Data-Driven Cycle Time Prediction of Fitting and Welding Stations in Steel Fabrication

Kamyab AGHAJAMALI[1], Alaeldin SULIMAN[1], Abdulaziz ALATTAS[1*], and Zhen LEI [1]

[1] *Offsite Construction Research Centre (OCRC), University of New Brunswick*
[*] *Corresponding author's e-mail:* aalattas@unb.ca

ABSTRACT

The construction industry's lack of materials, resources, and financial assets streamlined a shift toward using digital lean principles to obtain precise management over the limited resources. Steel fabrication companies rely heavily upon the enormous equipment to get promising results. However, implementing lean principles in the fabrication process is not straightforward due to the non-repetitive nature of steel construction products. Hence, the time-based modeling for such a process lacks accuracy and reliability, especially for manual steel fabrication processes. Accordingly, the current study aims to achieve a practical and accurate estimation of fabrication time aspects. This study targets modeling manual steel fabrication processes (fitting and welding workstations) in terms of processing times (cycle time and value-added time). The proposed approach builds a machine learning (ML) model to estimate the identified processing time aspects. For performance assessment, the typical correlation analysis and linear regression (LR) approach was used as a benchmark to quantify the ML model's pros and cons in terms of practicality and accuracy. The required data source for this study is a steel fabrication industry partner. The results of this study show ML superiority in accuracy over LR processing time predictive models, particularly when predictive parameters increase ML presents 13.2 % improvement in mean squared error compared to LR predictive model. LR models need fewer data and are not computationally expensive like ML models, making them more practical. Additionally, the study introduces a precise and practical time estimation approach. Such an approach provides precious input for simulation models which support evidence-based decisions and benefits quantification of plans.

KEYWORDS

Steel Fabrication; Fitting and welding workstations Machine learning; Linear regression; Process time estimation

INTRODUCTION

Steel Fabrication Shops are offsite construction factories that assist the construction industry during the construction phase by supplying steel assemblies (beam, columns, etc.) (Alvanchi et al., 2012). Offsite plants in steel construction projects aim to improve the project's deliverable by using specialized equipment to produce steel products in a controlled environment (Eastman & Sacks, 2008). The massive investment for offsite steel fabrication plants requires them to work in multiple shifts and keep the steel-based construction industry effective. A wide range of equipment serves different fabrication tasks, and several tools with various capacities exist in these plants. The customized nature of offsite steel fabrication plants requires a prudent plan for product flow in steel fabrication. Predicting the cycle time of workstations is of great importance through scheduling within offsite construction. The scope of this study is limited to improving the predictability in fitting and welding manual station. Unlike automated stations within the steel fabrication that perform a repetitive course of actions to process a product using a machine, manual stations (like fitting and welding) follow the distinctive design of each product to prepare it by human labor.

The fitting station starts when Computerized Numerical Control (CNC) coping machines provide the input product. A description of scheduled tasks in the fitting station is: (1) Locating the product on a fitting bench based on its complexity, (2) Reviewing the product design and gathering the needed pieces (stiffeners, end plates), and (3) Product preparation and then bolting, coping, and tack welding (tacking) of pieces onto the product, and inspection.

The next station for processed products in the fitting station is the welding station, but project requirements may skip some products from the welding process, and these products go to the next station (painting). In the welding station, the fitted assembly follows the following steps: (1) Locating the product on the welding bench, (2) Marking the product based on product design, (3) Welding preparation and welding the pieces to the product, and (4) Welding inspection.

Due to the high exposure of human resources, the stations that involve a manual process instead of a machine process are areas where scheduling and predictability are highly subjective. These manual stations experience bottlenecks because they cannot keep up with production coming from upstream automated stations. Considering the manual stations ' challenges, the low repetitive and diverse range of products in the steel fabrication makes the project's planning and predictability difficult (Hofacker & Gandhi, 2009).

