Proceedings of International Structural Engineering and Construction
Holistic Overview of Structural Design and Construction
Edited by Vacanas, Y., Danezis, C., Singh, A., and Yazdani, S.

ISSN: 2644-108X

FORECASTING CONSTRUCTION DELAY TIMES IN HIGH-RISE BUILDING PROJECTS

MUIZZ SANNI-ANIBIRE[1], ROSLI ZIN[2], and SUNDAY OLATUNJI[3]

[1]*Dammam Community College, King Fahd University of Pet. and Min., Dammam, Saudi Arabia*

[2]*School of Civil Engineering, Faculty of Engineering, Universiti Teknologi Malaysia, Johor Bahru, Malaysia*

[3]*Dept of Computer Science, College of Computer Science and Information Technology, Imam Abdulrahman Bin Faisal University, Dammam, Saudi Arabia*

High-rise buildings, which have become significant parts of the urban habitat, are particularly notorious for delayed completion times. Though there exist a plethora of studies on construction delays, the problem is insufficient research on prescriptive methods to mitigate delays. This study sought to employ Machine Learning (ML) techniques to learn from historical data on high-rise construction to forecast potential delay times. An input data containing 9 features and 12 cases were used. Initially, five feature sets were built based on the recursive feature elimination process. Further to that was the classification process that employs the following ML techniques: Multi-Linear Regression Analysis (MLRA), k-Nearest Neighbors (KNN), Artificial Neural Networks (ANN), and Support Vector Machines (SVM) to determine delay times. The predictive performance of these techniques was measured by co-efficient (R^2) and Root Mean Squared Errors (RMSE). The best three models were SVM with 2 independent variables (R^2 0.56, RMSE 1.6), ANN with 2 independent variables (R^2 0.49, RMSE 1.83), and KNN with all independent variables (R^2 0.46, RMSE 1.71). To improve the predictive performance of developed models, three best performing models were combined using fixed and trained rules. Results showed an improvement for a fixed rule based on minimum values with (R^2 0.59, RMSE 1.65). The study has significant implications to avoid delays in high-rise projects to avoid delays, which first employs ML in for the first time.

Keywords: Artificial intelligence, Machine learning, Hyperparameter optimization, Multi-classifier techniques.

1 INTRODUCTION

The 21st century is witnessing a rising complexity in buildings, observable in the rapid growth of tall buildings in urban centers globally. Despite the potential of tall buildings in the urban context to be a sustainable solution to an impeding housing crisis, such megastructures are subject to underperformance issues such as delays, which consequently results in time overruns, cost overruns, dispute, arbitration and litigation, total abandonment and dissatisfied stakeholders. Interestingly, CTBUH (2014) in its report "Dream Deferred: Unfinished Tall Buildings" noted the alarming rate of increase of "never completed" tall buildings. Though there exist a plethora of studies on the subject of construction delay in the research arena, the problem is that these studies tend to be descriptive and exploratory, and thus inadequate in providing the desired solution sought by the industry. Remarkably, the construction industry in its bid to solve its unproductivity issues is seeking the adoption of modern digital technology. The Construction research domain has featured studies adopting machine learning (Attal 2010, Czarnigowska and Sobotka 2014, Bayram 2017, Peško *et al.* 2017). These studies have been directed towards the estimation of costs and duration of construction projects. No study, to the best of the authors' knowledge, has sought to address the prediction of construction delay times, which may be attributed to the unavailability of relevant data. The effective prediction of delay times has the potential advantage to facilitate the construction planning and scheduling process by ensuring that the planned schedule is resilient to anticipated delays, and thus reduced risks. Thus, this study aims to develop a prediction model to predict delay time based on historical data, using common machine learning techniques.

