Winter 2-19-2025

Predicting the Indirect Cost of Construction Projects in Egypt: An Artificial Neural Network Approach

Aya Effat
ayaeffat@aucegypt.edu

School of Sciences and Engineering

Graduate Studies

Summer 2024

Predicting the Indirect Cost of Construction Projects in Egypt: An Artificial Neural Network Approach

A Thesis Submitted to the
Construction Engineering Department

in partial fulfillment of the requirements for
the degree of Master of Science in Construction Engineering

By
Aya Mohamed Effat

Under The Supervision Of

Dr. Ossama Hosny
Construction Engineering Department
The American University in Cairo

Dr. El Khayam Dorra
Construction Engineering Department
The American University in Cairo

Declaration of Authorship

I, [Aya Effat], declare that this thesis titled, [Predicting the Indirect Cost of Construction Projects in Egypt: An Artificial Neural Network Approach]" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

Aya Mohamed Effat

Date:

03/09/2024

Acknowledgments

First and foremost, I would like to express my sincere gratitude to God for giving me the strength, knowledge, and perseverance to undertake and complete this research.

I would like to thank my advisors, Dr. Ossama Hosny and Dr. ElKhayam Dorra, for their unwavering mentorship and support. Their invaluable expertise and insightful guidance have been instrumental in shaping this research.

I would like to extend my heartfelt thanks to my beloved parents and siblings for their unconditional love, encouragement, and unwavering belief in me. Their support has been my constant source of strength and motivation, enabling me to achieve this milestone.

I am grateful to my colleagues and managers at work for their generous support and cooperation throughout this research project. I am truly fortunate to have had the opportunity to work with such a dedicated and supportive team.

Finally, I would like to express my sincere appreciation to my friends for their encouragement, and support. Their belief in me has been a constant source of motivation.

Abstract

Cost estimation is one of the vital processes in construction management that needs to be done early in any project in order to determine the project's budget. The accuracy of the cost estimate is a key factor in the success of construction projects since it enables project managers to successfully control the project's expenses. Construction costs mainly consist of direct cost and indirect cost. Generally, indirect costs can be categorized into two types: site overheads and general overheads. In a construction project, overheads, particularly site overhead costs, make up a considerable portion of a contractor's budget. Accordingly, accurately estimating the site overheads of construction projects is a crucial task that needs to be done in order to manage projects efficiently. Therefore, the main objective of this research is to enhance the contractor's ability to accurately predict the site overhead costs of construction projects in Egypt through identifying and analyzing the key factors influencing site overheads in the Egyptian construction industry. This study proposes three-stage ANN approach for predicting site overheads of construction projects. The first ANN model estimates the total site overhead percentage while the second and third ANN models then utilize both the predicted total site overheads percentage and existing project data to forecast the breakdown of site overhead across its different subcategories and across the project's different construction phases, while incorporating both economic and non-economic variables. In order to form the model's database, the major factors affecting the site overheads were first identified through an extensive literature review. These factors were project type, project location, project duration, contract type, project direct cost, client type, class of contracting company and lastly macroeconomic indicators such as inflation rate, interest rate and currency exchange rates. In addition, cost data from 55 real-life projects executed during the past 10 years were obtained to be used as a database for the learning process of the ANN model. Cost data for 5 new projects were then used to test each model. Model 1 had 2.75% MAE for training set and 3.9% for testing set. Model 2 had 2.62% MAE for training and 2.83% for testing while Model 3 had 2.12% MAE for training and 2.31% MAE for testing data set. Overall, the models performed well and can be considered a useful tool for the predicting the percentage of site overheads as well as the percentage of site overheads allocated to each subcategory and each construction phase. Thus, these models offer a valuable tool for contractors to enhance cost estimation, improve decision-making and mitigate financial risks.

Table of Content

List of Tables

List of Figures

1. Introduction

1.1. General Background

The construction industry is among the most competitive ones as it constantly deals with several complex challenges across a number of distinct areas such as risk analysis, bidding, cost estimation, delays and disputes (Kulkarni et al., 2017). Not only is this field a competitive one, but it is also considered as one of the main sectors of the Egyptian economy, particularly with regard to real estate and commercial buildings (Idrees et al., 2019). The construction industry has a significant influence on the Gross Domestic Product as it accounts for about 14% of Egypt's GDP, which is considered the highest percentage compared to other sectors (Mordor Intelligence, 2022). However, many industries suffered as a result of the uncertain economic condition coupled with the political dangers in Egypt following the Egyptian revolution in 2011 and the floating exchange rate for the Egyptian pound in November 2016 which was later followed by a high rate of inflation (Idrees et al., 2019; Khedr et al., 2016). Taking into account all of these factors, precise estimates for project costs is highly required by construction companies in Egypt and managing project expenses has become more significant and impactful than it was before (Idrees et al., 2019).

Generally, the success of any construction project is mainly determined through its ability to balance the three major conflicting project constraints of cost, time and quality, which are predetermined by the stakeholders and eventually result in customer satisfaction (Rezaian, 2011). However, projects are rarely built according to the original plan and are frequently finished beyond budget, so if the expenses are not accurately assessed, the anticipated earnings might become losses (Cheng et al., 2010). Therefore, a crucial component of a successful construction project is performing precise cost estimation (Enshassi et al., 2013). In fact, cost estimating is defined by the Project Management Institute (PMBOK 2013) as the process of creating an approximation (estimate) of the costs of resources necessary to finish project tasks.

In order to arrive at precise and approximate estimates and resolve the issue of cost overruns, advanced cost estimating methodologies should be employed during the planning phases (Enshassi et al., 2013). Artificial Intelligence (AI) methods like artificial neural networks (ANN), case-based reasoning, fuzzy logic and genetic algorithms are being frequently utilized in construction management to overcome the challenges faced. In fact, the past two decades of the 20th century have seen an increase in publications in a number of construction management-related fields addressing artificial intelligence approaches, particularly ANN (Kulkarni et al., 2017).

The majority of construction organizations have no difficulty calculating the direct expenses of a project. In fact, the misestimating is often associated with overhead costs, leading to a discrepancy in costs between the budget and the actual cost, which can result in either cost overrun or cost savings (Hassouna et al., 2020). The impact of overhead cost estimates is critical to the construction company's financial condition (Othman, 2020). The overhead expenses incurred by the building contractor are frequently divided into two main categories: site overhead costs and company overhead costs. All expenses paid by the contractor on the construction site in order to finish the work, excluding direct costs, are referred to as site overhead costs (Lesniak and Juszczyk, 2018). General overhead costs are expenses that are necessary to operate a firm and cannot be allocated to a specific project, and therefore are allocated across all of the business's projects (Bakr et al., 2018). Generally, site overheads are more difficult to accurately estimate than general overheads and that is why recent researchers are more concerned with finding advanced techniques to calculate site overheads as the lack of systematic and precise techniques to evaluate site overhead costs for construction projects in Egypt exposes construction businesses to the possibility of inaccurate bid package estimates that might impact their profit margin (Khedr et al., 2014).

Accurately estimating site overhead expenses is crucial for the success of construction projects. Thus, this research aims to enhance contractors' ability to predict these costs by developing an Artificial Neural Network (ANN) model that predicts the percentage of site overheads as a percentage of the total direct costs of projects, with the model being applied in the Egyptian construction industry. This would enhance the companies' performance in forecasting overheads for upcoming projects. It will also help control factors that affect site expenditures, generate an information system and compile historical data for projects to be used to improve the predictability of site overheads in future projects and reduce the amount of time and energy used in estimating site overheads.

1.2. Problem Statement

Most companies in the Egyptian construction industry and worldwide find it challenging to accurately estimate the site overhead costs due to the complex nature of these costs and the presence of a wide range of distinct factors that have a direct influence on it (Othman, 2020). The scarcity of systematic and precise techniques that incorporate both economic and non-economic variables in order to evaluate site overhead costs for construction projects in Egypt exposes construction businesses to the possibility of inaccurate bid package estimates (Bakr et al., 2018). In fact, several contractors still utilize traditional estimation methods that mainly rely on historical data, expert judgment, and simplified cost models. These conventional methods of cost estimation lack adequacy due to their failure to effectively leverage the implicit knowledge gained from previous projects. Consequently, these estimation methods are slow, inaccurate, and exhibit high variability, impacting project profitability (Matel et al., 2019). Thus, it is important to focus on utilizing advanced artificial intelligence techniques to develop models capable of accurately forecasting site overheads.

1.3. Research Objectives

The main objective of this research is to enhance the contractor's ability to accurately predict the percentage of site overheads in construction projects in Egypt through the following:

- identifying and analyzing the key factors influencing site overheads in the Egyptian construction industry.
- developing a robust dataset containing historical cost data of projects executed in the past 10 years.
- developing ANN models capable of accurately predicting the total percentage of site overheads for construction projects in Egypt and subsequently allocating these costs across different site overheads subcategories and across the different project construction phases, incorporating both economic and non-economic variables to enhance the predictive accuracy

The developed models aim to provide project managers with a robust tool for predicting total site overheads, allocating these costs across different cost categories, and constructing an S-curve for overhead distribution throughout the project's construction phases.

1.4. Thesis Organization

This research paper consists of 7 chapters, which are organized as follows:

1. **Chapter 1: Introduction:** This chapter introduces the research topic, its significance, research objectives, and outlines the thesis structure.
2. **Chapter 2: Literature Review:** This chapter provides a comprehensive overview of existing research on the topic, identifying research gaps and justifying the need for the study. It also assists in identifying the main factors affecting the site overheads in construction projects in Egypt in order to be used as inputs/independent variables in the developed models.
3. **Chapter 3: Research Methodology:** This chapter describes the details the research framework. In addition, it explains the data collection process, which includes a questionnaire form sent to experts in the field based on the factors previously obtained through the literature review. It also includes obtaining macroeconomic data from online sources like the website of the Central Bank of Egypt. In addition, it provides a theoretical foundation for ANN models. It explains the necessary steps needed to develop any ANN models, like data preprocessing, design of model architecture, model training and selection and model testing and validation.
4. **Chapter 4: Data Analysis:** This chapter involves conducting data analysis and Pearson and Spearman correlations tests for the collected data.
5. **Chapter 5: Development of the ANN Models:** This chapter is concerned with using the steps explained in Chapter 3 to develop the 3 ANN models using Python Programming Language on Google Colab.
6. **Chapter 6: Results and Discussion:** It presents the findings of the study, analyzes the results, and interprets their implications. It also uses data from new projects that were not introduced to the model before to assess the model's performance by calculating the percentage error between the actual and predicted values.
7. **Chapter 7: Conclusion and Recommendations:** It summarizes the key findings, contributions of the research, and provides recommendations for future studies.

2. Literature Review

2.1. Introduction

This section provides a detailed literature review that mainly includes a cohesive summary of the extensive research and analysis done on the existing knowledge related to cost estimating in construction projects. This literature review mainly covers the definition and importance of cost estimation, definition of direct and indirect (overhead) costs, types of overheads, definition and importance of neural networks and their application in construction, previous work done on cost estimation using neural network models, factors affecting indirect cost percentage of projects, and lastly the definition of macroeconomic indicators and their effect on indirect cost.

2.2. Cost Estimation

Multiple researches have shown that cost estimation is a crucial component of the construction industry. As mentioned by the Project Management Institute (2013), cost estimation can be defined as “the process of developing an approximation of the monetary resources needed to complete project activities.” The Society of Cost Estimating and Analysis (SCEA) defines cost estimating as “the art of approximating the probable worth or cost of an activity based on information available at the time” (El-Sawalhi and Shehatto, 2014). The significance of the accuracy of the cost estimate has been clarified by several researchers. Hatamleh et al. (2018) clarified that cost estimation is crucial to the success of any construction project and needs to be considered at the beginning of the project. Furthermore, inaccurate cost estimates would lead to the project being over budget and behind schedule (Hatamleh et al., 2018). Enshassi et al. (2013) highlighted that the accuracy of the cost estimate is a key factor in the success of construction projects. Additionally, accurate cost estimates enable project managers to successfully manage the project’s expenses (Matel et al., 2019). In fact, overestimated costs would lead to significantly high tender price, which will affect the contractor’s ability to win bids (Enshassi et al. 2013), while underestimated costs will surely lead to contractors experiencing considerable losses (Avinash et al. 2018). Thus, the ultimate goal would always be to achieve highly accurate cost estimates.

Construction costs are composed of multiple different factors. As mentioned by the AACE International Recommended Practice (2017), construction costs are generally “the sum of all costs, direct and indirect, inherent in converting a design plan for material and equipment into a project ready for start-up.”. They further explained that construction costs do not only include items that are necessary in the production operation, they may also include other items like “the sum of field labor, supervision, administration, tools, field office expense, materials, equipment, taxes, and subcontracts.” Accordingly, construction costs mainly consist of direct costs and indirect costs (Stolz, 2010).

2.3. Direct Cost

Direct costs are generally any costs that can be directly attributable to the physical construction of the project on site. In other words, direct costs are costs that can be directly assigned to a particular activity within a project with a high degree of accuracy. In terms of construction work, the direct cost is mainly the cost of material, labor and equipment as well as the cost of subcontractors essential for the physical completion of the project (AACE, 2013).