The question to be addressed in this study is which of the Machine Learning (ML) and Linear Regression (LR) predictive models perform better in terms of accuracy and practicality. The significance of this study is improving predictability in the critical stations of steel fabrication. The predictive models presented in this study eliminate the subjectivity in time estimations by using data as a core steppingstone. The authors collected these data directly from industrial partners' steel fabrication plants. Promising results of data science have been drawing a substantial amount of attention in academics and media recently. Applying intelligent and computer-based tools and techniques to achieve meaningful information from data to help humans is pivotal in this era. This line of application of data is the content of the new field of knowledge discovery in databases (KDD). KDD maps a large number of preliminary data into understandable data. First, initial data excessively voluminous for a human to digest turns into concise and helpful information in this process. A practical example for this information could be a predictive model for estimating the value of future cases. The application of data-driven techniques for pattern discovery is the core of KDD (Fayyad et al., 1996).

KDD and data mining (DM) assist construction managers in planning projects by identifying functional and previously unknown patterns. In a given construction project, a project manager's responsibilities involve three primary efforts: (1) planning: Listing the project's activities, determining required resources, and estimating the time and cost for each activity,(2) scheduling: determining the sequence of listed activities and the required resource during the project stages, the planned date for each activity is the output of this effort, (3) controlling: monitoring the project progress during the construction phase and analyzing as-built projects with as-planned (Demeulemeester & Herroelen, 2002).

The focus of this study is predicting activities duration. This effort helps project managers in scheduling means "the determination of the timing and sequence of operations in the project and their assembly give the overall completion time" (Mubarak, 2015). Estimating the duration of the task's within a project incorporates expert judgment, analogous estimating, and other techniques (Project Management Institute, 2017). The customized nature of industrialized construction creates a unique atmosphere for scheduling the processes. Steel fabrication poses distinctive challenges because of the varying markets and the highly customizable product assemblies. The nature of construction projects carries risk due to the scheduling phase uncertainties. Several lines of research focused on production planning and control in the repetitive manufacturing environment. Meanwhile, limited investigations focused on improving the predictability of processing time in manual operations.

In the steel fabrication scheduling sector, scholars focused on the heuristic decision support system for scheduling steel fabrication projects; these works focused briefly on the total processing time of machinery operations and didn't consider manual processes in their study (Karumanasseri & AbouRizk, 2002). The processing time of each product is highly correlated with the product's work amount. For enlightening the scope of projects involved with non-repetitive operations like steel fabrication operations, some lines of research focused directly on quantifying the work amount for each product in steel fabrication. They automatically collected and used products category (e.g., a beam, truss. etc.) and complexity (e.g., number of fittings) data to quantify the needed effort for drafting a steel product. They called this quantity for a product a drawing unit (Song & AbouRizk, 2005).

The proposed methodology in this study borrows complexity features of products from this study to predict the processing time in both ML and LR models. The drawing unit quantifies the amount of work used by scholars to measure the productivity of steel drafting in steel drafting work. They implemented a neural network to analyze variation in productivity by using an abstract drawing unit measurement system, historical labor hours, and influencing factors of a given project (Song et al., 2012). Using historical labor hours for productivity analysis is applicable for completed projects and is applicable for future projects; the research in this study provides a data-driven approach for predicting activities time for productivity analysis. Multiple factors affect manual operations' processing time and vary considerably in industrial fabrication shops. Practitioners traditionally used statistical distribution from historical data or expert judgments to determine processing time. These methods are highly subjective or do not reflect influencing factors in determining processing time. Scholars implemented an artificial neural network (ANN) to predict processing time and thoroughly reflect these processing time factors. They integrated the ANN model into the simulation of steel fabrication operations (Song & AbouRizk, 2006).Using ANN for predicting processing time in other construction sectors shows promising results. Researchers integrated ANN and observation data in the earthmoving sector to predict excavation operations cycle time (Chao & Skibniewski, 1994; Karshenas & Feng, 1992). Moreover, ANN is used to model labor productivity in concrete formwork installation tasks (Portas & AbouRizk, 1997). Although many scholars have scrutinized production planning and control in industrialized construction, advanced data analytics is necessary to improve project planning and predictability in different customized industrial construction sectors. This study aims to enhance the predictability of two manual fitting and welding stations in steel fabrication. The proposed methodology evaluates the ML and LR predictive models' performance in terms of accuracy and practicality in manual stations.

