Exploring Deep Learning for Estimating Construction Costs during the Early Stages of Design

Xiaohua Jin[1,*], Robert Osei-Kyei[1], Varin Kant[1]
[1]Western Sydney University
[*]email: Xiaohua.Jin@westernsydney.edu.au

Abstract

Deep learning has unequivocally demonstrated far superior capabilities in solving problems than any other machine learning technique and has thus far been showcased in a wide variety of applications such as; Autonomous Cars and AlphaGo. Deep learning relies on large data sets to train models using the neural network architecture with many hidden layers. Existing studies have demonstrated deep learning as an effective alternative to standard methods of early stage cost estimation for construction building projects. These undertakings have however been limited to a small sample size of data observations and often from a singular source, which leaves an important question unanswered regarding the ability of deep learning to generalise when exposed to a large sample size. This research aims to understand how incremental increases in data size impact the performance of deep learning in estimating costs during the early stages of design. The design of this research is to collect volumes of data from project records and model these with a deep learning architecture designed from researched key literature. This research employs a quantitative approach to analysing the data. The mean square error and r-coefficient of correlation are the key markers by which the performance of the test data is to be measured against. As a model is retrained in each instance under a different set of conditions, it can be difficult to determine when a global minimum has been reached by the model and as such the best of 5 attempts will be used as the marker for analysis for each data internal. The research is at the stage of data collection and this paper reports on the literature review and research design.

Keywords: Estimating, Design, Deep Learning, Big data, Construction

1. Introduction

Technological transformative leaps eventually infiltrate and meaningfully impact most industries in the modern commercial world (Venturi, 2014). In the AEC (Architecture, Engineering and Construction) Industry, leaps in technological advancements at pivotal points in time have impacted the cumbersome ways by which projects are delivered from inception through to competition. These transformative leaps disrupt stalwart processes in the AEC Industry and historically occur at a moment of convergence when improvements in technological capabilities, barriers to utilisation and real reductions to cost occur to such an extent that the pendulum swings relatively quickly in favour of any particular technological adoption (Wynne, 2012). The next transformative leap to shape the AEC industry is pointing towards several different technological vehicles, such as; artificial intelligence (AI), the internet of things (IoT) and robotics (Kelly, 2016). The aggregation and use of data are at the heart of each of these technological vehicles in which large data sets, and machine-learned algorithms converge to drive optimal outcomes and efficiencies on order of magnitude previously unseen (Kelly, 2016). Machine learning uses algorithms that can learn from data without relying on rules-based programming. Many different forms of machine learning exist, such as multiple linear regression, decision trees and neural networks, of which many different subsets and types exist for each form (Pyle and San Jose, 2015).

Deep learning (Neural networks with deep layers) has recently separated itself from the forms of machine learning that exist today and poses as the most attractive. This is due to deep learning's ability to distinctly understand non-linear dependencies and complex relationships at a human level. The underpinning which drives the performance of these deep neural networks are large amounts of data, which can be properly trained in order to mimic the complex relationships that exist in the real world (Pyle and San Jose, 2015). Large data sets in the AEC industry exist in many forms such as drawings, schedules, quotations, applications and models and can contain many types of information such as costs, engineering details, methodologies and schedules (Martinez-Rojas *et al.*, 2015). As such there are extensively many permutations and combinations of data points that can be used for a variety of unique machine learned applications. One specific application which naturally fits well with the contextual information described above is cost estimation of building construction as there is a heavy reliance on historical data collected when formulating a predicted cost value (Skitmore and Picken, 2000; Smith *et al.*, 2016).

