MACHINE LEARNING-BASED CONSTRUCTION PLANNING AND FORECASTING MODEL

Ahmet E. Keser & Assoc. Prof. Dr. Onur B. Tokdemir

Construction Management, Civil Engineering Department (Istanbul Technical University, Turkey)

ABSTRACT: *Construction planning and scheduling are crucial aspects of project management that require a lot of time and resources to manage effectively. Machine learning and artificial intelligence techniques have shown great potential in improving construction planning and scheduling by providing more accurate insights into project progress and forecasting. This paper proposed a machine learning model that utilizes regularly updated site data to generate predictions of quantity variances from the plan and enable a more accurate forecast of future progress based on historical data on concrete activities. Also, the outputs of this model can be used when creating a schedule for a new project. New schedules created with the help of this model will be more consistent and reliable due to its vast data pool and ability to generate realistic forecasts from this data. The model utilizes data from completed and other ongoing projects to generate insights and provide a more accurate and efficient construction planning and scheduling solution. Within the scope of this study, different attributes of concrete pouring activities of different projects and locations were used as input data for a machine learning process, and then, using this model on test data, the forecast concrete quantities were obtained. This model provides a more advanced solution than traditional project management tools by incorporating machine learning techniques while significantly improving construction planning, scheduling accuracy, and efficiency, leading to more successful projects and increased profitability for construction companies.*

KEYWORDS: *Machine Learning, Planning, Scheduling, Forecasting, Data Visualizing, Construction, Business Intelligence*

1. INTRODUCTION

Developing project schedules is critical to all projects, including engineering, manufacturing, construction, and others (Faghihi, Reinschmidt & Kang, 2014). Creating a reliable schedule and then updating and monitoring it as the project progresses is crucial for project management. A continuous data flow from the site is necessary to monitor project progress correctly. This process creates a vast amount of data. The construction industry deals with significant data from various disciplines throughout the life cycle of a facility (Bilal et al., 2016). Despite the abundance of data generated, its utilization in construction projects is often overlooked, resulting in a staggering amount of unused information. It is postulated that 96% of the data collected during construction projects goes unused (Snyder et al., 2018). In order to harness the potential of this unused data, various techniques such as statistics, machine learning, and artificial intelligence can be employed. Statistics are already commonly studied and applied within the construction sector. However, the importance of machine learning (ML), more generally, artificial intelligence (AI), is mostly overlooked and not being applied by companies as necessary, despite the studies on the matter.

Machine Learning applications have proven to outperform existing techniques, methods, and human decision-making on construction sites (Hammad et al., 2014). These methodologies offer valuable tools for further processing the data, enabling applications such as forecasting, risk analysis, labor allocation, and defect analysis. By leveraging these processes, construction professionals can unlock insights and optimize decision-making throughout the project lifecycle. Exploiting the power of ML data analytics tools can result in significant corporate benefits by enhancing the time performance of construction projects—regarded as one of the critical indicators of a successful project (Gondia et al., 2019). The most important part of construction project scheduling is the selection of resources (e.g., workforce, machines) and harmonizing their work (Jaskowski & Sobotka, 2006). This study aims to create more accurate forecasts for concrete pouring activities for effective planning, such as resource allocation in power plant projects. Often, project planners lack detailed drawings and necessary quantities at the beginning of the project. Even if such information is available initially, these quantities frequently change the project due to various factors. These factors may include unexpected soil features, inexperienced workforce, supplier delays, adverse weather conditions, or suboptimal planning. With the help of Machine Learning, correlations were sought between planned and at-completion quantities for data obtained from construction

Referee List (DOI: 10.36253/fup_referee_list)
FUP Best Practice in Scholarly Publishing (DOI 10.36253/fup_best_practice)
Ahmet Esat Keser, Onur Behzat Tokdemir, *Machine Learning-Based Construction Planning and Forecasting Model*, pp. 711-717, © 2023 Author(s), CC BY NC 4.0, DOI 10.36253/979-12-215-0289-3.71

projects. Data was collected from various ongoing or completed projects a construction company undertook to accomplish this. This data served as valuable input for machine learning models, enabling us to obtain meaningful and actionable results.

2. MATERIAL AND METHODS

Firstly, data was anonymously collected from 4 different projects (all personal and company-related information was removed). The projects are power plants in the following locations:

- Tashkent, Uzbekistan,
- Ashgabat, Turkmenistan,
- Sulaymaniyah, Iraq (two projects).

