Future Directions of Cost and Productivity Estimating using Artificial Intelligence (AI) [1]

Danilo Arba

ABSTRACT

With an ever-increasing competitive world of engineering services, and with always thinner profit margins and decreasing market shares, the cost of a project is one of the significant criteria in decision making at the early stages of the construction industry.

To remain competitive in the market, companies must have an accurate estimate of their projects. With the rise of computing power, there is now a tendency to use Machine Learning (ML)-based methods, such as Artificial Neural Networks (ANNs), [2]for more accurate cost estimation that can remain reliable in the face of insufficient details during the tendering phase.

This technical paper will review an artificial neural network approach to the cost estimate of engineering services and construction activities. While developing the mentioned model, firstly, the influential factors that affect the costs of construction and engineering services are identified, after this, a model developed using data of multiple projects from the author experience.

Keywords: Artificial Intelligence, Artificial Neural Network, Construction Industry, Construction Management, Construction projects, Construction Estimating, Project cost modelling, GAPPS, CIFTER.

INTRODUCTION

While there are no written records from the Pyramids, credible research done by Damian Zimmerman[3] back in 1997 on the Great Wall of China explained:

"The fact is, that the cost of the wall's construction bankrupted Dynasty after Dynasty, and we should also remember that Herodotus[4], while not giving specific cost information, stated that it took 100,000 men around 20 to 23 years to construct the Great Pyramid of Khufu[5] .

Khufu or also known by his Greek name, Cheops, was the Egyptian pharaoh in the Fourth Dynasty, who was described by Herodotus as a cruel and strange figure, that prostituted his

---

[1] How to cite this paper: Arba, D. (2020). Future Directions of Cost and Productivity Estimating using Artificial Intelligence (AI), *PM World Journal*, Vol. X, Issue VIII, August.
[2] What are artificial neural networks? (n.d.). Bernard Marr. https://bernardmarr.com/default.asp?contentID=2126
[3] Great Wall of China. (n.d.). Mandala Projects. https://mandalaprojects.com/ice/ice-cases/wall.htm
[4] Herodotus. (n.d.). Encyclopedia Britannica. https://www.britannica.com/biography/Herodotus-Greek-historian
[5] Egyptian art and architecture - The pyramid of Khufu. (n.d.). Encyclopedia Britannica. https://www.britannica.com/art/Egyptian-art/Pyramid-of-Khufu

daughter as he runs out of money although the Westcar Papyrus[6] describes Khufu as a traditional oriental monarch: good-natured, amiable to his inferiors and interested in human existence and magic. Despite not being remembered as fondly as his father, the funerary cult of Khufu was still followed in the 26th Dynasty, and he became increasingly popular during the Roman period.

Figure 01: The Westcar Papyrus and the Miracle Stories of the Old Kingdom [7]

With the construction industry embarks on the much-touted journey to embrace leading-edge technologies like blockchain, digital twins, 3D printing, drones and laser scanning, it cannot lose sight of the fundamental responsibilities of meeting the critical project metrics of time and cost.

For almost seven decades the Project Management Triangle—also known as the Triple Constraint or Iron Triangle[8]—has been the base of a rubric for measuring project management success. Over the years, project teams have increased other constraints such as risk, safety, sustainability, to this list.

Considering the always increasing scrutiny from the public, project cost management comes to the focal point. While we consider this, how can the industry address the issue of cost overruns? Can we use technology to improve estimating, budgeting, managing, and controlling project

[6] The Westcar papyrus. (2017, April 5). Ancient History Encyclopedia. https://www.ancient.eu/The_Westcar_Papyrus/
[7] The Westcar papyrus and the miracle stories of the old kingdom. (2016, October 28). Ancient Origins | Reconstructing the story of humanity's past. https://www.ancient-origins.net/artifacts-ancient-writings/westcar-papyrus-and-miracle-stories-old-kingdom-006895
[8] Project management triangle | Overview of triple constraints. (n.d.). StarAgile Consulting | Certification Training Courses. https://staragile.com/blog/project-management-triangle

costs to reduce cost overruns? More specifically can artificial intelligence (AI)[9] improve the estimation of construction projects costs? This technical paper articulates an answer to these critical questions.

What is a construction estimate? The dictionary meaning of a cost estimate is to "An approximation of the probable cost of a product, program, or project computed based on available information."[10]. Considering a construction project, an estimate is the approximate cost of resources required to complete the project. Early construction estimates produced to determine the viability or affordability of the project whereas more accurate and detailed estimates developed to procure construction services. However, it has to be noticed the lack of standardisation in the levels of cost estimates.

Whit the project progressing the accuracy of the estimate increases with more information becoming available. The steps used to develop an estimate are similar between different standards, such as NASA, the US Accountability Office, the Guild of Project controls and AACE International[11].

Why are estimates not accurate? Experts mainly blame imprecisions on technical, psychological, and political-economic factors. Experts can mainly say that experts feel that psychological and political-economic factors play a dominant role and result in optimism bias and 'strategic misrepresentation'.

The author believes that data, data analytics, and artificial intelligence fortify the view from outside to overcome the optimism bias and the strategic misrepresentation in the construction estimates.

But how can artificial intelligence help cost estimators? Data analytics supported by Artificial Intelligence techniques can help the organisations develop meaningful insights industry-wide historical data.

---

[9] What is artificial intelligence (AI)? (2020, June 3). https://www.ibm.com/cloud/learn/what-is-artificial-intelligence
[10] What is the cost estimate? definition and meaning. (n.d.). BusinessDictionary.com. https://www.businessdictionary.com/definition/cost-estimate.html
[11] W20.1_Budgeting continuous improvements and process estimating. (2020, September 26). Danilo Arba Blog. https://www.daniloarba.com/post/copy-of-w20-1_budgeting-continuous-improvements-and-process-estimating

Scope Definition	Guild of Project Controls	AACE Current	ANZI Z94.0 (PMI)	AACE (Pre-1972)	Association of Cost Engineers (AcostE) UK	Norwegian Project Management Association	American Society of Professional Estimators (ASPE)
Less	Level 1	Class 5	Order of Magnitude ('-30% to + 50%)	Order of Magnitude	Order of Magnitude Estimate Class IV ('-30% to + 30%	Concession Estimate Exploration Estimate Feasibility Estimate	Level 1
	Level 2	Class 4	Budget Estimate ('-15% to +30%)	Study Estimate	Study Estimate Class III ('-20% to + 20%	Authorizaton Estimate	Level 2
	Level 3	Class 3		Preliminary Estimate	Budget Estimate Class II ('-10% to + 10%	Master Control Estimate	Level 3
More	Level 4	Class 2	Definitive Estimate ('-5% to +15%)	Definitive Estimate	Definitive Estimate Class I ('-5% to + 5%	Current Control Estimate	Level 4 Level 5
	Level 5	Class 1		Detailed Estimate			Level 6

Figure 02: Illustrating the lack of standardisation in Levels of Cost Estimates[12]

In general terms, AI can help develop company-specific and, even better, industry-wide benchmarks that can reduce optimism bias and strategic misrepresentation.

We classify and categorise unclassified cost data for organisations with a huge amount of granular cost data that normally is unclassified and unstructured, rendering it unusable. The data must be classified using a high-level cost classification system to utilise this type of historical data.