2 LITERATURE REVIEW

Delays can be defined as situations where an event occurs at a time later than expected, or to be performed later than planned, or not to take timely actions, or occurring beyond the agreed date specified in the contract (Trauner 2009). Delay studies are popular in the construction industry. Though these studies may provide guidelines and controls for delay mitigation (Kolb *et al.* 2017), they tend to be descriptive and exploratory in nature and are potentially inadequate in providing the desired solution. AlSehaimi et al. (2013) suggest that alternative research approaches are needed to tackle the inherent problem. Such research approaches should develop and implement prescriptive methods in the form of innovative tools and techniques. To buttress this point further, the current mantra of the construction industry is to leverage the capabilities of modern digital technology in solving its problems. This new philosophy has been described as the fourth industrial revolution (IR 4.0), while in the construction industry has been expressed as "Construction 4.0". Despite the widely acknowledged need for digitization in construction, it is still lagging behind in adopting modern technology when compared to other industries. There has been little effort in the research domain geared towards adopting modern technology to mitigate delays.

3 RESEARCH SCOPE AND METHODOLOGY

The research methodology is summarily described in the following sections.

3.1 Stage 1: Data Preparation

The data used in this study was originally published by Ogunlana *et al.* (1996). The data was gathered from building projects made in Bangkok, Thailand. The dataset consisted of 12 cases of high-rise buildings, including 3 residential, 4 office, 1 hotel, 2 hospital, and 2 academic buildings. Table 1 presents descriptive statistics of the numerical variables in the data set.

Table 1. Descriptive statistics of the dataset.

Features	**Mean**	**Median**	**Standard deviation**	**Maximum**	**Minimum**
Contract price (mill baht)	657.58	343	635.12	1800	156
Total floor area m2	56.70	40	39.88	140	16
Total no. of floors	25.64	19	15.60	52	9
No. of floors underground	1.38	1	1.19	3	0
Project time (months)	24.92	22	8.95	41	15
Time spent (months)	15.83	14	6.60	31	9
Delay (months)	2.95	2	1.98	7	1

3.2 Stage 2: Developing Forecasting Models by Recursive Feature Elimination

In developing the models for cost and time, the recursive feature elimination (RFE) procedure explained by (Akande *et al.* 2015) was employed. RFE is a feature selection method that identifies the most relevant attributes towards accurately predicting an outcome. In selecting the best features for the model, the correlation of the attributes/features to the intended output is computed and the attributes/features are ranked accordingly, as presented in Table 2. The forecasting models thus developed are presented and described in Table 3.

Table 2. Correlation between each attribute and the target attribute (delay time).

Features	**Correlation coefficient**	**Rank**
Building type	-0.38698	2
Contract price (mill baht)	-0.16305	5
Total floor area m^2	0.011507	8
Total no. of floors	-0.24784	4
No. of floors underground	0.146429	6
Project time (months)	-0.27055	3
Time spent (months)	-0.14314	7
Construction management	0.505379	1
Contractor	0.505379	1

Table 3. Forecasting models based on the correlation co-efficient of features.

Delay Forecasting Models	**Selected features**
DFM-1	All
DFM-2	Best 5 attributes (Construction management; Building type; Project time (months); Total no. of floors; Contract price (mill baht))
DFM-3	Best 3 attributes (Construction management; Building type; Project time (months))
DFM-4	Best 2 attributes (Construction management; Building type)
DFM-5	Best attribute (Construction management)

3.3 Stage 3: Hyperparameter Optimization

In this study, Weka (Witten 2011) was used to set various optimization parameters. While MLRA does not require this process, the optimization parameter for KNN is the k value, as well as the search and distance function, while ANN depends on the learning rate and hidden layers. SVM optimization depends on the regularization factor C, the type of kernel function, as well as the ε-insensitive loss function. The optimal parameters used are as described in Table 4.

Table 4. Optimization parameters for ML classifiers.