The primary data was obtained from the SAP Database ("SAP: Enterprise Application Software," n.d.) of the company, which is updated weekly for every project. SAP export data consisted of detailed weekly progress of the projects. Another data source is the Oracle Primavera ("Primavera P6 Enterprise Project Portfolio Management," n.d.) database. The company database has detailed L3 Updated Schedules and Baselines for each project. These schedules can provide planned and at-completion durations and start/finish dates if needed. At the date of this study, one of the projects was completed, and the other two were still ongoing, so data until the latest data date (30.06.2023) was used even though some of the activities were not completed.

Firstly, concrete pouring activities were filtered. The projects and schedules were taken from the same company and created according to the same procedures. Thus, all concrete pouring activities' wording and coding format are the same, as shown in Table 1.

Table 1: Activity ID and Name Structure of the Schedules

Activity ID	Activity Name
BZC-U-C-UBE-1800	Excavation of Soil - Foundation Level - Control Building
BZC-U-C-UBE-1810	Filling & Compaction - Foundation Level - Control Building
BZC-U-C-UBE-1860	Lean Concrete Pouring - Foundation Level - Control Building
BZC-U-C-UBE-1820	Installation of Formwork - Foundation Level - Control Building
BZC-U-C-UBE-1830	Installation of Rebar - Foundation Level - Control Building
BZC-U-C-UBE-1840	Concrete Pouring - Foundation Level - Control Building
BZC-U-C-UBE-1870	Installation of Formwork for Column - Ground Floor - Control Building
BZC-U-C-UBE-1880	Installation of Rebar for Column - Ground Floor - Control Building
BZC-U-C-UBE-1890	Concrete Pouring for Column - Ground Floor - Control Building
BZC-U-C-UBE-2040	Installation of Formwork for Beam & SLCA - Ground Floor - Control Building
BZC-U-C-UBE-2050	Installation of Rebar for Beam & Slab - Ground Floor - Control Building
BZC-U-C-UBE-2060	Concrete Pouring for Beam & Slab - Ground Floor - Control Building

The wording format of the data is as follows;

"Activity Description" – Level/Element – Building

The concrete activities start with "Concrete Pouring" as Activity Description. After the activity description, another attribute can be "level" or "element": foundation, column, slab, pedestal, wall, or trench. Moreover, there are different buildings of different sizes and floors. However, these projects mostly have concrete structures for mechanical and electrical equipment foundations. The data pool consisted of 263 activities; 180 were foundation concrete, and the other 83 were the other types of concrete activities.

RapidMiner (Mierswa & Klinkenberg 2018) was used as a tool for further processes. Rapidminer is a program that enables modifying the data and applying various Machine Learning techniques with a simple interface.

Raw data consists of planned quantities for each activity, weekly realized quantity for every data date, at completion quantity, project name, and project country for every activity. The data was manually transformed to distribute the attributes in the activity names to different columns for the machine-learning processes. In the early stages of the projects, planned quantity values were set to "1" for some of the foundation activities due to the unavailability of concrete quantity data during the baseline schedule development. As the concrete pouring activities progressed, these 'at completion' quantities for these specific activities were updated to reflect the actual volumes poured.

These activities need to be considered as outliers. The outlier is the data far from the average value of a statistics group. Outliers may affect the statistics and results substantially; therefore, they must be removed from the pool. Normalization is required to detect outliers in data pools with actual values, such as the one in this study, to ensure that variables with different scales are brought to a standard scale, preventing biased results.

In order to apply this process, the model in Fig. 1 was created.

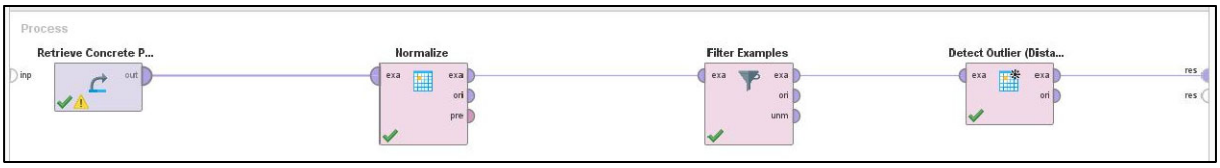

Fig.1: Outlier Detection Model

Due to the abundance of foundation concrete activities, the algorithm considered the "non-foundation" entries as outliers in a previous model. Thus, a "foundation concrete activities" filter was applied to detect outliers only among the foundation activities. The model detected five excessive values (which have value of 1 m^3 as Planned Quantity) as outliers, and these rows were deleted. After clearing the outlier entries, the new model in Fig.2 was created with clean data.