Unclassified cost data, in electronic formats, is normally available. AI tools that perform natural language processing[13] might be used to classify this type of data into a predefined cost classification system. A natural language classifier can, as a first step, be trained and tested with the use of historical cost data. Once we have trained the system, it can process vast volumes, form historical projects, cost data, to classify it into cost categories and cost groups.

Artificial neural networks (ANN)[14]inspired by the human nervous system, are a form of artificial intelligence, more precisely the neurons. This is for sure one of the main tools used in machine learning.

[12] What is a Cost Estimate? Guild of project controls compendium and reference (Car). (n.d.). Planning Planet | dedicated to Project Controls. https://www.planningplanet.com/guild/gpccar/introduction-to-managing-cost-estimating-budgeting
[13] What is natural language processing? (2020, July 2). IBM - United States. https://www.ibm.com/cloud/learn/natural-language-processing
[14] Seker, E. (2020, September 16). Artificial neural networks (ANN). Medium. https://towardsdatascience.com/artificial-neural-networks-ann-21637869b306

Once we have established a supervised learning and testing process, we can use it to predict the cost of our new building using the trained ANN.

It is vital for analogous estimating to identify a similar historical project that can estimate a new project's cost. (CBR) Case-based reasoning can be used to enhance to identify and recalling the comparator project.[15] , that is a different artificial intelligence form.

In CBR, we can use historical database of projects, to allow to bring back a project that is similar by project attributes using a metric called similarity index[16]. The information needed to estimate the project and knowledge from the retrieved project is 'reused' to arrive at a preliminary estimate, further improved via the 'revise' process, leading to the cost estimate's confirmed data. At this point, the new project itself is 'retained' in the historical project database as a learned case or project for future use.

In conclusion, the author envisions using historical cost data and analytics powered by AI can improve the estimating process significantly.

In new projects, these tools can inform in a very early stage, cost advice and guarantee accurate and precise cost estimates, help monitor and control project costs, and assist in forensic cost analysis.

For industry-wide benchmarks and related solutions, clients, in particular, public clients must lead the way.

As detailed before our scope of works to ensure the project's achievement for constructing two hospitals for our case study. We have available historical data from previous projects.

So how can we leverage the use of artificial intelligence to:

1. Classify and categorise unclassified cost data from previous projects.
2. Develop a statistical model for parametric estimation.
3. Identify reference class projects.

METHODOLOGY

STEP 1

In the last 20 years, statistics made it possible to develop predictive algorithms that are much more efficient, especially inaccuracy. In the field of cost estimation, which are the possible applications? While traditional analytical models based on the product or service's manufacturing processes still are primarily used in our society, statistical models are gradually

[15] What is case-based reasoning (CBR)? Definition from WhatIs.com. (2019, August 16). SearchEnterpriseAI. https://searchenterpriseai.techtarget.com/definition/case-based-reasoning-CBR
[16] Similarity and metrics in case-based reasoning. (n.d.). ResearchGate. https://www.researchgate.net/publication/27827429_Similarity_and_metrics_in_case-based_reasoning

imposing themselves for their formidable efficiency. Nevertheless, rather than an opposition, these two methods are enriched and complement each other for better cost modelling.

As a reminder, there are now three main methods used to estimate the cost of a product:

- *The analogical method*[17]: with this method we estimate, for a new product, the cost compared to similar products produced or purchased previously. The reliability of this method is not very high but can be used in extremely upstream phases (the study of opportunity) when the project's characteristics or the service are unknown.
- *The analytical method*[18]*:* by modelling the cost of a product in the process of industrial production.
- This method is based on the cost structure of the product of which it estimates each intermediate element, based on the materials & components involved, process costs (machine and labour), and related structural costs. By using this method, we estimate the cost of a product or service by statistical modelling. This method uses similar product or service histories to define equations or statistical laws that model the cost evolution according to specific parameters known as "cost drivers".
- *The parametric method*[19]*:* Whit this method estimates the cost of a product or service by statistical modelling. Using this method with similar product or service histories to define equations or statistical laws that models the cost evolution according to specific parameters known as "cost drivers".

These models are typically based on linear, multilinear, polynomial, or logarithmic regressions. These estimation methods have several advantages:

- The company characteristics make it possible to estimate the cost of a new product/service (weight, size, volumes, country of production, critical elements of the specification) without necessarily knowing the manufacturing process's details or external benchmarks. It is, therefore, a rapid and straightforward implementation method.
- On the other hand, basing our analysis on the observation of products/services manufactured or purchased in the past, the estimated cost is most probably more consistent and precise than a "theoretical" analytical model. This is sure of the quality of the history quality.
- These statistical methods are beneficial in the early phases of the life cycle (opportunity, feasibility, detailed design) because they make it possible to quickly make the right decisions for an optimised design and accelerate the time to market

[17] (PDF) Effort estimation using an analogy. (1996, April 25). ResearchGate. https://www.researchgate.net/publication/3625950_Effort_estimation_using_analogy
[18] What is analytical estimating - definition of analytical estimating by business training. (n.d.). https://www.businesstrainingcollege.com/business/Analytical-Estimating.htm
[19] Parametric estimating. (2008, January 15). Project Management Knowledge - Simply explained by a PMI-certified Project Manager. https://project-management-knowledge.com/definitions/p/parametric-estimating/

while still securing the margin.

- Consequently, they also make it possible to quickly analyse the consistency or the inconsistencies in the current prices, thanks to the predictive model's dispersion analyses. Thus, they reveal irrelevant products or services, at an abnormally high cost, for example, regarding the predictive model. This optimises leads for buyers (renegotiation, change of supplier) or R & D (redesign).

Also, we need to consider that these methods have several limitations:

- Traditional statistical models (based on regressions) hardly consider the qualitative parameters (except to reduce the database size).
- They do not manage the missing data properly and therefore, require very clean databases, which are not common in the construction industry.
- The price can have a linear, for example, behaviour over a specific range, then a radically different behaviour from a specific threshold, this since the manufacturing process may vary.
- All these elements directly affect the accuracy of traditional parametric models and therefore their use.

AI opens the way for a fourth model of cost modelling. In recent years, the advances made in algorithmic and machine learning mostly solve traditional parametric methods' disadvantages and improve their performance and application field.

So, what are the potential advantages if any? The principal advantages of this artificial intelligence algorithm are:

1. Ability to model a vast number of parameters ("cost drivers") and mainly qualitative or "symbolic" parameters.
2. Ability to process databases where the number of variables vastly exceeds the number of observations
3. Ability to identify and weight the essential parameters automatically, and thus the "cost drivers" that impact most the cost of the product
4. Ability to manage missing values / incomplete databases
5. Robustness to outliers
6. Ability to identify behavioural breaks in variables
7. Interpretation of the tree
8. Precision increased by 30 to 40% compared to traditional methods.

The application of random forests in cost estimation solves many of the disadvantages of traditional parametric approaches and opens new opportunities for companies interested in efficiency and competitiveness. A precise estimate of costs is now possible, even with a limited number of observations (a few dozen), limiting the resources used to collect and capitalise the data. On the other hand, complex systems' price can be modelled from easily accessible functional cost drivers, making encryption particularly simple and fast.