The outline of this paper is as follows: (1) a description of the proposed method and its steps, (2) the results of the study that shows the improvement gained by using ML-based predictive models and a discussion of the results and how we could be able to improve predictive models, and (3) a conclusion that summarises challenges, limitations, and importance of this study and presents ideas for future studies.

METHOD

Predicting assemblies' processing time through steel fabrication stations is essential for planning productivity. The proposed methodology herein evaluates the results of ML and LR prediction methods for processing time. A time study collects the input data in this study through manual tasks in the steel fabrication's fitting and welding stations. **Figure 1** presents an outline of the study in two data collection and processing sections. The data collection process starts with reviewing the welding and fitting historical data and getting information about required tasks at each station. Next, the authors meet experts and managers in steel fabrication to discuss the outline of the study, and then they develop the required format for data collection (See **Table 1.**). Finally, the data are collected from the industrial partner's fitting and welding station in saint john, New Brunswick, Canada.

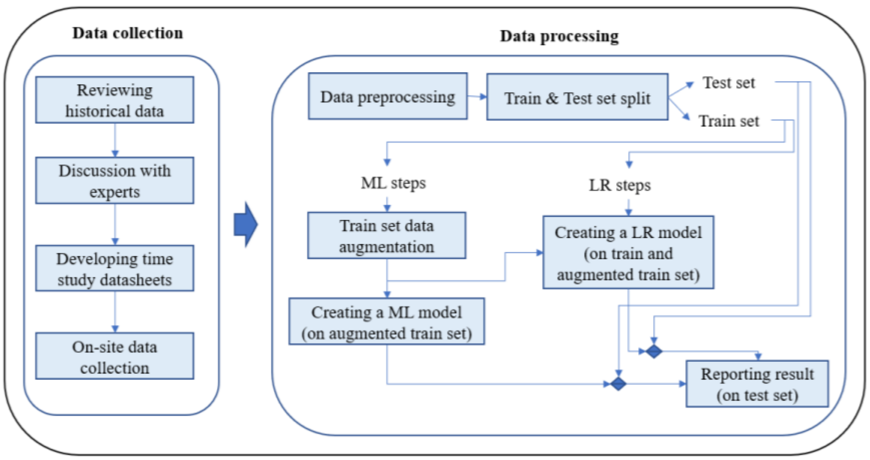

Figure 1. Outline of the proposed methodology

In the data processing section, the data preprocessing task prepares the data for data analysis and provides reliable and consistent data for the study. Second, the collected data is split into a train set (for creating the predictive) and a test set (for testing the models' performance). ML steps first augment the data for the predictive model, then implement the model based on the augmented train set, and finally report the model performance based on the test set. LR model works without a data augmentation process, and when this model is trained based on the train set (or augmented train set), the performance of the LR model on the test set is the output. Evaluating the LR model and ML model errors based on the same test datasets shows the accuracy of each predictive model. The practicality of each model depends on the ease of implementation of each model.

Table 1. shows the data collection formats for fitting and welding. The fitting station has six related features: (1) Fitter rating, which represents fitter's skill. (2) Parts tacked are the number of parts added to the assembly by welding. (3) Parts bolted are the number of parts that add to the assembly by bolting. (4) Bolts are the number of bolts at each assembly. (5) Copes represent the number of cutting that each assembly needs. (6) Drills are the number of holes for each assembly. The welding station has only one related feature, the weld length. **Table 1.** presents the required task at fitting and welding stations, and the authors collected these tasks' processing time by direct observation. The size of collected data for fitting and welding stations is 100 assembly.

Table 1. Datasheet's outline for time study.