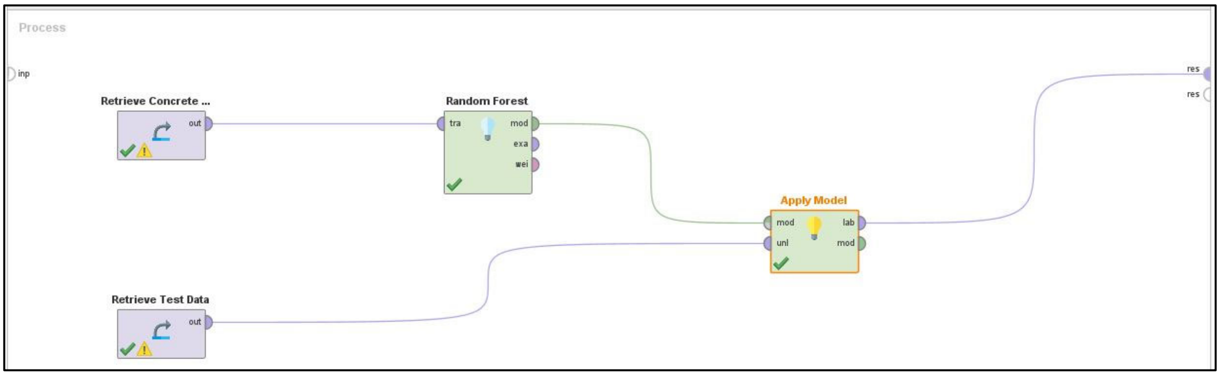

Fig.2: Random Forest ML Model

The input data consists of 5 different columns;

1- Activity ID: It is a unique ID for each activity.
2- Country: It includes the country project taking place.
3- Element Type: Element Type is the type of concrete element, which can be the foundation, column, slab, pedestal, wall, or trench.
4- Planned Quantity: It is the quantity planned at the beginning of the project, according to the baseline schedule.
5- At Completion Quantity: At Completion Quantity is the actual quantity on the site, which often differs from the planned quantity for various reasons.

Activity ID is unique for every row; the country column may include Iraq, Turkmenistan, or Uzbekistan. Element Type is the type of concrete element, which can be a foundation, slab, column, etc. Planned quantity is the quantity specified and planned at the beginning of the project, and At Completion, Quantity is the updated actual quantity throughout the project.

The model learns through the upper arm and applies the process to the Test Data (Table 2) below at the merging point (Apply model). Test Data consists of not started activities from the same three countries; thus, the table does not have at-completion quantities, leaving filling this column to the ML model.

Table 2: Structure of the Test Data

Activity ID	Country	Element Type	Planned Quantity
T_Act1	Iraq	Foundation	286.7
T_Act2	Iraq	Foundation	35.89
T_Act3	Iraq	Foundation	201.49
T_Act4	Turkmenistan	Foundation	31.25
T_Act5	Turkmenistan	Foundation	90.89
T_Act6	Turkmenistan	Foundation	219.19
T_Act7	Iraq	Slab	40.8
T_Act8	Turkmenistan	Slab	23.5
T_Act9	Uzbekistan	Slab	84.5
T_Act10	Iraq	Column	99.21
T_Act11	Turkmenistan	Column	137.69
T_Act12	Uzbekistan	Foundation	140.7
T_Act13	Uzbekistan	Column	175.55
T_Act14	Iraq	Pedestal	121.79
T_Act15	Turkmenistan	Pedestal	132.37
T_Act16	Uzbekistan	Pedestal	212.58
T_Act17	Uzbekistan	Column	90.05
T_Act18	Turkmenistan	Foundation	196.72
T_Act19	Iraq	Foundation	114.84
T_Act20	Turkmenistan	Foundation	108.51
T_Act21	Uzbekistan	Foundation	24.18
T_Act22	Turkmenistan	Wall	131.3
T_Act23	Uzbekistan	Wall	286.66
T_Act24	Turkmenistan	Pedestal	55.83
T_Act25	Uzbekistan	Pedestal	13.6
T_Act26	Iraq	Trench	272.11
T_Act27	Uzbekistan	Trench	155.09
T_Act28	Turkmenistan	Foundation	113.56
T_Act29	Turkmenistan	Foundation	91.76
T_Act30	Uzbekistan	Foundation	96.38

Random Forest regression was selected for the prediction process because more than two parameters affect the at-completion quantity: Country, Element Type, and Planned Quantity. Random Forest Regression is a widely used model in regression and classification problems. The accuracy of predictions increase when there are multiple decision trees.