2.4. Indirect Cost (Overheads)

Indirect costs, also commonly referred to as overheads, are costs incurred for a common or joint purpose and thus cannot be directly attributed to a certain activity within a project. These costs are typically assigned to all activities of a project on a predetermined basis. In construction projects, indirect costs are expenditures which do not become a part of the physical execution on site but are vital for the orderly completion of the project. Examples of indirect costs in construction projects include field administration, contractor’s fees, direct supervision, insurance and taxes (AACE, 2013).

In literature, overhead expenses are frequently discussed. There are generally main key research trends that may be used to categorize pertinent studies on overhead expenses. For example, some researchers concentrated on understanding and analyzing the concept of overhead costs, the relation between construction delays and overhead costs, the way companies distributed and allocated their overhead costs and how fixed expenses can be recovered (Lesniak and Juszczyk, 2018). In addition, several recent researches have focused on estimating the overhead cost of projects in different countries (i.e. USA, Canada, Egypt, Jordan, etc.) and through various ways, all of which will be explained in more details in the coming sections of this literature review (Al-Tawal et al., 2020; ElSawy et al., 2011; Lesniak and Juszczyk, 2018).

2.5. Types of Overheads

In general, researchers have categorized overhead costs of contractors into two different types: site overheads and general overheads (Bakr et al., 2018; Lesniak and Juszczyk, 2018; Patil and Bhangale, 2014). Site overhead costs are all expenses incurred by the contractor on the construction site for the purpose of completing the works, excluding direct costs (Lesniak and Juszczyk, 2018). In other words, they are site-related costs that are essential for activities to be completed but cannot be allocated to a certain activity within the project (Bakr et al., 2018). Examples of site overhead costs include staff salaries, site's safety provisions, transportation of site operatives, equipment, and site accommodation (Bakr et al., 2018). On the other hand, general overhead costs are ones that are needed for the purpose of running a business and keeping it in operation which cannot be assigned on one certain project, and thus are divided amongst all projects within the business (Bakr et al., 2018). These costs include office salaries, office rent, sales and marketing costs, software costs, office furniture and many other items that are related to running the head office of the company (Lorman, 2014).

2.6. Factors Affecting Indirect Cost of Construction Projects

The factors that have a direct effect on indirect cost of construction projects have been widely discussed in several researches worldwide. According to the study done by Awad (2017), the top major factors that are considered when estimating the project's overheads are the contractor's experience in executing similar projects, difficulty in obtaining materials, number of similar projects implemented in the same year, project size, payment method/schedule, firm's need for work, economic inflation and contracting company's system for cost control, monitoring and evaluation. Furthermore, Chan (2012) indicated that the most influential factors on indirect cost are project size, project duration, project complexity, contract type, tendering method, firm's need of work and inflation rate. According to Lesniak and Juszczyk (2018), two of the major critical factors that affect the project overhead costs are the method of work and the location of site. In another research done by Lesniak and Juszczyk (2019), it was stated that other major factors governing the indirect cost of projects are the class of the construction company, project type, project size, staff and equipment requirements, annual work volume, and the local economic conditions. Furthermore, Hesami and Lavasani (2014), in their study on factors influencing construction overhead costs in Iran, highlighted other effective factors that must be taken into consideration like the prevailing regional economic conditions and certain macroeconomic factors like inflation rate, interest rate, governmental laws and taxes of the country in which the project takes place.

The factors affecting the indirect cost of construction projects in Egypt specifically have also been widely discussed in literature. ElSawy et al. (2010) proposed a list of governing factors that contribute to the site overheads' percentage in construction sites in Egypt. These factors are as follows: "construction firm category, project size, project duration, project type, project location, type-nature of client, type of contract, contractor-joint venture, special site preparation requirements, project need for extra-man power." Additionally, after analyzing the data obtained, ElSawy et al. (2010) confirmed that the 5 factors that have the largest influence are the project duration, project type, total contract value, project location and special preparation needs for site. ElSawy et al. (2010) further added that factors like nature of the client and contractor-joint venture do have an effect on overheads percentage, but not as significant as the other factors.

Furthermore, the findings of another questionnaire done by Bakr et al. (2018), which was sent to experts with more than 10-year experience in the field, have shown that the major factors contributing to site overheads percentage of residential construction projects in Egypt are type of contract, class of company, project duration, project location, project direct costs and whether the company is a public or private one. Similarly, Idrees et al. (2023) have conducted a survey that was sent to a number of experts in order to identify the main factors influencing the percentage of site overheads in Egyptian construction projects and the results of this survey, along with the data they obtained through an extensive literature review, included the same factors previously reached through the research done by Bakr et al. (2018), except for one additional factor which is inflation rate. However, the inflation rate was not one of the factors considered in the model they created.

Also, Othman (2020) considered the same factors in his study on indirect cost estimation in Egypt's construction projects but had an additional factor taken into consideration, which is client type. However, it can be noted that most of the models previously developed in researches done on this topic in Egypt only considered factors related to the project characteristic and did not consider the effect of macroeconomic indicators, like inflation rate and interest rate, on the overheads percentage, which as recommended by Al-Tawal et al. (2020) would highly improve the reliability of the model if considered.

In conclusion, according to literature, there are numerous different factors that can be considered as being highly influential on the indirect cost percentage of construction projects, whether in Egypt or worldwide. However, according to the analysis done by ElSawy et al. (2011), several important factors are being considered worldwide and are not currently being accounted for in Egypt. In addition, he added that contractors in Egypt tend to combine more than one governing item into one major factor, which is considered by researchers as unprofessional and inaccurate. Also, as it was stated previously, some factors like macroeconomic indicators are considered highly influential yet were not utilized in most of the previous models related to this topic done in both Egypt and worldwide. Therefore, it was important to cross-match between the data collected from different researches in Egypt and worldwide and create a final list of factors to be used later on in this research that would accurately represent the major factors contributing to the site overhead costs in construction sites.

Table 1 shows a matrix that includes all the most prominent factors indicated by different scholars. This table will then be used to choose the factors with highest effect on indirect cost of construction projects to be used in this research's model.

Table 1: Factors Affecting Indirect Cost Concluded from Literature Review

	Reference / Factor	ElSawy et al. (2011)	Lesniak & Juszczyk (2018), (2019)	Bakr et al. (2018)	Idrees et al. (2023)	Othman (2020)	Awad (2017)	Hesami & Lavasani (2014)	Al-Tawal et al. (2020)	Chan (2012)	Frequ-ency
1	Class of Contracting Company	✓	✓	✓	✓	✓		✓			6
2	Project Size (Project Direct Cost)	✓	✓	✓	✓	✓	✓	✓		✓	8
3	Project Duration	✓		✓	✓	✓		✓	✓	✓	7
4	Project Type	✓	✓			✓		✓			4
5	Project Location	✓	✓	✓	✓	✓		✓	✓		7
6	Client Type	✓		✓		✓	✓				4
7	Contract Type	✓		✓	✓	✓		✓	✓	✓	7
8	Economic Indicators		✓				✓	✓	✓	✓	5
9	Contractors – Joint Venture	✓				✓		✓			3
10	Special Site Preparation Requirements	✓									1
11	Project Need for Extra Manpower	✓				✓					2
12	Project Complexity				✓			✓		✓	3
13	Payment Schedule				✓		✓	✓			3
14	Client's Strictness							✓			1
15	Tendering Method				✓			✓		✓	3

Table 1: Factors Affecting Indirect Cost Concluded from Literature Review (Continued)

		ElSawy et al. (2011)	Lesniak & Juszczyk (2018), (2019)	Bakr et al. (2018)	Idrees et al. (2023)	Othman (2020)	Awad (2017)	Hesami &Lavasani (2014)	Al-Tawal et al. (2020)	Chan (2012)	Frequ-ency
16	Project Management Method						✓	✓			2
17	Method of Performing the Work		✓					✓			2
18	Number of Competitors						✓	✓			2
19	Contractor's Cash Availability							✓			1
20	Assigning Work to Subcontractors				✓			✓			2
21	Country of Performing the Project							✓			1
22	Required Quality Level of the Projects							✓			1
23	Work Scope							✓			1
24	Stakeholders' Profits							✓			1
25	Site Layout							✓			1
26	Contractor's Designing Necessities							✓			1
27	Experience in Similar Projects						✓	✓			2

Table 1: Factors Affecting Indirect Cost Concluded from Literature Review (Continued)

		ElSawy et al. (2011)	Lesniak and Juszczyk (2018), (2019)	Bakr et al. (2018)	Idrees et al. (2023)	Othman (2020)	Awad (2017)	Hesami and Lavasani (2014)	Al-Tawal et al. (2020)	Chan (2012)	Frequ-ency
28	Volume of Work in Construction Market							✓			1
29	Annual Work Volume		✓					✓			2
30	Staff and Equipment Requirements		✓					✓			2
31	Governmental Laws and Taxes							✓			1
32	Firm's Need of Work						✓	✓		✓	3
33	Project Schedule						✓				1
34	Difficulty in Obtaining Materials						✓				1
35	Similar Projects Implemented in the Same Year						✓				1
36	Supervision and Consulting	✓									1
37	Special Site Preparation	✓				✓					2
38	Project Delays	✓									1
39	Project Cash flow	✓									1
40	Specialized Subcontractor	✓									1

As shown in the Table 1, the total number of factors obtained from literature was 40 different factors. To prioritize the most significant factors, a frequency analysis was conducted, focusing on those most commonly mentioned in the literature. Subsequently, the Pareto principle (80/20 rule) was applied to select 20% of these factors that were identified by 80% of the scholars. This approach, which states that around 80% of outcomes come from 20% of causes, ensured that the chosen factors were both highly influential and representative (Dunford et al., 2021). The factors identified through this process as the most influential on indirect costs are presented in Table 2.

Table 2: Most Influential Factors on Indirect Cost

1	Project Type
2	Project Location
3	Project Duration
4	Contract Type
5	Project Size (Total Contract Amount or Total Direct Cost)
6	Client Type
7	Class of Contracting Company
8	Macroeconomic indicators (inflation rate, interest rate, exchange rate)

The relation between each of these factors and the percentage of overheads in construction projects will be clearly explained in the coming section.

2.6.1. Project Type

In the construction field, there is a wide variety of project types. Generally, the main project types include residential buildings, commercial buildings, administrative buildings or infrastructure projects. According to ElSawy et al. (2010), project type can be considered as the third most significant factor affecting the overhead costs. Chan (2012) explained that the type of project has a direct influence on the overhead costs as it decides the amount of involved jobs in the project as well as the required coordination, supervision, safety and transportation which differs from one type of project to another. In other words, each type of project requires a different amount of resources, which accordingly means a different amount of overhead costs (Hesami and Lavasani, 2014). Furthermore, ElSawy et al. (2010) clarified that each type of project requires different construction methods, resources, architectural designs, quality management plans, safety regulations, overall construction requirements and client demands.

2.6.2. Project Location

Enshassi et al. (2008), in their study on overhead costs in construction projects, have stated that the project location could be considered as the most important factor influencing the overhead costs. Generally, researchers classify project locations into two main categories: inside the capital city or outside the capital city (rural areas) (Idrees et al., 2023; ElSawy et al., 2010; Bakr et al., 2018; Chan, 2012). The importance of this factor is mainly due to the fact that it is directly attributable to the amount of extra services and resources needed on site (ElSawy et al., 2010). According to Chan (2012), the location of the project has an effect on several components of the project's overhead costs like transportation costs, cost of temporary facilities, providing and maintaining offices, amount of importation and security costs. Bakr et al. (2018) further highlighted that projects outside the capital city tend to require more fuel consumption, higher accommodation rates and significantly higher cost of delivery for materials to site. Thus, it can be clearly said that projects located outside the capital city (in rural areas) usually tend to have a higher percentage of site overheads than those inside the city (Bakr et al, 2018).

2.6.3. Project Duration

The majority of studies consider the project duration as a significant factor influencing overhead expenses (Hesami and Lavasani, 2014; ElSawy et al., 2010; Idrees et al., 2023; Bakr et al., 2018; Chan, 2012). In the study done by ElSawy et al. (2010), it was proven that the relationship between the project duration and the percentage of site overheads in construction projects in Egypt is clearly directly proportional. In fact, Hesami and Lavasani (2014) mentioned that the duration-related expenditures generally account for over 45% of the project's overhead costs. Furthermore, the significant possibility of a project delay makes this aspect much more crucial in determining overhead expenses (Chan, 2012).

2.6.4. Contract Type

According to ElSawy et al. (2010), the project's contract type is considered a crucial element. In most of studies done on projects in Egypt, researchers were mainly concerned with 2 types of contracts as they are the most commonly used in construction projects in Egypt, and they are: fixed price contracts (lump-sum and unit price) and cost plus contracts (Idrees et al., 2023; ElSawy et al., 2010; Bakr et al., 2018). In the study done by ElSawy et al. (2010), it was shown that fixed price contracts are even more commonly in Egypt than the cost plus contracts. The study also demonstrated that there is a clear difference between the percentage of overheads in fixed price contracts and in cost plus contracts. In fact, ElSawy et al. (2010) mentioned that the percentage of site overheads in projects with fixed priced contracts in Egypt is lower than that of other types of contracts. In contrast, Bakr et al. (2018) claimed that projects with lump-sum contracts tend to have a higher percentage of site overheads than other types due to the need to cover the risk of cost overrun that usually accompanies this type of contracts.