The research into the viability of deep neural networks when purposed for early stage cost estimation has been carried out by academic's around the world and published in journals over the years (Sontag, 1998). It is clear that the research conducted into the viability of deep learning as an alternative for cost estimation in the building construction industry has been limited to comparatively small samples of data sets. It can be inferred that due to the sensitivity of the cost data and decentralisation of data locality, researchers have been physically constrained in the volume of data available. Deep learning inherently requires large amounts of data to optimise the performance of a model, in fact incremental increases of data added to a model should improve the overall performance of the model (Foody *et al.*, 1995). The size of the data based on the parameters and network design used commonly across the research studies have not been tested under a meaningfully large data set, which would by association cover a wider spectrum of projects and permutations. A deep neural network trained under the design principles specified in the studies with large amounts of data should be able to generalise the underlying relationships to such an extent that it can outperform the adhoc abilities of an experienced estimator in order for it to deemed as a viable alternative.

Therefore, the following research questions are raised and to be answered in this study. Firstly, how does incremental increases in data size impact the performance of deep learning in estimating costs during the early stages of design? And secondly, is deep learning a viable and practical alternative to traditional methods of estimating construction costs based on the current market requirements estimators? At the time of writing this paper, the research is at the stage of data collection and this paper reports on the literature review and research design.

2. Literature Review

2.1 Stages of Design

Buildings are intrinsically a complex product to design, produce and manage given the sheer scale of the tasks being undertaken, the bespoke nature by which the design is developed and the non-linear relationships that exist between work activities. The management of cost planning activities is by extension a difficult and complex activity during the development process of a project and is critical to achieving the overarching goals of any construction project of being; on budget, on programme and adhering to quality standards (Ostrowski, 2013). Seeley described approximate estimating as "*. . . techniques which attempted to give a forecast of the probable tender figure, although the basis of the computation often left much to be desired*" (Seeley, 1972). The latter end of this definition highlights the stigma around cost estimation which has carried forward nearly 50 years later, in that the computation behind approximate estimating is said to be somewhat of a "pseudo-science" or in other words a mix between the disciples of science and art (Sundaram, 2015). There have since been various nuanced definitions and understandings on what building construction cost estimation entails by corporate bodies, researchers, authors and industry leaders.

According to the Australian Institute of Quantity Surveyors (AIQS) (2017a), a project undergoes an evolution and transformation in succinct stages from an initial brief through to tender and then construction. A cost planner is required to reassess the construction cost budget at these pivotal stages in order to ensure that the intent of the brief is being met and to identify whether the design scope is increasing beyond the view formed during the very early initial stages of the preconstruction process or whether there has been a gross miscalculation in setting the budget at the brief or outline proposal stage. The critical challenges associated with cost estimation during the early stages of pre-construction stem from a lack thereof design documentation. The minimal design at this stage of the project is by intent, as it allows various stakeholders such as architects, engineers and project managers to make dynamic decisions that are comparatively cost-effective. Whilst empowering the project team, the cost planner is inhibited in formulating a robust view of the defining project characteristics and hence an accurate assessment of cost. The cost-estimate formulated by the cost planner during these early stages of pre-construction, feeds back into the decision-making process. A vicious cycle of misinformed decision making can begin, in which an inaccurate estimate is guiding the decision making of various stakeholders. The minimal design during the early stages of pre-construction, limits the cost planner to employing only a few traditional cost estimation methods based on an underlying principle of abstracting rates to measured quantities.

2.2 Cost Estimation Techniques and Expected Accuracy

The ***superficial method*** (or otherwise known as the floor area method) is a cost estimation technique that heeds to a simplistic arithmetic model. The nuances required in the application of such a rudimentary cost estimation technique lays in the interpretation of the variables, "Area" and "Rate". Gross Floor Area (GFA) has inadvertently provided the building construction industry with a consistent framework by which to benchmark project costs. Common alternate frameworks include Gross Building Area (GBA), Net Lettable Area (NLA) and Gross Lettable Area (GLA). On the other hand, the "Rate" variable is increasingly open to the cost planner's judgement in estimating a unit rate which reflects the actual cost. As a starting point, a cost planner traditionally filters and isolates historical projects based on a set of similar characteristics such as size, sector, building envelope, storeys, location and height, forming a basis for a $/m2 rate applicable for the given project to be estimated. At this stage, a cost-planner will have to draw upon a wide variety of internal and external sources as well as individual experiences in order to form a view on the adjustments required to the unit rate in order to reflect changes in factors such as project location, economic conditions, inflation, project design and risk (Smith *et al.*, 2016). The simplicity built-in to the technique gives the cost planner a high degree of flexibility and discretion which can leave room for individual bias, errors in judgement and cognitive dissonance, which edges the superficial method towards the non-scientific realm.

