5222. https://doi.org/10.1016/j.eswa.2011.11.018

Jumas, D., Mohd-Rahim, F. A., Zainon, N., & Utama, W. P. (2018). Improving accuracy of conceptual cost estimation using MRA and ANFIS in Indonesian building projects. *Built Environment Project and Asset Management*, *8*(4), 348–357.

Kim, G.-H., An, S.-H., & Kang, K.-I. (2004). Comparison of construction cost estimating models based on regression analysis, neural networks, and case-based reasoning. *Building and Environment*, *39*(10), 1235–1242.

Kim, S., & Shim, J. H. (2014). Combining case-based reasoning with genetic algorithm optimization for preliminary cost estimation in construction industry. *Canadian Journal of Civil Engineering*, *41*(1), 65–73. https://doi.org/10.1139/cjce-2013-0223

Latief, Y., Wibowo, A., & Isvara, W. (2013). Preliminary cost estimation using regression analysis incorporated with adaptive neuro fuzzy inference system. *International Journal of Technology*, *1*, 63–72.

Lu, Y., & Zhang, J. (2022). Bibliometric analysis and critical review of the research on big data in the construction industry. *ENGINEERING CONSTRUCTION AND ARCHITECTURAL MANAGEMENT*, *29*(9), 3574–3592. https://doi.org/10.1108/ECAM-01-2021-0005

Pike, J., & Grosse, S. D. (2018). Friction Cost Estimates of Productivity Costs in Cost-of-Illness Studies in Comparison with Human Capital Estimates: A Review. *Applied Health Economics and Health Policy*, *16*(6), 765–778. https://doi.org/10.1007/s40258-018-0416-4

Riquelme, P., & Serpell, A. (2013). Adding Qualitative Context Factors to Analogy Estimating of Construction Projects. *Procedia - Social and Behavioral Sciences*, *74*, 190–202. https://doi.org/10.1016/j.sbspro.2013.03.037

Schröer, C., Kruse, F., & Gómez, J. M. (2021). A systematic literature review on applying CRISP-DM process model. *Procedia Computer Science*, *181*, 526–534.

Shin, Y. (2015). Application of boosting regression trees to preliminary cost estimation in building construction projects. *Computational Intelligence and Neuroscience*, *2015*, 1–1. https://doi.org/10.1155/2015/149702

Sonmez, R. (2004). Conceptual cost estimation of building projects with regression analysis and neural networks. *Canadian Journal of Civil Engineering*, *31*(4), 677–683.

Wirth, R., & Hipp, J. (2000). CRISP-DM: Towards a standard process model for data mining. *Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining*, *1*, 29–39.

Received: 7th June 2023. Accepted: 5th September 2023