For this reason, random forests have begun to be used by some companies in the early phases of the product life cycle, including:

- Gain productivity on their encryption activities (saving time and resources that they can focus on technological innovation figures)
- Respond more quickly to their clients' tenders and primarily use this time saving to optimise their proposal better
- Secure and optimise their margin on new business

The second step was to use these algorithms to perform consistency or price inconsistency analyses by identifying products with large discrepancies between the actual price and the estimated price. Finally, once the model is perfectly calibrated, it becomes a cost control tool to validate the supplier's fair price. This reduces the bargaining process. Automated cost estimation aims to identify the correlations between the influential factors and the project cost using predictive models or algorithms.

We can identify four different types of AI-based approaches:

- namely machine-learning (ML),
- knowledge-based systems (KBS), [20]
- evolutionary systems (ES),[21]
- hybrid systems (HS).[22]

Researchers analysed three costs estimating model using:

- Artificial neural networks (ANNs), [23]
- Multiple regression analysis (MRA),[24]
- Case-based reasoning system (CBR)[25]

---

[20] Knowledge-based systems. (n.d.). https://www.journals.elsevier.com/knowledge-based-systems#:~:text=Knowledge-based%20Systems%20is%20an%20international%20and%20interdisciplinary%20journal,research%20in%20knowledge-based%20and%20other%20artificial%20intelligence%20techniques-based
[21] Evolutionary system. (n.d.). * Home. https://en.mimi.hu/artificial_intelligence/evolutionary_system.html
[22] Hybrid artificial intelligence system for the design of highly-automated production systems. (n.d.). ScienceDirect.com | Science, health and medical journals, full-text articles and books. https://www.sciencedirect.com/science/article/pii/S2351978918313684
[23] What are artificial neural networks? (n.d.). Bernard Marr. https://bernardmarr.com/default.asp?contentID=2126
[24] Multiple regression analysis. (n.d.). ScienceDirect.com | Science, health and medical journals, full-text articles and books. https://www.sciencedirect.com/topics/economics-econometrics-and-finance/multiple-regression-analysis
[25] What is case-based reasoning (CBR)? Definition from WhatIs.com. (2019, August 16). SearchEnterpriseAI. https://searchenterpriseai.techtarget.com/definition/case-based-reasoning-CBR#:~:text=%20case-based%20reasoning%20%28CBR%29%20%201%20Four%20step,to%20CBR.%20On

- and concluded that ANNs work more accurately then MRA and CBR estimating models[26]. Furthermore, ANNs represent the most frequently applied approach in estimating construction projects' duration and costs during the preliminary stages.

ANNs can self-learn, which saves a lot of development time. Also, ANNs can identify non-linear relationships between cost factors and project cost with no additional effort. With an ANN model, it is possible to obtain a reasonably accurate prediction, even when sufficient information is not available in the design process's early stages.[27].

ANNs are initially inspired by the study of processes in the human brain. ANNs consist of nodes (neurons in ANNs) grouped in interconnecting layers and sets of layers to form a network. There are three different types of layers, namely, input, hidden and output layers. The layout or architecture of a network is presented below in figure 03.

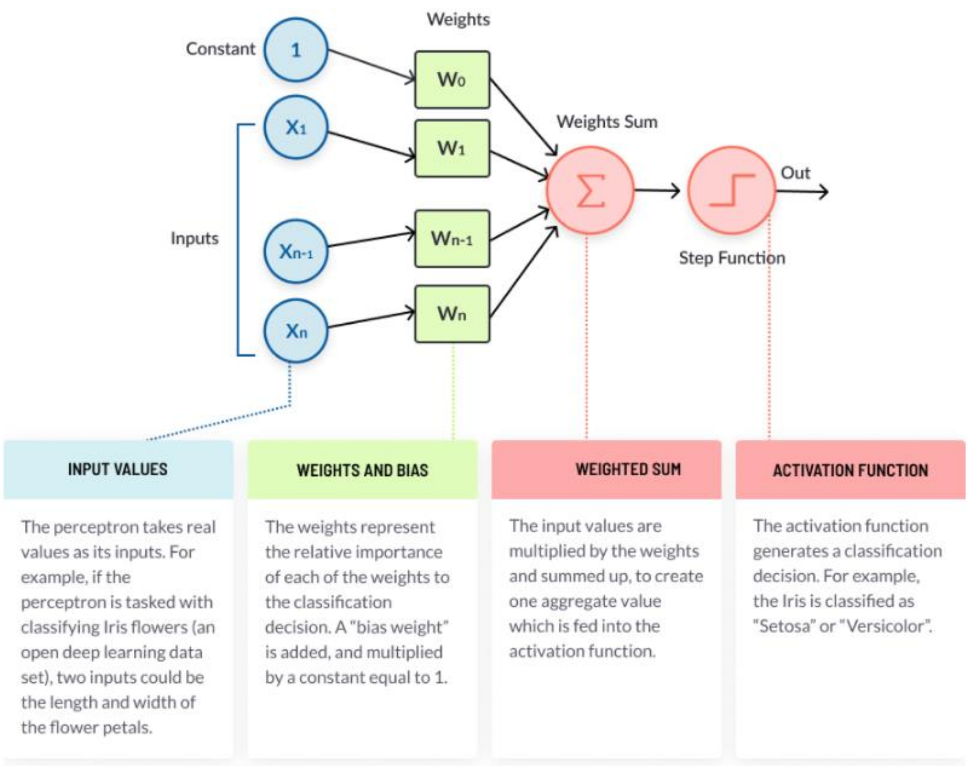


Figure 03: Structure of deep neural network or multi-layer perceptron.[28]

---

[26] Cheng MY, Tsai HC, Sudjono E. 2010. Conceptual cost estimates using evolutionary fuzzy hybrid neural network for projects in the construction industry

[27] Günaydin HM, Doğan SZ. 2004. A neural network approach for early cost estimation of structural systems of buildings.

[28] Perceptrons & multi-layer Perceptrons: The artificial Neuron - MissingLink. (n.d.). MissingLink.ai. https://missinglink.ai/guides/neural-network-concepts/perceptrons-and-multi-layer-perceptrons-the-artificial-neuron-at-the-core-of-deep-learning/

Conventionally, neural networks had a straightforward structure with only input and output layers. These were called single-layer neural network or external neural networks. Neural networks with multiple hidden layers are called multi-layer neural networks or deep neural networks. Every input node has a connection with all the nodes from the next hidden layer. This connection is illustrated by the arrow in Figure 3 and is corresponding to a particular weight. The training of the network consists of two different steps, namely, feedforward propagation and backpropagation. The training of a network begins with feedforward propagation, wherein the inputs and correct outputs from the training data are inputted to the neural network. Eventually, the neural network provides outputs based on the inputs and a random configuration of the weights. Subsequently, the neural network outputs are compared to the actual outputs, and the error is calculated. The backpropagation is when the weights are updated according to each node's error contribution and adjust the weights accordingly to reduce the error. These two steps are repeated for all the available training data. In this way, the neural network grows in the accuracy by learning from examples.

The quality and amount of training data are often the single most dominant factor in determining a model's performance. The amount of data needed for a machine learning algorithm depends on the complexity of the problem and the complexity of the chosen algorithm.

