2023 European Conference on Computing in Construction
40th International CIB W78 Conference
Heraklion, Crete, Greece
July 10-12, 2023

A COMPARATIVE EVALUATION OF CATBOOST DECISION TREES AND EXPERT INTUITION TO PREDICT DURATIONS IN THE PREDESIGN PHASE

Svenja Lauble[1], Dominik Steuer[1], and Shervin Haghsheno[1]
[1]Karlsruher Institute of Technology (KIT), Karlsruhe, Germany

.

Abstract

Construction projects are often subject to scheduling errors caused by uncertainty and systematic planning fallacies. In the research, different statistical and predictive models were tested to predict the duration of a construction project. However, the results of these predictions were not yet been compared with the estimations derived by a human expert. This paper evaluates the prediction accuracy of CatBoost and expert intuition to predict the duration of public construction projects in the predesign phase. The authors use a dataset of the city of New York (USA) with 367 projects. Both expert intuition and CatBoost are compared with the performance indicators Mean Absolute Error (MAE) and absolute preference. The results show high outliers in the expert intuition, while the CatBoost indicates more consistent predictions. From a practical perspective, especially in uncertain situations, the CatBoost has advantages.

Introduction

Time schedules of construction projects are planned according to the work break-down structure. With this, the project is hierarchically broken down in its subtasks, resulting in schedules of different granularities. The first schedule defining project phases is the basis for the following schedules and has therefore a significant importance to the project's success. In the predesign phase, not much information about the corresponding project is available. Consequently, experts often plan the duration based on their knowledge and experiences in individual project phases.

During their realization, many real-world projects show that the milestones of the planned construction phases cannot be met (Potts 2005; Magnussen and Olsson 2006). These deviations result in time pressure, increased costs, quality loss, conflicts, and claims (Braimah and Ndekugri 2008). To reduce the deviations between the planned and realized durations, analytical models are a highly investigated topic in research. However, the results of their prediction were not yet juxtaposed with the estimations derived by a human expert. This is necessary to demonstrate the added value of analytical models to practitioners and to show a targeted use of analytical models as well as expert intuition. According to Kahneman and Tversky (1977) experts often tend to underestimate durations. They merely use an 'internal view' rather than looking back at historic projects finalized in the past. The target of analytical models' is to predict the durations of project phases objectively based on historical data to reduce time deviations.

In the first project stage, when not much information is available, the question arises whether analytical models yield a higher prediction accuracy than experts' intuition. Based on a real-world dataset we compare the prediction accuracy of both expert intuition and CatBoost. CatBoost is a tool for gradient boosting on decision trees that can be used for machine learning and predictive analytics. Lauble (2022) shows the promising prediction accuracy of a CatBoost in comparison to other decision trees for predicting durations in construction projects. The higher accuracy and better performance are due to the ability to handle categorical features more efficiently, its implementation of an ordered boosting technique, and its effectiveness in dealing with imbalanced datasets. In general, decision trees use a tree-like model of decisions and their possible consequences to classify input data into one or more categories.

In the next step, we identify situations in which experts or analytical models should be preferred. With these situational drivers' managers and their schedulers can make use of the best decision-making process, expert intuition, or analytical model in the right situation. To compare and evaluate both decision-making processes, the following paper relies on five parts.

First, we summarize related work defining expert intuition and analytical models. Further on, relevant analytical models to predict construction project durations during the predesign phase are identified and categorized.

Second, we develop an approach that compares the prediction accuracy of expert and analytical models as well as identifies situational drivers.

Third, we execute this approach on a real-world dataset and analyze the results with two performance metrics: the mean absolute error (MAE) and the number of cases in which one of the two decision-making processes is preferred. We further train a classification model that distinguishes the two decision-making processes to identify situational drivers for both.

Fourth, we summarize the experimental results and derive implications for practitioners based on these results. Here, we first show the high number of outliers in the predictions by the experts. In contrast, when comparing the number of cases for a preference, we can show a much smaller gap between both. To better understand the situations in which expert intuition fails in comparison to the CatBoost model and in which the Catboost shows its advantages, we train and evaluate a classification model.

This model presents a preference for a CatBoost in cases of high project changes and in an unstable environment.
Fifth, we derive a conclusion. The results of the first part conclude a collaboration of model and expert to reduce the planning fallacies of the expert. Lastly, we address a higher weighting of the CatBoost for the mentioned specific situations.

