(article starts on next page)

SORTING THINGS OUT? MACHINE LEARNING IN COMPLEX CONSTRUCTION PROJECT

Christian Koch, May Shayboun
Chalmers University of Technology, Gothenburg, Sweden

Abstract

This research includes answers from 324 main contractor representatives and 256 clients for a survey in Sweden, 2014. The literature review covers project management success in construction projects. A statistical correlation method is used to select the features that are strongly correlated with three performance indicators: cost variance, time variance and client- and contractor satisfaction. A linear regression prediction model is presented. The conclusion is an identification of the most correlating factors to project performance, and that human related factors in the project life cycle have higher impact on project success than the external factors and technical aspects of buildings.

Introduction

Construction projects are usually affected by multiple and interrelated factors which have a direct impact on projects performance and productivity; for example, factors such as poor management practices, unclear goals, performance measure, and crises orientation, are all contributing to poor productivity and cost overruns (Forbes and Ahmed, 2010). Construction projects suffer from scheduling problems related to the best sequencing for activities and resource allocation (Wauters and Vanhoucke, 2014). There are also problems associated to uncertainties in design, construction management and decision making, and these can be dependent on mathematics, mechanical calculations, and practitioners' experience (Lu et al. 2012). Regardless of emerging new performance measures such as safety and environmental performance, cost and schedule are still considered the most important to evaluate construction project performance (Chan and Chan, 2004).

Artificial intelligence (AI) is currently considered a solution superior to traditional systems used in studying complex systems such as construction project cost and time performance, (Lu et al. 2012). AI is a branch of computer science involved in the research, design and application of intelligent computers, and a comprehensive discipline, which was developed based on the interaction of several kinds of disciplines, such as computer science, cybernetics, information theory, psychology, linguistics, and neurophysiology. It is now used in different fields, such as expert systems, knowledge base system, intelligent database system,

and intelligent robot system and it solves problems such as knowledge management and decision making, (Lu et al. 2012). However, Walker (2016) argues that despite the innovative advancement in the construction industry in terms of knowledge management and organization learning, the full potential of high-capacity computing power is not fully utilized. Walker's perspective is that intelligent machines are to be the new construction management means to be utilized along with knowledge management and organizational learning, promising the combination of human learning with intelligent machines to solve construction problems and improve performance.

The aim of this paper is to develop prediction models for cost variance, time variance and satisfaction of the contractor and the client. With the utilization of a literature review of the subject of cost and schedule performance in the construction industry and empirical results utilizing a machine learning-based prototype. The literature review covers the possible causes for cost and time overrun. Then, this knowledge is transformed to empirical survey factor selection and linear regression results.

The structure of the article commences with the framework of understanding, the literature review for project success and performance indicators and predicting performance, and then moves to the sections containing method, limitation, the experimental results, the discussion, the conclusions, recommendations for future research, and the contribution.

Framework of Understanding

This paper follows on the model presented by Josephson (2013), see (Figure 1). This productivity report undertakes the debate of measuring productivity and elaborates on that by measuring productivity in cost per square meter for different building types. The model visualizes the relationship between input and output. In Josephson's report (2013), and in Koch and Lundholm (2018), project input depends on project conditions, and then the production stage continues causing costs to finalize construction work. Productivity is the output and it is measured by SEK (Swedish Krona) per square meter of total gross area, and by hours per square meter of total gross area. The construction project is also dependent on the performance of project organization, which represents

the most significant stakeholders in the project. The framework presented by Josephson (2013) is used in this paper to categorize factors behind project success or failure in the body of literature that discuss the subject. According to Tangen (2005), productivity is part of the broader term of performance, the assessment of which is aimed at. Therefore, it is suggested to replace productivity by performance, and then use this as an output. Time, cost, and satisfaction are the three performance measures.
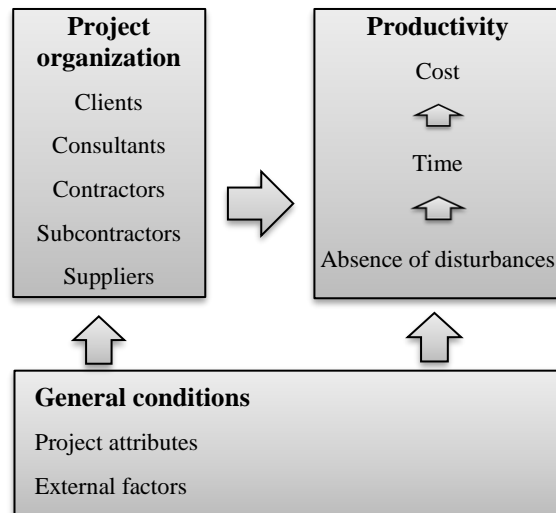

Figure 1: Productivity measurement basic model. Presented by Josephson (2013)

Literature review

Project success and performance indicators

Projects success can have more than one definition for different stakeholders, (Ogunlana and Toor, 2009; Chan and Chan, 2004). According to Frödell et al. (2008), success is when project-planned goals are met, and projects accomplished are on budgeted cost and schedule, as well as keeping high quality and technical performance. Sanvido et al. (1992) refer to project success as when the expectations for all parties (e.g. client, contractor, engineer or end user) are met or exceeded. De Wit (1988), suggests that a distinction should be made between project success and success of project management effort. The term "project success" is when the project objectives are met, judged from the perspective of all possible stakeholders, whereas project management effort is usually measured using cost, time and quality.