In the supervised learning process of an ANN, the learning process is based on datasets that provide input and output values. While the output value is usually determined by the purpose of the model (e.g. cost or time), the selection of inputs (or features) to be considered in the model is at the modeller's discretion. Some features in the training data may have outside elements. For example, when the potential input vector's dimension is substantial, it can be beneficial to eliminate redundant or irrelevant features.

STEP 2

In this section, the model and the development methodology are described. The methodology is adapted from[29] and consists of the following process:

1. Data collection: The data collection concerns the development of the dataset. In this phase, the input variables of the model are determined. Furthermore, the data that is needed to train the model is collected.
2. Network training: Subsequently, the network training covers the development of the actual model. The training phase consists of creating an ANN model and improving its performance by carrying out a heuristic optimisation strategy developed for this study and will be explained in detail in section 'Training'. Briefly, this methodology consists of three iterative phases. The first iterative phase determines the best training algorithm

[29] Hagan MT, Demuth HB, Beale MH, Orlando DJ. 2014. Neural network design. 2nd ed. Natick, Massachusetts, United States: MathWorks.

and best network architecture using the complete dataset. The second iterative phase determines whether the model can have better performance by using fewer input variables. The third iterative phase in the optimisation strategy to identify the most relevant scope of the input variables, i.e. different proposal value ranges.

3. Validation: is about the internal validation of the model.

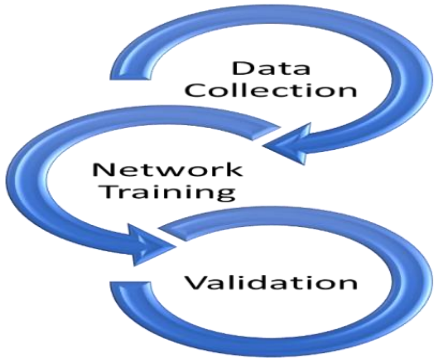


Figure 04: Model development[30]

This step is concerned with the selection, gathering and pre-processing of data. Given that ANNs are not efficient at handling extrapolation, the training data must be as comprehensive as possible to cover the model application's entire range.

As shown in Table 5, 116 projects, ranging in value, were selected for this study. The full dataset was divided into 12 categories based on the value. Due to the confidentiality agreement with the company, the value ranges must remain undisclosed. Because the model's accuracy depends on the input data's size, three different scenarios were considered to identify ranges at which an accurate estimator can be trained, given the uneven distribution of project values. These scenarios are shown in Table 5. Given the insufficiency of data and company experts' advice who pointed out the rarity of projects within specific ranges, the last four ranges were excluded from the analysis. In Scenario 1, all but the last three ranges are considered. In Scenarios 2 and 3, based on the expert opinion, four consecutive ranges, which accommodate most of the more recent projects, were chosen. It should be noticed that we are highlighting the range of projects included in each scenario, as in figure 5.

[30] Model development, by author

		Inclusion in differente Scenarios		
Range of Project Value (V inKEURO)	**Number of Projects**	**Scenario 1**	**Scenario 2**	**Scenario 3**
Category Range 1	2	✓		
Category Range 2	12	✓		
Category Range 3	23	✓		
Category Range 4	21	✓		✓
Category Range 5	1	✓	✓	✓
Category Range 6	15	✓	✓	✓
Category Range 7	15	✓	✓	✓
Category Range 8	11	✓	✓	
Category Range 9	8			
Category Range 10	5			
Category Range 11	1			
Category Range 12	2			

Figure 05: Project value range, selection of the range.[31]

The application of these scenarios will be explained in the section 'Phase 3: Determining the model scope/range' below.

As for the input variables, the 16 input variables identified are used as the basis for the model development. Input variables were of both qualitative and quantitative natures. Quantitative values were represented using positive real numbers and can be directly used in the model. The qualitative data were quantified using a mapping scheme. For instance, the type of client and his/her requirements can be categorised into very high demands, high demands, standard demands, low demands, and shallow demands.

[31] Project value range, selection of range, by author

Influencing factor	Description	Unit	Rank (by experts)
Scale of work	The costs of the total construction	Category Value in €	1
Project phases	The level of detail of the design	1 = Masterplan 2 = Conceptual design 3 = Basic design 4 = Detailed design 5 = Basic + detailed	15
Project duration	Number of weeks the project will take	Positive real number	3
Scope of work	The activities that are included in the contract	1 = Engineering (E) 2 = Engineering, Procurement, Construction (EPC) 3 = Engineering, Procurement, Construction Management (EPCm)	4
Type of work	The extent to which the project is a brownfield (modification) or greenfield (new construction) project	1 = 100% GF - 0% BF 2 = 75% GF - 25% BF 3 = 50% GF - 50% BF 4 = 25% GF - 75% BF 5 = 0% GF - 100% BF	5
Level of experience on clients side	The level of experience on the client-side	1 = Shallow level of experience 2 = Low level of experience 3 = Moderate level of experience 4 = High level of experience 5 = Very high level of experience	7
Scope definition	The extent to which the scope is defined	1 = Very poor scope definition 2 = Poor scope definition 3 = Moderate scope definition 4 = Good scope definition 5 = Very good scope definition	8
Size project team	Number of team members	Positive real number	9
Multidisciplinarity	Number of disciplines involved	Positive real number	10
Type of client and requirements	How demanding the client is towards standards and documentation	1 = Shallow demands 2 = Low demands 3 = Standard demands 4 = High demands 5 = Very high demands	11
Main market type	The main market in which the project takes place	1 = Oil & Gas 2 = Chemicals 3 = Energy & Environment 4 = Health and Nutrition 5 = Infrastructure 6 = Industrial 7 = Property 8 = Public sector 9 = Pharma	12

Attitude towards design changes	The attitude of the client toward design changes	1 = Very high level of cooperation	13
		2 = High level of cooperation	
		3 = Average level of cooperation	
		4 = Low level of cooperation	
		5 = Shallow level of cooperation	
Project manager experience	The number of hours of experience the selected project manager has	1 = Project manager D (<2000 hours)	14
		2 = Project manager C (<10.000 hours)	
		3 = Project manager B (<25.000 hours)	
		4 = Project manager A (<100.000 hours)	
		Five = Project director (>100.000 hours)	
Pre-contract design	The extent to which the pre-contract design is complete	1 = To a small extent	15
		2 = To some extent	
		3 = To a moderate extent	
		4 = To a great extent	
		5 = To a very great extent	
Contract type	The type of contract in which the project is carried out	One = Fixed Price	6
		2 = Reimbursable	
Intensity	The average hours a team member spend per week	1 = 8 hours/team member/week	2
		2 = 16 hours/team member/week	
		3 = 32 hours/team member/week	
		4 = >32 hours/team member/week	

Figure 06: Final input variables

We should also refer to the GAPPS Complexity Rating Tool.[32] Moreover, apply their metrics to establish the results as per the following table:

[32] Complexity rating for projects & programs: GAPPS. (2016, September 13). Global Alliance for the Project Professions. https://globalpmstandards.org/tools/complexity-rating/

Crawford-Ishikura Factor Table for Evaluating Roles (CIFTER)