3. RESULTS AND DISCUSSION

Although the amount is enough for consistency and prediction, usage of a broader data pool enables all the attributes to show effects on the results more clearly. For example, the projects' countries have many hidden variables affecting the quantities; however, this effect could not be seen clearly with only three countries. Also, most of the entries are for foundation concrete activities, so wall or column quantities did not affect the results as intended. Then, using the ML model with the input data in the test table, quantities of the activities were calculated at completion. The Final Table is given in Table.3 Forecasted "At Completion Quantities" are shown in the Prediction Column.

Table 3: Predictions on the Testing Data

Activity ID	Country	Element Type	Planned Quantity	Prediction
T_Act1	Iraq	Foundation	286.7	494.71
T_Act2	Iraq	Foundation	35.89	59.04
T_Act3	Iraq	Foundation	201.49	226.17
T_Act4	Turkmenistan	Foundation	31.25	34.41
T_Act5	Turkmenistan	Foundation	90.89	91.79
T_Act6	Turkmenistan	Foundation	219.19	269.12
T_Act7	Iraq	Slab	40.8	60.73
T_Act8	Turkmenistan	Slab	23.5	31.13
T_Act9	Uzbekistan	Slab	84.5	112.65
T_Act10	Iraq	Column	99.21	159.52
T_Act11	Turkmenistan	Column	137.69	130.16
T_Act12	Uzbekistan	Foundation	140.7	160.31
T_Act13	Uzbekistan	Column	175.55	208.09
T_Act14	Iraq	Pedestal	121.79	206.10
T_Act15	Turkmenistan	Pedestal	132.37	124.42
T_Act16	Uzbekistan	Pedestal	212.58	250.10
T_Act17	Uzbekistan	Column	90.05	122.94
T_Act18	Turkmenistan	Foundation	196.72	227.03
T_Act19	Iraq	Foundation	114.84	221.01
T_Act20	Turkmenistan	Foundation	108.51	66.39
T_Act21	Uzbekistan	Foundation	24.18	154.24
T_Act22	Turkmenistan	Wall	131.3	121.42
T_Act23	Uzbekistan	Wall	286.66	322.51
T_Act24	Turkmenistan	Pedestal	55.83	74.38
T_Act25	Uzbekistan	Pedestal	13.6	32.94
T_Act26	Iraq	Trench	272.11	422.71
T_Act27	Uzbekistan	Trench	155.09	208.74
T_Act28	Turkmenistan	Foundation	113.56	77.91

T_Act29	Turkmenistan	Foundation	91.76	91.79
T_Act30	Uzbekistan	Foundation	96.38	130.45

After inspecting the results and the graph in Fig.3, it was seen that they are consistent. Activities with Iraq in the Country column tend to differ the most from the baseline plan because projects in Iraq suffered from substantial design changes until their completion. On the other hand, the difference is lower in Turkmenistan activities because the baseline plan for the Turkmenistan project was closer to the realized work. Therefore, even if the graph has some more significant gaps, they are because of country and project differences. However, a more extensive data pool would enable predictions with less error if available. The rows with trench, pedestal, and walls do not have as much input data as foundation concrete activities; thus, these predictions may not be as accurate as foundation concrete activities. This study used country and element types as supplementary features to the planned quantities dataset. However, it is worth noting that including more comprehensive variables, such as detailed weather conditions, workforce experience, and material strength, which are known to impact quantities at project completion significantly, can further enhance the predictive accuracy of the model.