2.6.5. Project Size (Project's Total Contract Amount)

El Sawy et al. (2010) clarified that the relationship between the total contract value and percentage of overheads in construction projects is directly proportional. This means that as the total contract value increases, the percentage of overheads also increases. However, he further highlighted this is generally applicable until a certain contract amount is reached after which the percentage of site overheads does not experience a considerable increase. Furthermore, Hesami and Lavasani (2014) explained that reason behind this directly proportional relationship is increasing the size of project will definitely lead to requiring more time, resources and staff do the job, which as a result will increase the overhead costs.

2.6.6. Class of Contracting Company

According to the Egyptian Federation for Construction and Building Contractors, contracting companies in Egypt can be classified into 7 grades. Although this factor counts as an important factor which affects the percentage of overheads, ElSawy et al. (2010) ranked it as the tenth most effective factor on overheads in Egypt. The study done by ElSawy et al. (2010) also showed that grade A contracting companies tend to have more overhead costs than grade B companies, and this is due to the former having longer duration projects, more quality management expenses, and larger sizes of projects than the latter. Furthermore, it is evident that organizations with higher grades require larger workspaces, equipment, and repositories in addition to better and more knowledgeable personnel, which results in higher overhead expenses (Hesami and Lavasani, 2014). In support of this, the analysis of the study done by Idrees et al. (2023) also showed that as the grade of the contracting company improves, its overhead costs increase. However, in the study done by Bakr et al. (2018), it was stated that construction companies in the second group have the highest average percentage of site overheads due to the first category companies' strict control system, as opposed to the second category, which has a somewhat weaker control system.

2.6.7. Client Type

Client type is considered as an important factor that governs the percentage site overheads, as mentioned by ElSawy et al. (2010). In general, according to ElSawy et al. (2010), the type of clients dealt with in the Egyptian construction industry can be categorized into two types: public entities or private companies. It was found that the percentage of site overheads in projects where the client is a public entity (the government) is often lower than in cases where the client is a private entity. The reason behind this is usually that projects with private entities tend to usually demand more quality control measures, technical engineering requirements and strict project management plans.

2.6.8. Macroeconomic Indicators

Macroeconomic indicators are figures or data values that indicate the state of the economy in a specific nation, area, or industry. Analysts and governments use them to evaluate the state of the economy and financial markets both now and in the future. The macroeconomy has a significant impact on the construction industry's performance, making it susceptible to macroeconomic fluctuations. This may, in some circumstances, lead to this industry being insolvent (Puci et al., 2023). In fact, a country's economic standing and the success of its construction sector are closely correlated, and any fluctuations in the world economy will surely lead to an increase in the uncertainties in the construction sector (Fan et al., 2010). The construction industry is affected differently by economic shifts; some construction projects are finished at a higher cost than anticipated, while others are canceled because they are not financially feasible (Shiha, 2019).

According to research outcomes of Puci et al. (2023), the leading macroeconomic indicators that have a considerable effect on the construction industry are gross domestic product (GDP) growth, interest rate, inflation rate and foreign currencies exchange rate. Warsame (2006) indicated that inflation and interest rates can be considered as the most significant factors influencing the cost of construction projects. In addition to that, Enshassi et al. (2008) confirmed that inflation is mainly the most important factor that leads to the increase in overhead costs. According to the research done by Musarat et al. (2020), inflation can be considered as a major problem that negatively impacts the construction industry. That is mainly because inflation, which is usually ignored in most of the construction projects' budgeting, leads to cost overruns and deviations from initial project budget since inflation causes increases in material prices, labor wages and equipment rates on a yearly basis. Musarat et al. (2020) also added that although inflation was not paid attention to before, it is now starting to be highly involved due to technological developments, large-scale construction projects, the complexity of the client, and several additional variations to the original project in order to meet the client's requirements. Furthermore, extreme fluctuations in interest rates have the potential to threaten any construction company's profits, increase expenses, and affect the worth of its assets and future cash flows (Puci et al., 2023).

2.7. Neural Networks Definition, Properties and Comparison with Other Types

Through modeling and simulations, the behavior of a project's lifetime and its patterns may be predicted before the project actually starts (Mackenzie and Briggs, 2006). It has been demonstrated that using artificial intelligence, such as expert systems and neural networks, is helpful in solving problems related to prediction and estimation (Cheng et al., 2010). An artificial neural network (ANN) is a paradigm for information processing that is inspired by how the brain and other biological nervous systems handle data. It is made up of several densely interconnected processing units, or neurons, that work together to solve particular problems and produce a certain output. It is a feature of a computer system that emulates the data analysis and processing capabilities of the human brain (Dastres and Soori, 2021). Artificial neural networks are primarily used for prediction, estimation, control, association, categorization, and recognition of patterns. They are also used for analysis of data, optimizing, and data association. ANNs started being utilized for the purpose of managing construction projects in the early 90s of the past century (Lesniak and Juszczyk, 2018). Numerous attempts have been made up until today to employ artificial neural networks in engineering construction processes to address problems like implementation time analysis, productivity, and efficiency of construction projects (Dikmen and Sonmez, 2011), forecasting the cost of construction equipment maintenance, predicting whether a new technology can be adopted, and simplifying the processes of decision-making in construction projects (Lesniak and Juszczyk, 2018).

ANNs and conventional computing vary in their pattern recognition mechanism. When faced with a particular problem, a conventional programming algorithm follows a series of guidelines or instructions to find the answer. These guidelines or directives are converted into signals that the computer software can understand and use to get the answer. As a result, the problem's interpretation and resolution by the computer are restricted to the programmer's methods. Conventional computers execute jobs sequentially using an algorithmic method, and they are unable to solve a problem unless they are aware of the precise steps that need to be taken. ANNs, on the other hand, are not deterministic nor generally sequential; instead of having a single complicated central processor, they contain a number of simpler ones that accept the input's weighted total from other processors and alter it in accordance with patterns using a kind of learning rule (Al-Tawal et al., 2020). Another significant distinction between ANNs and traditional computers is how they operate. Conventional computers solve problems using cognitive methods, which requires that the problem be understood and expressed in brief, clear instructions. The answer can only be obtained by following the stages or sequences of the algorithm. The orders that are received are then converted into a computer-understandable coding language. In contrast, the neurons in ANNs are the components that collaborate to solve problems. Neurons are not programmed to do a specific job, they rather learn from examples to be able to respond to the provided information in along with the input patterns (El-Sawalhi, 2014).

2.8. Previous Work on Cost Estimation Using Neural Networks

In fact, artificial neural networks (ANNs) have been used in a number of earlier researches in various countries to provide construction estimates, with differing degrees of success. Thus, this section will mainly be discussing some of the ANN models developed in past years for the purpose of cost estimation.

Hegazy and Amr (1998) used ANNs in order to efficiently handle construction cost data and create a parametric cost estimating model for highway projects. The cost data of this model was derived from 18 real examples of highway projects built in Newfoundland, Canada.

Later on, numerous other ANN models were developed for the purpose of cost estimating.

Gunaydin and Dogan (2024) developed an ANN model to forecast, at an early stage of the design process, the cost per square meter for buildings using reinforced concrete structural systems. Thirty construction projects comprise the set of data used in the network development. The created ANN architecture comprised of 1 output layer and 8 input layers. The model that was produced had a 93% accuracy in calculating the price per square meter of residential constructions with reinforced concrete structural elements.

In 2005, Georgy and Barsoum created an ANN model to estimate the cost of constructing schools in Egypt. They employed neural network and statistical models to perform the cost estimation and their findings indicated that a neural network with a hidden layer that consists of an number of neurons equal to two-thirds of the input layer's neurons would be enough.

Luu and Kim (2009) developed an ANN model to calculate the total cost of construction projects in Vietnam. MATLAB software was used in the creation of the ANN. In order to use the neural network in real-world applications, Visual C++ was used to construct the program. The results showed that neural networks are an effective tool for cost modeling and that it is possible to anticipate the overall building costs for residential developments in Vietnam using ANN.

Zheng, Chen and Yan (2010) gathered data from road projects in particular regions developed between 2000 and 2002 and suggested an ANN model to predict the construction costs of highway projects. The model produced outputs with less than 5% is the relative error between the estimated and actual values. With this relatively small error, the model has proven that predicting the construction costs of highway projects and meeting the criterion for cost estimation are practically achievable using ANNs. It also demonstrates the ANN models have a high potential for accuracy and generalization. In addition, the results of the model also prove ANN technology can be used in cost estimation of highway construction to reach precise outcomes.

Arafa and Alqedra (2011) created an artificial neural network model in the Gaza Strip to forecast building costs early in the project. The learning database comprised 71 building projects from the Gaza Strip. The final ANN model architecture consisted of 1 hidden layer and 7 neurons. Without requiring in-depth project information, the developed model has shown to be quite successful in forecasting construction costs early on.

Roxas and Maximino (2014) created an AAN based on six parameters in order to calculate the overall structural cost of building construction projects in the Philippines. Data from thirty distinct building construction projects make up the data utilized for the learning process. The final model design consisted of a network of 1 output neuron (the total structural cost), 6 input neurons and 1 hidden layer of 7 neurons.

Yadav et al. (2016) created an ANN model that estimates the construction cost of residential projects through using distinctive parameters as inputs. The database of this model includes projects executed throughout the 23 years prior to the year 2016. After testing the model, the selected model architecture consisted of 8 input variables. Also, the ANN model was able to accurately estimate the total construction costs with R2 equal to 0.9905, which further proves the high accuracy of ANN models in cost estimation.

Lesniak and Juszczyk (2018) created a regression model, using ANN, that predicts the site overhead costs of construction projects in Poland. A database including 143 cases of finished construction projects was utilized to create the model. Several multilayer perceptron artificial neural networks, each with different architectures, activation functions, and training techniques, were used in the modeling process. The neural network that was chosen to be the central component of the created model is able estimate site overhead costs in the early phases of a building project with a reasonable degree of precision.

Al Tawal et al. (2020) established a model that utilized ANN in the cost estimation of construction projects in Jordan. To create, train, and evaluate ANN models, cost and design data from 104 projects completed in Jordan during the five years prior to 2020 were utilised. The first ANN model was developed at the detailed design stage using 53 design elements; the factors were subsequently reduced to 41 and used to construct the second prediction model at the schematic design stage. Ultimately, the third ANN model made use of 27 design parameters that were accessible during the concept design phase. The findings of this model have shown that in the stages of detailed, schematic, and concept design, the models' average cost estimation accuracy was 98%, 98%, and 97%, respectively.

2.9. Previous Work on Indirect Cost Estimation Using Neural Networks in Egypt

When looking at Egypt specifically, few researches covered the idea of predicting indirect costs of construction projects using ANN models. Firstly, ElSawy et al. (2011) developed a parametric cost estimating model using ANN to predict the percentage of site overheads in Egypt. In order to develop this model, 52 actual cases of construction projects that took place between 2002-2009 in Egypt were used. In addition to this, the main factors affecting the site overheads were identified using questionnaires send to academicians and experts in the field. The model mainly comprised of 1 output layer, 1 input layer with 10 neurons and 1 hidden layer with 13 hidden nodes and a sigmoid transfer function. The model was tested using 5 new projects and the testing's outcomes showed an accuracy of 80%.

Bakr et al. (2018) established a neural network model to predict the percentage of site overheads for residential projects in Egypt. Data related to the major factors affecting site overheads were identified using a structured questionnaire send to experts in the field. Additionally, data for 55 distinct projects that took place between 2011 and 2018 was gathered, each with varying requirements regarding company class, project location, project duration, total direct cost and contract type. This data mainly served as a database for the neural network's learning process. The developed model consisted of one output layer (percentage of site overheads), 6 input neurons, 1 hidden layer with 6 neurons and another hidden layer with 5 neurons. The developed model was tested and the results of this testing showed that the model is accurate by 83.3%.

Othman (2020) developed an ANN that predicts and estimates the percentage of site overheads for construction projects in Egypt. To create this model, the author mainly used data from 40 projects of different types (residential, administrative, banks, hotels, schools). The data collected was related to seven different factors that the author found were most effective on site overheads through an extensive literature review, these factors were mainly project type, project location, project duration, project budget, company ranking, contract type and client type. The model's architecture comprised of 1 output variable, 7 input variable, 1 hidden layer with 5 neurons and another hidden layer with 3 neurons. After developing the model, he tested it using 5 projects that were not used beforehand in creating the model. The linear regression analysis's results revealed a correlation coefficient (R^2) of 0.888.

Idrees et al. (2023) have recently developed an ANN to predict and estimate the percentage of site overheads in commercial projects in Egypt. Major factors affecting site overheads of commercial projects in Egypt were identified using structured data collected from experts. Furthermore, the model was developed using data from 55 different projects that took place in the period from 2017-2023 and that had distinct conditions in terms of company ranking, project location, project duration, total direct cost and contract type. Similar to the model developed by Bakr et al. (2018), the developed model contained one output layer (percentage of site overheads), 6 input neurons, 1 hidden layer with 6 neurons and another hidden layer with 5 neurons. The developed model was tested using 6 new projects and the results also showed that its percentage of accuracy is 84%.

2.10. Gap in Literature Review

As it can be noted from earlier parts of this literature review, most research done in this area in Egypt and worldwide only took into consideration internal factors related to the project's specific characteristics and did not take into account any external factors such as the leading macroeconomic factors (inflation rate, interest rate and exchange rate) which have been proven to be highly influential on the project's site overheads and would highly improve the reliability of the model if considered. Not only this, but also most of the previously created models in this area had only one output, which is the percentage of site overheads, and did not predict the percentage allocated to each subcategory of site overheads nor the distribution of the total site overheads over the duration of the project. Hence, the main aim of this research is to develop an ANN model that takes into consideration internal and external factors in order to predict the percentage of site overheads in construction projects, with the model being applied in the Egyptian construction industry. This research also aims at reaching results with higher accuracy by using a database of more recent data and developing a model that produces more than one output: percentage of site overheads and the percentages of site overheads allocated to its different subcategories and to the different phases of the project's construction process.