Project Management Complexity Factor	Descriptor and Points			
1. Stability of the overall project context	Very high (1)	High (2)	Moderate (3)	Low or very low (4)
2. Number of distinct disciplines, methods, or approaches involved in performing the project	Low or very low (1)	Moderate (2)	High (3)	Very high (4)
3. Magnitude of legal, social, or environmental implications from performing the project	Low or very low (1)	Moderate (2)	High (3)	Very high (4)
4. Overall expected financial impact (positive or negative) on the project's stakeholders	Low or very low (1)	Moderate (2)	High (3)	Very high (4)
5. Strategic importance of the project to the organisation or organisations involved	Very low (1)	Low (2)	Moderate (3)	High or very high (4)
6. Stakeholder cohesion regarding the characteristics of the product of the project	High or very high (1)	Moderate (2)	Low (3)	Very low (4)
7. Number and variety of interfaces between the project and other organisational entities	Very low (1)	Low (2)	Moderate (3)	High or very high (4)

Figure 07: Crawford-Ishikura Factor Table[33]

This would result in the following result for our scenarios:

	Project Management Complexity Factor							
Sample	**1. Stability**	**2. No of Methods**	**3. Implications**	**4. Financial Impact**	**5. Strategic Importance**	**6. Stakeholder Cohesion**	**7. Project Interfaces**	**Total Score**
Category Range 1	1	1	2	2	2	1	1	10.00
Category Range 2	3	2	2	2	2	4	2	17.00
Category Range 3	5	1	5	4	4	5	4	28.00
Category Range 4	4	1	5	5	5	5	5	30.00
Category Range 5	2	2	2	2	2	2	2	14.00
Category Range 6	5	4	5	5	5	4	5	33.00
Category Range 7	5	3	4	4	4	5	5	30.00
Category Range 8	4	3	3	3	3	4	3	23.00
Category Range 9	1	2	1	2	2	2	2	12.00
Category Range 10	2	1	1	1	1	2	1	9.00
Category Range 11	1	1	1	1	1	1	1	7.00
Category Range 12	1	2	1	1	1	1	1	8.00

Figure 08: Crawford-Ishikura Factor Table evaluation of case projects, by author

When the required input and output criteria were determined, the data were gathered from various sources. Most of the required data were available in the database of the company. The data that were not available in the database were gathered using an online survey. In this survey,

[33] Complexity rating for projects & programs: GAPPS. (2016, September 13). Global Alliance for the Project Professions. https://globalpmstandards.org/tools/complexity-rating/

the responsible tender managers or project managers were asked to provide the missing data. Upon the completion of data gathering, a database with only the relevant data was built. Then, the data was cleaned to ensure homogeneity in the data. During this step, incomplete data sets were eliminated.

STEP 3

At this point, we can establish three different phases to optimise the strategy, as per below:

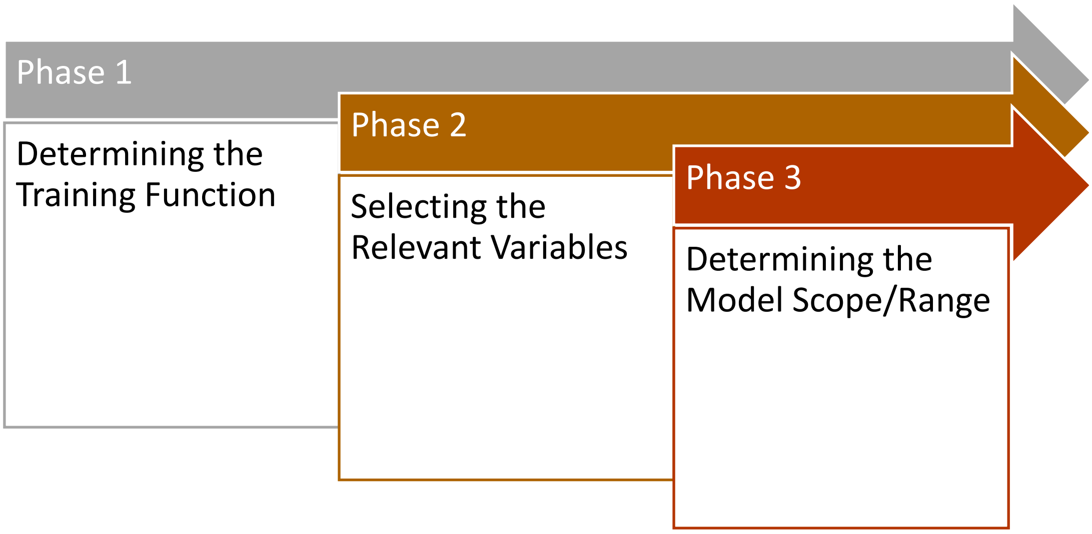


Figure 09: Phases of Strategy Optimisation

Phase 1: determination of training function: In this phase, the best training function that can model the data was determined. In this research, the three most common training functions were considered, namely, the Levenberg-Marquardt (LM), [34]Bayesian Regularization (BR), [35]and Resilient Backpropagation (RB) [36]functions. The Correlation Coefficient (R) and Mean Absolute Percentage Error (MAPE) [37]were used for the assessment of the models' performances. The model was considered a good fit when (1) R was close to 1, which indicates the goodness of the fit and (2) the difference between R values of the training and test sets was small, which indicates the model is stable and generalisable. Also, the MAPE value was used to estimate the error of the test predictions. When the model was determined as a good fit, the model with the least MAPE value was selected.

[34] Levenberg-Marquardt backpropagation - MATLAB trail - MathWorks Italia. (n.d.). MathWorks - Creatori di MATLAB e Simulink - MATLAB e Simulink - MATLAB & Simulink. https://it.mathworks.com/help/deeplearning/ref/trainlm.html
[35] Bayesian regularisation backpropagation - MATLAB trainer. (n.d.). MathWorks - Makers of MATLAB and Simulink - MATLAB & Simulink. https://www.mathworks.com/help/deeplearning/ref/trainbr.html;jsessionid=e4358f31
[36] How to use resilient backpropagation to train neural networks -- Visual studio magazine. (2015, March 1). Visual Studio Magazine. https://visualstudiomagazine.com/articles/2015/03/01/resilient-back-propagation.aspx
[37] Mean absolute percentage error (MAPE). (2020, September 5). Statistics How To. https://www.statisticshowto.com/mean-absolute-percentage-error-mape/

First, a training function was selected, and the network architecture was optimised using the growing method. In this method, the training is initiated by a single hidden neuron, and then neurons are added one at a time until a threshold is reached on the overfitting of the model.

Given that a multi-layer neural network's training involves two stochastic elements, every training run may result in slightly different models. These two elements are (1) the initial weights and biases and (2) the random selection of training, testing and validation sets. To get a robust estimate of the stochastic model, this variance must be considered. This was done by applying the Bootstrapping method. In this method, the distribution of an estimator or test statistic is captured by resampling the data several. We trained each configuration of the model 100 times, and the performance was calculated. For each configuration, the best model was identified and selected.

When the network architecture was optimised, and the best model was identified, the next training algorithm was selected until all the training algorithms were tested. At the end of this phase, the best training algorithm and the best network architecture explain the total dataset.

The best results for the best network architectures for the three different training algorithms are shown in Table 4. It can be discerned that with 16 input variables and the complete dataset, the Bayesian Regularization (BR) training algorithm with one hidden layer and four hidden neurons performed the best. For this architecture, R-value is 0.98, and the difference between training and the testing dataset is 0.03. Besides, this architecture has the lowest MAPE score, which means that it has the lowest mean average percentage error for both the training and the testing set.