Related Work

In the following, we elaborate on relevant work regarding the advantages of expert intuition and analytical models in general, as well as analytical models for predictions in the predesign phase.

Expert Intuition and Analytical Models

In general, it can be differentiated into two types of decision-making processes, intuitive and analytical. Intuition follows unknown and not controllable thinking behaviors based on experts' experiences and knowledge (Braun and Benz 2015). Analytical models, in contrast, result from known and rational structures (Kahneman 2003).
With intuition, experts can deal with ambiguity (Scherm et. al 2016). Multiple authors plead for using intuition, especially in situations with high uncertainty, to get a rough understanding of the solution with minimal effort (Huang 2019, Jakoby 2019).
However, intuitive decisions are subject to the risk of situational bias. The 'self-serving bias' states, that past achievements are overestimated. According to the 'confirmation trap' information that is not consistent with one's beliefs is excluded from the decision. Lastly, individuals cannot assess whether their experience is based on a small sample with unreliable data. They would rather trust small samples with unanimous data than small sample sets with non-unanimous data. This results in a wrong understanding of patterns in data (the 'overconfidence effect'). (Kahneman 1977)
Therefore, multiple authors choose intuitive decisions only if there is not enough information available and quick decisions have to be made (Kahneman 2003, Bonabeau 2003, Davenport 2007, Huang 2019). In situations where a lot of information is available, individuals often do not have the capabilities to optimally analyze this information. However, computational capabilities can execute this and assist in an objective decision-making especially in complex situation. This research area is summarized under data-driven analytics. For data-driven analytics, good data management structures are necessary. The better the database is, the better the resulting solutions are. Davenport and Harris (2007) differentiate between four sequential levels of data analytical models: statistical analysis, extrapolations, predictive models, and optimizations. In all these levels, correlations are made to gain an objective result and thereby reduce the situational bias. With its levels, more insights are gained, which results in a higher competitive advantage for the company using these analytical methods.

Analytical Models to Predict Durations in the Predesign Phase

In Table 1, we summarize the identified analytical methods to predict durations for the predesign phase with exemplary references. The references are categorized according to the levels of Davenport and Harris (2007).
Optimizations are left out in Table 1 (e.g., Zheng (2004) with a genetic algorithm, Kalhor (2011) with an ant colony model, Jung (2016) with a tabu search, or Kumar (2011) with simulated annealing). These models are based on more detailed information that was not available in the predesign phase.

Table 1: Analytical methods to predict construction project durations (Lauble and Haghsheno 2022)

Level	Method and Reference
Statistical analysis	Relativ Importance Index (Meng 2012)
	Correlations (Walker 1995)
Extrapolations	Linear Regressions (Bromilow 1969)
	Fuzzy systems (Wu 1994)
	Box Jenkins (Agapiou 1998)
	Monte Carlo Simulation (Moret 2016)
Predictive models	Artificial Neural Networks (Lam 2016, Bhoka 1999)
	Decision Trees (Lauble 2022)
	AI ensemble (Erdis 2013)

Zheng (2004) exemplary work concentrates on time-cost optimizations for the process steps for the structural works. Following the level of the highest competitive advantage are predictive models.
The Australian researcher Bromilow developed in 1969 one of the first models to predict the construction duration. His linear regression equation is:

$$T = K * C^B \qquad (1)$$

T describes the time (in working days), K is a constant describing the general level of time performance, C are the costs of the project according to the contract (in a million USD), and B is a constant reflecting the sensitivity

of time to cost. Linear regression in general is a statistical modeling technique that aims to find the linear relationship between a dependent variable and one or more independent variables. Based on this equation, other models with limited variables were developed. These models are therefore easy to simulate and interpret.
In contrast, Artificial Neural Networks (ANNs) show better results by analyzing even weak correlations in big datasets compared to linear regressions (Chen and Huang 2006; Dissanayaka et al. 1999; Lam and Olalekan 2016; Petruseva et al. 2012; Wang et al. 2016). ANNs are a type of machine learning model that is inspired by the structure and function of biological neurons and can be used to learn complex patterns and relationships in data. However, the results of ANNs are generally not comprehensible to the scheduler and his clients, as they follow a 'black box' approach.
Lauble (2022) shows with an experiment on two real-world datasets that decision trees as part of machine learning perform better in comparison to linear regressions and ANNs. Especially the prediction accuracy of the CatBoost model can be highlighted. Here, for a dataset of residential buildings in San Francisco (USA), the MAE (see formula 2) is for the linear regression 16 days and for the ANN 150 days higher as the MAE of the CatBoost model. The decision tree not only performs better, but it also shows advantages regarding explainability and opening the 'black box' to its users. In general, decision trees can handle missing attributes well (Sheh 2017), include categories in the prediction, and show good results even with limited data. Therefore, in the following, the authors train the model with CatBoost.