3. Research Methodology

3.1. Introduction

The main aim of this research is to enhance the contractor's ability to accurately predict the percentage of site overheads in construction projects in Egypt through the following:

- identifying and analyzing the key factors influencing site overheads in the Egyptian construction industry.
- developing a robust dataset containing historical cost data of projects executed in the past 10 years.
- developing ANN models capable of accurately predicting the total percentage of site overheads for construction projects in Egypt and subsequently allocating these costs across different site overheads subcategories and across the different project construction phases (initiation, growth, maturity and decline), incorporating both economic and non-economic variables to enhance the predictive accuracy

In order to be able to do this, the methodology illustrated in Figure 1 was followed.

Phase 1: Literature Review

An extensive literature review was conducted to establish a comprehensive understanding of indirect cost, particularly site overheads, in the construction industry. The review focused on identifying key factors influencing indirect costs both globally and within the Egyptian construction industry. These factors were subsequently used as input variables for the ANN model. In addition, the review involved a thorough examination of existing studies to understand the complexities of indirect cost management and assess the applicability of Artificial Neural Networks (ANNs) in indirect cost prediction. Furthermore, the review sought to identify research gaps, thereby providing a clear rationale for this research and contributing to further advancements in this field. Through a systematic analysis of relevant literature, this phase laid the groundwork for the subsequent data collection and model development stages.

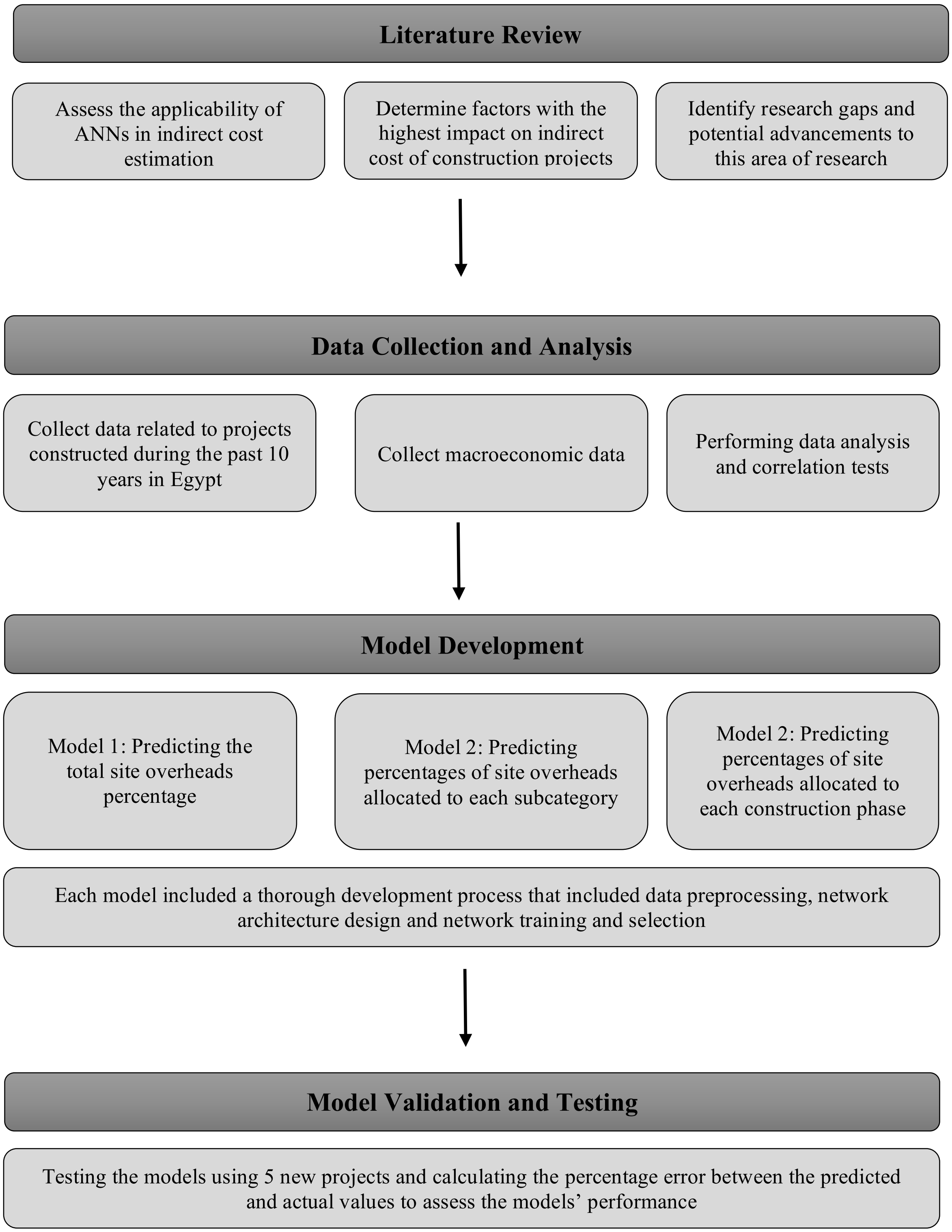

Figure 1: Research Methodology

Phase 2: Data Collection and Analysis

In order to form the model's learning database, data was collected through a structured questionnaire distributed to industry experts. Respondents were asked to provide comprehensive data on existing construction projects completed within the past 10 years. The questionnaire first focused on capturing detailed information related to the factors that have the highest impact on site overheads, which were primarily identified in the preceding literature review. These factors included project type, project size, project location, project duration, client type, contract type, class of contracting company. Additionally, respondents were requested to provide cost data, which included total actual contract amount, direct cost and indirect cost for each project as well as the allocation of the total actual site overheads across various indirect cost categories (e.g., salaries, equipment, transportation) and across the project's construction phases (initiation, growth, maturity, decline). To enhance data quality and reliability, a rigorous data cleaning process was implemented. Outliers and inconsistencies were identified and addressed to ensure the accuracy and validity of the dataset.

To incorporate macroeconomic factors, monthly data of interest rates, inflation rates, and exchange rates were collected from the online website of the Central Bank of Egypt. These 3 indicators were specifically selected based on the literature review findings which identified them as highly influential factors on site overheads. For each project collected from experts, average values for these macroeconomic indicators were calculated over the project duration. These computed averages were subsequently integrated into the dataset as additional input variables for the ANN models.

The next step was the analysis of the collected data, which was performed to examine the effect of each factor (independent variable) on the total site overheads percentage as well as on the percentage allocated to each subcategory of site overheads and to each project construction phase. The last step in this phase was conducting correlation and multicollinearity tests among the independent and dependent variables to identify key factors influencing site overheads, eliminate variables with negligible impact, and address potential multicollinearity issues to ensure model robustness.

Phase 3: Model Development

The third phase of the research involved the development of three Artificial Neural Network (ANN) models using Python Programming Language on Google Colaboratory (2024). These models were developed using a thorough model development process, including data preprocessing (data coding and normalization), architecture selection, and training using the prepared dataset. The first model was designed to predict the overall percentage of site overheads based on the collected dataset of project characteristics, project cost data and macroeconomic indicators.

The second model utilized the predicted total site overhead from the first model as an input, along with other project-specific data, to forecast the distribution of site overheads across different site overheads subcategories. These subcategories are:

1. Salaries and wages: salaries for project staff and labor wages
2. Site facilities: water and sewage network for site, site electricity, office furniture, etc.
3. Accommodation: any costs related to staff and labor accommodation
4. Mobilization/Demobilization: costs incurred during the initial setup and final dismantling of the project site
5. Communication and IT: expenses related to communication systems, internet access, and computer equipment.
6. Site equipment: costs of renting or purchasing heavy equipment, tools and scaffolding
7. Quality and safety expenses: costs for ensuring quality and safety standards are met
8. Material and personnel transportation: costs of transporting materials and personnel to and from the project site
9. Engineering fees: fees paid to external engineering consultants for design
10. Owner and consultant expenses: costs incurred by the project owner or consultant

These 10 subcategories of site overheads selected for this model were identified and chosen based on their frequent appearance in literature. In addition, their selection was also based on consulting experts in the field who further confirmed that these were the main subcategories constituting site overheads in Egyptian construction projects. The specific definition and scope of each category were also determined through expert consultation.

The third model followed a similar approach to the second, but focused on allocating the total site overheads over the duration of the project by predicting the allocation of site overheads across the different project construction phases that make up the S-curve of any project. According to the Project Management Institute (PMBOK 2013), the S-curve is generally a graphical display of the project's cumulative costs plotted against time. The S-like shape of the curve resembles the typical behavior of projects during the construction phases, in which they start slowly, accelerate then tail off as the project reaches completion (PMBOK 2013). Forbes and Riso (2024) further explained that the S-curve can be segmented into four distinct phases, each phase representing critical stages of the project's construction process. These phases are:

1. Initiation: the phase at which project begins slowly, with minimal activity and spending.
2. Growth: the phase at which the project accelerates rapidly, with significant resource allocation and construction.
3. Maturity: the phase at which the project's pace stabilizes as it approaches completion.
4. Decline: the phase at which the project winds down, with activities and expenses decreasing.

Accordingly, the third model will be predicting the percentages of site overheads allocated to each of these 4 phases.

Phase 4: Model Validation and Testing

The final phase of the research involved model validation and testing. To assess the predictive accuracy of the developed ANN models, data from five construction projects, excluded from the training dataset, were utilized. The models were tested using these new projects to generate predictions for total site overheads percentages, percentages allocated to each subcategory, and percentages allocated to project construction phases. The accuracy of these predictions was evaluated by comparing them to the actual values. Percentage error calculations were obtained to quantify the models' performance, providing insights into their predictive capabilities and reliability.

3.2. Data Collection

3.2.1. Questionnaire Design

To form the learning database of the model, a questionnaire form was designed specifically to gather information in order to meet the intended outputs of the model. The questionnaire form was prepared and sent to experts in the field asking them to provide data related to projects started and completed in the past 10 years. Each participant was asked to fill the questionnaire with data for a minimum of 1 project. Analysis of the questionnaire's respondents is provided later in Chapter 4.

The questionnaire consisted of 5 main sections. Section 1 asked the participant to provide their personal information like the department they currently work at, the contracting company they currently work for and their years of experience.

Section 2 asked the participants to provide general data about the project like its name, type, location, duration, start and finish dates, contract type, client type and category of contracting company.

Section 3 asked the participants to provide some cost data related to the project like the total actual contract amount, total actual direct cost and total indirect actual cost.

Section 4, the asked the participants to provide the actual costs for the indirect subcategories which are: salaries and wages, site facilities, accommodation, mobilization and demobilization, communication and IT, site equipment, material and personnel transportation expenses, quality and safety expenses, engineering fees, and owner and consultant expenses.

Section 5 asked the participants to distribute the actual indirect cost over the duration of project by providing the total actual percentage indirect cost incurred during the 4 main phases of any project's construction cycle or "s-curve" which are: initiation, growth, maturity and decline phase. The data collection questionnaire form sent to the experts is shown in Figure 2.

Data Collection Questionnaire Form	
Section 1: Personal Information	
Name:	Email:
Years of Experience:	Current Position:
Section 2: Project General Data	
Project Name:	Name of Contracting Company:
Category of Contracting Company: O A O B O C O D	
Project Type: O Residential O Administrative O Commercial O Infrastructure O Industrial O Landscape O Marine O Other	
Project Location:	Project Duration:
Project Start Date:	Project End Date:
Client Name:	Client Type O Private(National) O Private (International) O Public (Army) O Public (Government)
Contract Type: O Lump sum O Re-measured O Cost Plus O Other	
Section 3: Project Cost Data	
Total Actual Contract Amount:	Total Actual Direct Cost:
Total Indirect Actual Cost:	
Section 4: Project Indirect Cost Sub-Categories	
Salaries and Wages Actual Cost:	Site Facilities Actual Cost:
Site Equipment Actual Cost:	Personnel and Material Transportation Actual Cost:
Communication and IT Expenses Actual Cost:	Engineering Fees Actual Cost:
Accommodation Actual Cost:	Mobilization and Demobilization Actual Cost:
Quality and Safety Actual Cost:	Client/Consultant Expenses Actual Cost:
Section 5: Distribution of Indirect Cost Over Project Duration	
Total Actual Cost of Initiation Phase:	Total Actual Cost of Growth Phase:
Total Actual Cost of Maturity Phase:	Total Actual Cost of Decline Phase:

Figure 2: Data Collection Questionnaire Form

Gathering the above-mentioned data about real life projects executed during the past 10 years in Egypt was a difficult process as contracting companies consider such data confidential and find it hard to share with others. Thus, the approach used to gather this data was based on reaching personal contacts with Egyptian contracting companies.

Fundamental presumptions and criteria were implemented to address flaws in the collected data:

- Projects need to be built and completed entirely.
- The projects must be executed during the period from 2014-2024
- All missing, erroneous, or incomplete information were eliminated.
- Similar projects with the same values, or duplicate data, were removed.