Forecasting models	**KNN (Euclidean Distance)**	**ANN (Hidden layers, learning rate)**	**SVM (RBF Kernel)**
	K		C
DFM-1	3	1(4 nodes), 0.3	15
DFM-2	3	1(4 nodes), 0.3	9
DFM-3	1	2(4,4 nodes), 0.1	110
DFM-4	1	1(4 nodes), 0.3	7
DFM-5	1	1(4 nodes), 0.3	1

3.4 Stage 4: Performance Measurement

In measuring the performance of the techniques employed, the correlation coefficient and Root Mean Squared Error (RMSE) have been employed in line with previous studies (Akande *et al.* 2015, Peško *et al.* 2017).

3.5 Stage 5: Combining Classifiers

To improve the performance of the techniques used, a multi-classifier system also known as ensemble methods may be used. This is an approach that combines the prediction outcomes of a set of classifiers with the same of different sets of features. This can be achieved through fixed and trained rules (Xia *et al.* 2011), as illustrated in Figure 1.

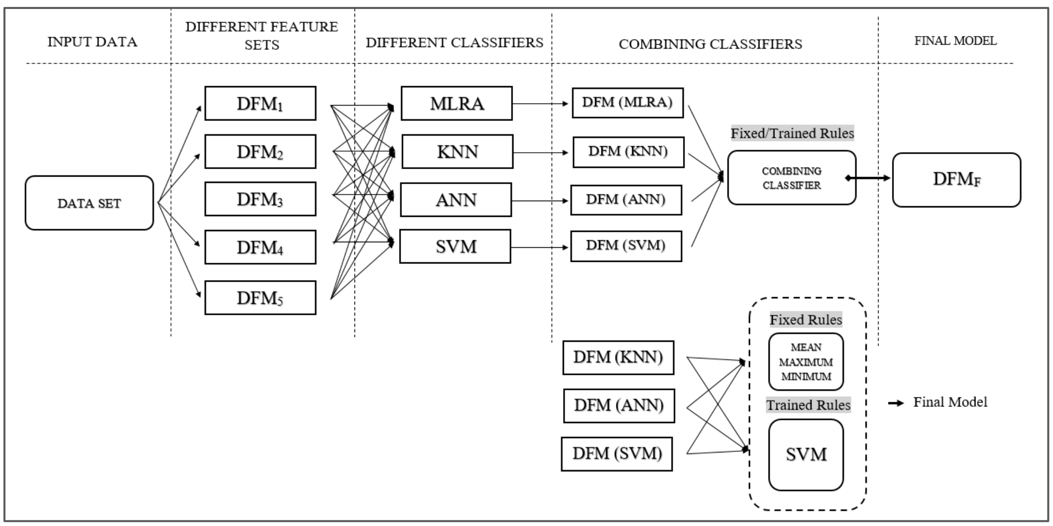

Figure 1. Architecture of multi-classifier system for delay forecasting.

4 RESULTS AND DISCUSSION

As highlighted previously, numerous studies exist in the domain of construction delays; however, few studies have been directed towards the adoption of modern digital technology in delay mitigation. Remarkably, the capabilities of Artificial Intelligence in solving construction-related problems have become popular in the last few decades. Despite that, the authors are not aware of any study that seeks to forecast the delay times that may occur in construction projects. The ability to learn from historical data to forecast the potential for delay in high-rise building projects could be considered a potential solution. This has the implication of properly assessing the risks of delays in construction schedules, allowing adequate planning for float times, buyouts, and padding of the schedule. It is worthy to note that the lack of relevant historical data is a potential barrier to the existence of such studies. Due to the foregoing, this study may be considered a prime study in the forecasting of construction delay times. The study has adopted four common machine-learning techniques (MLRA, KNN, ANN, and SVM). Firstly, feature selection was executed using the recursive feature elimination process. Furthermore, hyperparameter optimization was carried out and the machine learning techniques were deployed. The best models were selected according to each machine learning technique, as shown in Table 5. These were MLRA (R^2 0.172, RMSE 2.17), KNN (R^2 0.461, RMSE 1.71), ANN (R^2 0.492, RMSE 1.83) and SVM (R^2 0.564, RMSE 1.6). The best performing models were further combined, as shown in Figure 1, illustrating the architecture of the multi-classifier system. Fixed rules and trained rules were investigated. The results show that fixed-rule based on the minimum predicted values exhibited the highest correlation, and thus the best classifier system. Remarkably, all classifiers have correlation coefficients less than 0.7, which may be considered a benchmark for a satisfactory prediction outcome, this can be attributed to the limited number of the sample size (12), and thus, an opportunity for further study is presented in light of a more robust dataset. However, this study shows that construction delay times can be forecasted with the aid of machine learning techniques.