The ***functional area method*** has its foundations built-up on the principles of the superficial method but deviates by clustering spatial areas based on its functional use and applying specific $/m2 rates to each of the compartmentalised areas. The numerical variable represents any number of functional areas such as; residential, commercial, basement and lobby spaces. The clustering of areas by functional use, gear the cost planner to apply realistic $/m2 rates which can reflect the design intent of the project veraciously unlike the superficial method, which essentially is assessing a project as a single monolithic item. Whilst the functional area method is more accurate at estimating costs comparatively to the superficial method during the early stages of pre-construction, its challenges lay in the lack thereof a framework which is consistent, valid, reliable and reproduceable. The superficial method has an area framework which is singularly based and allows for easy translations between a wide variety of projects unlike the functional area method that is multiplicity based and can be difficult to translate between different spatial intents of projects. The effective utilisation of this technique requires a consistent framework by which projects are measured and rated (Smith *et al.*, 2016).

The margin of error demonstrated and expected in the AEC industry between estimated costs against the actual costs of a construction building project gives credence to the pass/fail nature of when an estimate is within an acceptable tolerance and when it is not. The performance of deep learning for early stage cost estimation against markers expected by most professionals gives an insight into the real-life use case for such a method. Cheung *et al.* (2008) examined the attitudes of clients and estimators towards estimating errors by way of a questionnaire survey in Hong Kong. In reviewing the acceptable errors held by clients for different sectors and market conditions, it is evident that based on any given set of conditions the acceptable error percentage could range from as low as 7.06% for overestimates in a good market for industrial projects to as high as 23.84% for overestimates in a bad market for school projects. In averaging the acceptable percentage tolerance across all the different permutations and combinations of conditions, clients averaged an acceptable tolerance error of 13.74% whilst estimators an error of 13.55%. These averages suggest that on average an error of 14% is viewed as perceptually, the acceptable tolerance between estimated and actual costs. It is important to acknowledge that this bias is not a reflection on actual discrepancies but rather on perception which may or may not be influenced by quantitative facts.

2.3 Deep Learning for Cost Estimation

Deep learning exists as a class of machine learning algorithms that exhibit features which include a cascading of multiple layers of nonlinear processing units for feature extraction and transformation, learning environments to be within supervised or unsupervised manners and the ability to abstract from levels within a hierarchy of concepts (Deng and Dong, 2014). At this time, the most popularised and documented form of deep learning to exist is in the form of deep neural networks, which in itself is an artificial neural network (ANN) but distinguished due to the number of hidden layers in the network, thus giving the artificial neural network, the distinction of being a deep neural network (Bengio, 2009). An artificial neural network is a machine learned system in which data dictates the framework of any given application instead of explicit logics scripted. The conceptual architecture of an ANN encapsulates a framework in which a number of inputs are feed forward into a neuron which maintains an activation function. Each neuron activates in the presence of certain observations and can generalise theoretically the class of inputs which form the path to a highly accurate prediction. An activation map is trained at each neuron as it is re-modelled, and a fully connected network exists between the inputs and neurons, through to the corresponding output. Activation functions which can be used at each of the neurons varies between Sigmoid, RELU, tanh and many other variants. A deep neural network layers neuron in hidden layers between the inputs and outputs and this is the distinction that gives a neural network its deepness (Olafenwa, 2018).