Most popular KPIs are cost, time, and quality, commonly known as the "iron triangle", (Ogunlana and Toor, 2009; Chan and Chan, 2004). Client satisfaction, the number of defects and the accuracy of the predicted project outputs are also presented as possible KPIs (Ogunlana and Toor, 2009; Alias et al., 2014). According to Xiong et al. (2014), to facilitate good collaboration among the members of the project team, the satisfaction of the contractor is especially important.

KPIs can be categorized into *subjective* and *objective* measures. *Objective* KPIs are, for example, time and speed of construction, rate of accidents, time variation, and cost variation. *Subjective* ones are quality, client and construction team satisfaction (Chan and Chan, 2004). Cox et al. (2003) categorize key performance indicators into *qualitative* and *quantitative*. The *quantitative* ones cover construction cost, units per man hour, cost per unit, on-time completion, and amount of rework. The *qualitative* ones are related, for instance, to safety, turnover, or workers motivation. The research effort and analysis presented in this paper considers evaluating the project in terms of satisfaction of clients and contractors, as well as the percentage of cost and time variance. The interpretation of cost and time over- and underrun is made based on Freeman's and Beale's (1992) description of the relationship between the sponsor (client) and the project manager and if the contract is based on fixed price. From the client's perception, if the project is finished before the due date, it creates revenues, while it is considered a failure if the project exceeds the timeframe. Also, cost overrun will not affect the sponsor, but for the project manager, cost overrun is regarded as a loss, whereas cost underrun is a gain. Therefore, schedule and cost underruns are considered success criteria in construction projects.

General project conditions

As shown in Figure 1, general conditions consist of project attributes and external factors. Project attributes are characteristics specific for the project; size, material, type of building, prefabrication, structure type, and contract and procurement route (Josephson, 2013; Chan et al., 2004; Koch and Lundholm, 2018). According to findings by Josephson (2013) and Al-Momani (2000), site conditions also play a role in project performance; for instance, lack of space affects delivery of goods and adequate storage space on site, thus making production more complicated. Technical complexity is mentioned by Lam and Wong (2009), project performance in terms of cost, time, quality and safety is highly affected by construction buildability. Ground conditions are mentioned by Josephson (2013), as the location of the project indicates the type of ground work to be performed, such as blasting work or piling.

External factors that affect the results of a project are presented in the framework by Chan et al. (2004), which categories the external environment of a project to its social, physical, economic and political dimensions. Weather is also seen as an external factor causing delays in a project, Jha and Iyer (2006), and Al-Momani (2000). For Josephson (2013) and Koch and Lundholm (2018), the Swedish weather is one of the most frequently mentioned disturbance by surveyed specialists. Moreover, geographical location,

market conditions such as material price fluctuations, competition, and aggressive tendering, are seen to impact the production. For example, Jha and Iyer (2006) found that aggressive tendering and high competition can have a negative influence on the quality performance of the project.

Project organization attributes

Project organization in the model (see Figure 1) is represented by the clients, the contractors, the consultants, and the subcontractor. This organization contributions in the project has an impact on its outcome. By the distinction made by Chan et al. (2004) and Alias et al. (2014), human related factors are represented by knowledge, skills, relationships and commitment with the project organization. The conclusion made by Jha and Iyer (2006) is that human-related factors are very important to achieve good quality in a construction project.

Frödell et al. (2008), Chan et al. (2004), agree on that client decision-making is a factor behind project success. Josephson (2013), found that in the main contractor's view, clients lack the experience to give clear messages and to plan the projects, although they can indeed create a collaborating environment and make decisions. Odeh and Battaineh (2002), show in a survey that slow decision making by the client is a common cause to delays. Another aspect of the client role are changes made by the client to the design and scope of the project, which are found to be considerably prominent in causing cost and time overruns (Al-Momani 2000; Attalla and Hegazy 2003). The client is required to have the ability to collaborate, manage the relationships within the project organization, such as the construction management and the design team (Walker 1995). Findings made by Koch and Lundholm (2018) show that the type of client has an impact on the productivity of the construction project. The procurement criteria that the client priorities in the contractor offer are also presented by the authors; whether the client emphasis on low price, high quality, time limit or special competence.

