Article

Bim Machine Learning and Design Rules to Improve the Assembly Time in Steel Construction Projects

Mathieu Fokwa Soh, David Bigras, Daniel Barbeau, Sylvie Doré and Daniel Forgues *

Construction Department, École de Technologie Supérieure, Montreal, QC H3C 1K3, Canada;
mathieu.fokwa-soh.1@ens.etsmtl.ca (M.F.S.); davidbigras@davidbigras.ca (D.B.);
cc-daniel.barbeau@etsmtl.ca (D.B.); Sylvie.Dore@etsmtl.net (S.D.)
* Correspondence: Daniel.Forgues@etsmtl.net

Abstract: Integrating the knowledge and experience of fabrication during the design phase can help reduce the cost and duration of steel construction projects. Building Information Modeling (BIM) are technologies and processes that reduce the cost and duration of construction projects by integrating parametric digital models as support of information. These models can contain information about the performance of previous projects and allow a classification by linear regression of design criteria with a high impact on the duration of the fabrication. This paper proposes a quantitative approach that applies linear regressions on previous projects' BIM models to identify some design rules and production improvement points. A case study applied on 55,444 BIM models of steel joists validates this approach. This case study shows that the camber, the weight of the structure, and its reinforced elements greatly influence the fabrication time of the joists. The approach developed in this article is a practical case where machine learning and BIM models are used rather than interviews with professionals to identify knowledge related to a given steel structure fabrication system.

Keywords: assembly time; BIM models; construction industry; design rules; machine learning; Steel Joists

1. Introduction

The fabrication phase of structural steel projects represents 30 to 40% of the overall building cost [1]. Yet, the decisions taken during the design phase affect 88% of the steelworks' costs and time of execution [2]. However, the highest interest of the designers remains the compliance with standards [3] and the appropriate choice of structural elements for the resistance of the structures [4]. Not much time, a few minutes to a few hours are devoted to evaluating models for cost and time reduction and the search for alternative solutions during the design phase [5]. These evaluations are made without considering the particularities of the manufacturing plant where the work will be carried out [6]. This situation leads to a sub-optimal design [7–9]. In the traditional Design Bid Build (DBB) procurement, where a project is carried out in a linear and fragmented process, manufacturing specialists often intervene at the end of the design phase [10,11]. At this moment, the modifications they make cause delays and additional costs in completing the projects [12]. This situation is similar to the situation in Product Development Engineering (PDE) in the 1990s.

In PDE, designers and manufacturers collaborate formally through different methods and design rules, such as design for manufacturing and assembly (DFMA) [13]. These rules provide designers with the essential knowledge to reduce the cost, the time, the tools, the number of operations, the quantity of material, and the number of workers during projects while improving quality during the manufacturing and assembly of parts from the design phase [14], and for a specific workshop [15].

DFMA consists of identifying and considering manufacturing and assembly constraints during design. This process leads to design rules and tools, which help obtain

simplified and standardized products suitable for the manufacturing and assembly process [13,14]. The DFMA methodology also improves the manufacturing and assembly process by integrating structural changes that promote essential design criteria [16]. A case study from Douglas Commercial Airlines demonstrates the significant benefits of using DFMA in a manufacturing process. Notably: 51% reduction in the number of parts, 37% reduction in the cost of manufacturing parts, 50% faster time to market, 68% improvement in the quality and reliability of the final product, 62% reduction in assembly time, and 57% reduction in manufacturing time [17]

DFMA identifies design factors with a high impact on manufacturing and assembly processes [18]. One approach in identifying these factors is to hold meetings with designers, manufacturers, and assemblers with extensive knowledge and experience to assess the design factors available for a designed product [11,13]. The identified factors help to establish criteria that will allow the evaluation of different product designs [19]. This approach, which seems to be possible to implement in the PDE, is difficult to apply in the construction industry because of the context of Design Bid Build (DBB), where there is real fragmentation between project phases [6]. However, recent work in machine learning (ML) shows that it is possible to extract knowledge from the digital data of a process. Therefore, Building Information Modeling (BIM) offers relevant data for ML in the construction industry.

This paper addresses the research question: Is it possible to identify design rules such as DFMA from BIM models of previous projects and machine learning algorithms? As an answer to this question, this paper aims to propose an approach to identify design rules from BIM models of previous projects and ML algorithms:

To achieve this goal, this paper proposes to validate the possibility of extracting design factors from BIM models of steel structures and ML, the possibility of establishing some design rules to reduce the fabrication time from the obtained design factors.

For that, this article suggests a literature review to justify the choice of methodology, a methodology, and a case study with 55,444 BIM models of steel joists. The BIM models are from a major North American steel structure manufacturer.

2. Literature Review

2.1. Choice of a Knowledge Extraction Method

The extraction of knowledge specific to construction processes is one of the main motivations of industrial and scientific organizations related to the construction industry. Among these organizations, the Construction Industry Institute (CII) [20] and the Independent Project Analysis (IPA) [21] consider that the extraction of knowledge from processes is essential for verifying constructability and seeking efficiency during projects.

Two classes of methods can be used to extract knowledge from an industrial process: qualitative and quantitative.

The qualitative method analyzes speech and texts from experts, from their experiences, from conferences or brainstorms [22]. The quantitative method provides knowledge based on the statistical and historical characteristics of the available data. This method mainly uses mathematical models in scientific logic to propose a probabilistic form, which may occur in a given process with identified input data [23].

In the context of large amounts of data, one of the significant limitations of using qualitative methods in knowledge extraction is the limit of analysis of the human brain [23]. Data has grown exponentially since the advent of BIM and machine controllers [1]. The traditional nature of contracts in the construction industry forces a separation between the design and construction phases [24]. This separation is accentuated by the increasing complexity of projects and the growing level of client requirements, which requires specialization of activities in the construction industry [25]. Another barrier to using the qualitative method is the difference in academic training between design and construction professionals. Indeed, the organization of brainstorming between professionals in these two professions can lead to costly and unproductive discussions caused by the difference in perception [26,27].

Given these difficulties, this article proposes using a quantitative method based on machine learning (ML) to bypass human limitations related to knowledge extraction.

2.2. Machine Learning to Identify and Extract Knowledge

The construction of steel structures requires complex fabrication and assembly operations [1]. Complex fabrication systems require sophisticated and accurate prediction systems, such as those proposed by machine learning (ML) [28]. ML is the science of giving computers the ability to learn and act as humans do and improve their learning over time autonomously by providing data and information from a real process. ML is also defined as a process that extracts models automatically from historical data [29]. ML belongs to the domain of Artificial Intelligence (AI). AI is a field of research that aims to reproduce, through artificial systems, the different cognitive capacities of human beings [30]. One of the most targeted objectives of AI use is its ability to solve complex problems that are beyond human competence [28,30–32], as well as to develop programs capable of learning from data [33]. Numerous applications of ML are found in finance, insurance, and medicine [28,34]. ML is also found in manufacturing production management [35], and in construction [36–38]. The observed benefits of ML in construction are widely appreciated in the industry [28,39]

2.3. Choice of the Type of Learning and the Type of Algorithm

ML's two main approaches: supervised learning and unsupervised learning [31]. Supervised learning applies to processes with known output data. The objective here is to understand the relationship between input and output data. Unsupervised learning is applied to a series of data that is not understood. The goal is to find a natural link between these data.