Kim and An (2004) compared a number of non-traditional cost estimation models for construction projects namely, case-based reasoning, neural networks and multiple regression analysis. A sample of 530 projects are used in this study with 40 projects out of the 530 used to test the performance of each of the non-traditional methods. Neural networks outperformed case-based reasoning which outperformed multiple linear regression, however the time to reach global minima posed as an issue as the varied results could not be extrapolated in definition. The input variables used for this study include, year of construction, gross floor area, storeys, total units, duration of construction, roof type, foundation type, usage of basement and finishing grades. On average 90% of the test values fell within 10% of the actual cost when modelled as a neural network that is optimised. The paper did not publish, disclose or further explain the rationale behind the number of trials required to achieve global minima in the neural network and hence an optimised result. A sample of 10% (or 40 projects) is on the lighter end but given the constraints of data it is within reason why more projects are used to train the model instead of testing its performance to further validate the model's framework. The variables used in this model are far reaching in terms of information that may be available to be considered during the early stages of design. There is question mark regarding the inputs concerning the construction duration, roof type, foundation type and finishing grades for which exact information is quite limited, as in some cases design is required to be further developed in order to flush out the exact details of each input (Smith *et al.*, 2016).

Chandanshive and Kambekar (2014) developed a neural network model based on MATLAB's neural net fitting tool, using the Levenberg-Marquadt algorithm. 58 projects are used in the modelling of construction costs based on 9 parameters, which include; site area, ground floor area, typical floor area, height of a building, quantity of shear wall, quantity of exterior wall, number of columns, foundation type, number of householders and total structural skeleton cost. The artificial neural network used for training, validation and training phase 70%, 15% and 15% respectively. The model is trained for between 1 and 10 hidden layers, demonstrating progressively higher overall regression's as the number of hidden layers is increased up to 10. The study did not address the accuracy of the model in terms of the practical implications in the industry and only viewed the results from the lens of statistical analyses such as mean square error and regression values. The research paper did not delve into the performance of the model from a standpoint that is digestible for analysis that could draw back to real life implications or comparatives. Using 58 projects for a study of this nature, is extremely small as well, the sample size does not dictate the use of artificial neural networks, especially with 10 hidden layers, as the framework is exposed to being over-fitted due to the small number of observations.

Arafa and Alqedra (2011) conducted a study into the use of artificial neural networks for early stage cost estimation of build projects in Gaza, Palestine. This study followed a very similar approach to that of which has been described in the two studies above except that 71 projects are along with a different set of parameters. These 7 variables include; ground floor area, typical floor area, number of storeys, number of columns, type of footing, number of elevators and number of rooms. The model had one hidden layer and tested 12 projects against its trained model. The correlation of coefficient is 0.97 demonstrating a positive correlation between actual and predicted values, however this did not address the performance of the model from a standpoint that is digestible for analysis that could draw back to real life implications or comparatives. In a sensitivity measure conducted, they found that no. of stories, ground floor and no, of elevators are the most influential in dictating the output of the mode. Using 71 projects for a study of this nature, is extremely small as well, the sample size does not dictate the use of artificial neural networks, especially with 10 hidden layers, as the framework is exposed to being over-fitted due to the small number of observations.

3. Research Methodology

The methodology employed in this study is tailored to meet its design requirements, which mainly requires the size of the data to be abnormally large in comparison to studies carried out of a similar nature. This abnormality requires a unique approach in that the data are collected from a secondary source, from which many options are examined. As a result, a large portion of the findings and analysis are limited by the scope of the data collected and not necessarily by the intent, as would be suffice when primary sources for data are utilised for collection(Oluwatosin Ajayi, 2017). The study employs a quantitative approach to discerning results and analyses for discussion(Creswell, 2013).