Network architecture	R train	R test	R all	MAPE train (%)	MAPE test (%)	MAPE all (%)
LM-16-6-1	0.9997	0.9168	0.9684	57.05	100.18	77.98
BR-16-4-1	0.9998	0.9645	0.9796	37.25	50.36	39.24
RP-16-6-1	0.9966	0.7509	0.8666	59.48	88.68	89.51

Figure 10: Results of the first phase, by author

Phase 2: feature selection: To find the simplest model that explains the data, it could help eliminate redundant or irrelevant input variables. This process is also known as feature selection. By calculating the relative importance of input variables of the network with the highest performance, redundant input variables can be removed and thus make the model more generic. A method called Connection Weights Algorithm (CWA) [38]was used to calculate the relative

[38]Optimising connection weights in neural networks using the whale optimisation algorithm. (n.d.). Soft Computing. https://link.springer.com/article/10.1007/s00500-016-2442-1

importance of the input variables of the model. This approach is based on the estimation of the network's final weights obtained by training the network.

Once the relative importance of the input variables was identified, the next step was to eliminate the variables that have a low impact from the best model identified in the previous phase. The elimination started by removing the variable with the lowest relative importance. The elimination of the variables continued until the model's performance decreased regarding the previous configuration. Since the number of input neurons decreased, the number of neurons in the hidden layers also needed readjustment. Therefore, the training entered the network architecture optimisation module again. Eventually, the simplest model that explains the data was determined.

Input variables	Variable relative importance based on CWA (%)	Variable ranking by experts	Variable significance based on MLR
Scale of work	7	1	12
Project phases	6	15	11
Project duration	3	3	1
Scope of work	14	4	10
Type of work	12	5	16
Level of experience on client's side	10	7	9
Scope definition	16	8	15
Number of project team members	2	9	4
Collaborating disciplines	4	10	3
Type of client and requirements	13	11	14
Main market type	8	12	7
Client's attitude towards design changes	11	13	13
Project manager experience	15	14	8
Pre-contract design	9	15	6
Contract type	5	6	5
Intensity	1	2	2

Figure 11: Comparison of variable importance using different methods.

In this technical paper, to ensure that the most accurate model is achieved, in addition to CWA, two other methods were used to filter the input variables, namely, multiple linear regression analysis [39]and expert opinion. MLR analysis is a suitable method to identify which variables have a significant influence on the proposal price. It can help determine whether there is a linear association between the independent variables and proposal price. The relative importance of the independent variables is determined by the unit drop in R2 when a variable is deleted from the sample. R2 is the coefficient of determination and shows the percentage of variation in a dependent variable explained by all the independent variables together. The larger the drop in R2 when a variable is removed from the sample, the more critical it is assumed. Also, expert estimators can be interviewed to rank the importance of the input variables. These two methods were applied, and the results were compared to the model formed by CWA.

Using CWA, it was observed that the performance decreased significantly after eliminating the 12th variable (Contract type). As shown, the best performances occurred when the model was trained using between 5 to 9 top variables. Within this range, the differences between performances were relatively minor. The phase has proven effective in improving the accuracy of the model. To put this into perspective, the lowest MAPE with all 16 variables was 50.36%, i.e. the best performance in Phase 1.

Nevertheless, when only the top 5 variables were used, the MAPE plummet to 27.41%. This indicates a 45.6% improvement inaccuracy. It could be concluded that the most dominant variables for the estimation of engineering services based on CWA method are intensity, several project team members, project duration, collaborating disciplines, contract type, project phases, the scale of work, primary market type, and pre-contract design.

Variable selection method	Network architecture	R train	R test	R all	MAPE train (%)	MAPE test (%)	MAPE all (%)
Connection weights algorithm	BR-9-5-1	0.9992	0.964	0.9813	48.27	51.32	48.73
	BR-8-6-1	0.9995	0.9556	0.9979	55.07	42.26	53.13
	BR-7-6-1	0.9994	0.9648	0.9952	46.73	37.41	45.32
	BR-6-8-1	0.9997	0.946	0.9985	35.19	32.83	34.83
	BR-5-7-1	0.9996	0.9952	0.9991	33.15	27.41	32.28
Multiple linear regression	BR-7-6-1	0.9998	0.8419	0.9921	46.68	52.7	47.87
	BR-6-6-1	0.9996	0.9784	0.9939	38.83	42.56	39.56
	BR-5-7-1	0.9999	0.9065	0.9806	23.56	42.47	27.28
Expert opinion	BR-7-2-1	0.9583	0.9449	0.9519	167.69	121.04	158.5
	BR-6-7-1	0.9824	0.9405	0.9668	210.61	105.18	189.84
	BR-5-6-1	0.753	0.8609	0.7664	274.98	93.25	239.19

Figure 12: Results of the second phase.

[39] Multiple linear regression analysis. (n.d.). ScienceDirect.com | Science, health and medical journals, full-text articles and books. https://www.sciencedirect.com/topics/medicine-and-dentistry/multiple-linear-regression-analysis#:~:text=%20Multiple%20Linear%20Regression%20Analysis%20%201%20Individualizing,Techniques.%20G.%20Hanrahan%2C%20...%20%20...%20More%20

When the MLR method was used, the best performances were achieved when the top 5 to 7 variables are used. As shown in Table 6, the best results based on MLR ranking was achieved when 5 top variables (project duration, intensity, collaborating disciplines, number of project team members and contract type) are considered. In this case, the MAPE of the model is 42.47%. While this indicates a slight improvement of accuracy compared to the best model from Phase 1, it is evident that CWA performs better in capturing the relative importance of variables for this case. This can be justified because MLR assumes linear relationships between independent variables and a dependent variable. Some of the variables could have a non-linear relationship with the dependent variable, which cannot be captured by MLR. CWA-based variable selection, on the other hand, is capable of handling non-linearity in the data.

Finally, the training's best results based on the ranking of variables by expert opinion were achieved when the top 5 variables (scale of work, project phases, project duration, the scope of work and type of work) are used. In this model, MAPE was 93.25% when five hidden neurons are used. Based on comparing MAPE values of different methods, expert opinion performed poorly to capture the essential variables. This can be partially attributed to the fact that the final ranking was based on all the interviewees' average scores. It is conceivable that some interviewees were ranking the factors from their standpoint and lacked a global view of how different factors impact the firm's full service.

Phase 3: determining the model scope/range: In the last phase of the training, the best scope/range for the input data was determined. As stated earlier in the section 'Data collection', because ANN is not efficient in handling extrapolation outside the range of the training set and fitting an accurate model with insufficient data, it is crucial to determine which scope/range of the input data the model performs the best. For this purpose, the three scenarios presented in figure 6 were used to investigate the impact of scope reduction/adjustment on the model's accuracy. As mentioned in section 'Data collection', these scenarios were formed based on the amount of data available for the different project value categories. It should be noted that the elimination of specific input data from the dataset will improve the overall accuracy of the model at the expense of sacrificing the generic-ness of the model. First, it is decided to proceed with the training with the five different network architectures (top 9 to 5 variables determined by the CWA method). However, when a data selection was made, the data's underlying function's complexity could be different compared to the full database. Therefore, the growing technique was used again, and the number of hidden neurons was changed for every network in each training set.