The study proposed in this article is a simple application case of ML. Supervised learning with a regression algorithm is suitable for this study [31,40].

Regression algorithms consist of building a prediction model and training it with available data to respond accurately to new data belonging to the process to be studied [29]. Several regression-based operating time prediction cases exist in the literature [28]. In each of these cases, a comparison between algorithms identifies the algorithm that best fits the study [28].

2.4. Ensemble Learning

In the field of prediction with regression algorithms, there is increasing interest in Ensemble Learning (EL), a method that combines predictions from several algorithms and aggregates their results to obtain a higher accuracy than any individual algorithm [41–44]. The two main methods used in EL are boosting and bagging [43].

In boosting, successive prediction trees make incremental contributions to improve the predictions of previous trees. In the end, a weighted vote is taken for the final prediction. One of the techniques used for boosting is the Gradient Boosting Regressor GBR.

In Bagging, prediction trees do not depend on previous prediction trees. All trees are individually constructed. In the end, a simple majority vote is taken for the prediction [41]. The Random Forest Regressor (RFR) is one of the techniques used in Bagging. The GBR and RFR are both used and compared in this article for knowledge extraction.

To better appreciate the performance of GBR and RFR techniques in knowledge extraction, this study proposes to use another linear regression technique: Lasso (least absolute shrinkage and selection operator).

Lasso is a regression method that performs variable selection and regulation to improve the accuracy of predictions and the interpretation of the statistical model it produces [45]. The method was popularized by Robert Tibshirani in 1996 and is widely used today in ML, in the specific case of linear regressions. The objective of the method is to minimize the prediction error. For that, the Lasso method imposes a constraint on the sum of the absolute values of the model parameters. The sum must be less than a fixed value.

The imposition of constraints is done by applying a process of narrowing (regulation), or it penalizes the regression variable coefficients by reducing certain to zero [46].

The use of Lasso has many advantages, including the accuracy of predictions by reducing the coefficients. This is particularly useful when the number of observations is small, and the number of characteristics is large. Lasso also allows researchers to increase the interpretation of the models by removing irrelevant variables [46].

2.5. Evaluation of Prediction Quality

It is essential to evaluate the quality of the prediction results. Lantz (2015) suggests the use of Mean Absolutes errors (MAEs), and Relative Absolute Errors (RAEs).

MAEs consider how far, on average, the prediction is from the real value [31]

$$MAE = \frac{1}{N} \sum_{i=1}^{n} |x_i - y_i| \tag{1}$$

RAE is a performance metric that compares the actual forecast error to a very simple forecasting model [47].

$$RAE = \frac{\sum_{i=1}^{n} |x_i - y_i|}{\sum_{i=1}^{n} |y_i - \overline{y}|} \tag{2}$$

We add the Gap between predicted and real-time (GBP) to these measurements. This measure represents the percentage difference between predicted and actual values.

$$GBP = \left| \frac{y_i - x_i}{y_i} \times 100 \right| \tag{3}$$

where:

x_i is the predicted value for the individual sample i,

y_i is the real value for the individual sample i,

$\overline{x}$ is a mean value of x, with $\overline{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$

$\overline{y}$ is a mean value of y, with $\overline{y} = \frac{1}{n} \sum_{i=1}^{n} y_i$

n is the sample size.

As MAE and GBP get closer to 0 and RAE gets closer to 1, the quality of the prediction improves.

2.6. BIM for Data Extraction

The success of a prediction depends on the quality of the data used [40]. It is essential to pay special attention to the data quality coming from a process. BIM offers dedicated technologies and processes for better information management in a construction project [24].

BIM technology offers high-quality data through the BIM models [24], and makes it possible to gather and classify information [48,49] specific to different disciplines in a single 3D model. In steel construction, the information extracted from these models can be used for constructability and quantitative estimation [48,50]. This information can also be used for cost and time estimation [51]. These data can be extracted automatically and reduce the time extraction [52,53]. The method proposed by this article will use the BIM models as a data source.

2.7. BIM Is an Asset for the Success of DFMA in the Construction Industry

Three main characteristics of DFMA are the component-based approach, modularization, and standardization [54]. BIM can be used as an object-oriented collaborative process to integrate information representing the fabrication and assembly phases of steel structures. BIM application in the DFMA approach has allowed professionals to simulate construction virtually to identify potential constraints that could increase project costs [54]. We believe that BIM can bring to the steel construction industry the benefits that Computer-

Aided Design (CAD) has brought to DFMA. These include a more systematic analysis of fabrication and assembly options to produce a structural design that is more suited to the available processes [55] and fabrication process information to allow for multiple fabrication and assembly simulations [56]. BIM is used in the construction industry as a tool and process to improve the way buildings are designed and constructed [24].

3. Methodology

According to the Cross-Industry Standard Process for Data Mining (CRISP-DM), the main steps in data prediction are system understanding, data understanding, data preparation, data modeling, and outcome evaluation [57]. This paper adds pattern identification and design rules to these steps (Figure 1).
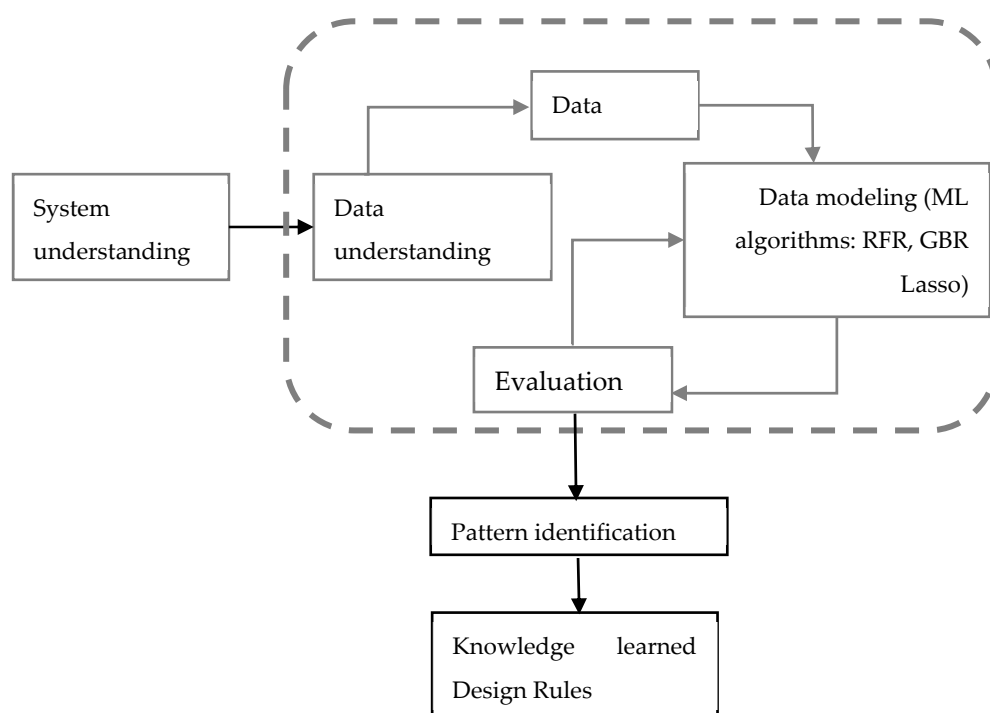
Figure 1. Illustration of the proposed method.