3.1 Design Requirements

The design requirements of the data that need to be collected for this study can be divided into two parts. The first part is concerned with the number of data observations that needed to be collected. As the basis of this study is concerned with the performance of deep learning for cost estimation at scale, it is critical that the choice of sourcing is able to account for the large number of data observations required. The second part is concerned with the variables that are to be available for choice when modelling the neural network. As mentioned in the literature review, there are many inputs which need to be considered from previously published research papers. As studies previously carried out of a similar nature had demonstrated, it is critical that inputs that are at the project level are available to be modelled. The number of data observations used in previous studies ranged from 34-530 and as such it is determined that at least double the amount of data observations are required in order for this study to meaningfully research the impact of a larger data set when used for this application. The variables (or inputs) required for the design would ideally have been as many as possible, given the structure of a neural network is to disseminate the complex correlations that may or may not exist between variables and to essentially remove the human dissemination of ideologue and thus allowing the model to make its own determine on these relationships (Aamodt and Plaza, 1994). As such, there exist a mutual number of inputs that would be required for this study. These variables include floor area, storeys and functional use for the inputs and the value of the project as the output. Any additional inputs are to be treated as secondary but still important to achieving the level of accuracy required. A dataset of a size greater than 1000 observations and that contains data about floor area, storeys and functional use of projects at a minimum form the minimum design requirement needed to be achieved for this study. If several different sources and collection methods meet this criterion, the selection criteria will then be based on a combination of factors including number of data observations, data quality, reliability and number of secondary variables.

3.2 Data Collection

The design requirements of the data to be collected immediately nullify the option to source data from primary sources. Primary sources would include methods such as interviews, questionaries or direct data from the source such as a government entity voluntarily disclosing the financials of a project they had undertaken (Oluwatosin Ajayi, 2017). As the fundamental requirements of the study require a significant amount of data observations, primary sources which would pose a challenge from a practical standpoint as that the collection of mass amounts of data would require the collaboration of many participants. The decentralised market forces that exist in the AEC industry means that many companies only hold a small portion of information of the total construction market and thus gathering at least 1000 data observations would require the participation of too many organisations (Flanagan *et al.*, 2007). Additionally, the type of data requested, in this case the costings of a project is a very sensitive form of data and from a confidentiality point of view, it would pose as an additional challenge if such information is to be sourced directly from stakeholders (Smith *et al.*, 2016). The same principles described can be viewed from the lens of other stakeholders such as clients or quantity surveyors, in both these cases there exist varying levels of decentralisation and sensitivities that would generally be too profound to be selected as a source for data collection.

It is due to the constraints described above, that it is determined a secondary source is needed which can not only meet the design objectives but reasonably be objectified as being reliable, valid and accurate (Oluwatosin Ajayi, 2017). Two types of secondary sources are considered for this study, the first being that which exists in public records and the second being that which is published by the private sector. Public records seemed to be the logical path for this study as the records can reasonably be determined as being reliable, valid and accurate due to the regulatory and legislative nature of the records. The secondary sources of public records considered are that which are published by the Australian Bureau of Statistics (ABS), Data61 and Local Government Councils. The ABS publish a wide range of quantitative figures related to the building industry covering a spectrum of categories and formats which determine key economic indicators. At the project level however, the ABS due to its sensitiveness does not disclose project details and thus could not be used for this study (2017b). Data61 is a data initiative by the government which encourages the use of government data by opening up records held by the government for the wide benefit of the community. However, Data61 does not have the required data published and needs a request to be lodged in order to be disclosed. As there is no guarantee on if this information would be disclosed, again due to the issue of sensitivity and as well the amount of time needed to process a request, this option is also rendered null. The last option looked at using process automation to extract information from development applications lodged with councils. Each local council generally has a development application portal available for the public to access as public development notifications. Although the information varies from council to council in the way that it is presented and what is available to view, generally one could access some project details such as project location and cost, but any further information would require manual data entry from associated documents and drawings. This option is feasible if there is enough manpower or automation to classify the required information from council websites and index data accordingly.