Approach

In the following, we outline our approach which consists of a distinction between expert intuition and a machine learning model, the CatBoost. Figure 1 displays an overview of the entire process pipeline to identify cases and drivers for a preference.
To determine the error by the expert and CatBoost the raw dataset is filtered. To determine the quality of expert intuition, the first project version displays the prediction by the expert as starting point of the planning. We assume this estimation is done by humans. Then the first project version is compared with the final project version to measure the error of prediction by the expert. To train the CatBoost model, the last version of each project is used. This data is splitted with a ratio of 80/20. 80 percent of the data serves to train the model with the target of regression. The test dataset (20 percent) serves as the basis for analyzing the prediction accuracy of CatBoost. Further on, a cross-validation of k=3 was chosen. Cross-validation is done to assess the generalizability and robustness of a machine learning model by testing its performance on multiple independent subsets of the data. The prediction error of CatBoost is the difference between the predicted duration of the training model and the realized duration. Now, the prediction error of the CatBoost model and expert intuition can be compared.

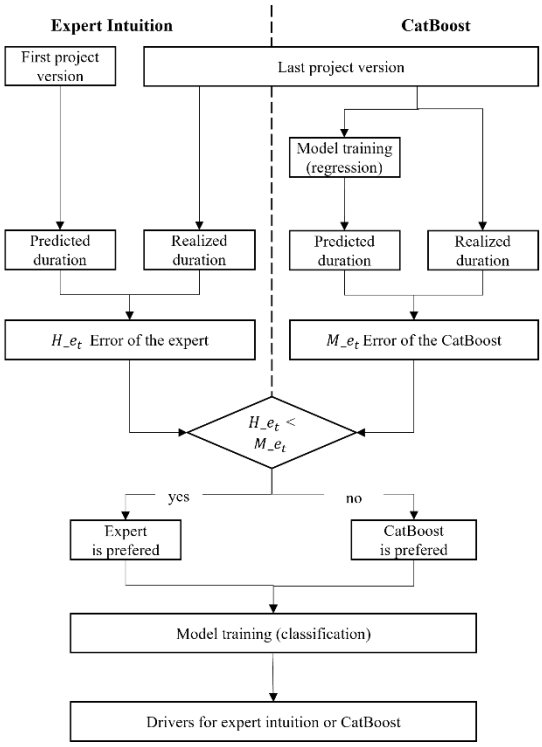

Figure 1: Pipeline of the experimental setup

Further on, now situations can be distinguished in which expert intuition or the CatBoost model is preferred. By documenting this per prediction, a classification model is trained to automatically differentiate in both situations. With this model, a prediction should be made about which mode (expert or CatBoost) should be chosen to select the prediction with the smallest error. Based on this model, drivers are identified that support the classification. These identified drivers can serve as assessment standards to consult a CatBoost model and weight predictions of expert intuition and the CatBoost.

Experimental Setup

In the following, we describe the experimental setup with the used dataset, evaluation metrics, and details on the implementation.

Data

Our experiment is based on real-world industry data for the city of New York, United States of America (USA). The full data set is available online (City of New York). The dataset consists of 2,400 public construction projects with 14 comparable project features, including the start of design and the realized or expected finalization of the project. The projects are documented at different time stamps, starting at the design phase and ending at the handover to the client. The first project version was documented in 1993, and the last expected finalization is in 2032. Therefore, the first and last project versions are filtered. The first project version reflects the expert's intuition, and the last version serves as the basis for the

model's training. After filtering, the data set consists of 367 projects. The 14 project features are:

- Documentation date (date)
- Project category (name, e.g., streets and roadways, schools, parks)
- Municipality of construction site (name)
- Administrative authority (name)
- Client agency (name)
- Phase at the beginning (name)
- Current phase (name)
- Design start (date)
- Budget prediction (in USD)
- Last budget change (in USD)
- Total budget changes (in USD)
- Predicted realization (date)
- Last change of the predicted duration (in days)
- Total duration changes (in days)

We further present the main properties of the feature "total duration" in Table 2.