The proposed research approach could be described as follows:

3.1. System Understanding

The objective of system understanding is to understand the goals of the prediction and the requirements necessary to achieve these objectives. In construction projects, cost and schedule are generally the performance criteria sought to realize the project. They can be defined as prediction objectives. Costs and schedules depend on the geometrical and functional decisions made during the design phase. These decisions will be identified to serve as prediction criteria during system understanding. During this step, technologies and tools are also identified for the prediction set.

3.2. Data Understanding

Data understanding is the next step after system understanding. This step consists in collecting and analyzing the data necessary for the prediction objectives. To do this, it is necessary to group and classify the data related to the prediction criteria sought with data analysis tools during the data understanding. The collection can be done with MS Excel or Google sheet spreadsheets. After data collection, the next step is to explore the data to ensure its quality. One method is to eliminate data that is either too large or too small to fit the criteria being analyzed.

3.3. Data Preparation

Data preparation consists of preparing the data for modeling. This is done in several steps, including:

- data cleaning, which consists of removing, correcting, or deleting erroneous values,
- data construction which consists of determining additional attributes that will be useful for data modeling, and
- data integration consists of combining data from various sources.

3.4. Data Modelling

Modeling consists of building and evaluating various models based on different modeling techniques. Here we select the algorithms to try; we assess the competing models based on the results obtained and the performance criteria sought.

The implementation of the approach proposed in this article requires the use of an interpretive programming language and a programming package for the ML. In this article, the tool proposed as an interpretive programming language is Python, and the package for ML is Tensor Flow. The prepared data will be used in modeling with the RFR, GBR, and Lasso techniques. The prediction results with these techniques are presented at the end of the modeling.

3.5. Data Evaluation

This step compares the results of the three techniques used for modeling. The algorithm with which MAE and GBP will be close to 0 and RAE get closer to 1 has the best performance.

3.6. Pattern Identification

This step consists of identifying the variables that have the most impact on steel structure fabrication and assembly time.

3.7. Knowledge Learned & Design Rules:

This part consists of understanding the results of the pattern identification, establishing design rules, and formulating recommendations to improve the fabrication and assembly line.

4. Case Study

The case study in this article concerns the assembly of steel joists for a major manufacturer of steel structures in North America

4.1. System Understanding

In the steel construction industry, steel joists are lightweight steel structures that support roofs and floors and transfer the loads they receive directly to the steel structures that support them.

Joists are mainly composed of top and bottom chords, webs, and seats. See Figure 2.
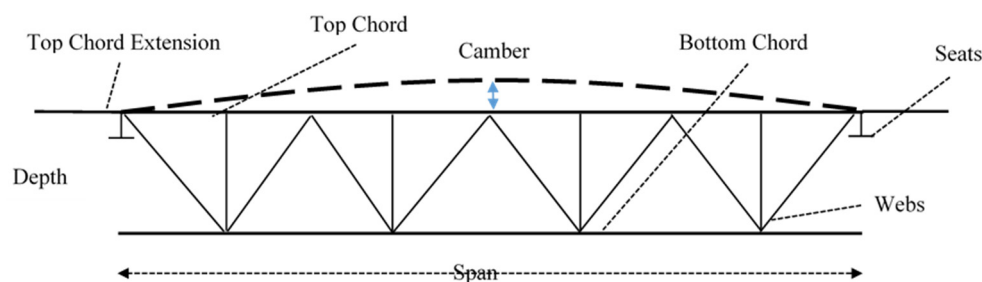

Figure 2. Part of a joist.

The joist assembly operation consists of:

- Reading and understanding the plans,
- Identification of the steel elements that will build the joist,
- Identification of the positions of the elements on the joist,
- Lifting and placing the top and bottom chords on the assembly table,
- Lifting and positioning of joist components,
- Welding of joist elements,
- Final check,
- Removing the assembled joist.

An automatic device is installed on the joist assembly table to measure the assembly time. Each time the beam elements arrive on the assembly table for the first time, the automatic device starts counting the assembly time. The counting will stop when the assembly table becomes empty again. This feature reduces human intervention in starting the countdown. However, the device does not stop when, leaving the joists on the assembly table, the workers go on a break or weekend.

4.2. Data Understanding

The data for this study come from the main variables that characterize the joists.

- **Depth** (mm) is the vertical distance between the axis of the top and bottom chords
- **Span** (mm) is the distance between the two positions of the joist seats
- **Camber** (mm) is the distance between the highest point before, and after bending; the camber is imposed on the upper chords to keep the joist horizontal when loaded
- **ComponentCount** (qty) is the number of elements
- **Memb_Lgth** (mm) is the total length of all its elements
- **weight** (Weight(kg)) is the total mass of the features that make up the joist
- **ChordsDis** (binary) checks if the top and bottom chords are different in profile types and dimensions
- **ReinfWeight** (mm) is the weight of the additional materials used to reinforce the joist elements
- **Tcs_Lgth_l** (mm) is the extension length to the left of the top chord.
- **Tcs_Lgth_r** (mm) is the extension length to the right of its top chord
- **Tcx_Depth_l** (mm) is the depth of the profile used for the left extension of the top chord
- **Tcx_Depth_r** (mm) is the depth of the profile used for the extension to the right of the top chord
- **TcxType_x** (from type1 to type 8) is the type of top chord used to manufacture the joist
- **Dsg_type** (from type 1 to type 4) is the type of joist design used

4.3. Data Preparation

Data cleaning: More than 170,000 assembly times were recorded, corresponding to more than 170,000 joists. However, these data contain noise. The noise is mainly due to inattention errors by workers, joists remaining on tables during breaks, weekends, and holidays. To reduce the noise in the data set, this study proposes to obtain the maximum and minimum time to assemble a joist on the assembly table. This technique consists of excluding the study, the times too short or too much to correspond to the time for assembling the beams. A maximum time of 25 min and a minimum time of 5 min is obtained from professionals. These times will be retained as the minimum and maximum limits of the assembly times retained for the study. This technique will reduce the data by 170,000 for 55,444 see Figure 3.
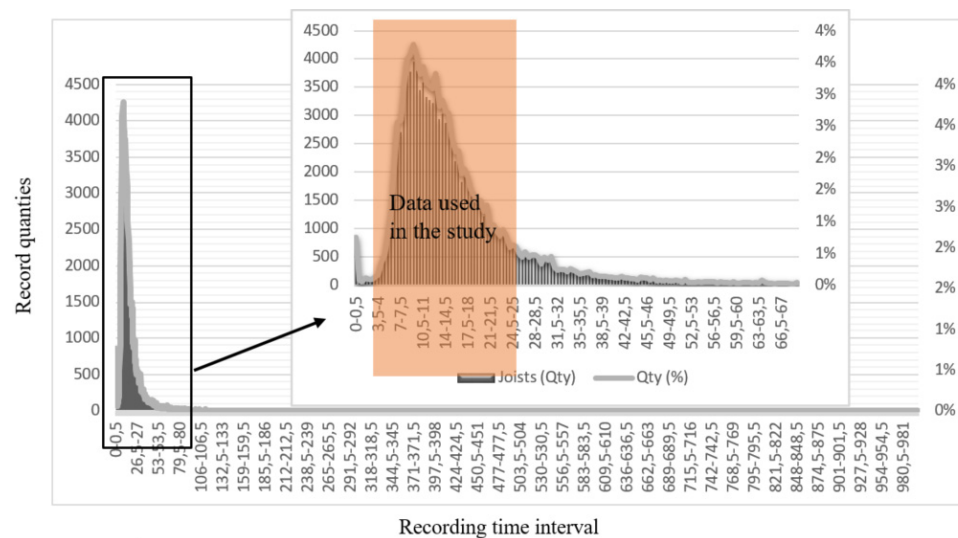
Figure 3. Data distribution for the study.