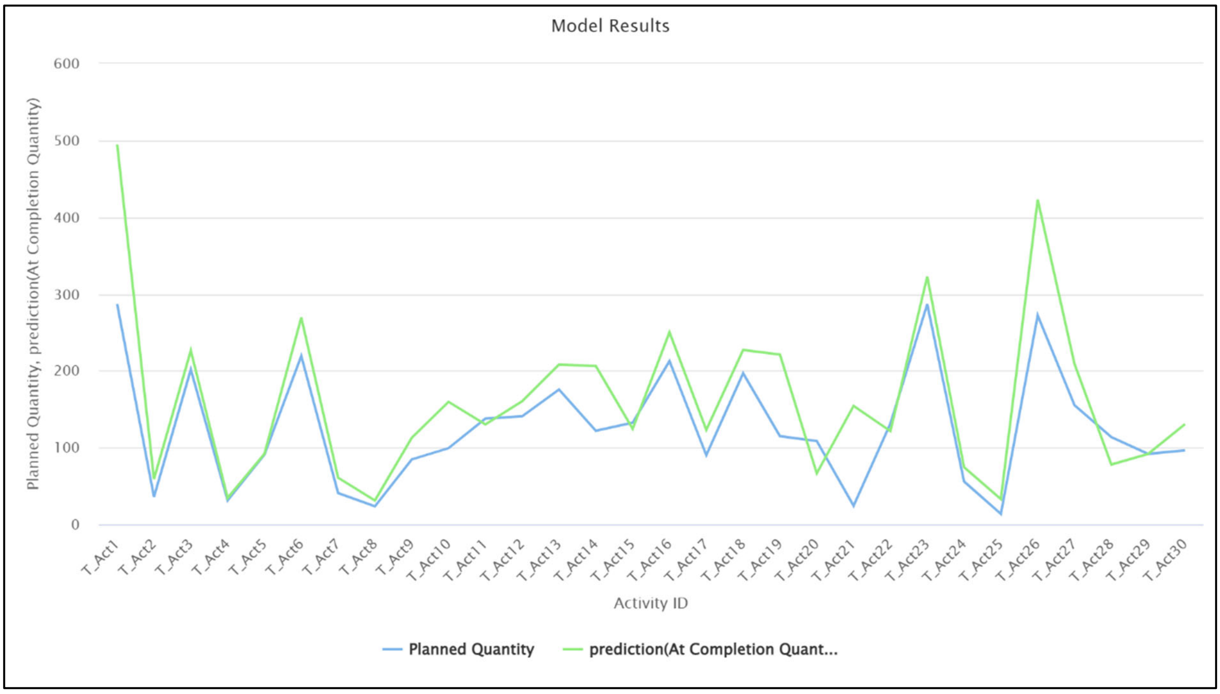

Fig.3: Graph of the Predictions

4. CONCLUSIONS

This study collected data from various ongoing and completed construction projects a construction company undertook, encompassing parameters related to concrete pouring activities. Utilizing advanced supervised learning algorithms, Machine Learning models were trained to establish correlations between planned and at-completion concrete quantities. While the model shows promising potential, it is important to note that future research could obtain more realistic and accurate results with more extensive and diverse data. The optimization of quantity forecasting is a key outcome, and the integration of Machine Learning-based forecasting offers powerful decision support for project management, enabling proactive measures to minimize delays and resource shortages. The successful implementation of Machine Learning underscores the importance of data analytics tools in the construction industry, which, with further exploration and expanded data availability, can lead to improved project management practices, resource utilization, and more profitable projects. It is essential for future research to address data limitations and consider real-time data integration to enhance the reliability and effectiveness of Machine Learning applications in the construction sector.

REFERENCES

Bilal, M., Oyedele, L. O., Qadir, J., Munir, K., Ajayi, S. O., Akinade, O. O., Owolabi, H. A., Pasha, M. (2016). Big Data in the construction industry: A review of present status, opportunities, and future trends. *Advanced engineering informatics*, *30*(3), 500-521.

Faghihi, V., Reinschmidt, K. F., & Kang, J. H. (2014). Construction scheduling using genetic algorithm based on building information model. *Expert Systems with Applications*, *41*(16), 7565-7578.

Gondia, A., Siam, A., El-Dakhakhni, W., & Nassar, A. H. (2020). Machine learning algorithms for construction projects delay risk prediction. *Journal of Construction Engineering and Management*, *146*(1), 04019085.

Hammad, A., AbouRizk, S., & Mohamed, Y. (2014). Application of KDD techniques to extract useful knowledge from labor resources data in industrial construction projects. *Journal of Management in Engineering*, *30*(6), 05014011.

Jaśkowski, P., & Sobotka, A. (2006). Scheduling construction projects using evolutionary algorithm. *Journal of Construction Engineering and Management*, *132*(8), 861-870.

Mierswa, I.; Klinkenberg, R.: RapidMiner Studio (9.2) [Data science, machine learning, predictive analytics] (2018)

Primavera P6 Enterprise Project Portfolio Management. (n.d.). Retrieved from https://www.oracle.com/tr/construction-engineering/primavera-p6/

Snyder, J., Menard, A., & Spare, N. (2018). Big data= big questions for the engineering and construction industry. *White Paper. First Myanmar Investment (FMI). Raleigh, US.*

SAP: Enterprise Application Software. (n.d.). Retrieved from https://www.sap.com/about.html