Headings								
The Fitting Station					The Welding Station			
Assembly's Features		Assembly's Required tasks			Assembly's Feature	Assembly's required tasks		
Fitter rating	#Of Copes	Crane use	Sweeping	Bolting	Weld length	Crane use	Grinding	Move to next station
#Of parts tacked	#Of Drills	Reviewing drawings	Grinding	Drilling		Reviewing drawings	Gouging	
#Of Parts Bolted		Gathering pieces	Coping	Inspection		Marking spots	Welding	
#Of Bolts		Marking spots	Tacking	Move to next station		Sweeping	inspection	
Notes: * # present: number or quantity of the next term.								

First, the authors scrutinized the collected data from stations and removed inconsistent data for further analysis for data preprocessing. The inconsistency rate through the dataset was 10 %, and finally, we had 90 data points for analysis. We randomly separated these 90 pieces of data into training and test sets, and hence we had 63 data points for training and 27 data points for the test (Rácz et al., 2021).
This study utilized sci-kit-learn for implementing LR model in Python because of its strong and efficient tools for data analysis. Due to the limited size of the training set in our problem, sci-kit-learn directly solves the linear regression based on Eq.1. Next, when θ matrix weight is available, the cycle time is the product of θ and the test set.

$$\begin{aligned} If\quad X^T.X \text{ non}-\text{invertible:}\ \ \theta &= (X^T X)^{-1} X^T y \\ Else\text{:}\ \ \theta &= (X^T X + Z)^{-1} X^T y \end{aligned} \tag{1}$$

In Eq.1, X is the whole training set's matrix, and each row of it presents features of an assembly, and y is the training set's cycle time. Z is a non-negative matrix that that its diagonal values are λ (except the first row and first column) and other values of this matrix are zeros.
The same train and test set was used to implement a neural network and evaluate its performance with linear regression. The collected data for this study was insufficient for implementing a neural network, and hence the data was augmented to improve its performance. The authors used the augmented trainset to implement the LR prediction model and ensure that the augmentation process doesn't affect the comparison between LR and ML models. **Figure 2.** illustrates the data augmentation process in this study (Hu et al., 2019). The idea for this augmentation process is to generate several instances from each data in our train set. Each instance i., from our train set, is a parent for M children in our augmented data. Each child j inherits the same features of its parent i. The processing time for each child j is based on a normal distribution. The mean of this distribution is the average parent's processing time. Its standard deviation is the product of the train set's standard deviation, and K (a small, non-negative value that will be trained by architecting the neural network). **Table 2.** presents fitting stations augmentation process.

A neural network model trained in TensorFlow framework and Keras API to efficiently predict the cycle time in each fitting and welding station. We used Adam optimizer (Kingma & Ba, 2017) in Keras that provides an efficient ground for running gradient descent. These hyperparameters are learning_rate, layers and units in each layer, epochs, batch_size, and dropout regularization.

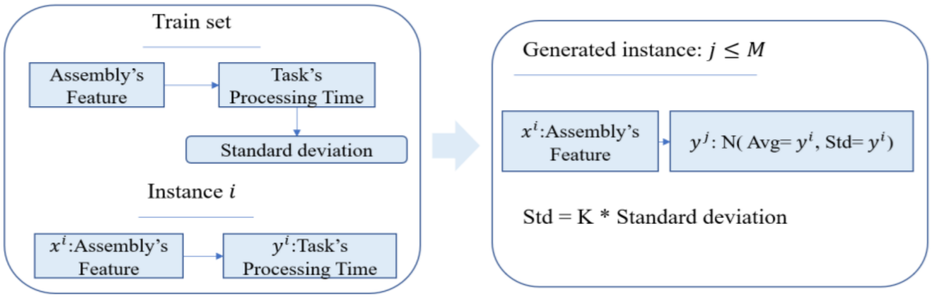

Figure 2. Data augmentation for ML model

Table 2. Fitting station sample collected data and augmentation process

ID	# Parts Tacked	# Parts Bolted	Bolts	Copes	Drills	Cycle Time
1	0	6	8	0	1	23
2	4	0	0	0	0	45
3	2	4	4	4	0	30
4	6	2	4	0	0	28