The category of contractors, consultants and sub-contractors, is also perceived to have a contribution to the project's performance. A common reason for bad performance is the contractor's relevant experience for a project, which can cause both time and cost overruns (Odeh and Battaineh, 2002). Managerial skills are highly valued throughout the literature as being important for achieving cost and time goals, solve problems with scheduling, (Attalla and Hegazy, 2003), and inadequate planning (Sambasivan and Soon, 2007). The good planning aspect can be associated with good planning for the specific design and construction phases; the skills of the designer are important for the project performance (Alias et al., 2014). The design team's communication with the

construction management and client's representative teams is one of the most important factors (Walker, 1995). The support that the contractors get from their own organizations is also highlighted (Josephson, 2013; Koch and Lundholm, 2018). In Koch's and Lundholm's (2018) report, support such as providing enough workforce, the priority of the project, and the degree the site-manager's involvement in choosing subcontractors and suppliers, were presented.

It is also found in Jha's and Iyer's (2006) work, that the high competence of the project manager has a positive impact on the quality of a project. The ability to organize, give feedback, plan, communicate and coordinate the project, as well as motivating the staff, are key to success (Frödell et al., 2008; Chan et al., 200). On the construction site, management and supervision, (Sambasivan and Soon, 2007); contractor's productivity during construction (Odeh and Battaineh, 2002), and workforce competence, commitment and participation in the project (Frödell et al., 2008), are important factors for project success.

Sambasivan and Soon (2007) present that problems with subcontractors, supply of material and labor, are some of the most common causes behind delays. In the report made by Koch and Lundholm (2018), the results of the study show that, when asked, the site managers claim that the performance of the ventilation and piping subcontractors are not satisfying. Conversely, the subcontractors are not fully satisfied with the site managers either. In specific, it is the main contractors' time planning that the subcontractors are not satisfied with, according to Koch and Lundholm (2018).

Predicting project performance

As stated by previous authors, for instance Alias et al. (2014), and Chan et al. (2004), it is possible to predict the level of a project's performance if the relationships between such a performance and the factors influencing it are established. Lowe et al. (2006), Kim et al. (2004) and Attalla and Hegazy (2003), used regression and neural networks in attempts to predict project cost at early project stages. Cost has been the center of attention for authors who have built different models for prediction. In these efforts, the focus had been on facts like duration, building type, finishing grades, type of installation, floor area, total units, procurement type, building function and type of client, which were turned into input variables.

Chi et al. (2012), proposed a decision support framework using data mining strategies, and utilizing a case study consisting of 193 projects submitted from 2008 to 2010. The data used reflect the degree of utilized technology and cost performance in construction projects. Chi et al. (2012) used four different ranking algorithms for project work functions, and chose the highest-ranking attributes showcasing the highest capability for predicting cost

performance. Attributes are; planning and execution, estimating, process and architectural engineering and definition, project scoping, detailed engineering and specifications, supplier qualification, warehouse and laydown yard, quality assurance (QA)/quality control (QC) systems, subcontractor's management, and customer satisfaction.

Another approach in the literature utilizes the earned value management methodology and support vector machines in forecasting estimate at completion and budget at completion, (Cheng, 2012; Wauters and Vanhoucke, 2014). This type of forecasting is based on data collected from different stages of the project; such as the schedule performance index, the earned duration and the earned schedule. Both studies conclude that forecasting models using AI outperform traditional methods used in estimate at completion calculations. However, so far, such models do not provide deeper understanding for project performance.

Previous models fail to fulfil the appraisal of the impact of the management performance on project success. Studying the impact of project management on project success lacks mathematical evidence. Therefore, it is necessary to find the link between project management performance and project success (Mir and Pinnington, 2014) which this study effort is attempting to provide with the utilization of machine learning as a method to study the assumptions about construction projects success. Mir and Pinnington (2014) concluded that project management (PM) performance in terms of PM leadership, PM staff, and PM life cycle management processes, has great impact on project team. More significantly, PM KPIs were proven to have great correlation with project success. In this research effort, the collected literature knowledge is transformed to tangible relationships between the factors contributing to project performance and project success.

Method

The data in Koch and Lundholm (2018) was collected through telephone interviews supported by questionnaires. In this way, answers are based on the respondent's own perspective as is always the case with questionnaires (Koch and Lundholm, 2018). The buildings included in the survey are primarily premises and group-built houses. Premises are buildings (such as day care centers, schools, office buildings, administrative buildings, sports and recreation facilities, hospitals and long-term care, church buildings, and elderly homes), stores, industrial properties, and others. In addition, civil engineering construction projects within the survey include pipelines, roads / streets / squares, railways, bridges, landfills, port facilities, fencing, plank and water

pipelines.

In each project, the client's project manager and the site manager of the contractor were asked within October-November 2014 to answer questions. The questions covered aspects about the product, such as the relative processes, the organizational structure, costs, times, the way the work progressed, and the way the project teams performed (Josephson, 2013). The clients received 23 questions and the site managers 21, in the respective questionnaires. The surveys were sent to 1000 individuals, with the valid responses amounting to 580, indicating an overall valid response rate of 58%. The survey was answered by 324 contractor representatives, displaying a 72% response rate, and 256 clients with a 62% response rate. The data used in this analysis contain information regarding the project number, location, structure type, client type, type of the construction method, the area of the building, and the phone numbers of respondents. Also, questions regarding the performance of clients, consultants, main contractor and subcontractors, are included. The datasets are also separated into two categories: the pre-construction phase consisting of the project program, pre-design phase, design phase, procurement and contracting of main contractor, and the construction phase covering external factors, construction site aspects, and the performance of the main contractor and the consultants.