3.2.2. Collection of Egypt's Macroeconomic Indicators

The main source for the 3 major macroeconomic indicators that have the highest impact on indirect cost of projects (inflation rate, interest rate and USD to EGP exchange rate) was the Central Bank of Egypt (CBE). Monthly data for the 3 indicators from 2014-2024 was obtained and used in the model's database. Samples of the collected data are shown in Figures 3, 4 and 5.

البنك المركزي المصري
CENTRAL BANK OF EGYPT

Weighted Average Interest Rates *

	EGP Deposits			EGP Loans
Month	> 1 Month <= 3 Months	> 3 Months <= 6 Months	> 6 Months <= 1 Year	<= 1 Year (Corporate)
Jan - 2024	14.7%	13.9%	13.7%	19.3%
Dec - 2023	14.2%	14.1%	13.6%	19.5%
Nov - 2023	14.4%	14.0%	13.5%	19.2%
Oct - 2023	13.9%	13.9%	12.9%	19.2%
Sep - 2023	13.9%	13.4%	12.7%	19.1%
Aug - 2023	13.4%	13.4%	12.3%	18.7%
Jul - 2023	13.3%	13.1%	12.2%	18.2%
Jun - 2023	12.2%	12.5%	11.9%	18.1%
May - 2023	12.3%	12.4%	11.8%	18.0%
Apr - 2023	11.7%	11.4%	11.5%	17.6%
Mar - 2023	11.0%	10.7%	10.8%	15.9%
Feb - 2023	10.8%	10.6%	9.8%	15.2%
Jan - 2023	10.7%	9.7%	9.5%	14.4%

Figure 3: Sample of Collected Interest Rates

Date	Currency	Buy	Sell
31/12/2023	US Dollar	30.8414	30.9386
28/12/2023	US Dollar	30.8414	30.9386
27/12/2023	US Dollar	30.8414	30.9386
26/12/2023	US Dollar	30.8414	30.9386
25/12/2023	US Dollar	30.8414	30.9386
24/12/2023	US Dollar	30.8386	30.9357
21/12/2023	US Dollar	30.8414	30.9386
20/12/2023	US Dollar	30.8414	30.9386
19/12/2023	US Dollar	30.8386	30.9357
18/12/2023	US Dollar	30.8414	30.9386
17/12/2023	US Dollar	30.8414	30.9386
14/12/2023	US Dollar	30.8414	30.9386
13/12/2023	US Dollar	30.8414	30.9386

Figure 4: Sample of Collected USD to EGP Exchange Rates

Date	Headline (y/y)	Core (y/y)	Regulated Items (y/y)	Fruits and Vegetables (y/y)
Jan 2024	29.800%	29.010%	24.020%	67.390%
Dec 2023	33.700%	34.180%	21.730%	78.720%
Nov 2023	34.550%	35.860%	21.580%	73.260%
Oct 2023	35.800%	38.100%	17.000%	89.220%
Sep 2023	38.000%	39.700%	17.880%	107.710%
Aug 2023	37.415%	40.380%	18.370%	86.330%
Jul 2023	36.460%	40.730%	17.470%	66.690%
Jun 2023	35.710%	41.000%	17.020%	46.450%
May 2023	32.750%	40.310%	12.730%	19.650%
Apr 2023	30.596%	38.575%	12.505%	7.065%
Mar 2023	32.665%	39.509%	11.302%	34.554%
Feb 2023	31.932%	40.262%	10.315%	18.206%
Jan 2023	25.834%	31.241%	10.136%	24.615%

Figure 5: Sample of Inflation Rates

Figures 6, 7 and 8 show the behavior of each of the 3 macroeconomic indicators in the period from 2014-2024 in Egypt, according to the data obtained from the online website of the Central Bank of Egypt.

Figure 6: Interest Rate from 2014-2024 (Central Bank of Egypt)

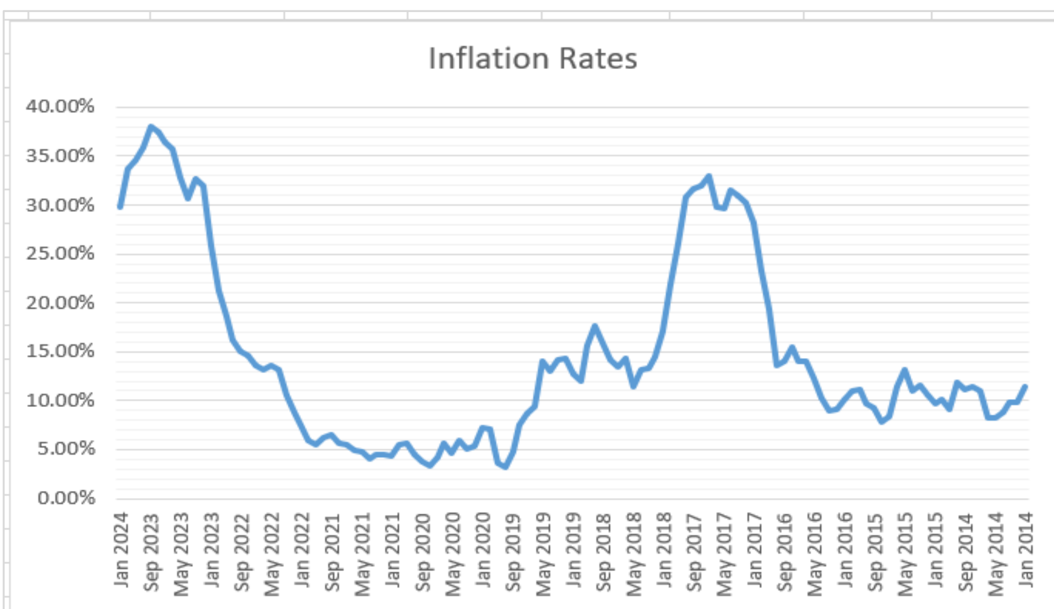

Figure 7: Inflation Rates from 2014-2024 (Central Bank of Egypt)

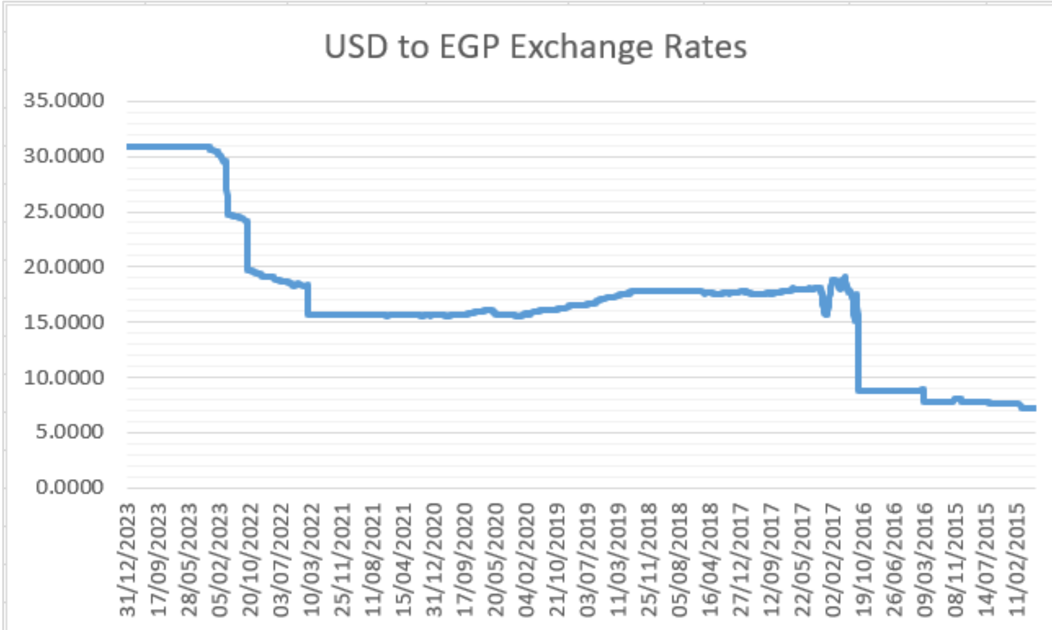

Figure 8: USD to EGP Exchange Rate from 2014-2024 (Central Bank of Egypt)

As shown in Figures 6, 7 and 8, the period from 2014 to 2024 was marked by significant fluctuations in Egypt's economic landscape, with interest rates, inflation, and the exchange rate of the Egyptian pound experiencing notable changes. Regarding the interest rate, a relatively stable interest rate environment prevailed prior to 2016. However, the subsequent years witnessed a sharp increase in interest rates as the Central Bank of Egypt sought to stabilize the economy and control inflation following the currency devaluation. Subsequently, there was a gradual decline in interest rates as inflationary pressures eased followed by a gradual increase again in late 2022. Regarding inflation rate, Egypt experienced a sharp surge in inflation rates starting in 2016. This was primarily attributed to the floating of the Egyptian pound. The inflation rate then gradually declined over the following years due to government interventions and economic stabilization efforts. However, the global economic landscape, particularly the COVID-19 pandemic and the Russia-Ukraine conflict, introduced new challenges, causing sharp escalations in Egypt’s inflation rate. Lastly, regarding the currency exchange rate, it can be noted that the Egyptian pound underwent significant fluctuations during the analyzed period. Prior to 2016, the Central Bank maintained a managed float exchange rate system, intervening to stabilize the currency. However, a substantial devaluation in the exchange rate took place in late 2016. Following this adjustment, the Egyptian pound exhibited relative stability until the year 2022 when it started experiencing significant depreciation in its value against the US dollar.

3.2.3. Sample Size

The number of construction projects collected during the data collection phase was 55 projects. In order to ensure that this number is considered a representative sample, the sample size was calculated based on Equation 1 (Israel, 1992).

$$n = (\frac{z * \sigma}{E})^2$$

Equation 1

where:

- n represents the sample size (the number of projects in this case)
- z is the z-statistic corresponding to your chosen confidence level. For 90% confidence level, z = 1.645.
- σ represents the standard deviation of site overheads of construction projects in Egypt. Since this number is difficult to determine due to the scarcity of data related to it, the standard deviation used in the sample size calculation was based on a combination of the findings of Chao (2008) and the results of this study. While Chao's study reported a standard deviation of 0.0428 for site overheads in construction projects, this value might not be entirely representative of the Egyptian context. Therefore, the standard deviation of site overheads for Egyptian construction projects was calculated using historical data from this study and found to be 0.0457. For this sample size equation, an average between both values was used, which was equivalent to 0.0443.
- E signifies the margin of error. For +-10% margin of error, E=0.1.

$$n = (\frac{1.645 * 0.0443}{0.1})^2 = 53 \, projects$$

The number of collected projects is 55 projects, which exceeds the required sample size.

3.3. Artificial Neural Network Models Guide

3.3.1. Reasoning for the Selection of ANN

Given the intricate nature of predicting site overheads and the absence of a linear relationship between the variables, ANNs were selected as the most appropriate machine learning technique for this research. ANNs, renowned for their ability to handle complex, nonlinear relationships and large datasets, are well-suited for the models developed in this research. Their capacity to learn from historical data and identify complex patterns, and adapt to changing conditions makes them particularly valuable in this context. Additionally, the backpropagation algorithm, a core component of ANNs, enables them to effectively learn from data, refine their predictions and generate outputs with high levels of accuracy (Nielsen, 2015).

3.3.2. Model Design Steps

To develop any ANN model, the steps shown in Figure 9 need to be followed (Hatem, 2009; ElSawy et al., 2011; Bakr et al., 2018; Idrees et al., 2023). Developing the model requires following these steps in an iterative manner of trial and error till reaching the model with the optimum architecture (ElSawy et al., 2011; Bakr et al., 2018; Idrees et al., 2023).

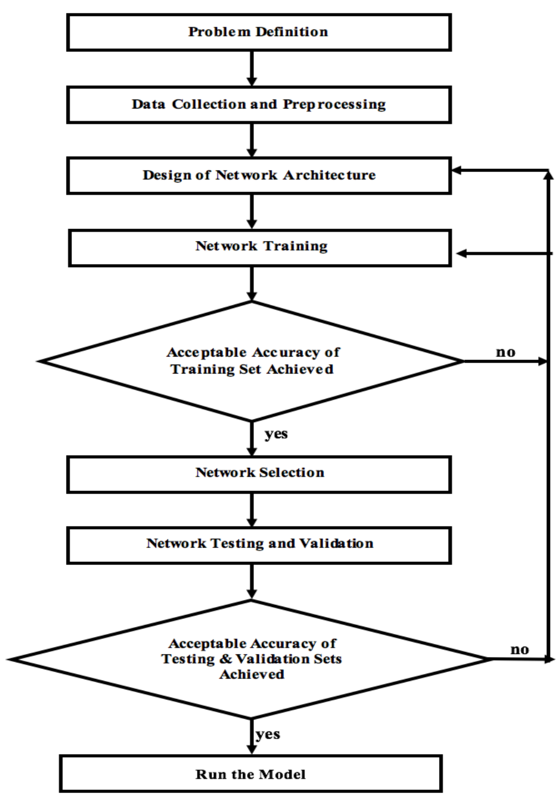

Figure 9: Steps for ANN Model

Problem definition and data collection were explained in the previous sections. Thus, the next step is data preprocessing.

3.3.3. Data Preprocessing

3.3.3.1. Coding of Categorical Data

Data sets usually contain all the information needed to build the model. Data sets usually consist of columns, representing variables, and rows, representing the projects. Variables are classified into independent (inputs) and dependent variables (outputs). In order to make sure that the projects data were being collected from a representative sample, sample size calculations should be done. After collecting all the necessary information and before importing it to any neural networks software, categorical data must be coded in order to be properly interpreted and utilized on the selected software. This is mainly because ANNs only deal with values in a numerical form (Kshirsagar and Rathod, 2012).