Table 2: Description of the predicted duration

	Initial prediction	Final realization
Average duration	2.291 days	2.723 days
Standard deviation of the duration	1.153 days	1.250 days
Minimum	476 days	656 days
0.25-Quantil	1.584 days	1.902 days
Median	2.241 days	2.447 days
0.75-Quantil	2.558 days	3.172 days
Maximum	8.830 days	10.049 days

By comparing the initial prediction with the realized duration, planning fallacies can be identified. In all measuring points, the initial prediction is underestimated. The average initial duration differs, by 431 days from the realized duration. Also, the standard deviation gets higher with the ongoing project status.

The dataset is further enriched with 65 external data points each. These data points were selected with the goal of describing the economy and politics in New York, USA. Among other things, these include key figures on inflation, corruption, the innovation index, and the number of building permits. The full list of features is displayed in Table 3.

Table 3: Integrated external features to describe the economic and political situation for the dataset (F-1 to F-12 from OECD and F-13 to F-65 from Global Economy)

Nr.	Feature
F-1	Investment in fixed assets % of gross fixed capital formation (GFCF)
F-2	Real GDP forecast annual growth rate (%)
F-3	Employment in construction thousand persons
F-4	Hours worked hours/worker
F-5	Inflation (CPI) annual growth rate (%)
F-6	Prices of houses long-term average = 100
F-7	Built-up area square meters per capita
F-8	Consumer confidence index (CCI) (0-200)
F-9	Business confidence index (BCI) (0-200)
F-10	Population millions
F-11	Employed population millions of people
F-12	Price level index OECD = 100
F-13	Capital investment billion USD
F-14	Exchange rates Units of local currency per USD
-F-15	Unemployment rate percent
F-16	Employment rate percent
F-17	Government spending billion USD
F-18	Investment forecast Ratio of total invest to GDP
F-19	Competitiveness - WEF Index (1) 1-7
F-20	Competitiveness - WEF index (2) 0-100
F-21	Shadow economy Percent of GDP
F-22	Control of corruption -2.5 weak; 2.5 strong
F-23	Political stability index -2.5 weak; 2.5 strong
F-24	Short-term political risk 1=low, 7=high
F-25	Medium/long-term political risk 1=low, 7=high
F-26	Internet users, percent of population percent
F-27	Quality of roads 1=low, 7=high
F-28	Innovation index 0-100
F-29	Information technology exports, % of exports
F-30	Bank loans % of GDP
F-31	Number of listed companies Number
F-32	Innovation index 0-100
F-33	Index of property rights 0-100
F-34	Freedom from corruption index 0-100
F-35	Fiscal freedom index 0-100
F-36	Entrepreneurial freedom index 0-100
F-37	Labor freedom index 0-100
F-38	Monetary freedom index 0-100
F-39	Trade freedom index 0-100
F-40	Trade freedom index 0-100
F-41	Financial freedom index 0-100
F-42	Economic freedom, total index 0-100
F-43	Health expenditure per Kop USD per inhabitant
F-44	Death rate per 1000 persons
F-45	Poverty Percent of Population

F-46	Public expenditure on education percent of GDP
F-47	Globalization index 0-100
F-48	Economic globalization index 0-100
F-49	Political globalization index 0-100
F-50	Social globalization index 0-100
F-51	Percentage of world population
F-52	Percentage of world GDP Percent
F-53	Percentage of world exports Percent
F-54	Percentage of world imports Percent
F-55	Value added by industry billion USD
F-56	Value added by services billion USD
F-57	Quantity index of local suppliers 1=low, 7=high
F-58	Happiness index 0=unhappy, 10=happy
F-59	Human development index 0 - 1
F-60	Land area square kilometers
F-61	House price index % change; base year = 100
F-62	Building permits Number
F-63	Real residential property prices, annual % change
F-64	Consumer price index (CPI) percentage change
F-65	Government spending

These features are integrated for the year the design of the construction project started to include the economic and political starting situations. It is assumed that this will be included in a consideration. In total, this results in a total of 79 features per project.