In this study the same framework of understanding used by Josephson's report (2013), and in Koch and Lundholm (2018), was utilized, but productivity was replaced with cost variance and time variance as outputs. If cost variance and time variance are reduced, this implies savings in both cost and schedule. In other words, fewer costs and less time taken to complete the project, translate into fewer Swedish kronor (SEK) and fewer working hours per square meter.

Limitations

The limitations in this study are mainly due to the survey was pre-designed by others, and the data was not gathered by the authors of this research effort. It is possible that there were entry mistakes and missing values in cases that were not known or answered by the respondents. Such mistakes were tackled by checking and understanding the context and the ranges of the available gathered data and deleting incomplete answers. Moreover, the questions asked in the survey focus on cost and time, as well as performance of project teams.

Extreme projects that show high cost overrun or extreme savings are minority in the data. This implies that they are harder to have their outcome predicted. If that is the typical distribution of the data, then increasing the data size will actually not result in improvements in the prediction in future developed models. Such projects should be individually studied

further. There is also a possibility that there are factors that we are not aware of, which caused unlikely result for a project.

Experiment and results

The experiment includes the utilization of feature selection, linear regression, n-fold cross-validation, and Quantile-Quantile plot analysis. The tool that used in this study is Weka (Waikato Environment for Knowledge Analysis), which is a data mining and machine learning software, (Frank et al, 2016). The software contains a variety of methods for common data mining problems such as regression, classification, clustering, and attribute selection. Weka provides tools for data processing, feeding data into learning schemes and analysing the results, without writing any code.

Understanding the data is trying to know what it features contains, range of attributes, what the values represent, investigate output variables and detect outliers, (Han et al., 2011). Clearing the data off of noisy instances such as features that lack more than 20% of its values and projects lacking to be answered. The main contractor dataset consists of 17 input variables and 2 dummy variables in the pre-construction phase. The construction phase dataset consists of 25 input variables. And, the client data set consist of 58 input variables (of which 21 dummy variables) related to pre-construction and 11 input variables related to construction phase. The features selected with the CfsSubsetEval "correlation-based feature selection" method. It is a method to reduce the number of input variables used to build a prediction model, and to find the attributes with the most distinctive predictive capability in relation to the output, (Witten et al., 2011). The output variables used in this study are project cost-variance, time-variance and satisfaction of clients and contractors. Literature reviews about KPI:s, for instance by Chan and Chan (2004), cost- and time variance as well as participants' satisfaction, are important measures for project performance and still represent the most popular in the industry. Therefore, these three performance indicators are suitable as performance measures in this report. Also, in the data set used in this report, cost- and time variance and satisfaction of the client and the contractor are the only variables that can be identified as relevant ''outputs'', and it was decided to follow on the data availability.

The regression analysis is made using the selected features and resulted in 10 models. The models are validated with n-fold cross-validation for n=10 folds of the data sets. The 10-fold cross-validation is, by a large number of experiments, agreed to be the most appropriate for a good error estimate, according to Witten et al. (2011). The final error estimate is the average of the 10 error estimates (Witten et al., 2011). The error estimate used in this analysis is the root mean square error (standard deviation of residuals), which is defined as the root of the square value of the difference between the predicted value and the actual value, (Witten et al., 2011). If the standard deviation is low, then the data observations are closer to the ground truth value.

The regression results are presented in Table 1. The RMSE results show that the best prediction model is the cost variance of the construction phase with an error of 9.8%. On the other hand, time variance and satisfaction are harder to predict with error rates up to 16.0% for pre-construction time variance and 16.6% for predicting satisfaction of the contractor based on pre-construction data.

Table 1: Testing the adequacy for prediction model

	Prediction model	**RMSE**
Pre-construction	Contractor - cost variance	10.1%
	Contractor - time variance	16.0%
	Contractor - satisfaction	16.6%
	Client - cost variance	12.8%
	Client - satisfaction	13.5%
Construction	Contractor - cost variance	9.8%
	Contractor - time variance	15.3%
	Contractor - satisfaction	15.1%
	Client -cost variance	13.1%
	Client - satisfaction	12.5%

The plot to plot analysis shows that the data is normally distributed, which is a good indicator since the error is centred around 0 standard deviation. However, there also appear heavy tails in the error distribution (large error). Such projects were not explained by the input; for example, a project might have good indicators such as good collaboration and high quality, but still be overbudgeted or have low satisfaction and vice versa.