According to Potdar et al. (2017), the two most accurate coding techniques for categorical data in ANNs are label coding and one-hot encoding. Label coding involves assigning an integer to each category, which is a technique that does not add any new columns to the data but may imply an order to the variables that may not exist. One-hot encoding incudes representing categorical data as binary vectors, where each category is assigned a unique binary vector. Potdar et al. (2017) further explained that although the latter is more widely used but it could greatly increase the dimensionality of the dataset, especially if it contains a large number of distinct categories, which may impact the model's training time. Given the presence of several different categories in the dataset and the existing complexity of the models due to the large number of variables present within each model, label encoding was chosen for this research in order to avoid the increased data dimensionality associated with one-hot encoding which could potentially lead to longer training times and more complex models.

As shown below, Table 3 illustrates the coding technique used to code all the categorical inputs of this research's models in order for them to be understandable by any ANN software.

Table 3: Coding Scheme

Category of Contracting Company	
A	1
B	2
Project Location	
Inside the city/urban	1
Outside the city/rural	2
Project Type	
Residential	1
Administrative	2
Commercial	3
Infrastructure	4
Industrial	5
Landscape	6
Marine	7
Client Type	
Private Sector (National)	1
Private Sector (International)	2
Public Sector (Army)	3
Public Sector (Ministry/Government)	4
Contract Type	
Lump sum	1
Re-measured	2
Cost Plus	3

3.3.3.2. Normalization of Numerical Data

Regarding the numerical input data, each input has different measuring units (months, percentages, millions and billions). Thus, data normalization/scaling should be done in order to ensure that all inputs contribute proportionally during training, preventing inputs with larger scales from dominating the learning process. Without normalization, features with large scales can lead to exploding gradients in some layers and vanishing gradients in others. This disrupts the training process and hinders the model's ability to learn effectively.

There are several data normalization techniques that can be used, depending on the data and the specific needs of the machine learning model. For ANN models, the most common technique is min-max scaling. This technique scales each feature to a specific range, typically between 0 and 1 (or -1 and 1). Accordingly, the Equation 2 was used for each of the numerical inputs:

$$Scaled\ value = \frac{x - \min(x)}{\max(x) - \min(x)}$$

Equation 2

where x is the unscaled value of input, min(x) is the minimum value within the entire dataset related to this input, max(x) is maximum value within the entire dataset related to this input. In order to reverse the scaled values back to the original values, the following equation can be used:

$$Unscaled\ value = (scaled\ value * (\max(x) - \min(x))) + \min(x)$$

Equation 3

3.3.4. Design of Network Architecture

ANN model architecture refers to the general arrangement and structure of the network. It outlines the structure of the network, including the number of layers, the number of neurons within each layer, and how these layers are connected. The key components of an ANN architecture include input layer, hidden layers, output layer and the number of neurons within each layer. The input layer represents the data that will be fed into the network.

The hidden layers generally handle the network's central processing. They are not directly linked to the input or output and can have multiple layers on top of each other. In fact, the model's capacity and complexity are mostly determined by the number of hidden layers and neurons in each layer. In order to reach the optimum number of hidden layers and neurons in each layer, it is advisable to start with simple architecture that includes a single hidden layer with a moderate number of neurons then increasing complexity till reaching optimum performance of the model. According to Heaton (2017), there is no specific rule for determining the exact number of hidden layers that should be used within the ANN. In fact, in order to determine the most suitable number of hidden layers or the number of neurons within each layer, it is advised to follow one of two techniques. The first technique is trial and error till reaching the most optimum network architecture while the second technique involves utilizing optimization methods like genetic algorithm (Heaton, 2017).

However, Heaton (2017) suggested using some rule-of-thumb methods to determine the number of neurons within each layer to start with the trial and error process, and they are as follows:

1. "The number of hidden neurons should be between the size of the input layer and the size of the output layer.
2. The number of hidden neurons should be 2/3 the size of the input layer, plus the size of the output layer.
3. The number of hidden neurons should be less than twice the size of the input layer" (Heaton, 2017).

Regarding the output layers, they generally refer to the model's predictions. A representation of a typical ANN architecture is shown below in Figure 10.

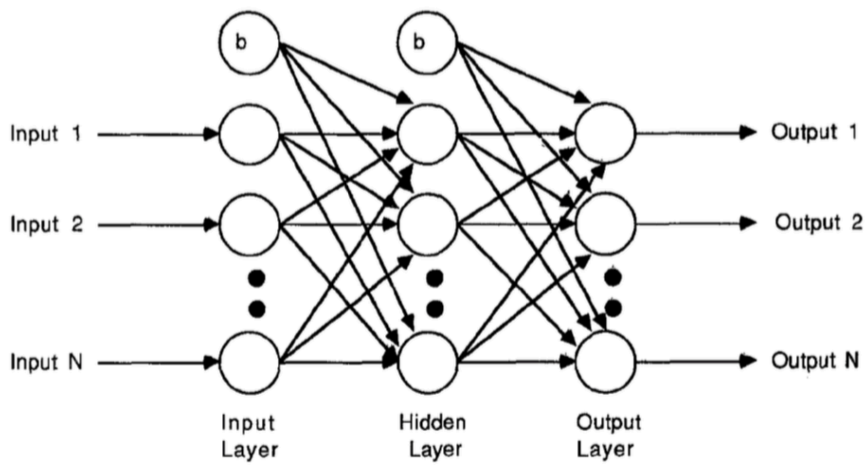

Figure 10: ANN Model Architecture (Williams, 1994)

Another main component of the ANN architecture is the activation function. This function mainly introduces non-linearity into the network and enables it to learn intricate correlations between predictions and features. Common activation functions include linear, sigmoid, softmax, tanh and reLU. The linear function simply outputs the original input value without any changes. The sigmoid function is an s-shape function that produces outputs that range from 0 to 1. Tanh function, also referred to as hyperbolic tangent, is function which is similar to sigmoid but produces values that range from -1 to 1. The reLU (rectified linear unit) function outputs the input itself if it's a positive value and outputs zero of the input is a negative value. Furthermore, using softmax function, a vector of input values (z) is transformed into a vector of output values (y) ranging from 0 to 1, reflecting the probability for each category. Since it guarantees that the total of all outputs equals 1, it may be used to illustrate a probability distribution over several classes (Sharma and Athaiya, 2020). Table 4 summarizes the advantages and disadvantages of each of the above-mentioned activation functions.

Table 4: Advantages and Disadvantages of Activation Functions (Srinivasan et al., 2019)

Function	**Advantages**	**Disadvantages**
Linear	Simple to process	Lacks non-linearity (limits learning complex patterns)
	Easy to understand	Only suitable for specific output layers (continuous scale)
Sigmoid	Outputs range between 0 and 1 (useful for probabilities)	Suffers from vanishing gradients in deep networks
	Smooth function (easier for gradient descent)	
Tanh (Hyperbolic Tangent)	Outputs range between -1 and 1	Suffers from vanishing gradients in deep networks
	Zero centered (avoids gradient saturation in some cases)	
ReLU (Rectified Linear Unit)	Computationally efficient (faster training)	Limited output range (0 and positive values)
	Mitigates vanishing gradients	
Softmax	Guarantees that the outputs sum to 1	In rare cases, with significantly large or small input values, it can encounter numerical stability issues.
	Can be used for multi-class classification	

3.3.5. The Training Strategy

The most important step in creating a neural network model is selecting the appropriate training algorithm. The process of training the network involves altering weight, or connection strengths, using one of many learning techniques. Every trial model used in this study was trained using a back-propagation learning technique. The network is given a training data set as inputs, and calculations are made for the outputs. The network's weights are then modified to lessen the discrepancies by evaluating the differences between the predicted outputs and the actual target output. The network's weights are continually modified during training until the estimated output error converges to a reasonable level (ElSawy et al., 2011). The back propagation algorithm involves gradually reducing the root mean squared error (RMSE) between the model output and the actual output. As soon as the RMSE becomes constant, the training procedure should end (Idrees et al., 2023). Equation 4 shows the formula used to calculate the RMSE.

$$RMSE = \frac{\sqrt{\sum_{i=1}^{n}(xi - E(i))^2}}{n}$$

Equation 4

Where n is the number of projects being evaluated during the training stage, Xi is the predicted output of the model, E is the actual target output.

3.3.6. Model Testing and Validation

In order to evaluate the accuracy of the training and testing processes, two main parameters need to be obtained: mean absolute error (MAE) and mean squared error (MSE).

- Mean Absolute Error (MAE): the average absolute difference between the predicted and actual values.
- Mean Squared Error (MSE): a statistical measure of how close a predicted value is to the actual value. It is calculated by taking the average of the squared differences between the predicted and actual values.

To validate the generated model, it must be tested using data that was not previously included in the model's training set. By comparing the predicted value with the actual, real-life value and computing the difference between the two values, the prediction accuracy can be assessed. For the projects not used in the model’s learning database, the absolute percentage error between the predicted and actual values is calculated in order to determine the model’s accuracy. Equation 5 shows the formula used to calculate the absolute percentage error.

$$\text{Absolute \% Error} = \left| \frac{(\text{Actual value} - \text{Predicted Value})}{\text{Actual Value}} \text{ x } 100 \right|$$

Equation 5

4. Data Analysis

This chapter first includes analysis of the questionnaire respondents, indicating their years of experience, their positions and the department in which they work. This chapter also involves conducting data analysis and Pearson and Spearman correlations tests for the collected data.

4.1. Analysis of Questionnaire Respondents

To ensure thorough and accurate results, experts for this research were carefully selected. These experts included experienced engineers working in the department of cost estimation and control in well-known contracting firms in Egypt. The questionnaire was sent to 64 experts in the field, of whom 55 responded.

All the collected data was from engineers working in Grade A companies only, which resulted in having all the projects in the model's database relating to only one grade. Furthermore, as shown in Figures 11 and 12, the cost data was collected from engineers with different levels of experience including seniors, team leaders, section heads, managers and senior managers.

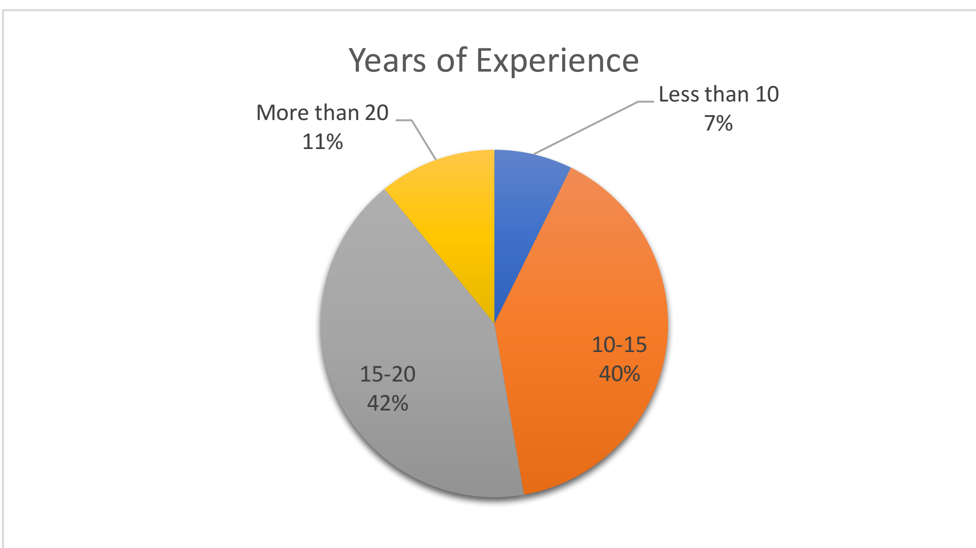

Figure 11: Questionnaire Respondents' Years of Experience

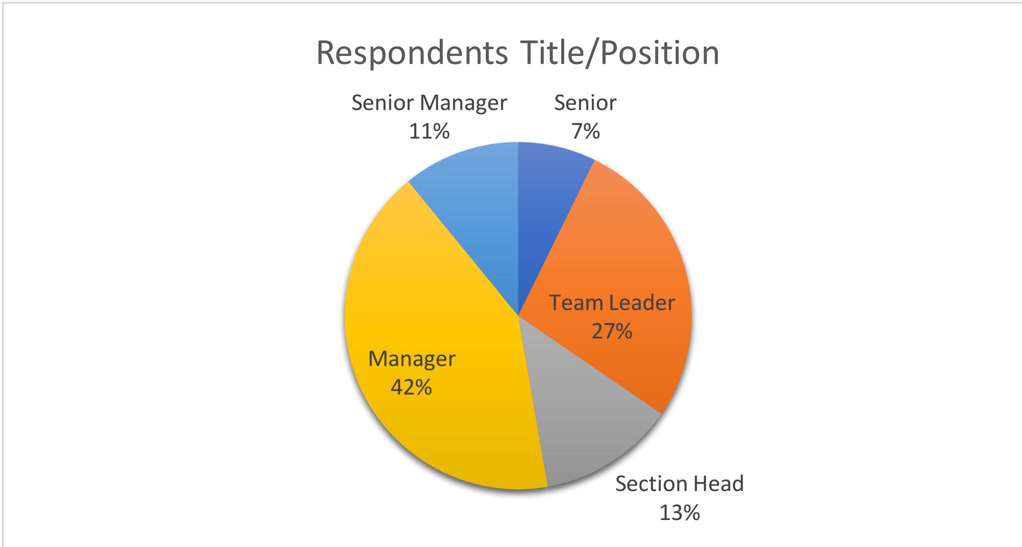

Figure 12: Questionnaire Respondents' Titles/Positions

As shown in Figures 11 and 12, the majority of the respondents were cost managers with 15 to 20 years of experience, followed by team leaders and section heads with 10 to 15 years of experience. 11% of the respondents were senior managers with more than 20 years of experience. The minority of the respondents (only 7%) were seniors with less than 10 years of experience.