Data splitting for a better balance in prediction, the data were divided into four equivalent groups with similar statistical characteristics. Tables 1–4 present the organization of the configured groups with their respective features. The features were considered according to the prediction criteria Depth, Span, Camber1 ComponentCount Memb_Lgth Weight, and RealTime. These criteria were organized according to the mean, the minimum (min), the first quartile (25%), the median (50%), the third quartile (75%), the maximum (max), and the standard deviation (std), as seen in Tables 1–4. These measures ensure that the data is well distributed among the four groups. Three of these groups were used for training the learning algorithms and the fourth for testing them.

Table 1. Statistical characteristics of the first quarter of data.

	Depth	Span	Camber1	ComponentCount	Memb_Lgth	Weight	RealTime
count	13,861.0	13,861.0	13,861.0	13,861.0	13,861.0	13,861.0	13,861.0
mean	634.8	9084.4	16.2	22.0	53,595.9	152.3	13.1
std	145.8	2803.9	7.2	5.3	18,161.9	84.4	4.4
min	400.0	2120.0	0.0	8.0	12,045.5	25.8	5.0
25%	508.0	7080.0	11.5	17.0	40,344.6	92.7	9.7
50%	609.6	8737.6	15.9	23.0	51,156.2	129.9	12.4
75%	750.0	10,845.8	20.8	26.0	65,284.1	189.7	15.8
max	915.0	18,694.4	114.6	68.0	218,027.3	1063.8	25.0

Table 2. Statistical characteristics of the second quarter of data.

	Depth	Span	Camber1	ComponentCount	Memb_Lgth	Weight	RealTime
count	13,861.0	13,861.0	13,861.0	13,861.0	13,861.0	13,861.0	13,861.0
mean	634.8	9088.5	16.2	22.0	53,596.5	152.5	13.1
std	145.6	2763.4	7.0	5.2	17,795.0	82.0	4.4
min	355.6	1917.7	0.0	8.0	11,600.9	22.5	5.1
25%	508.0	7112.0	11.5	17.0	40,427.4	93.6	9.7
50%	609.6	8763.0	15.9	23.0	51,228.6	131.9	12.4
75%	750.0	10,850.0	20.7	26.0	65,255.6	190.8	16.0
max	915.0	18,694.4	87.3	68.0	218,027.3	1042.0	25.0

Table 3. Statistical characteristics of the third quarter of data.

	Depth	Span	Camber1	ComponentCount	Memb_Lgth	Weight	RealTime
count	13,861.0	13,861.0	13,861.0	13,861.0	13,861.0	13,861.0	13,861.0
mean	634.0	9089.9	16.1	22.0	53,613.5	153.0	13.1
std	146.8	2798.6	7.2	5.3	18,165.2	84.0	4.4
min	400.0	2104.0	0.0	8.0	12,638.5	23.7	5.0
25%	508.0	7061.2	11.5	17.0	40,132.6	93.0	9.7
50%	609.6	8755.0	16.0	23.0	51,191.5	130.8	12.3
75%	750.0	10,782.0	20.6	26.0	65,156.9	192.3	15.9
max	915.0	18,694.4	127.0	56.0	181,796.0	1137.5	25.0

Table 4. Statistical characteristics of the fourth quarter of data.

	Depth	Span	Camber1	ComponentCount	Memb_Lgth	Weight	RealTime
count	13,861.0	13,861.0	13,861.0	13,861.0	13,861.0	13,861.0	13,861.0
mean	634.3	9097.5	16.2	22.1	53,640.2	152.9	13.1
std	145.8	2788.6	7.1	5.3	17,969.3	83.6	4.4
min	400.0	2105.0	0.0	8.0	11,707.6	23.5	5.0
25%	508.0	7105.7	11.5	17.0	40,469.4	93.1	9.7
50%	609.6	8790.0	15.9	23.0	51,380.4	131.4	12.4
75%	750.0	10,750.0	20.5	26.0	64,843.2	190.9	15.9
max	915.0	18,440.4	90.0	56.0	181,796.0	786.0	25.0

4.4. Modeling

Three algorithms (GBR, RFR, and Lasso) are used to predict the manufacturing time. The results of this modeling are as illustrated in Tables 5–7.

Table 5. Prediction results with GRB.

Categories	Number of Items	Real-Time	Prediction	GBP	Rae	Mae
0–15	1502	16,153	16,073	0.5%	0.94	3.14
15–20	5672	61,873	62,109	−0.4%	0.92	2.89
20–25	11,129	127,439	127,217	0.2%	0.93	2.97
25–30	12,827	157,715	158,419	−0.4%	0.92	2.97
30–35	9650	129,735	129,484	0.2%	0.92	2.97
35–40	6212	91,654	91,743	−0.1%	0.93	2.85
40–45	4726	73,934	73,734	0.3%	0.93	2.71
45–50	2159	37,050	36,993	0.2%	0.94	2.57
50–55	1172	21,542	21,411	0.6%	0.92	2.44
55–60	322	6507	6433	1.1%	1.01	2.30

Table 6. Prediction results with RFR.

Categories	Number of Items	Real-Time	Prediction	GBP	Rae	Mae
0–15	1502	16,153	16,258	−0.7%	0.98	3.27
15–20	5672	61,873	62,136	−0.4%	0.94	2.96
20–25	11,129	127,439	127,970	−0.4%	0.95	3.06
25–30	12,827	157,715	158,221	−0.3%	0.94	3.02
30–35	9650	129,735	129,902	−0.1%	0.94	3.03
35–40	6212	91,654	91,731	−0.1%	0.96	2.95
40–45	4726	73,934	73,700	0.3%	0.95	2.77
45–50	2159	37,050	37,057	0.0%	0.96	2.62
50–55	1172	21,542	21,423	0.6%	0.94	2.49
55–60	322	6507	6438	1.1%	1.06	2.40

Table 7. Prediction results with Lasso.

Categories	Number of Items	Real-Time	Prediction	GBP	Rae	Mae
0–15	1502	16,153	19,644	−22%	1.27	4.27
15–20	5672	61,873	74,182	−20%	1.26	3.96
20–25	11,129	127,439	145,548	−14%	1.16	3.74
25–30	12,827	157,715	167,755	−6%	1.06	3.40
30–35	9650	129,735	126,209	3%	0.99	3.19
35–40	6212	91,654	81,243	11%	1.03	3.17
40–45	4726	73,934	61,807	16%	1.16	3.38
45–50	2159	37,050	28,236	24%	1.60	4.35
50–55	1172	21,542	15,328	29%	2.05	5.46
55–60	322	6507	4211	35%	3.20	7.28

For Tables 5–7,

- *Categories* represent the range of lengths in feet to which the items belong.
- *The number of items* represents the number of registered joists belonging to a category
- *Real-time* is the sum of the recorded assembly times of the joists corresponding to a given category.
- *Prediction* is the sum of the predicted times of the beams corresponding to a given category.
- The *GBP*, *Rae*, and *Mae* measurements are the measures that allow the evaluation of the joists by category according to the prediction technique used.