Discussion

General project conditions

Since the attributes presented in each model, (see Table 2) are those with the highest correlation with the output (cost variance, time variance, satisfaction), the observation is that general project conditions do not have a correlation with the satisfaction of the client or the contractor. On the other hand, cost variance in the client survey is influenced by project attributes such as "main procurement criteria - reference project", "Main procurement criteria - environmental aspects" and ''Type of contract = trades contract''. Cost variance in the contractor statistical analysis is influenced by "Contract type = traditional contract". Naoum and Egbu (2016), argue that the design process in traditional contracts is more influenced by the designer, hence the contractor's task is to do the

construction without being involved in the design. The fragmentation between the design and the building process cause rework and changes and produce higher costs for the projects.

It is important to investigate the underlying reason for criteria such as the "earlier collaboration with the contractor" or the "specific competence of the contractor" do not show any significant impact on project cost underrun. It could be explained by that the contractor with specific competence will not be identified in the correlation-based feature selection if such a contractor is more expensive. According to Alzahrani and Emsley (2013), contractors with relevant experience from previous projects can have positive impact on the success of a project.

This effort analysis finds that ''environmental aspects'' as a procurement criterion, have a positive impact on the project cost variance. Naoum and Egbu (2016), show that the emphasis on procuring a contractor according to environmental criteria can increase project performance because the environmental requirements call for more collaboration which lead to an overall performance gain. It is even argued that design- and build contracts are better organizational environments to achieve sustainability requirements.

Moreover, cost variance in the contractor statistical analysis is influenced by, "Gross area" and" project technically challenging and advanced". Time variance

is influenced by the "Percentage of prefabrication", ''extensive blasting works'' and "project technically challenging and advanced". It is observed that a technically challenging and advanced project induce time- and cost overrun. As stated by Lam and Wong (2009), construction buildability is highly affecting project performance (time, cost, quality and safety). Table 2 shows that the amount of prefabrication influences time variance but not cost variance. It is therefore reasonable to believe that prefabricated constructions save time but does not lead to lower costs due to being more expensive when compared to situ-built structures, (Josephson, 2013).

The client survey shows that ''Design phase - absence of disturbances'' is important factor to the cost variance, it has a decreasing effect. ''Absence of disturbances'' is also shown to be correlated to the decrease of the construction phase cost variance and the increase the satisfaction of the client and contractor. Weather conditions and ground conditions appearing to be some of the most important factors behind project performance according to Swedish studies as well as international studies, (Kock and Lundholm, 2018). However, these factors are not distinguished in the statistical analysis of this study effort. The reason is that the output (cost-, time variance or satisfaction) is not changing accordingly with the complication of groundwork or bad weather.

Table 2: Selected features (input variables)

Pre-construction phase	Construction phase
Cost variance – Client data set	Cost variance – client dataset
x_1: *Main procurement criteria - reference project*	x_1: *Construction phase - good collaboration within the project team at site*
x_2: *Main procurement criteria - environmental aspects*	x_2: *Construction phase - absence of disturbances and problems*
x_3: *Pre-design phase - investigation of plot*	x_3: *Construction phase - good results*
x_4: *Design phase - absence of disturbances and problems*	x_4: *Main contractor - good cooperation*
x_5: *Design phase - follow time plan weekly*	Satisfaction – Client data set
x_6: *Design phase - performance of architect*	x_1: *Construction phase - good collaboration within project team at site*
x_7: *Type of contract - trades contract*	x_2: *Construction phase - absence of disturbances and problems*
Satisfaction – Client data set	x_3: *Construction phase - good results*
x_1: *Project program - follow time plan weekly*	x_4: *Main contractor - innovative production method*
x_2: *Project program - good result*	Cost variance – Contractor data set
x_3: *Pre-design phase - conducting risk analysis*	x_1: *Client - giving clear message on time*
x_4: *Pre-design phase - well thought-through description of goals*	x_2: *Construction phase - good collaboration within project team at site*
x_5: *Pre-design phase - analysis of suitable type of contract*	x_3: *Construction phase - absence of disturbances and problems*
x_6: *Design phase - good collaboration within project team*	x_4: *Construction phase - follow time plan weekly*
x_7: *Design phase - good results*	x_5: *Quality of documents - piping consultant*
x_8: *Design phase - performance of ventilation consultant*	Time variance – Contractor survey
x_9: *Design phase - performance of architect*	
Cost variance – Contractor data set	
x_1: *Client - planning the project well*	
x_2: *Client - good decision-making*	
x_3: *Contract type - traditional contract*	
x_4: *Own company - production people involved in choosing*	