Secondary sources from private entities seem to be the next logical source for secondary data that fit the design requirements of this study. In exploring sources from the private sector, it is determined that the viable options include CoreLogic's Cordell Connect and Macromonitor’s Key Project List. Although being project specific in terms of identifying the project, its cost and when it would be constructed, Macromonitor’s Key Project List does not provide any further details that are required as per the design criteria. On the other hand, Core Logic’s Cordell Connect is an online platform that published projects in Australia with a lot of information. This information includes many project specific details along with ancillary information such as consultants and contact numbers. Cordell Connect has a large database of projects that are researched and indexed by a team of 60 in Australia. These projects have their information verified by council minutes taken at meetings and by industry personnel. This poses as the most suitable option for this study as it can be reasonably extrapolated that the data is reliable, valid and accurate based on the through research and processes being executed at Core Logic. As the data exist online in a platform it is critical that the data stored within each of the projects record can be extracted in a manner that is usable for MATLAB to model. In this case the data is required to be extracted into a .csv file that follows the required standard principles of data design in which each row is a data observation (a single project) and each of the rows represents an input (or variable) (Wickham and Grolemund, 2017). Core Logic has data exported in a certain way to a .csv file with a maximum capacity of 500 projects to be exported at a time. As such a R-Script is compiled which can transform and merge the data from these raw files into a single .csv file, ready to be modelled.

3.3 Modelling and Analysis

The neural network modelling of data will be carried out in MATLAB, a popular programming language used for a variety of purposes particularly for science and engineering. Its use for deep learning is particularly popular and powerful due to the development of proprietary tool boxes which minimise the technical programming skills required to model data for deep learning. The tool box used for this study is the neural net fitting app, which is designed to map a data set of numeric inputs against a set of numeric targets. In this case the variables are required to be numeric and the data which exist as factors (i.e. categories) are converted to individual variables with 0 representing no and 1 represent yes. The study uses the very fundamental selection of parameters that can be amended for such a use case. In this case the Levenberg-Marquardt backpropagation algorithm is used along with a threshold of 70% of data to be used to train the model and 15% to validate the model to prevent the model from over fitting. The remaining 15% of the data is used to test the trained model. The splits can be amended and adjusted as required but the split used for this study is to be kept as a constant variable using the default values in MATLAB’s tool box. The number of layers used for this model is 10 with sigmoid hidden neurons. The variable which is being assessed for this study is the number of data observations to be used for the training of the neural network model. For the purposes of this study the number of data observations will begin by training a model at 100 observations and double each time over as a new interval is introduced. This means that the next set of data observations to be assessed is 200 and then 400 and so on. The data observations which will be used to train the model and then be tested are randomly assigned and kept consistent over 5 different trials with the best trial used as the basis for assessment. The reason the best trial is used as neural networks aim to achieve a global minima and it is very common for training of a data set to begin down a particular path and achieve only local minima(Birui, 1989).

The neural net fitting tool outputs the results as a regression plot between the targeted values and the values outputted by the trained model. The strength of the correlation is then assessed by the coefficient of correlation of the line of regression. This value is between -1 and 1 with a value of -1 meaning that the two values pitted on the x and y axis are moving perfectly not in unison whilst a value of 1 means that the two values pitted on the x and y axis are moving perfectly in unison. A value of 1 in this study would indicate that the model has perfectly generalised and modelled the complex relationships that exist between the inputs of the deep neural network and is able to predict exactly the value of a construction project based on the inputs used to train the model. The coefficient of correlation is one of the key markers used in this study to compare the performance of the deep neural network at several different data set size intervals. The second marker used to assess the performance of the model is the RMSE, which is the absolute value of the mean difference between the targets and outputs. This is a holistic number that is comparable to the values of the projects used to train the model and given an insight into the average bias between targets and outputs. The third marker used to assess the performance of the model is the “E-Value”, which is a value to give insight into the practicality of the model in the real world by assessing its performance against the expected performance of an estimator by industry standards. The bias expected to be achieved by estimators is between 10% and 14%, when considering actual and perceptual thresholds of acceptability. As a result, the E-Value is to be analysed against projects which predict values within 10% of the actual value and is to be referred to as E-10 for this study, whilst projects which predict values within 20% of the actual value known as E-20. The results of each of these markers give insight to a unique set of analyses.