Table 5. Performance of classifier and multi-classifier techniques.

Classifiers	Best Forecasting Models	R2	RMSE
MLRA	DFM 4	0.172	2.17
KNN	DFM 1	0.461	1.71
ANN	DFM 4	0.492	1.83
SVM	DFM 4	0.564	1.60
Fixed Rule (Mean)	-	0.537	1.55
Fixed Rule (Max.)	-	0.452	1.78
Fixed Rule (Min.)	-	0.594	1.65
Trained Rule (SVM, Polykernel, C = 3)	-	0.415	1.82

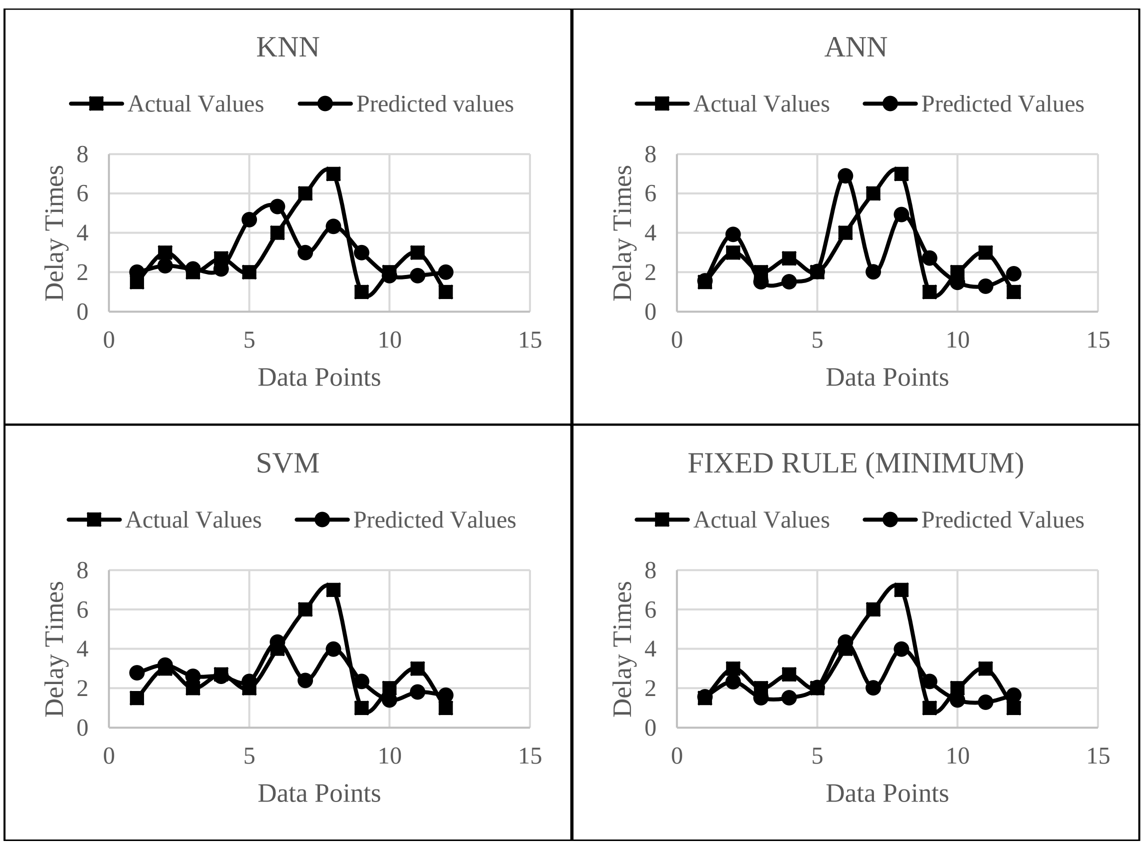

Figure 2. Cross plots of actual vs. predicted values for delay times.