The data resulting from Tables 5–7 give the difference of the values GBP, Rae, and Mae of the algorithms Lasso, GBR, and RFR. Indeed, the order of magnitude of the GBP, Rae, and Mae values are very close for the GBR and RFR algorithms. However, these values are more significant for the Lasso algorithm.

4.5. Evaluation

The following observations are made from Tables 5–7 and Figures 4–6.
For the choice of the best algorithm.

- The GBP obtained from the prediction results with the Lasso algorithm range from −22% to 35%, while the GBP obtained from the prediction results with the RFR and GBR algorithms range from −0.7% to 1.1% for the RFR and −0.4% to 1.1% for the GBR respectively, see Tables 5–7, and Figure 4. Considering that the GBP of ideal prediction results is very close to 0%. Thus, from the GBP point of view, the GBR and RFR algorithms provide more accurate prediction results than the Lasso algorithm.
- The Rae obtained from the prediction results with the Lasso algorithm range from 0.99 to 3.20, while the Rae obtained from the prediction results with the RFR and GBR algorithms range from 0.94 to 1.06 for RFR and 0.92 to 1.01 for GBR respectively, see Tables 5–7, and Figure 5. Considering that the Rae of the ideal prediction results is very close to 1. Thus, from Rae's point of view, the GBR and RFR algorithms provide more accurate prediction results than the Lasso algorithm.
- Finally, the Mae obtained from the prediction results with the Lasso algorithm vary between 3.17 and 7.28, while the Mae obtained from the prediction results with the RFR and GBR algorithms vary between 2.40 and 3.27 for the RFR and 2.30 and 3.14 for the GBR, respectively see Tables 5–7, and Figure 6. Considering that the Mae of the results of an ideal prediction is very close to 0. Thus, from Mae's point of view, the GBR and RFR algorithms provide more accurate prediction results than the Lasso algorithm.
- The GBR and the RFR present results with almost identical GBP, Rae, and Mae.
- The prediction times from the GBR and RFR are so close to real-time that their representative lines are overlayed see Figure 7.

Thus, The Lasso shows poor results compared to the other algorithms used in this article. The GBR and the RFR will be used in this article for pattern identification.

- For the RFR, below 1502 items, the GBP are higher than 0.65%, while for the GBR, GBP greater than 0.61 are observed below 1172 items. This may indicate that prediction with the GBR technique does not require large amounts of data to provide accurate results.
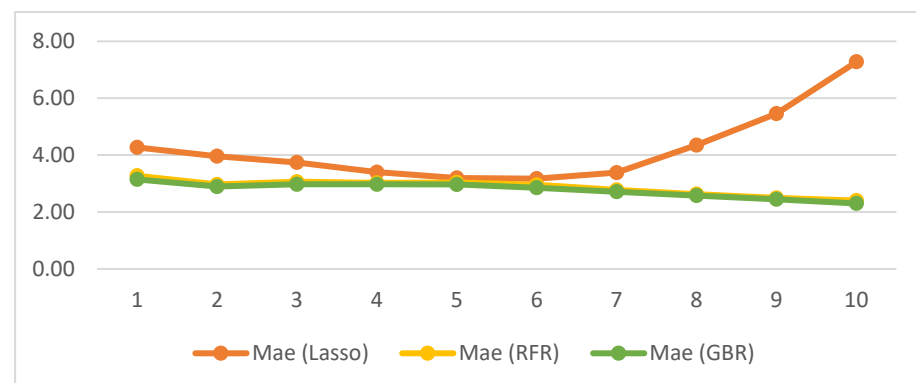
The highest RAEs (-1.05 for RBR and -1.14 for GBR) correspond to the lowest number of items category. This can be explained by ML prediction results being more accurate when more data are available [31,32].
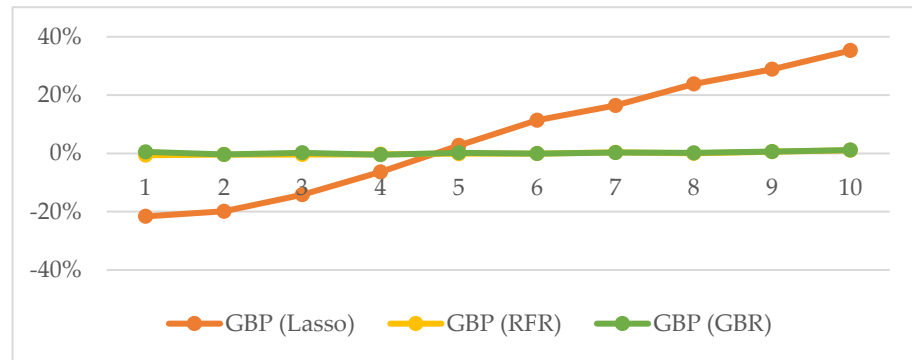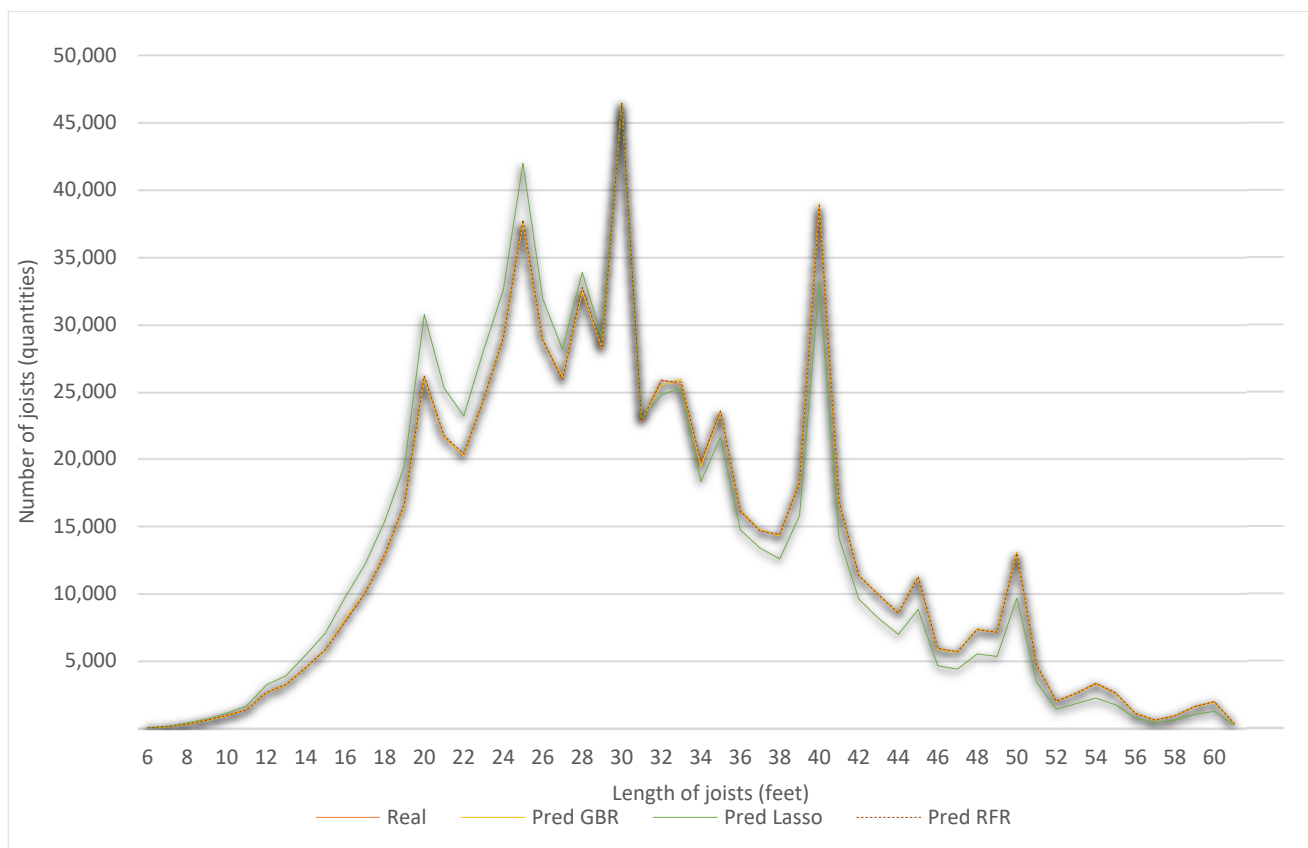