4. Summary

This study aims to extend upon previous research focusing on the utilisation of deep learning as a non-traditional method in estimating the construction costs for construction projects during the early stages of design when project information is limited. The research methodology used for this study was formed around two central design requirements of the data set, the first being the need for the number of data observations to be at least doubled the number of projects than previous similar studies. As such an arbitrary number of 1000 of projects was deemed as the minimum threshold of projects required to be modelled for this study, which focused on the scalability of deep learning for estimating costs during the early stages of design. The second design requirement of the data was concerned with the variables that would be available to be modelled in the neural network. In assessing similar studies, a mutual pool of project characteristics is evident as the most sensitive to cost. These parameters included floor area, number of stories and functional use. Secondary inputs or variables would be critical in disseminating and modelling the complex interplay between each of these variables but could be given some latitude in terms of the make-up of each of these parameters. Floor area, storeys and functional use were essential details of a project required to be modelled for the purposes of this study.

The data source selected which could meet the design objectives best were from CoreLogic’s Cordell Connect, a secondary private source that could reasonably be considered to be accurate, reliable and valid. The variables of the model including the algorithm used, randomised data sample and validation/testing threshold are kept constant in order to test the variable of data size on the performance of the neural network. Data intervals of 100, 200, 400, 800 and 1600 are the intervals measured, from a best of 5 trials. From a practicality point of view, the model’s performance is pegged against a percentage of projects which are able to meet the industry thresholds of 10% and 20%.

This paper has demonstrated the challenges that are present in utilising deep learning for estimating construction costs at scale and has presented the gaps that exist in current research as it relates to the application of this method from a practicality point of view in that the lack of project information makes it difficult for deep learning to generalise and model the complex behaviour of a wide spectrum of projects. Innovative approaches which aim to tackle this problem with an added caveat or a different approach may hold the key to successfully deploying deep learning for estimation of construction costs during the early stages of design at scale.

References

Aamodt, A. & Plaza, E., 1994. Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches. AI Communications, 7(1), pp. 39-59.

Aibinu Ayodeji, A. & Pasco, T., 2008. The accuracy of pre-tender building cost estimates. Construction Management and Economics, 26(12), pp. 1257-1269.

AIQS, 2017. 2016-2017 Annual Report, Sydney: s.n.

Arafa, M. & Alqedra, M. A., 2011. Early Stage Cost Estimation of Buildings Construction Projects using Artifical Neural Networks. Journal of Artifical Intelligence, 4(1), pp. 63-75.

Australian Bureau of Statistics, 2017. Privacy, Confidentiality and Security. [Online] Available at: http://www.abs.gov.au/websitedbs/censushome.nsf/home/privacy [Accessed 12 June 2018].

Bell, E. & Bryman, A., 2015. Business Research Methods. 4th ed. Oxford: Oxford University Press.

Bengio, Y., 2009. Foundations and Trends in Machine Learning. 1st ed. Ishington: s.n.

Birui, B., 1989. A new approach for finding the global minimum of error function of neural networks. Neural Networks, 2(5), pp. 367-373.

Chandanshive, V. B. & Kambekar, A. R., 2014. Prediction of Early Stage Construction Cost of Building Projects Using Artifical Neural Netork. International Journal of Scientific & Engineering Research, 5(7), pp. 453-463.

Cheng, Y.-M., 2014. An Exploration into Cost-Influencing Factors on Construction Projects. International Journal of Project Management, 32(1), pp. 850-860.

Creswell, J. W., 2013. Research Design: Qualitative, Quantaitative, and Mixed Method Approaches. 4th ed. s.l.:SAGE Publications Inc.

Deng, L. & Dong, Y., 2014. Deep Learning: Methods and Applications. 1st ed. Ishington: s.n.