Evaluation Metrics

We evaluate the experiment with two metrics, the number of preferred cases for a prediction by expert intuition or the CatBoost model and the mean absolute error (MAE) to measure the prediction error.

The MAE serves as a metric to compare the absolute prediction error. It is more robust to outliers than other metrics such as Mean Squared Error (MSE) or Root Mean Squared Error (RSME). This is because MAE measures the average absolute difference between predicted values $\hat{y}$ and actual values y (Hyndmann et al. 2006), rather than the squared or square root of the difference, which can give more weight to larger errors.. The MAE measures absolute quality.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \qquad (2)$$

The number of preferred cases shows, in general, which of the two prediction methods is preferred. It is presented as a percentage of total projects. This percentage indicates whether the prediction should be performed with the CatBoost model or by the expert.

Implementation Details

For programming, we used Google Colab and the following libraries: TensorFlow (1.14.0), keras-applications (1.0.6), CatBoost (0.24.3), and shap (0.37.0). The data set is split into a training and test set with a ratio of 80/20. The variable k=3 is chosen for cross-validation. To optimize the hyperparameters, a grid search is done. The results show a training with a maximum of 100 trees and a maximal depth of three.

Experimental Results and Discussion

In this section, we first compare the performance of predictions for durations in the predesign phase by experts and a trained CatBoost model. Second, we show relevant drivers to select the situation with the best outcomes.

Evaluation of Expert Intuition and CatBoost

In the following, we present the results of our model when comparing the prediction accuracy of expert intuition with that of CatBoost. For a detailed comparison, we further divided the dataset for the CatBoost in four subsets. These subsets contain the features of changes (total budget and duration change) and the external data describing the economic and political situation:

- Sub-dataset 1: with information about project changes and without external data
- Sub-dataset 2: with information about project changes and with external data
- Sub-dataset 3: without information about project changes and without external data
- Sub-dataset 4: without information about project changes and with external data

By comparing the prediction accuracy with and without these features, the preference of the CatBoost in the initial planning as well as using it for updates during the project execution can be analyzed. Second, the influence of external data on the prediction accuracy can be identified. Table 4 demonstrates the results of the prediction error using the MAE.

Table 4: Prediction Error (MAE) in days of Expert Intuition and CatBoost for the project duration (1: with information about project changes and without external data, 2: with information about project changes and with external data, 3: without information about changes and without external data, 4: without information about project changes and with external data; best results in bold for each row)

	Expert Intuition	CatBoost			
		1	2	3	4
Average	528	290	**212**	486	253
Standard deviation	653	427	278	491	**271**
Minimum	**0**	1	**0**	2	2
0.25-Quantil	**0**	94	73	171	94
Median	365	191	**157**	357	184
0.75-Quantil	833	336	**263**	595	319
Maximum	4.528	3.976	3.986	3.625	**3.182**

The results show the high-level planning fallacies of the expert's intuition. Each of the four CatBoost models performs regarding the average prediction error better than the expert. Here, especially the models containing external data indicate better results (CatBoost models 2 and 4). The information about changes also seems to influence the prediction accuracy slightly when comparing CatBoost models 1 and 2 to 3 and 4. Still, also the expert intuition shows, with the minimal and its 0.25quantile prediction error, better results in these categories than the CatBoost model. This high mismatch, once of very high planning errors and once of very good prediction, we also demonstrate in Table 5.
Table 5 represents the percent of cases in which the expert or the CatBoost should be preferred.

Table 5: Preferred Cases in percent of Expert Intuition and CatBoost for the project duration (1: with information about project changes and without external data, 2: with information about project changes and with external data, 3: without information about changes and without external data, 4: without information about project changes and with external data; preference in bold)

	1	2	3	4
CatBoost	**57,49**	**59,12**	44,96	40,05
Expert Intuition	42,50	40,87	**55,05**	**59,95**

Even if Table 3 demonstrates the high planning fallacies of expert intuition, Table 4 shows the close results when distinguishing the preferred cases for each decision-making process. As a maximum, the difference in dataset 4 between the preferred cases of expert intuition and CatBoost is only 19.90 %. And the number of cases in which CatBoost is preferred is not that clear. Even though the results are close together, we can detect a preference for CatBoost when changes about project information are available. In contrast, the expert's intuition is preferred for the initial prediction when project changes are so far unknown.
Comparing the results of Table 3 with Table 4, experts show higher outliers than the CatBoost where the 'internal view' distorts the understanding of the situations. These situations must be better understood to support in these the expert by the CatBoost.