In this phase, a model is developed for each scenario. Given that the input dataset's change may affect the architecture of the model, a new model was trained and optimised for each scenario. Ultimately, the performance of the models was compared and analysed.

While the model's performance has improved for all scenarios, scenario two has shown the most significant improvement compared to the best models from phase 2. In this scenario, projects with a value range category two until five are chosen. The best model for this input dataset had seven input variables and four hidden neurons. In this case, MAPE of 13.65% was achieved, which indicates 50.2% and 72.9% improvement over the best model of Phases 1 and 2, respectively. Very important is that the R-value for both training and test sets are very similar and only differed by 0.0013, which indicates high generalizability. Furthermore, both sets' R values were very close to 1, indicating high goodness of the fit. Finally, the performances of the training MAPE, test MAPE, and overall MAPE were very similar.

Scenario	Network architecture	R train	R test	R all	MAPE train (%)	MAPE test (%)	MAPE all (%)
1	BR-5-7-1	0.9989	0.9705	0.9819	24.04	23.612	23.971
2	BR-7-4-1	0.9957	0.9944	0.9954	13.648	13.648	13.648
3	BR-5-6-1	0.9915	0.9581	0.9832	16.246	21.87	17.13

Figure 13: Best results third iterative process.

STEP 4

The last step that we need to consider is the validation of the model, and this was validated by investigating the model's performance outside the training sample.

Also, the Bootstrapping method,[40]was again used to reduce the impact of randomly selected elements in training. At the end of each iteration, the model's performance was analysed, and the MAPE of the model was calculated. Subsequently, the mean MAPE and the standard deviation were calculated for all the models combined. By doing so, a more robust estimate of the variance of the model can be acquired.

The regression plot shows that both the training and testing results were auspicious, with the respective R-value of 0.99575 and 0.99438. For this model, the MAPE of the entire set was 13.65%. With a maximum error of 62% and a minimum error of 0.32%. About 66% of the predictions had a relative error of lower than 10% for the test set. Besides, 37% of the predictions, both on training and test data, had a relative error of less than 5%.

FINDINGS

STEP 5

Therefore, the opportunities offered in cost estimation and optimisation are enormous and far from being fully exploited. Beyond cost optimisation, the algorithm's self-learning on companies and their suppliers' data allows it to consider intelligent contributions such as the automatic preparation of negotiations (objectives, levers arguments), the proposing optimised designs or

[40] A gentle introduction to the bootstrap method. (2019, August 8). Machine Learning Mastery. https://machinelearningmastery.com/a-gentle-introduction-to-the-bootstrap-method/#:~:text=%20Specifically%2C%20you%20learned

redesigns, recommending the most adapted purchasing strategies anticipating supplier behaviour.

The results showed that ANN could be used to obtain a reasonably accurate cost estimate, even with minimum data that is available during the tendering phase. The accuracy obtained with the ANN model is well within the range achieved by models developed in comparatives studies in the literature, see Table 1. However, these results were achieved for cost estimation from the contractors' perspective and not for engineering services. When looking at the ML-based cost estimation model for engineering services, i.e., Hyari et al. (2016), the model developed in this study showed a 14.5% improvement in the model's accuracy considering MAPE.

One of this study's main contributions is analysing the relevant and essential variables for the cost estimation of engineering services. It is shown that by decreasing the number of variables and excluding the less critical variables, the model's performance can improve. Furthermore, it is shown that the cost of the engineering services can be accurately estimated using the following seven input variables: (1) intensity, (2) several project team members, (3) project duration, (4) collaborating disciplines, (5) contract type, (6) project phases, (7) scale of work. It is discovered that the variables that are found more prominent for the cost estimation based on CWA and MLR methods are different from those identified by the expert opinion.

However, there are several limitations to the present research. First, 132 individual data points were collected, and the best neural network was developed using 60 data points. Due to the split-sample technique that is used, the test results were based only on 12 cases. This is considered a small sample and the model needs to be validated with a more extensive set of data. Second, the model also needs to be externally validated by applying it to new projects and comparing the results with the experts' actual estimate. Finally, more research needs to be done on the adoption of ML-based cost estimation in practice. Given the black-box nature of ANNs, building trust in the model within an organisation seems challenging. Neural networks are accurate predictors. However, it is a challenge to offer a justification for the structure and behaviour of the model.

Step 6

This paper aimed to investigate the possibility of developing an accurate ML-based cost estimation method for tendering and subsequent application of engineering services. This was done by systematically developing and optimising a neural network model to estimate engineering services' initial costs. The research applied a systematic methodology that provides a guideline for developing and optimising an artificial neural network for cost estimation.

The development of the neural network included measures to remove the nuisance from the data. The systematic methodology applied for optimising the network proved to be very efficient in improving the model's performance.

STEP 7

So how can we leverage the use of artificial intelligence to:

1. Classify and categorise unclassified cost data from previous projects
2. Develop a statistical model for parametric estimation
3. Identify reference project class.

Subsequently, a heuristic method is developed to improve and fine-tune the performance of the model systematically. The findings eventually show that artificial neural networks (ANNs) can obtain a reasonably accurate cost estimate, even with small datasets. The model proposed in this paper performed better than those proposed in other similar works. The model developed in this study showed a 14.5% improvement in the model's accuracy, considering MAPE.

CONCLUSION

As we have concluded, it does not make sense that we oppose, in cost estimation, the analytical and statistical methods because they complement each other

, giving a better result.

On one side the statistical method, which is more consistent because it is based on observing what the actual data says, allows us to have a precise and quick evaluation that will support us in making the correct decisions in the product design or redesign processes. Simple to implement, it allows modelling many families of products and services in a non-intrusive way and without needing to acquire advanced technological expertise.

The analytical method allows obtaining encryption precisely reflecting the reality (or the simulation) of a process. On the other side, even if more tedious to implement, allows defining cost targets that we have to reach with explanatory factors using industrial parameters and benchmarks. In this sense, it is more appropriate to quantify technological breakthroughs and to lead industrial suppliers progress plans to bring them to the target. It is also more relevant to quantify technological innovations on which the company does not have a history.

Anyway, both deep learning and self-learning algorithms open new horizons and fields of study and applications that we can use in statistical models, particularly by sharing information between companies or between them and their suppliers.