Average	Standard Deviation	K	M
26	16.78	100	6

Augmented Cycle Time ID = 1
23.04
23.1
23.15
22.3
22.5
22.01

RESULT AND DISCUSSION

The neural network architecture achieved by the experiment for both stations is shown in **Table 3.** The number of layers for both stations is three, and more layers cannot improve the predictability of processing time. The fitting station has seven units at each layer compared to three units for the welding station. The learning rate for both stations is close to Adam optimizer's default value (0.001) and for the fitting and welding station is 0.0007 and 0.0011, respectively. We used the dropout regularization method to avoid overfitting in our train set. These values based on an experiment (trial and error) for fitting and welding station are 0.0086 and 0.0009, respectively. The epochs, number of passes through the whole train set, and the batch size, the size that updates the optimization algorithm's parameters at each step are the same for both stations. The M value that augments the train set M times for both stations is 100 for each station, and the K value for both stations is 0.01 to limit the variance of processing time.

Table 3. Neural network architecture for fitting and welding stations

	Fitting Station	Welding Station
Number of layers	3	3
Units of each layer	7	3
Learning rate	0.0007	0.0011
Dropout	0.00086	0.0009
Epochs	30	30
Batch size	32	32
M (augmentation rate)	100	100
K (Std modifier)	0.01	0.01

Table 4. presents the mean squared error for welding and fitting stations using LR and ML. ML results in both stations have a minor error compared to the LR model. In this study, we used five features for predicting the processing time of the fitting and one feature for the welding station.

Due to the higher number of input features in the fitting station, the ML model's performance in the fitting station is higher than in the welding station. This study shows that ML models improve the simple LR model's error by around 13.2 percent when the number of predictors is five. When there is only one predictor ML model shows 2.3 % improvement compared to LR.

Table 4. Mean squared error (MSE) results for ML and LR in fitting and welding stations

	MSE of LR	MSE of ML	Improvement %
Fitting Station	68.58	59.56	13.2
Welding Station	86.84	84.59	2.3

The lack of reliable and sufficient data makes practitioners reluctant to choose ML methods for predictive models. The augmentation process discussed herein eliminates data insufficiency. The authors implemented another LR model based on the augmented data, and the results for linear regression were the same as LR before augmentation. Hence the better performance of the ML model over LR is not concerned with the augmentation process. During the error analysis, the authors realized that they should have used two features for parts that needed to be added to each assembly. Some parts in steel fabrication processes are liftable by workers, and others need cranes to be lifted. Hence, we should have defined two types of parts before data gathering: (1) liftable parts and (2) non-liftable parts.

CONCLUSION

Predicting the processing time of tasks is of high importance in scheduling construction projects. Precise estimation of the project's completion date leads to customer satisfaction and increased profit in the construction industry. Due to the many operational and context-level parameters relating to the work condition in the construction industry, the process of estimation of work duration is highly subjective. This study utilizes data to present ML and LR predictive models and eliminate subjectivity in the time estimation of fitting and welding manual stations. ML predictive models compared to LR present better results in terms of accuracy in fitting and welding stations by around 13.2 % and 2.3 % improvements in predictive model error, respectively. ML model superiority in accuracy over LR is more prominent when the number relating parameters to the work condition (input features for predicting) is higher. In terms of practicality, LR models are better. ML models suffer from a lack of data and are computationally expensive and complicated, and these two drawbacks make construction managers reluctant to use these models. This study eliminated subjectivity in time estimations within steel fabrication scheduling and enabled fabrication shop managers to predict the required time to complete manual fitting and welding station tasks and improve project scheduling and delays. The limitations of this study are in the size and the collection method of input data. The collection method in this study was based on a one-month direct observation and was prone to human error. Future studies can focus on improving the data collection method. An automatic data collection framework enhances the reliability and size of the data and improve the model's predictability.

ACKNOWLEDGEMENTS

MITACS accelerate program (Appl.# IT**26865**), and Ocean Steel Ltd support is appreciated.