Flanagan, R., Lu, W., Shen, L. & Jewell, C., 2007. Competitiveness in construction: a critical review of research. Construction Management and Economics, 25(9), pp. 989-1000.

Foody, G., Mcculloch, M. & Yates, W., 1995. The effect of training set size and composition on artificial neural network classification. International Journal of Remote Sensing, 16(9), pp. 1707-1723.

Franco K. T. Cheung, Margie W. L. Wong & Martin Skitmore (2008) A study of clients' and estimators' tolerance towards estimating errors, Construction Management and Economics, 26:4, 349-362

Genpact Research Institute, 2014. The Impact of Technology on Business Process Operations. s.l., Genpact.

Gunner, J. & Skitmore, M., 1999. Comparative Analysis of Pre-bid Forecasting of Building Prices based on Singapore Data. Construction Management and Economics, 17(5), pp. 635-646.

Jin, R., Han, S., Hyun, C. & Kim, J., 2014. Improving Accuracy of Early Stage Cost Estimation by Revising Categoical Variable in a Case-Based Reasoning Model. Journal of Construction Engineering and Management, 140(7).

K. Y. Cheung, F., W. L. Wong, M. & Skitmore, M., 2010. A study of clients' and estimators' tolerance. Construction Management and Economics, 26(4), pp. 349-362.

Kelly, K., 2016. The Inevitable : Understanding the 12 Technological Forces That Will Shape Our Future. 1st ed. New York: Random House USA.

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

Kunce, J. & Chatterjee, S., 2017. A Machine-Learning Approach to Parameter Estimation, Virgina: CAS .

Martinez-Rojas, M., Marin, N. & Vila, A., 2015. The Role of Information Technologies to Address Data Handling in Construction Project Management. Journal of Computing in Civil Engineering 3, 30(4).

MATLAB, n.d. Fit Data with a Shallow Neural Network. [Online] Available at: https://au.mathworks.com/help/deeplearning/gs/fit-data-with-a-neural-network.html

Newman, I. & Benz, C. R., 1998. Qualitative-Quantitative Research Methodology. 1st ed. s.l.:SIU Press.

Olafenwa, J., 2018. Introduction to Neural Networks, s.l.: Medium.

Oluwatosin Ajayi, V., 2017. Primary Sources of Data and Secondary Sources of Data, Makurdi: s.n.

Ostrowski, S. D., 2013. Estimating and Cost Planning Using the New Rules of Measurement. 1st ed. Somerset: Wiley.

Pacey, A., 2007. Medieval Architectural Drawing. s.l.:Tempus Publishing.

Pasco, T. & Aibinu, A. A., 2008. The Accuracy of Pre-Tender Building Cost Estimates in Australia. Construction Management and Economics, 26(12), pp. 1257-1269.

Pyle, D. & San Jose, C., 2015. An executive's guide to machine learning. McKinsey Quartely, September.

Seeley, I., 1972. Building Economics: Appraisal and Control of Building Design Cost and Efficiency. 1st ed. London: Macmillan Educaktion.

Skitmore, M. & Picken, D. H., 2000. The Accuracy of Pre-tender Building Price Forecasts: An Analysis of USA Data. Australian Institute of Quantity Surveyors Referred Journal , 4(1), pp. 33-39.

Smith, J. et al., 2016. Building Cost Planning for Design Team. 3rd ed. London: Taylor and Francis Group.

Sontag, E. D., 1998. VC Dimension of Neural Networks. Neural Networks and Machine Learning, pp. 69-95.

Sundaram, R., 2015. Cost Estimating: Art, Science or Both?. 20 08.

Venturi, A., 2014. Impact of Technology on Mangement and Organisations. 08 12.

Wickham, H. & Grolemund, G., 2017. R for Data Scuebce. 1st ed. s.l.:O'Reily.

Wynne, A., 2012. CAD : The Greatest Advance in Construction History. 05 09.