Finally, cross plots of the actual and predicted values for the best three classifiers and the best-combined classifier are presented in Figure 2. It can be observed that the classifiers with greater correlation coefficients exhibited close matching of the data points.

5 CONCLUSIONS

High-rise buildings have been subjected to underperformance issues that plague the construction industry for many years, such as delays. Identifying the causes of delay is inadequate in solving the inherent problem, and thus, prescriptive research methods should be employed in delay studies. This study shows that a potential solution to delay mitigation could be in leveraging the capabilities of artificial intelligence/machine learning in forecasting delay times in a project. Based on the dataset used in this study, it is shown that combining various machine learning techniques in a multi classifier system has the potential to improve the predictive capability. In this study, the fixed rule minimum value was the best multi classifier system with R^2 of 0.59 and RMSE 1.65. It is worthy to note that sufficient historical data is lacking for the purpose of developing innovative tools to solve the problem of delays with the aid of machine learning. Thus, further research could seek to develop a robust database of relevant information to build machine-learning models as prescriptive tools to mitigate the delay. Ultimately, the development of such models is of potential benefit to engineers, contractors, and developers in controlling the risks of delays in high-rise projects.

References

Akande, K. O., Owolabi, T. O., and Olatunji, S. O., *Investigating the Effect of Correlation-Based Feature Selection on The Performance of Support Vector Machines in Reservoir Characterization*, Journal of Natural Gas Science and Engineering, Elsevier, 22, 515–522, 2015.

AlSehaimi, A., Koskela, L., and Tzortzopoulos, P., *Need for Alternative Research Approaches in Construction Management: Case of Delay Studies*, Journal of Management in Engineering, 29(4), 407–413, 2013.

Attal, A., *Development of Neural Network Models for Prediction of Highway Construction Cost and Project Duration*, 2010.

Bayram, S., *Duration Prediction Models for Construction Projects: In Terms of Cost or Physical Characteristics?*, KSCE Journal of Civil Engineering, 21(6), 2049–2060, 2017.

CTBUH, *Dreams Deferred: Unfinished Tall Buildings*, CTBUH Journal, 4, 46-47, 2014.

Czarnigowska, A., and Sobotka, A., *Estimating Construction Duration for Public Roads During the Preplanning Phase*, 4(1), 26–35. 2014.

Kolb M. H. A., Dief, M. I. A. D., El Beheiry, H. S., Kafafi, A. S. M., *Guidelines for Delay Control in Construction Projects*, PM World Journal Guidelines for Delay Control in Construction Projects, VI(II), 1-15, 2017.

Ogunlana, S., Promkuntong, K., and Jearkjirm, V., *Construction Delays in a Fast-Growing Economy: Comparing Thailand with Other Economies*, International Journal of Project, 1996.

Peško, I., Mucenski, V., Seslija, M., Radovic, N., Vujkov, A., Bibic, D., Krkljes, M., *Estimation of Costs and Durations of Construction of Urban Roads Using ANN and SVM*, Artificial Neural Networks and Fuzzy Neural Networks for Solving Civil Engineering Problems, 2450370, 2017.

Trauner, T., *Construction Delays: Understanding Them Clearly, Analyzing Them Correctly*, 2009.

Witten, I. H., *Data Mining Pratical Machine Learning Tool and Tecniques*, 2011.

Xia, R., Zong, C., and Li, S., *Ensemble of Feature Sets and Classification Algorithms For Sentiment Classification*, Information Sciences, Elsevier, 181(6), 1138–1152, 2011.