Evaluation of relevant drivers

To identify relevant drivers for a preference of expert intuition or CatBoost, the results are further used for a classification model. If the prediction error for expert intuition is lower than the CatBoost a new target column is inserted with '0' (expert is preferred). Otherwise, if the prediction error for expert intuition is higher than the CatBoost, the target column is '1' (CatBoost is preferred). With this new target column, a SHAP framework is trained. The SHAP (SHapley Additive exPlanations) framework is a unified approach to explain the output of any machine learning model. It uses Shapley values, a concept from cooperative game theory, to assign importance scores to the input features of a model and explain how they contribute to the final prediction (Shapely 1953). The framework can be used for both global and local feature importance analysis, and has been shown to provide reliable and consistent explanations across different models and datasets. Therefore, SHAP support the interpretability of AI models (Lundberg 2018) and can be used to explain complex black-box models.
We conclude with the SHAP three main insights. First, CatBoost is preferred to predictions during the project's realization. This insight is already displayed with the datasets 1 and 2. For datasets 3 and 4 in contrast, this trend can also be confirmed when analyzing the columns 'phase at the beginning' and 'current phase'. For early documented projects (phase at the beginning: 'design') and a late current phase (phase 'construction realization'), also prefer a CatBoost as this gap also indicates changes.
Second, in an instable market environment, CatBoost is preferred. This instable environment is displayed in the datasets with a high unemployment rate, fewer capital investments, or high fiscal freedom. Kahneman and Klein (2009) as well as Agor (1986) confirm this fact. They argue that experts should make predictions in a stable environment, in which they have a good understanding of existing dependencies and influencing factors.
Third, we identify single features that are relevant for a distinction. These should be determined individually for each dataset. Exemplary, for the project category 'water supply', CatBoost is favored. Further on, the involved organizations or the area of the construction site can be relevant features to be analyzed.

Implications for Practitioners

Our work confirms the fact that there are planning fallacies in construction projects. With situational biases, experts show on average higher prediction errors than analytical models. However, comparing the number of the cases in which analytical models should be preferred is not statistically significant. Based on these results, we conclude with suggestions for research and practice.
Researchers are so far concentrating on the stand-alone use of analytical models. They should rather concentrate in the future on the collaboration of experts and an analytical model. Therefore, interesting topics are the interpretability of those models and the consequent reaction of the expert based on the gained insights. Documenting this reaction can serve as the basis for a continuous improvement of the analytical model.
Also, for construction purposes, collaboration is recommended. With the analytical model, the expert can check his intuitive predictions and adjust them. Especially, in an instable market environment and during project realization, analytical models should be weighted higher for prediction. Further on, to build up this collaborating system, a high focus on data management practices should be set. For the regression model as well as the classification model, relevant drivers must be identified. These drivers should have a central role in the

database and be documented accurately. Also, the scheduler's programming skills must be increased to develop and evaluate analytical models. The management must understand the general benefits of supporting the schedulers in the use of analytical models, but also know about the barriers to not holding schedulers accountable for prediction errors.

Conclusions

In this paper, we evaluate and compare the prediction performance of expert intuition and a CatBoost for the project duration in the predesign phase for a targeted use of both. CatBoost is trained on an exemplary real-world dataset of historical construction projects. We compare with two evaluation metrics, the expert's intuition to the trained CatBoost. These are the prediction errors of each decision-making process as mean absolute error (MAE) and the percentage of cases in which each is preferred. The comparatively higher prediction error of the expert displays the fallacies of planning. Still, comparing the number of cases, one should be preferred; a collaboration is recommended rather than a clear stand-alone use. We further analyze drivers that are relevant for weighting the decision power of both. Here, an instable market environment, updating predictions during the project realization, and single - or location-specific features support predictions by the CatBoost.
Thus, our work sets the basis for a new research area and has implications for schedulers in practice. With these implications, planning fallacies should be reduced in the future to generate high benefits for the AEC industry.

Acknowledgments

We express our gratitude to the city of New York (USA) for providing the public dataset. The authors acknowledge the financial support provided by the Federal Ministry for Economic Affairs and Climate Action of Germany in the project Smart Design and Construction (project number 01MK20016F).