subcontractors and suppliers x_5: *Subcontractors - office workers duration* x_6: *Gross area* x_7: *Project technically challenging and advanced* x_8: *Limited time to prepare production* Time variance – Contractor data set x_1: *Client - planning the project well* x_2: *Client - encouraging innovation* x_3: *Project technically challenging and advanced* x_4: *Percentage of prefabrication* x_5: *Own crew - craftsmen duration* Satisfaction – Contractor data set x_1: *Client - planning the project well* x_2: *Client - encouraging innovation* x_3: *Client - good decision-making* x_4: *Client - clearly presenting goals of the project* x_5: *Own company - production people involved in choosing subcontractors and suppliers* x_6: *Own company - support of administrative work* x_7: *Subcontractors - office workers duration*	x_1: *Quality of documents - architect* x_2: *Ventilation subcontractor - innovative production methods* x_3: *Construction phase - extensive blasting work* x_4: *Construction phase - absence of disturbances and problems* x_5: *Construction phase - follow time plan weekly* Satisfaction – Contractor survey x_1: *Client - good cooperation in project* x_2: *Ventilation subcontractor - work to minimize disturbances and problems* x_3: *Ventilation subcontractor - result according to expectations* x_4: *Construction phase - good collaboration within project team at site* x_5: *Construction phase - absence of disturbances and problems* x_6: *Construction phase - follow time plan weekly*

Project organization attributes

By examining the results in Table 2, it is observed that ''client planned the project well'' has a decreasing influence on cost variance and time variance as well as increasing influence on contractor's satisfaction. Moreover, ''client showing good decision making'', ''encouraging innovation'' and ''clearly presenting the goals of the project'' are all contribute to project success in the pre-production phase. Construction phase is affected most by ''client achieved good cooperation in project'' and ''giving clear message on time''. These findings align with Walker (1995) and Chan et al. (2004) as they argue that the client's ability to collaborate with the design- and construction and contribution are factors behind project success. Boyd and Chinyio (2008) explain the complexity in understanding the client requirements as the building exceeds being only technical structure.

The variable ''pre-design phase - investigation of plot'' has decreasing effect on cost variance if well performed. The client's ability to contribute to and manage the design stage is highlighted in the theory, (Chan et al.,2004). The performance in the design phase is represented several times by different variables. ''Design phase - compliance with time plan'', ''Design phase - good collaboration within project team'' and ''Design phase - good result''. This is in compliance with the findings made by Koch and Lundholm (2018), which show that problems with the design are sources behind disturbances in construction projects.

The result of the statistical analysis of the client survey show ''Project program - compliance with time plan'', ''project program - good result'' have correlation with the client satisfaction and will increase the client satisfaction if fulfilled.

Poor site management is one of the most common factors behind unsatisfactory performance in construction projects (Jha and Iyer, 2006; Sambasivan and Soon, 2007; Odeh and Battaineh, 2002). This is compatible with the results found in the statistical analysis in terms of good production on site and good collaboration on site aspects. It is also observed that the factor "compliance with time plan" increases satisfaction as well as it lowers cost and time. Satisfaction of a contractor is affected by "own company production people involved in choosing subcontractors and suppliers" and "own company support administrative work". It is generally known that site managers have a large load of administrative work that consumes time. When the company provides support for administrative work, site managers seem to work more effectively. Chan et al. (2004) suggest that one of the factors behind project success is the support and provision of resources that the project team leaders get from the own organization. This argument is aligned to the findings when the duration of contractors' own craftsmen is higher, the time variance is lower. The last aspect to be discussed about the contractor factors is "own company production people involved in choosing subcontractors and suppliers". This aspect has a positive influence on the site managers' satisfaction, as well as it can increase cost variance. Hartmann et al. (2009) finds that contractors tend to value the lowest price when choosing subcontractors and suppliers regardless of their previous trusted relationships. One explanation for the finding is that involvement of production team might

bring trust criteria back to decision making and value it over price.

It can be argued that the architect is one important actor in the project brief and the design phase. The regression analysis of the client survey, shows that a well performing architect in the design phase, contribute to lower cost variance as well to higher client satisfaction. Also, the results of the contractor survey show that the quality of the architect's documents lower time variance. But it must be noted that it might be not the skills what hinders the architect's performance. Boyd and Chinyio (2008), state that the design flaws are the result of hard competition caused by clients' focus on lower cost and therefore the architect provide incomplete drawings to keep the cost down. It is found in Table 2 that the performance of the ventilation consultant in the design phase is important for the client satisfaction. And the piping consultant documents is one of the influential factors behind lower cost variance. Thereafter, the consultant's performance is reflected in the quality of drawings, which can also be a source of disturbances at the construction site. In Table 2, the results show that the performance of the ventilation subcontractor is influential to contractor satisfaction. Ventilation subcontractor worked to minimize interruptions and result was as expected are such factors that are highly correlated with contractor satisfaction. Ventilation subcontractor using innovative production methods is one factor that is correlated to construction time variance.

Predicting project performance

The purpose of the prediction models developed with machine learning is to improve performance by increasing predictability of important KPIs. But, it is difficult to know exactly how the models can be deployed in the industry, especially if it is used as a consultancy tool for decision making. Also, if undesired project outcome is predicted and corrective actions are taken, the project outcome has the possibility to improve and therefore it will be hard to assess the predictability of the models. One way is to test the model on unseen data and use the data for new training. Cross validation is a valid method to develop the most generalized model (Witten et al., 2011). The lowest prediction error is the one associated with predicting contractor cost variance of construction (9.8%). It has been shown that predicting satisfaction is the hardest; RMSE 16.6%. Satisfaction is subjective, and each person have different criteria for being satisfied about a project. (Chan and Chan, 2004).