Figure 4. GBP comparison of Lasso, RFR, and GBR.
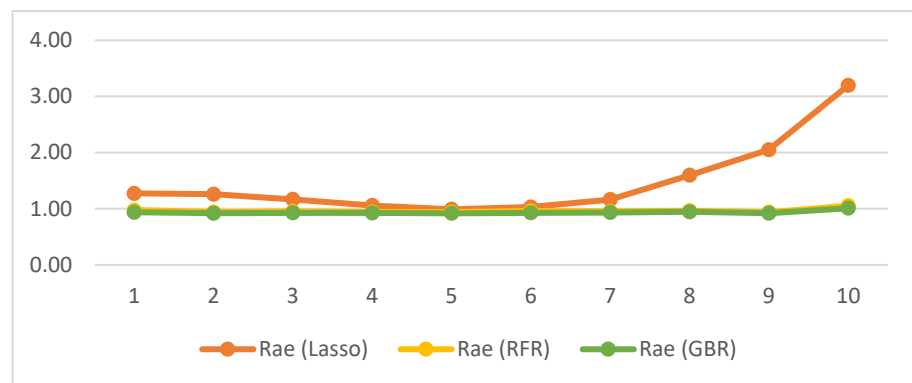
Figure 5. Rae comparison of Lasso, RFR, and GBR.

Figure 6. Mae comparison of Lasso, RFR, and GBR.

Figure 7. Comparison of prediction models.

4.6. Pattern Identification

Once the modeling is done, GBR and RFR allow for processing the pattern identification. Pattern identification proposes to identify variables that substantially impact the prediction results. This functionality is available on both the GBR and the RFR see Figure 8.
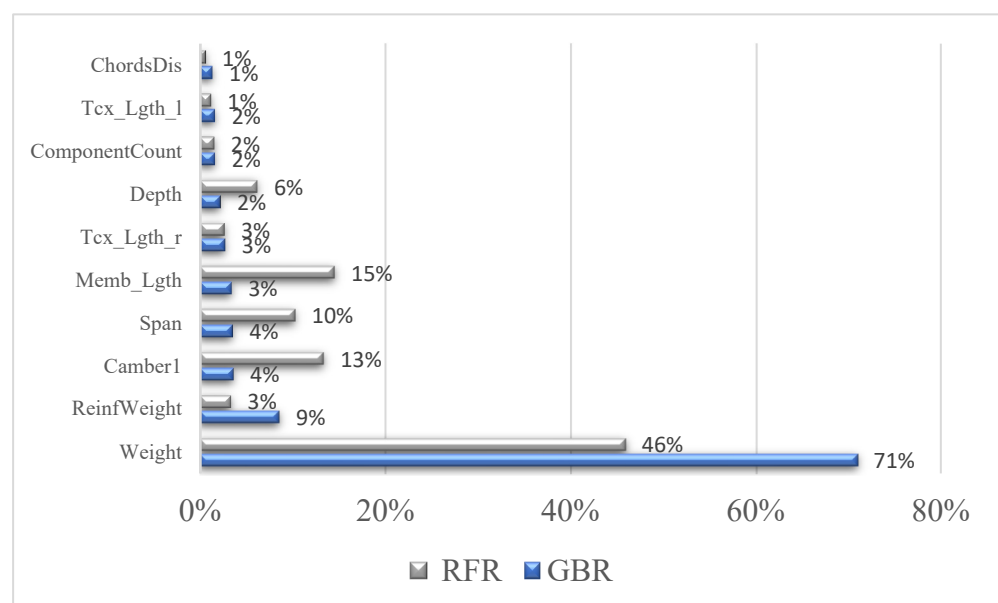
Figure 8. Feature importance.

According to Figure 5, the following remarks are made:

- For both prediction techniques, the variables Weight, ReinfWeight, Camber, Span, Memb_lgth, Tcx_Lgth_r, and Depth, are the variables that have the most impact on the joist assembly time.
- For both prediction techniques, joist weight is the variable with the most significant effect on assembly time.

4.7. Knowledge Learned and Design Rules

After the pattern identification, the following information is retained:

- The joist weight greatly influences the joist assembly time. This may indicate difficulty lifting and handling heavy elements on the assembly table. Improving the assembly line by installing additional lifting equipment could considerably reduce joist assembly time.
- The length of the members and the height of the joists also significantly impact the assembly time of the joists. This may indicate difficulty in maneuvering the long bars on the assembly table. A system for handling thin bars can considerably reduce joist assembly time.
- Extensions on the right side of the joist top chords have a more significant impact on joist assembly time than on the left side. This may indicate a difficulty in symmetrical work on the joist assembly table. Symmetrically equipping the assembly table can reduce joist assembly time.

Some design rules can be derived from this pattern identification:

- The ReinfWeight (the weight of the additional materials used to reinforce the joists elements) has a significant impact on the assembly time of the joists. For example, avoiding the use of reinforced bars by replacing them with larger profiles can significantly reduce joist assembly time.
- Joist length impacts joist assembly time, but the number of joist components does. Designing joists that can be assembled in subassemblies will undoubtedly increase the number of components but may reduce joist assembly time.
- The length of joist members has a considerable impact on assembly time. Fragmenting the length of joist parts such as top and bottom chords during design can reduce joist assembly time.

Top chord extensions to the right have more impact on joist assembly time than extensions to the left. Matching joists to make top chord extensions to the left during joist assembly before being flipped to the right can reduce joist assembly time.

5. Discussion and Interpretation of Results

The objective of this paper was to propose an approach to identify design rules such as DFMA from BIM models and ML algorithms. Thus:

Quantitative analysis of 55,444 BIM models by ML algorithms identified that the factors "steel component weight", "number of cambers", "component lengths", and "component depth" are the factors with the most significant impact on the fabrication time of steel structures. Thus, the analysis of BIM models can identify the factors with high impact on the fabrication time of steel structure components.