4.2. Analysis of Collected Data

Cost data for 55 projects constructed in Egypt executed during the past 6 years (from 2018 to 2024) was collected. The gathered project data was analyzed and examined in order to determine how each factor affected the overall site overhead percentage. A comparative analysis of each factor and how it affects the site overhead percentage will be provided in this section to determine the factors that have the highest and lowest impacts. It is important to note that throughout the entire research, the site overheads percentage refers to the percentage of site overheads from the total direct cost.

Further analysis was done to investigate the relationship between each factor and the percentage of site overheads allocated on each main subcategory. Not only this, but the relation between each factor and the percentage of site overheads incurred during the 4 main project construction phases (initiation, growth, maturity and decline) was also studied.

4.2.1. Effect of Category of Contracting Company

Due to the inability to access data from category B contracting companies, the collected data was from grade A companies only. Accordingly, the model developed will be limited to grade A companies only. However, several other researches who have done a comparative analysis on data collected from different categories of contracting companies have confirmed that the category of contracting company has a high impact on site overheads percentage and that as the grade of the company improves, its site overheads significantly increase (Othman, 2020; Bakr et al., 2018; Idrees et al., 2023).

4.2.2. Effect of Project Type

The 55 collected projects include the major types of projects that usually take place in Egypt as the data was collected from 7 different types of projects, which are: residential, commercial, administrative, infrastructure, industrial, landscape and marine. As shown below in Figure 13, the data set consists of 35% residential projects, 15% infrastructure projects, 13% landscape projects, 13% administrative projects, 13% commercial projects, 9% marine projects and 4% industrial projects.

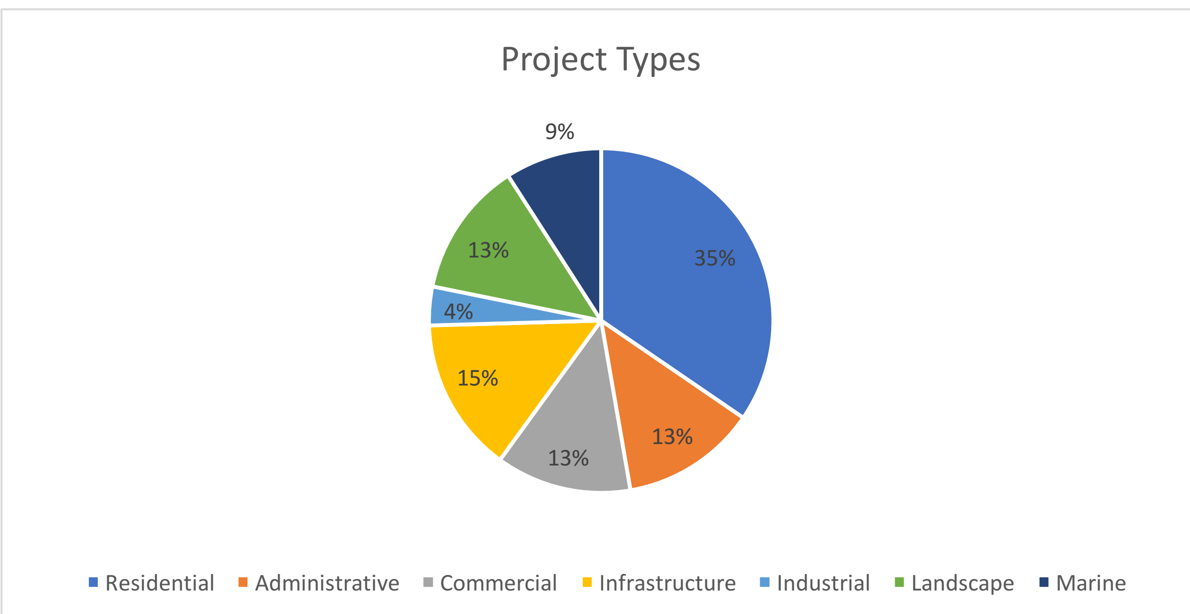

Figure 13: Classification of collected projects according to type

Effect of Project Type on Total Site Overheads Percentage

Through the analysis, it was found that the relationship between the project type and site overheads percentage is non homogeneous and cannot be represented using an equation. However, it can be noted that the average percentage of site overheads varies with different project types, as shown in Table 5 and Figure 14, which demonstrate the minimum, maximum and average values for site overheads percentage for each project type.

Table 5: Data Analysis of Project Type

Project Type	Min. Value	Max. Value	Average Value
Residential	7.00%	31.43%	17.92%
Administrative	8.00%	28.10%	17.58%
Commercial	7.32%	32.00%	18.03%
Infrastructure	7.74%	24.67%	12.54%
Industrial	10.43%	17.06%	13.74%
Landscape	7.89%	32.52%	22.68%
Marine	7.76%	15.74%	11.33%

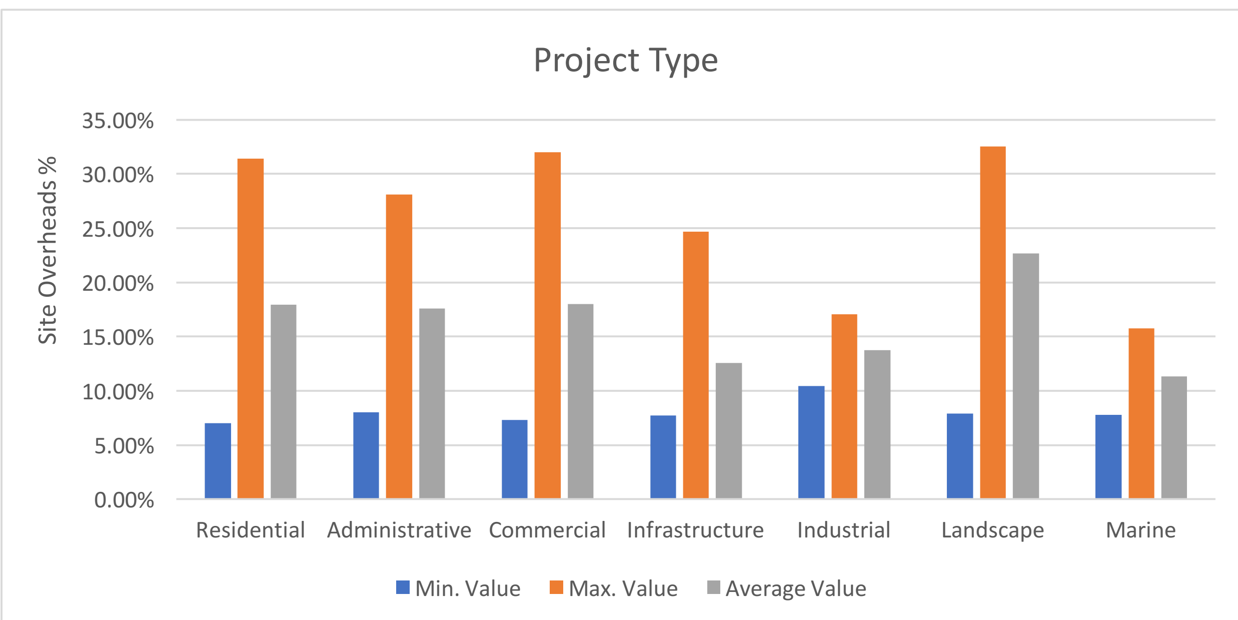

Figure 14: Site Overheads Percentage vs. Project Type

As shown in Figure 14, landscape projects have the highest average site overheads percentage (22.68%). This could be due to several factors like the fact that, compared to other types of construction projects, landscape projects usually entail lower individual project sizes. This may result in a larger overheads percentage as the fixed indirect costs needed for any project (like equipment and salaries) are now dispersed across a smaller project budget. In addition, the project types that also have a relatively high percentage of site overheads (from 17-18%) are commercial, residential and administrative projects and the project type which has the lowest average percentage overheads is marine projects (11.33%). The reason behind this could be that, compared to custom-built structures on land, certain marine projects may use prefabricated components that need less on-site assembly. Also, large-scale marine projects often have large overall budgets. Accordingly, distributing fixed expenses, such as equipment rental and mobilization, over a large project value could result in a reduced total overhead percentage.

Effect of Project Type on Site Overheads Subcategories

Furthermore, through analyzing the effect of project type on the main subcategories of site overheads, the following data were obtained and reported in Table 6. The data represents the average percentage of each category of site overheads for each project type.

Table 6: Average Percentage of Site Overheads Allocated to each Subcategory for each Project Type

	Residential	**Administrative**	**Commercial**	**Infrastructure**	**Industrial**	**Landscape**	**Marine**
Salaries & Wages	49.9%	47.7%	47.2%	38.4%	59.3%	44.5%	50.7%
Site Facilities	6.6%	5.4%	8.7%	4.8%	6.1%	4.5%	4.2%
Accommodation	2.0%	2.8%	2.6%	4.5%	3.0%	0.7%	3.9%
Mobilization & Demobilization	3.1%	3.2%	5.6%	5.1%	3.7%	4.6%	5.9%
Communication & IT	0.7%	1.5%	1.3%	1.2%	1.1%	1.7%	0.7%
Site Equipment	25.1%	25.9%	20.0%	27.7%	15.4%	22.0%	19.2%
Personnel & Material Transportation	6.5%	5.9%	6.7%	11.6%	7.6%	10.3%	8.0%
Quality and Safety	2.0%	3.1%	2.2%	1.3%	3.9%	1.6%	1.3%
Engineering Expenses	2.0%	2.0%	2.5%	5.3%	0.2%	3.0%	4.3%
Client/Consultant Expenses	1.8%	1.1%	1.3%	2.1%	0.0%	0.6%	1.8%

According to the data shown above, it can be noted that changing the project type relatively affects the percentage of site overheads allocated to each subcategory. For example, industrial projects followed by marine projects tend to spend the most on salaries and wages. This could be because maritime and industrial initiatives are usually undertaken in isolated or offshore sites. This may make it harder to find employees and call for more pay to make up for the inconvenience and moving expenses. Therefore, employees and labor may be offered higher salaries and wages to compensate for this inconvenience. Industrial, marine and infrastructure projects tend to also have the highest average percentages of personnel and material transportation and this could be due to most of these projects being located in isolated or offshore site. The far location of these types of projects makes them also have the highest percentage of accommodation costs as it requires them to provide temporary or permanent camps for labor as well as rentals for engineers during the entire duration of the project. In addition, industrial projects tend to have the highest average percentage of quality and safety procedures as environments that are industrial tend to be dangerous by nature, and thus, specialized safety procedures, such as operating underwater or at heights, add complexity and demand more safety and quality procedures. Also, the complex nature of industrial projects may be the main reason why they tend to have the highest percentage of site equipment. Furthermore, it can be noted that commercial projects tend to have the highest percentage of site facilities and mobilization and demobilization costs. These projects tend to spend a lot on site facilities, mobilization and demobilization as commercial projects are often located in urban areas that tend to have limited space for permanent facilities. The need for temporary offices, storage containers, and sanitation necessitates a higher investment in these facilities. The limited space provided for this type of projects also sometimes leads to them have to relocate the temporary mobilization several times throughout the duration of the project, increasing the mobilization and demobilization costs.

Effect of Project Type on Project's Construction Phases

In order to examine the effect of the project type on the projects' distinct construction phases, the data shown in Table 7 were obtained and used. Table 7 shows the average percentage of site overheads incurred during each construction phase.

Table 7: Average Percentage of Site Overheads During Each Construction Phase for each Project Type

	Residential	Administrative	Commercial	Infrastructure	Industrial	Landscape	Marine
Initiation	8.7%	8.7%	10.5%	11.0%	11.6%	10.3%	11.6%
Growth	66.2%	67.4%	68.5%	68.8%	67.3%	65.2%	70.4%
Maturity	15.4%	15.1%	14.6%	13.3%	17.3%	12.4%	13.7%
Decline	6.5%	5.9%	6.3%	6.7%	7.2%	5.8%	6.9%

As shown in Table 7, infrastructure, industrial and marine tend to have the highest average percentage of site overheads disbursed during initiation, growth, maturity and decline phases. During initiation, these projects may incur substantial engineering, legal, and administrative expenditures. For infrastructure and marine projects, obtaining the required clearances and permits from many government authorities may be a time-consuming and costly procedure. Also, these projects require putting together a group of skilled engineers and architects to manage the initiation phase. Furthermore, these types of projects may see an increase in overhead as they progress through the growth phase because of the need to mobilize sophisticated machinery and establish preliminary safety procedures on the building site. In addition, these projects frequently have substantial site overheads during the maturity phase because of the continuous expenses associated with running specialized machinery and upholding strict safety regulations on intricate construction sites. Regarding the decline phase, in order to securely dismantle these kinds of facilities, complex demolition procedures may be required. Furthermore, scaffolding and specialized equipment used may need to be disassembled, cleaned and transported appropriately.