The outputs of the analysis, cost variance and time variance have high variation. For example, time variance ranges from +100% to -55.55%, This makes instances highly deviating from the average and thus

standard deviation (error) will be higher. As a result, the prediction is approximate and not absolute.

Finally, challenges in inference. Just because time or cost variance is smaller for some projects, it does not automatically mean that they perform better or have higher productivity than other projects. A low variance sometimes tells a project's ability to meet the planned cost and time or the result of inaccurate estimations of the project cost in the early phases. Some projects have low cost variance by nature such as design and build contracts. It can be that the cost in these projects, by nature, are more easily predicted. However, it does not mean that these projects perform better. However, it is generally believed that construction contractors do not particularly want to give unnecessary high price offers, since it is often competition during tendering. Therefore, low cost variance does not equal good project performance unless the original tender was low.

Conclusion

In conclusion, a statistical evidence for what factors that are important to consider is provided. The identification of the most correlating factors can be used to build prediction models. It is proved that, the obtained models in this report can provide predictions when testing the formulas on the originate dataset. It can be learned that machine learning concepts are suitable method for understanding project performance.

This paper has identified the most influential factors behind project performance in Swedish construction projects. External factors, such as weather conditions, are broadly mentioned in the theory as causes behind bad performance and project failures. In this paper, this type of factor is not found to be highly correlated with project performance. The technical aspects of a building are found to be of importance and highly suitable for predictions. It can be concluded that prefabrication saves time, but not necessarily cost. Extensive blasting work is one factor that is shown to be correlated to time overrun. Projects, in which contractor feel that the client has not planned the project well and that the time for preparing production is limited, are most likely to show unsatisfactory results.

Although external factors and technical aspects of a building are important for project success, the most recurring factors behind project performance can be derived to human related factors. It can be concluded that construction projects tend to be more expensive when traditional contracts are used and there are presumably many reasons behind this finding. The work behind this. A contract or criteria that facilitates collaboration and communication between parties will

per se contributes to fulfilling or solving several of the other factors behind project performance. Added to this, a common goal the parties can strive for, such as environmental requirements, makes performance likely to be satisfactory for the involved project team.

Factors that are recurrent in the statistical result of the contractor's survey are client's ability to plan the project, make decisions and communicate as well as to encourage innovation. It can be concluded that the client is a very important actor for project success, not least in the early stages of a project where the client should be clear about the project scope and to select a suitable type of contract for carrying out the project. Disturbances in the construction phase are often caused by design flaws. Architects and consultants are often blamed for this. However, it is the client's role to address what is wanted and to pay for the design.

Recommendations for future research

It is suggested to conduct research that focus on one particular aspect beside the more general studies. Because it was found that it is challenging to establish a full understanding of the single practice or interaction attribute unless it is fully defined or put into detailed lens of objective measures. Especially ''collaboration'', '', ''client decision making'' and ''satisfaction''. Moreover, approaches of machine learning in this research are helpful for quantifying soft aspects and making them more measurable. Focusing on one process such as design, project brief, site management, are subject for machine learning model's development. It is also possible to focus on a certain type of contract to study its features.

Contribution

The contribution to the construction research and practice is the illustration of machine learning as a method to study project performance. The causes of delays and cost overruns continue to be of interest in the construction industry when aiming to optimize and increase efficiency. The factors behind success or failure are elaborated in the literature can be more comprehensively understood with feature selection methods and while being used in prediction using machine learning methods (such as linear regression, in the case of the current paper). Contractor satisfaction is a new addition to the literature as it is not a traditional performance indicator and it has been used as an output for projects in this effort.

References

Alias, Z., Zawawi, E. M. A., Yusof, K., & Aris, N. M. (2014). Determining critical success factors of project management practice: A conceptual framework. *Procedia-Social and Behavioral Sciences, 153*, 61-69.

Al-Momani, A. H. (2000). Construction delay: a quantitative analysis. *International journal of project management, 18*(1), 51-59.

Alzahrani, J. I., & Emsley, M. W. (2013). The impact of contractors' attributes on construction project success: A post construction evaluation. *International Journal of Project Management, 31*(2), 313-322.

Attalla, M., & Hegazy, T. (2003). Predicting cost deviation in reconstruction projects: Artificial neural networks versus regression. *Journal of construction engineering and management, 129*(4), 405-411.

Boyd, D., & Chinyio, E. (2008). *Understanding the construction client*. John Wiley & Sons.

Chan, A. P., & Chan, A. P. (2004). Key performance indicators for measuring construction success. *Benchmarking: an international journal, 11*(2), 203-221.

Chan, A. P., Scott, D., & Chan, A. P. (2004). Factors affecting the success of a construction project. *Journal of construction engineering and management, 130*(1), 153-155.