Quantitative analysis of BIM models by ML algorithms can provide information on the knowledge of the limits of the equipment available in fabrication and assembly plants. Indeed, the factors "weight of steel components", "number of cambers", "component lengths", and "component depth" are related to the capabilities of the equipment available in the fabrication and assembly plants.

The "weight of steel components" factor is more significant than the "number of bends" factor, which in turn is greater than the "depth of components" factor. Thus, quantitative analysis of BIM models by ML algorithms can enable the classification of the weight of

fabrication factors on the assembly time of steel structures. This can allow the formulation of design rules and the judicious selection of which rules to apply in case of rule conflicts.

Quantitative analysis of BIM models by ML algorithms can allow steel structure component manufacturers to identify deficiencies in the equipment available in the production facilities. Consideration of these deficiencies can allow fabricators to initiate modifications in a way that considers the limitations of their equipment.

6. Conclusions

This work proposes an approach to identify design rules from BIM models of previous projects and ML algorithms. To achieve this, this research suggests extracting and classifying data from BIM models of joists and using a predictive regression model to predict assembly time. Ensemble learning algorithms (RFR and GBR) proved to be better predictors than non-ensemble learning (Lasso). Furthermore, both ensemble learning algorithms were able to identify the most input variables. Based on these variables, it was possible to formulate recommendations concerning the assembly line and formulate design rules. A case study with 55,444 steel Joists demonstrates the feasibility of this method. Variables are classified according to their impact on the fabrication time. The study also proposes a series of relevant variables that could inspire future work in predicting manufacturing duration in steel joists projects in a specific workshop. The methodology proposed in this study can also be adapted to other productive construction industry sectors, such as steel structures installation, glass fabrication, and installation. For each of these applications, it will be necessary to get BIM models of previous projects and the duration of operations of these projects. The data must also come from a single production unit. A practical perspective for this study will be to apply the design rules developed on the design of new joists to be realized in these same workshops to appreciate the impact of these rules on the manufacturing time of the structures.

Author Contributions: Software, D.B. (David Bigras); Supervision, S.D. and D.F.; Validation, D.B. (Daniel Barbeau); Writing—original draft, M.F.S. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Carter, C.J.; Schlafly, T.D. Ave More Money. *Mod. Steel Constr.* **2008**, 55–59.
2. Evers, H.G.A.; Maatje, I.R.F. Cost based engineering and production of steel constructions. In *Steel Design Codes—Fourth International Workshop on Connections in Steel Structures*; American Society of Civil Engineering (ASCE): Fort Collins, CO, USA, 2000; pp. 14–22.
3. Schmidt, J.; Borsato, M.; Hinckel, E.; Storrer, P.; Onofre, E.; Maccari, F. A framework for capturing and applying design knowledge in new product development. *Int. J. Agil. Syst. Manag.* **2018**, *11*, 23–40. [CrossRef]
4. Heinisuo, M.; Laasonen, M.; Haapio, J. BIM based manufacturing cost estimation of building products. In Proceedings of the eWork and eBusiness in Architecture, Engineering and Construction, Cork, UK, 14–16 September 2010; pp. 53–59.
5. Barg, S.; Flager, F.; Fischer, M. An Analytical Method to Estimate the Total Installed Cost of Structural Steel Building Frames during Early Design. 2017. Available online: https://stacks.stanford.edu/file/druid:yb503ws4475/TR220.pdf (accessed on 28 November 2021).
6. Soh, M.F.; Barbeau, D.; Doré, S.; Forgues, D. Qualitative analysis of Request For Information to identify design flaws in steel construction projects. *Organ. Technol. Manag. Constr. Int. J.* **2020**, *12*, 2083–2094.
7. Sir, M.L. *Constructing the Team: Final Report of the Government/Industry Review of Procurement and Contractual Arrangements in the UK Construction Industry*; HMSO: London, UK, 1994.