References

Agor, W.H. (1986) How top executives use their intuition to make important decisions. Business horizons, 29.1, pages 49–53.

Bonabeau, E. (2003) Don't Trust Your Gut. Harvard Business Review, 81.5, pages 116–23.

Braimah, N. & Ndekugri, I. (2008) Factors influencing the selection of delay analysis methodologies. International Journal of Project Management, 26.8, pages 789–799.

Braun, F. & Benz, P. (2015) Genese natürlicher Entscheidungsprozesse und Determinanten kluger Entscheidungen [Genesis of natural decision processes and determinants of wise decisions]. Springer Gabler, Wiesbaden, page 15.

Bromilow, F.J. (1969) Contract time performance expectations and the reality. In Building forum, 1.3, pages 70–80.

Chen, W.T. & Huang, Y.H. (2006) Approximately predicting the cost and duration of school reconstruction projects in Taiwan. Construction Management and Economics, 24.12, pages 1231–1239.

City of New York (2020) Capital Projects by the Mayor's Office of Operations (OPS) [Data set opened on 05/06/2020]. https://data.cityofnewyork.us/City-Government/Capital-Projects/n7gv-k5yt

Davenport, T.H. & Harris, J.G. (2007) Competing on analytics: the new science of winning. Harvard Business School Press, Boston, Mass, pages 7-13.

Dissanayaka, S. & Kumaraswamy, M. (1999) Evaluation of factors affecting time and cost performance in Hong Kong building projects. Construction and Architectural Management, 6(3), pages 287–298.

External data (2020) External features describing the economic and political situation [Data set opened on 05/06/2020]. https://data.oecd.org/ and https://theglobaleconomy.com/

Huang, L. (2019) When It's OK to Trust Your Gut on a Big Decision. Harvard Business Review.

Hyndman, R.J. & Koehler, A.B. (2006) Another look at measures of forecast accuracy. International Journal of Forecasting, 22.4, pages 679–688.

Jakoby, W. (2019) Projektmanagement fur Ingenieure [Project Management for Engineers]. Springer Vieweg, Wiesbaden, 4, pages 176.

Kahneman, D. (2003) Maps of bounded rationality: Psychology for behavioral economics. American economic review, 93.5, pages 1449–1475.

Kahneman, D. & Klein, G. (2009) Conditions for intuitive expertise: a failure to disagree. American psychologist, 64.6, page 515.

Kahneman, D. & Tversky, A. (1977) Intuitive prediction: Biases and corrective procedures. Technical report, Decisions and Designs Inc Mclean Va.

Lauble, S. & Haghsheno, S. (2022) Predicting the construction duration in the predesign phase with machine learning decision trees. European Conference of Process and Product Modelling (ECPPM), Trondheim Norway.

Lam, K. & Olalekan, O. (2016) Forecasting construction output: A comparison of artificial neural network and box-jenkins model. Engineering, Construction & Architectural Management, 23, pages 302–322.

Lundberg, S.M. (2018) Interpretable machine learning with xgboost. Towards Data Science.

Magnussen, O.M. & Olsson, N.O.E. (2006) Comparative analysis of cost estimates of major public investment projects. International Journal of Project Management, 24.4, pages 281–288.

Petruseva, S. & Zujo, V. & Pancovska, V.Z. (2012) Neural network prediction model for construction project duration. International Journal of Engineering Research and Technology, 1.

Potts, K (2005) The new scottish parliament building a critical examination of the lessons to be learned. In Third International Conference on Construction in the 21st Century (CITC-III) Advancing Engineering, Management and Technology, pages 15–17.

Scherm, E. & Julmi, C. & Lindner, F. (2016) Intuitive versus analytische Entscheidungen Überlegungen zur situativen Stimmigkeit [Intuitive versus analytical decisions considerations of situational coherence], Springer, pages 299-318.

Shapely, L.S.A. (1953) Value for n—pemon game.

Sheh, R. (2017) Why Did You Do That? Explainable Intelligent Robots, page 631.

Wang, C.Y. & Yu, Y.R. & Chan, H.H (2012) Predicting construction cost and schedule success using artificial neural networks ensemble and support vector machines classification models. International Journal of Project Management, 30.4, pages 470 – 478.