Cheng, M. Y., Hoang, N. D., Roy, A. F., & Wu, Y. W. (2012). A novel time-depended evolutionary fuzzy SVM inference model for estimating construction project at completion. *Engineering Applications of Artificial Intelligence, 25*(4), 744-752.

Chi, S., Suk, S. J., Kang, Y., & Mulva, S. P. (2012). Development of a data mining-based analysis framework for multi-attribute construction project information. *Advanced Engineering Informatics, 26*(3), 574-581.

Cox, R. F., Issa, R. R., & Ahrens, D. (2003). Management's perception of key performance indicators for construction. *Journal of construction engineering and management, 129*(2), 142-151.

De Wit, A. (1988). Measurement of project success. *International journal of project management, 6*(3), 164-170.

Frank, E., Hall, M.A., & Witten, I.H. (2016). The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques", *Morgan Kaufmann*.

Freeman, M. & Beale, P. (1992). Measuring project success. *Project Management Journal, 23*(1), 8–17.

Frödell, M., Josephson, P. E., & Lindahl, G. (2008). Swedish construction clients' views on project success and measuring performance. *Journal of Engineering, Design and Technology, 6*(1), 21-32.

Forbes, L. H., & Ahmed, S. M. (2010). *Modern construction: lean project delivery and integrated practices*. Crc Press

Han, J., Pei, J., & Kamber, M. (2011). *Data mining: concepts and techniques*. Elsevier.

Hartmann, A., Ling, F. Y. Y., & Tan, J. S. (2009). Relative importance of subcontractor selection criteria: evidence from Singapore. *Journal of construction engineering and management, 135*(9), 826-832

Jha, K. N., & Iyer, K. C. (2006). Critical factors affecting quality performance in construction projects. *Total Quality Management and Business Excellence, 17*(9), 1155-1170.

Josephson, P. E. (2013). *Produktivitetsläget i svenskt byggande 2013, nybyggnad flerbostadshus och kontor*. Sveriges Byggindustrier.

Kim, G. H., An, S. H., & Kang, K. I. (2004). Comparison of construction cost estimating models based on regression analysis, neural networks and case-based reasoning. *Building and environment, 39*(10), 1235-1242.

Koch, C., Lundholm, M. (2018). *Produktivitetsläget i svenskt byggande 2014. Lokaler, Grupphus och Anläggning*. Sveriges Byggindustrier.

Lam, P. T., & Wong, F. W. (2009). Improving building project performance: how buildability benchmarking can help. *Construction Management and Economics, 27*(1), 41-52.

Lowe, D. J., Emsley, M. W., & Harding, A. (2006). Predicting construction cost using multiple regression techniques. *Journal of construction engineering and management, 132*(7), 750-758.

Lu, P., Chen, S., & Zheng, Y. (2012). Artificial intelligence in civil engineering. *Mathematical Problems in Engineering, 2012*.

Mir, F. A., & Pinnington, A. H. (2014). Exploring the value of project management: linking project management performance and project success. *International journal of project management, 32*(2), 202-217.

Naoum, S. G., & Egbu, C. (2016). Modern selection criteria for procurement methods in construction: A state-of-the-art literature review and a survey. *International Journal of Managing Projects in Business, 9*(2), 309-336.

Odeh, A. M., & Battaineh, H. T. (2002). Causes of construction delay: traditional contracts. *International journal of project management, 20*(1), 67-73.

Ogunlana, S. O. (2010). Beyond the 'iron triangle': Stakeholder perception of key performance indicators (KPIs) for large-scale public sector development projects. *International journal of project management, 28*(3), 228-236.

Sambasivan, M., & Soon, Y. W. (2007). Causes and effects of delays in Malaysian construction industry. *International Journal of project management, 25*(5), 517-526

Sanvido, V., Grobler, F., Parfitt, K., Guvenis, M., & Coyle, M. (1992). Critical success factors for construction projects. *Journal of construction engineering and management, 118*(1), 94-111.

Tangen, S. (2005). Demystifying productivity and performance. *International Journal of Productivity and performance management, 54*(1), 34-46.

Walker, D. H. (2016). Reflecting on 10 years of focus on innovation, organisational learning and knowledge management literature in a construction project management context. *Construction Innovation, 16*(2), 114-126.

Walker, D. H. (1995). An investigation into construction time performance. *Construction Management and Economics, 13*(3), 263-274.

Wauters, M., & Vanhoucke, M. (2014). Support vector machine regression for project control forecasting. *Automation in Construction, 47*, 92-106.

Witten, I. H., Frank, E., Hall, M. A., & Holmes, G. (2011). *Data mining: Practical machine learning tools and techniques with java implementations*. San Francisco: Elsevier Science & Technology

Xiong, B., Skitmore, M., Xia, B., Masrom, M. A., Ye, K., & Bridge, A. (2014). Examining the influence of participant performance factors on contractor satisfaction: A structural equation model. *International Journal of Project Management, 32*(3), 482-491.