REFERENCES

Alvanchi, A., Azimi, R., Lee, S., AbouRizk, S. M., & Zubick, P. (2012). Off-Site Construction Planning Using Discrete Event Simulation. *Journal of Architectural Engineering*, *18*(2), 114–122. https://doi.org/10.1061/(ASCE)AE.1943-5568.0000055

Chao, L.-C., & Skibniewski, M. J. (1994). Estimating construction productivity: Neural-network-based approach. *Journal of Computing in Civil Engineering*, *8*(2), 234–251. Scopus. https://doi.org/10.1061/(ASCE)0887-3801(1994)8:2(234)

Demeulemeester, E., & Herroelen, W. (2002). *Project Scheduling: A Research Handbook*.

Eastman, C. M., & Sacks, R. (2008). Relative Productivity in the AEC Industries in the United States for On-Site and Off-Site Activities. *Journal of Construction Engineering and Management*, *134*(7), 517–526. https://doi.org/10.1061/(ASCE)0733-9364(2008)134:7(517)

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From Data Mining to Knowledge Discovery in Databases. *AI Magazine*, *17*(3), 37–37. https://doi.org/10.1609/aimag.v17i3.1230

Hofacker, A., & Gandhi, D. (2009). *Evaluation and comparison of different Simulation-Software for the analysis and optimization of production processes at steel fabricators*. https://www.semanticscholar.org/paper/Evaluation-and-comparison-of-different-for-the-and-Hofacker-Gandhi/d848aace6f5ba7bffd7d4cb939931dfe21391db7

Hu, T., Tang, T., & Chen, M. (2019). Data Simulation by Resampling—A Practical Data Augmentation Algorithm for Periodical Signal Analysis-Based Fault Diagnosis. *IEEE Access*, *7*, 125133–125145. https://doi.org/10.1109/ACCESS.2019.2937838

Karshenas, S., & Feng, X. (1992). Application of Neural Networks in Earthmoving Equipment Production Estimating. *Undefined*. https://www.semanticscholar.org/paper/Application-of-Neural-Networks-in-Earthmoving-Karshenas-Feng/452ef4cdea8cbf938b091f4fbd9ca98d7da76d2a

Karumanasseri, G., & AbouRizk, S. (2002). Decision Support System for Scheduling Steel Fabrication Projects. *Journal of Construction Engineering and Management*, *128*(5), 392–399. https://doi.org/10.1061/(ASCE)0733-9364(2002)128:5(392)

Kingma, D. P., & Ba, J. (2017). Adam: A Method for Stochastic Optimization. *ArXiv:1412.6980 [Cs]*. http://arxiv.org/abs/1412.6980

Mubarak, S. A. (Saleh A. (2015). *Construction project scheduling and control* (Third edition, Vol. 1–1 online resource). Wiley. http://site.ebrary.com/id/11030124

Portas, J., & AbouRizk, S. (1997). Neural network model for estimating construction productivity. *Journal of Construction Engineering and Management*, *123*(4), 399–410. Scopus. https://doi.org/10.1061/(ASCE)0733-9364(1997)123:4(399)

Project Management Institute. (2017). *A guide to the project management body of knowledge (PMBOK guide)* (Sixth edition, Vol. 1–1 online resource). Project Management Institute. http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=1595320

Rácz, A., Bajusz, D., & Héberger, K. (2021). Effect of Dataset Size and Train/Test Split Ratios in QSAR/QSPR Multiclass Classification. *Molecules*, *26*(4), 1111. https://doi.org/10.3390/molecules26041111

Song, L., & AbouRizk, S. M. (2006). Virtual Shop Model for Experimental Planning of Steel Fabrication Projects. *Journal of Computing in Civil Engineering*, *20*(5), 308–316. https://doi.org/10.1061/(ASCE)0887-3801(2006)20:5(308)

Song, L., & AbouRizk, S. M. M. A. (2005). Quantifying Engineering Project Scope for Productivity Modeling. *Journal of Construction Engineering and Management*, *131*(3), 360–367. https://doi.org/10.1061/(ASCE)0733-9364(2005)131:3(360)

Song, L., Allouche, M., & AbouRizk, S. (2012). *Measuring and Estimating Steel Drafting Productivity*. 1–9. https://doi.org/10.1061/40671(2003)9