8. Egan, J. *The Egan Report-Rethinking Construction*; Report of the Construction Industry Task Force to the Deputy Prime Minister: London, UK, 1998.
9. Soh, M.F.; Barbeau, D.; Dore, S.; Forgues, D. Toward a Qualitative RFIs Content Analysis Approach to Improve Collaboration between Design and Construction Phases. In Proceedings of the Creative Construction Conference, Budapest, Hungary, 29 June–2 July 2019.
10. Boton, C.; Forgues, D. The Need for a New Systemic Approach to Study Collaboration in the Construction Industry. *Procedia Eng.* **2017**, *196*, 1043–1050. [CrossRef]
11. Soh, M.F.; Forgues, D.; Doré, S. Integrating concepts and principles from lean thinking and the design for X to BIM. *Can. Soc. Civ. Eng.* 2017. Available online: https://www.csce.ca/elf/apps/CONFERENCEVIEWER/conferences/2017/pdfs/GEN/FinalPaper_62.pdf (accessed on 28 November 2021).
12. Jeong, W.; Chang, S.; Son, J.; Yi, J.-S. BIM-integrated construction operation simulation for just-in-time production management. *Sustainability* **2016**, *8*, 1106. [CrossRef]
13. Therman, A. DFMA in Product Development. Ph.D. Thesis, ARCADA University of Applied Sciences, Helsinki, Finland, 2020.
14. Mesa, J.; Maury, H.; Arrieta, R.; Corredor, L.; Bris, J. A novel approach to include sustainability concepts in classical DFMA methodology for sheet metal enclosure devices. *Res. Eng. Des.* **2017**, *29*, 227–244. [CrossRef]
15. Staub-French, S.; Poirie, E.A.; Caldero, F.; Chikhi, I.; Zadeh, P.; Chudasma, D.; Huang, S. *Building Information Modeling (BIM) and Design for Manufacturing and Assembly (DfMA) for Mass Timber Construction*; BIM TOPiCS Research Lab University of British Columbia: Vancouver, BC, Canada, 2018.
16. Selvaraj, P.; Radhakrishnan, P.; Adithan, M. An integrated approach to design for manufacturing and assembly based on reduction of product development time and cost. *Int. J. Adv. Manuf. Technol.* **2009**, *42*, 13–29. [CrossRef]
17. Ashley, S. Cutting costs and time with DFMA. *Mech. Eng.* **1995**, *117*, 74.
18. Halevi, G.; Weill, R. *Principles of Process Planning: A Logical Approach*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2012.
19. Harik, R.F.; Sahmrani, N. DFMA+, a quantitative DFMA methodology. *Comput. Aided Des. Appl.* **2010**, *7*, 701–709. [CrossRef]
20. Song, L.; Mohamed, Y.; AbouRizk, S.M. Early contractor involvement in design and its impact on construction schedule performance. *J. Manag. Eng.* **2009**, *25*, 12–20. [CrossRef]
21. McCuish, J.D.; Kaufman, J.J. *Value Management & Value Improving Practices*; Royal Institution of Chartered Surveyors (RICS): London, UK, 2002; pp. 1–16.
22. Elliott, V. Thinking about the coding process in qualitative data analysis. *Qual. Rep.* **2018**, *23*, 2850–2861. [CrossRef]
23. Stockemer, D.; Stockemer, G. *Quantitative Methods for the Social Sciences*; Springer: Berlin/Heidelberg, Germany, 2019; Volume 50.
24. Sacks, R.; Eastman, C.; Lee, G.; Teicholz, P. *BIM Handbook: A Guide to Building Information Modeling for Owners, Designers, Engineers, Contractors, and Facility Managers*; John Wiley & Sons: Hoboken, NJ, USA, 2018.
25. Banks, C.; Kotecha, R.; Curtis, J.; Dee, C.; Pitt, N.; Papworth, R. Enhancing high-rise residential construction through design for manufacture and assembly—A UK case study. *Proc. Inst. Civ. Eng. Procur. Law* **2018**, *171*, 164–175. [CrossRef]
26. Lawson, B. *How Designers Think: The Design Process Demystified*; Routledge: Oxfordshire, UK, 2006.
27. Elvin, G. *Integrated Practice in Architecture: Mastering Design-Build, Fast-Track, and Building Information Modeling*; John Wiley & Sons: Hoboken, NJ, USA, 2007.
28. Lingitz, L.; Gallina, V.; Ansari, F.; Gyulai, D.; Pfeiffer, A.; Monostori, L. Lead time prediction using machine learning algorithms: A case study by a semiconductor manufacturer. *Procedia CIRP* **2018**, *72*, 1051–1056. [CrossRef]
29. Kelleher, J.D.; Namee, B.M.; D'arcy, A. *Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples, and Case Studies*; MIT Press: Cambridge, MA, USA, 2015.
30. Rajagopal, A. The Rise of AI and Machine Learning in Construction. 2017. Available online: https://medium.com/autodesk-university/the-rise-of-ai-and-machine-learning-in-construction-219f95342f5c (accessed on 8 November 2018).
31. Lantz, B. *Machine Learning with R*; Packt Publication Ltd.: Birmingham, UK, 2015; Volume 2.
32. Raschka, S.; Mirjalili, V. *Python Machine Learning: Machine Learning and Deep Learning with Python, Scikit-Learn, and TensorFlow 2*; Packt Publication Ltd.: Birmingham, UK, 2019.
33. Michalski, R.S.; Carbonell, J.G.; Mitchell, T.M. *Machine Learning: An Artificial Intelligence Approach*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2013.
34. Musumeci, F.; Rottondi, C.; Nag, A.; Macaluso, I.; Zibar, D.; Ruffini, M.; Tornatore, M. An overview on application of machine learning techniques in optical networks. *IEEE Commun. Surv. Tutor.* **2018**, *21*, 1383–1408. [CrossRef]
35. Esmaeilian, B.; Behdad, S.; Wang, B. The evolution and future of manufacturing: A review. *J. Manuf. Syst.* **2016**, *39*, 79–100. [CrossRef]
36. Soh, M.F.; Barbeau, D.; Dore, S.; Forgues, D. Design rules to improve efficiency in the steel construction industry. In Proceedings of the Creative Construction Conference, Ljubljana, Slovenia, 30 June–3 July 2018.
37. Wu, X.; Liu, H.; Zhang, L.; Skibniewski, M.J.; Deng, Q.; Teng, J. A dynamic Bayesian network based approach to safety decision support in tunnel construction. *Reliab. Eng. Syst. Saf.* **2015**, *134*, 157–168. [CrossRef]
38. Sarkar, S.; Vinay, S.; Maiti, J. Text mining based safety risk assessment and prediction of occupational accidents in a steel plant. In Proceedings of the 2016 International Conference on Computational Techniques in Information and Communication Technologies (ICCTICT), New Delhi, India, 11–13 March 2016.

39. Rainer, C. Data mining as technique to generate planning rules for manufacturing control in a complex production system. In *Robust Manufacturing Control*; Springer: Berlin/Heidelberg, Germany, 2013; pp. 203–214.

40. Thanaki, J. *Python Natural Language Processing*; Packt Publishing Ltd.: Birmingham, UK, 2017.

41. Dong, X.; Yu, Z.; Cao, W.; Shi, Y.; Ma, Q. A survey on ensemble learning. *Front. Comput. Sci.* **2020**, *14*, 241–258. [CrossRef]

42. Sagi, O.; Rokach, L. Ensemble learning: A survey. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2018**, *8*, e1249. [CrossRef]

43. Lee, K.; Laskin, M.; Srinivas, A.; Abbeel, P. Sunrise: A simple unified framework for ensemble learning in deep reinforcement learning. *Int. Conf. Mach. Learn.* **2021**, *139*, 6131–6141.

44. Liaw, A.; Wiener, M. Classification and regression by randomForest. *R News* **2002**, *2*, 18–22.

45. Tibshirani, R. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B* **1996**, *58*, 267–288. [CrossRef]

46. Fonti, V.; Belitser, E. Feature selection using lasso. In *VU Amsterdam Research Paper in Business Analytics*; Business Analytics Master: Amsterdam, The Netherland, 2017; Volume 30, pp. 1–25.

47. Hill, A.V. *The Encyclopedia of Operations Management: A Field Manual and Glossary of Operations Management Terms and Concepts*; FT Press: Upper Saddle River, NJ, USA, 2012.

48. Monteiro, A.; Martins, J.P. A survey on modeling guidelines for quantity takeoff-oriented BIM-based design. *Autom. Constr.* **2013**, *35*, 238–253. [CrossRef]

49. Shen, Z. *BIM-Assisted Construction*; Higher Education Press: Beijing, China, 2010.

50. Mohsenijam, A.; Lu, M. Achieving sustainable structural steel design by estimating fabrication labor cost based on BIM data. *Procedia Eng.* **2016**, *145*, 654–661. [CrossRef]

51. Shen, Z.; Issa, R.R.A. Quantitative evaluation of the BIM-assisted construction detailed cost estimates. *J. Inf. Technol. Constr. ITcon* **2010**, *15*, 234–257.

52. Plebankiewicz, E.; Zima, K.; Skibniewski, M. Analysis of the first polish BIM-based cost estimation application. *Procedia Eng.* **2015**, *123*, 405–414. [CrossRef]

53. Hu, X.; Lu, M.; AbouRizk, S. BIM-based data mining approach to estimating job man-hour requirements in structural steel fabrication. In Proceedings of the 2014 Winter Simulation Conference, Savannah, Georgia, 7–10 December 2014; pp. 3399–3410.

54. Wah, L. The Singapore BIM Roadmap. In *Government BIM Symposium*; Building and Construction Authority (BCA): Singapore, 2014; Volume 2014.

55. Brennan, L.; Gupta, S.M.; Taleb, K.N. Operations Planning Issues in an Assembly/Disassembly Environment. *Int. J. Oper. Prod. Manag.* **1994**, *14*, 57–67. [CrossRef]

56. Alfieri, E.; Seghezzi, E.; Sauchelli, M.; di Giuda, G.M.; Masera, G. A BIM-based approach for DfMA in building construction: Framework and first results on an Italian case study. *Archit. Eng. Des. Manag.* **2020**, *16*, 247–269. [CrossRef]

57. Chapman, P.; Clinton, J.; Kerber, R.; Khabaza, T.; Reinartz, T.; Shearer, C.; Wirth, R. *CRISP-DM 1.0: Step-by-Step Data Mining Guide*; SPSS Inc.: Chicaco, IL, USA, 2000; Volume 9, p. 13.