BIBLIOGRAPHY

[1] What are artificial neural networks? (n.d.). Bernard Marr. https://bernardmarr.com/default.asp?contentID=2126

[2] Great Wall of China. (n.d.). Mandala Projects. https://mandalaprojects.com/ice/ice-cases/wall.htm

[3] Herodotus. (n.d.). Encyclopedia Britannica. https://www.britannica.com/biography/Herodotus-Greek-historian

[4] Egyptian art and architecture - The pyramid of Khufu. (n.d.). Encyclopedia Britannica. h5tps://www.britannica.com/art/Egyptian-art/Pyramid-of-Khufu

[5] The Westcar papyrus. (2017, April 5). Ancient History Encyclopedia. https://www.ancient.eu/The_Westcar_Papyrus/

[6] The Westcar papyrus and the miracle stories of the old kingdom. (2016, October 28). Ancient Origins | Reconstructing the story of humanity's past. https://www.ancient-origins.net/artifacts-ancient-writings/westcar-papyrus-and-miracle-stories-old-kingdom-006895

[7] Project management triangle | Overview of triple constraints. (n.d.). StarAgile Consulting | Certification Training Courses. https://staragile.com/blog/project-management-triangle

[8] What is artificial intelligence (AI)? (2020, June 3). https://www.ibm.com/cloud/learn/what-is-artificial-intelligence

[9] What is the cost estimate? definition and meaning. (n.d.). BusinessDictionary.com. https://www.businessdictionary.com/definition/cost-estimate.html

[10] W20.1_Budgeting continuous improvements and process estimating. (2020, September 26). Danilo Arba Blog. https://www.daniloarba.com/post/copy-of-w20-1_budgeting-continuous-improvements-and-process-estimating

[11] What is a Cost Estimate? Guild of project controls compendium and reference (Car). (n.d.). Planning Planet | dedicated to Project Controls. https://www.planningplanet.com/guild/gpccar/introduction-to-managing-cost-estimating-budgeting

[12] What is natural language processing? (2020, July 2). IBM - United States. https://www.ibm.com/cloud/learn/natural-language-processing

[13] Seker, E. (2020, September 16). Artificial neural networks (ANN). Medium. https://towardsdatascience.com/artificial-neural-networks-ann-21637869b306

[14] What is case-based reasoning (CBR)? Definition from WhatIs.com. (2019, August 16). SearchEnterpriseAI. https://searchenterpriseai.techtarget.com/definition/case-based-reasoning-CBR

[15] Similarity and metrics in case-based reasoning. (n.d.). ResearchGate. https://www.researchgate.net/publication/27827429_Similarity_and_metrics_in_case-based_reasoning

[16] (PDF) Effort estimation using an analogy. (1996, April 25). ResearchGate. https://www.researchgate.net/publication/3625950_Effort_estimation_using_analogy

[17] What is analytical estimating - definition of analytical estimating by business training. (n.d.). https://www.businesstrainingcollege.com/business/Analytical-Estimating.htm

[18] Parametric estimating. (2008, January 15). Project Management Knowledge - Simply explained by a PMI-certified Project Manager. https://project-management-knowledge.com/definitions/p/parametric-estimating/

[19] Knowledge-based systems. (n.d.). https://www.journals.elsevier.com/knowledge-based-systems#:~:text=Knowledge-based%20Systems%20is%20an%20international%20and%20interdisciplinary%20journal,research%20in%20knowledge-based%20and%20other%20artificial%20intelligence%20techniques-based

[20] Evolutionary system. (n.d.). * Home. https://en.mimi.hu/artificial_intelligence/evolutionary_system.html

[21] Hybrid artificial intelligence system for the design of highly-automated production systems. (n.d.). ScienceDirect.com | Science, health and medical journals, full-text articles and books. https://www.sciencedirect.com/science/article/pii/S2351978918313684

[22] What are artificial neural networks? (n.d.). Bernard Marr. https://bernardmarr.com/default.asp?contentID=2126

[23] Multiple regression analysis. (n.d.). ScienceDirect.com | Science, health and medical journals, full-text articles and books. https://www.sciencedirect.com/topics/economics-econometrics-and-finance/multiple-regression-analysis

[24] What is case-based reasoning (CBR)? Definition from WhatIs.com. (2019, August 16). SearchEnterpriseAI. https://searchenterpriseai.techtarget.com/definition/case-based-reasoning-CBR#:~:text=%20case-based%20reasoning%20%28CBR%29%20%201%20Four%20step,to%20CBR.%20On

[25] Cheng MY, Tsai HC, Sudjono E. 2010. Conceptual cost estimates using evolutionary fuzzy hybrid neural network for projects in the construction industry

[26] Günaydin HM, Doğan SZ. 2004. A neural network approach for early cost estimation of structural systems of buildings.

[27] Perceptrons & multi-layer Perceptrons: The artificial Neuron - MissingLink. (n.d.). MissingLink.ai. https://missinglink.ai/guides/neural-network-concepts/perceptrons-and-multi-layer-perceptrons-the-artificial-neuron-at-the-core-of-deep-learning/

[28] Hagan MT, Demuth HB, Beale MH, Orlando DJ. 2014. Neural network design. 2nd ed. Natick, Massachusetts, United States: MathWorks.

[29] Model development, by author

[30] Project value range, selection of range, by author

[31] Complexity rating for projects & programs: GAPPS. (2016, September 13). Global Alliance for the Project Professions. https://globalpmstandards.org/tools/complexity-rating/

[32] Complexity rating for projects & programs: GAPPS. (2016, September 13). Global Alliance for the Project Professions. https://globalpmstandards.org/tools/complexity-rating/

[33] Levenberg-Marquardt backpropagation - MATLAB trail - MathWorks Italia. (n.d.). MathWorks - Creatori di MATLAB e Simulink - MATLAB e Simulink - MATLAB & Simulink. https://it.mathworks.com/help/deeplearning/ref/trainlm.html

[34] Bayesian regularisation backpropagation - MATLAB trainer. (n.d.). MathWorks - Makers of MATLAB and Simulink - MATLAB & Simulink. https://www.mathworks.com/help/deeplearning/ref/trainbr.html;jsessionid=e4358f31

[35] How to use resilient backpropagation to train neural networks -- Visual studio magazine. (2015, March 1). Visual Studio Magazine. https://visualstudiomagazine.com/articles/2015/03/01/resilient-back-propagation.aspx

[36] Mean absolute percentage error (MAPE). (2020, September 5). Statistics How To. https://www.statisticshowto.com/mean-absolute-percentage-error-mape/

[37] Optimising connection weights in neural networks using the whale optimisation algorithm. (n.d.). Soft Computing. https://link.springer.com/article/10.1007/s00500-016-2442-1

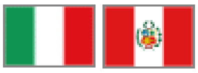
[38] Multiple linear regression analysis. (n.d.). ScienceDirect.com | Science, health and medical journals, full-text articles and books. https://www.sciencedirect.com/topics/medicine-and-dentistry/multiple-linear-regression-analysis#:~:text=%20Multiple%20Linear%20Regression%20Analysis%20%201%20Individualizing,Techniques.%20G.%20Hanrahan%2C%20...%20%20...%20More%20

[39] A gentle introduction to the bootstrap method. (2019, August 8). Machine Learning Mastery. https://machinelearningmastery.com/a-gentle-introduction-to-the-bootstrap-method/#:~:text=%20Specifically%2C%20you%20learned

About the Author

Danilo Arba

Milan, Italy
Lima, Peru



Danilo Arba is a project controls & management enthusiast, with 20 years of experience. He is a Certified Cost Engineer with an Executive MBA from Politecnico di Milano. With a thorough understanding of EPC (Engineering, Procurement, and Construction) industry, he has a verifiable track record of planning multimillion/billion-dollar construction projects worldwide. He lived & worked all his life around the world from South America, Africa, South East Asia to Europe. He is adept at building and leading cross-functional teams from project conception to completion, optimising performance, contractual, and financial deliverables. Currently he is furthering his education by way of a distance learning mentoring course, under the tutorage of Dr Paul D. Giammalvo, CDT, CCE, MScPM, MRICS, GPM-m Senior Technical Advisor, PT Mitrata Citragraha, to attain Guild of Project Controls certification.

Danilo lives in Milan, Italy and Lima, Peru and can be contacted at danilo.arba@mip.polimi.it