4.2.3. Effect of Project Location

The location of a project is considered a crucial factor as it impacts various aspects such as labor, transportation, accommodation and thus needs to be considered when predicting the indirect cost of new projects. Accordingly, this research's dataset includes projects that are located within the city (urban areas) and others which are located outside of the city (rural areas, deserts, new living zones, countryside, etc.). As shown in Figure 15, the dataset contains 49% projects inside the city and 51% outside the city.

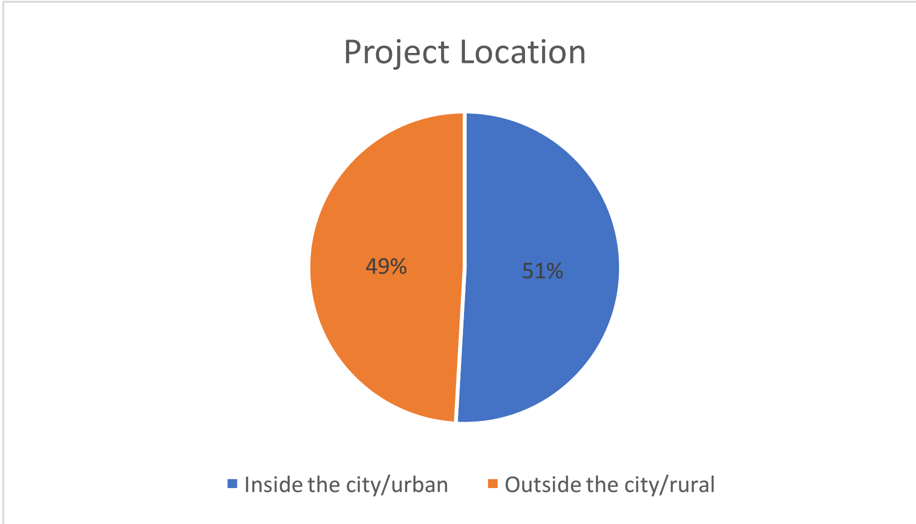

Figure 15: Classification of collected projects according to location

Effect of Project Location on Total Site Overheads Percentage

Table 8 and Figure 16 below demonstrate the minimum, maximum and average values for site overheads percentage for each project location.

Table 8: Data Analysis of Project Location

Project Location	Min. Value	Max. Value	Average Value
Inside the city/urban	7.0%	29.2%	16.9%
Outside the city/rural	7.2%	32.5%	17.1%

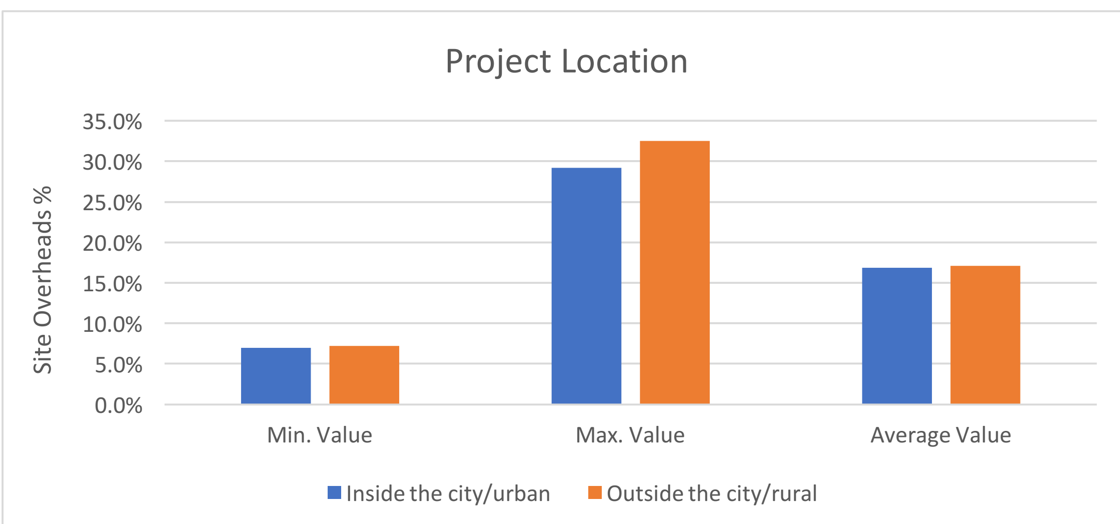

Figure 16: Site Overheads Percentage vs. Project Location

The results of the analysis show that projects located within the city boundaries have an average percentage of site overheads equal to 16.9% while the projects located in areas outside of city boundaries have an average percentage of site overheads equal to 17.1%. Thus, it is evident that projects located outside of capital cities often have a larger percentage of site overhead than projects located within them. There are numerous variables that contribute to this, including increased fuel usage, high accommodation costs, high cost of material delivery and increased need for site preparation.

Effect of Project Location on Site Overheads Subcategories

Table 9 below shows the average percentage of each category of site overheads for each project location.

Table 9: Average Percentage of Site Overheads Allocated to each Subcategory for each Project Location

	Inside the city/urban	**Outside the city/rural**
Salaries & Wages	46.0%	48.0%
Site Facilities	6.3%	9.0%
Accommodation	2.0%	3.1%
Mobilization & Demobilization	4.2%	6.0%
Communication & IT	1.2%	1.0%
Site Equipment	23.3%	25.5%
Personnel & Material Transportation	7.3%	9.0%
Quality and Safety	2.1%	2.0%
Engineering Expenses	3.0%	2.6%
Client/Consultant Expenses	1.5%	1.5%

As shown in Table 9, almost all of the subcategories tend to have higher percentage of site overheads in projects located outside of the city (rural areas). This is because projects located in rural areas have access to limited infrastructure and require extra logistical considerations. They also require higher costs due to their need for temporary facilities like portable units. They also need to provide accommodation and higher salaries and wages to attract workers to remote locations. Lastly, they also have increased transportation costs due to the need to provide transportation to personnel, materials and equipment needed on site.

Effect of Project Location on Project's Construction Phases

Table 10 below shows the average percentage of site overheads incurred during each construction phase for each project location.

Table 10: Average Percentage of Site Overheads During Each Construction Phase for each Project Location

	Inside the city/urban	**Outside the city/rural**
Initiation	10.8%	10.9%
Growth	68.1%	67.8%
Maturity	14.8%	14.2%
Decline	6.3%	7.1%

Similarly, projects located outside the city (rural areas) tend to have higher average percentage of site overheads during the projects' initiation, growth and decline phases. This is because of their increased need for mobilizing equipment, materials, personnel and setting up temporary facilities due to limited infrastructure. Attracting skilled labor to remote locations often requires accommodation and wage premiums, and materials cost more due to transportation distances and limited local suppliers. These factors combine for a higher percentage of site overheads throughout initiation, growth, and decline phases. However, the data in Table 10 show that urban projects tend to have a higher percentage of site overheads during the maturity phase. Urban projects can have high maturity costs even if their overheads are typically lower due to limited spaces forcing logistical challenges and higher labor expenses as there is no space for labor accommodation on site. In addition, the maturity phase expenditures in a city are increased by the need for specialized equipment, stronger safety precautions, and negotiating permissions due to complex urban infrastructure with existing structures and regulations.

4.2.4. Effect of Project Duration

The collected projects included durations ranging from 3 months to 57 months. Thus, in order be able to properly analyze this factor, the projects were categorized into 5 groups: duration less than or equal to 12 months, duration greater than 12 months and less than or equal to 36 months, duration greater than 36 months and less than 48 months and lastly duration greater than 48 months. The amount of collected projects relating to each group is shown in Figure 17.

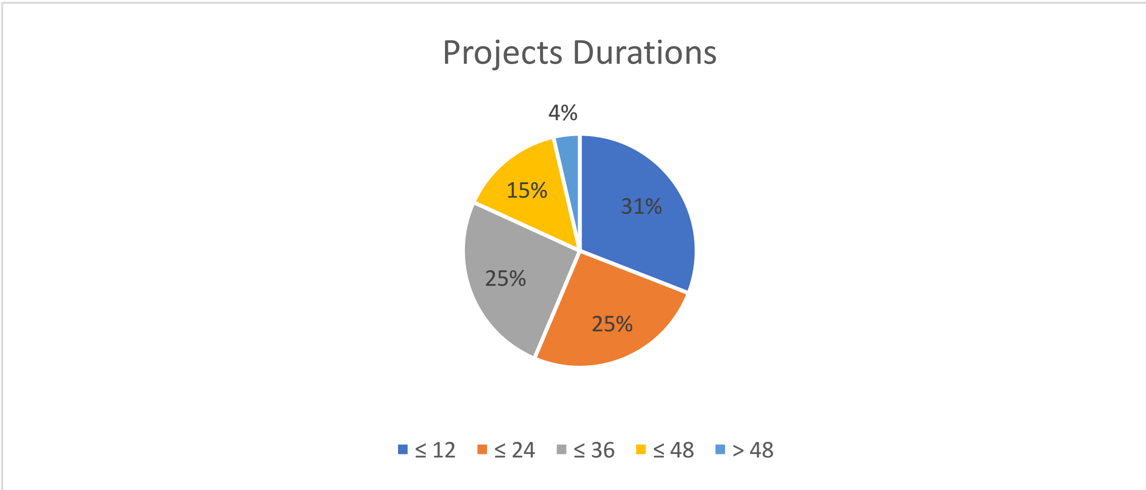

Figure 17: Classification of collected projects according to duration

Effect of Project Duration on Total Site Overheads Percentage

Table 11 and Figure 18 below illustrate the minimum value, maximum value and average value of site overheads percentage for each project location.

Table 11: Data Analysis of Project Duration

Duration (months)	Min. Value	Max. Value	Average Value
≤ 12	7.0%	29.2%	13.4%
≤ 24	7.1%	32.2%	14.7%
≤ 36	7.3%	32.8%	19.7%
≤ 48	7.5%	33.0%	21.7%
> 48	17.2%	33.5%	24.6%

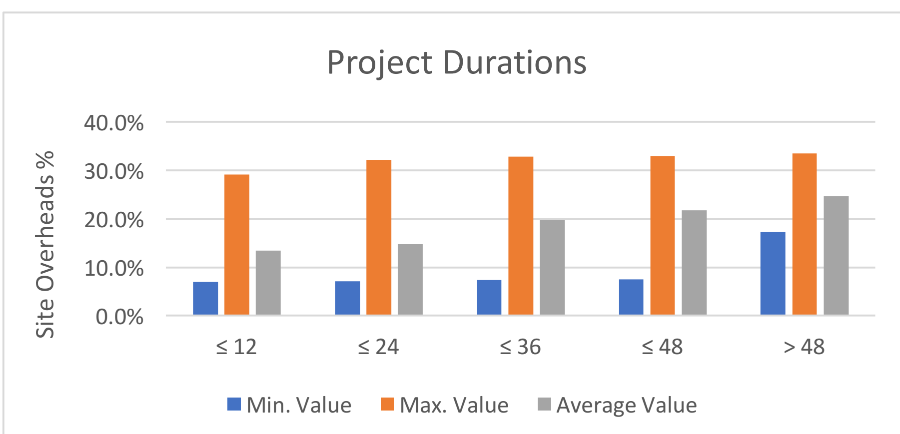

Figure 18: Site Overheads Percentage vs. Project Duration

As shown in Figure 18, the projects with durations less than 12 months have the lowest site overheads percentage while the projects with duration more than 48 months have the highest site overheads percentage. Thus, it can be concluded that the project direct and site overheads percentage are directly proportional (have a linear relationship) since the percentage of site overheads increases as the duration of the project increases. However, it can be noted that the rate of increase of site overheads is not always constant. In fact, it tends to decrease as the project duration increases.

Effect of Project Duration on Site Overheads Subcategories

Table 12 below shows the average percentage of each category of site overheads for different project durations.

Table 12: Average Percentage of Site Overheads Allocated to Each Subcategory for Different Project Durations

	Project Duration (months)				
	≤ 12	≤ 24	≤ 36	≤ 48	> 48
Salaries & Wages	44.3%	47.3%	48.3%	54.9%	57.2%
Site Facilities	5.5%	5.8%	6.0%	6.9%	7.1%
Accommodation	2.3%	2.8%	3.0%	3.1%	6.3%
Mobilization & Demobilization	3.8%	4.2%	4.0%	4.1%	4.3%
Communication & IT	0.8%	0.9%	1.0%	1.3%	1.6%
Site Equipment	22.5%	23.9%	26.5%	26.6%	26.7%
Personnel & Material Transportation	6.2%	6.8%	7.1%	8.9%	9.0%
Quality and Safety	2.1%	2.1%	1.9%	1.8%	2.0%
Engineering Expenses	3.4%	1.9%	4.1%	1.5%	1.0%
Client/Consultant Expenses	1.0%	2.3%	2.0%	2.2%	2.0%

Through analyzing the data in Table 12, it can be concluding that as the duration of the project increases, the percentage of site overheads allocated to each category also increases.

Effect of Project Duration on Project's Construction Phases

Table 13 shows the average percentage of site overheads incurred during each construction phase for different project durations.

Table 13: Average Percentage of Site Overheads During Each Construction Phase for Different Project Locations

	Project Duration (months)				
	≤ 12	≤ 24	≤ 36	≤ 48	> 48
Initiation	10.4%	10.0%	11.7%	11.8%	11.8%
Growth	66.2%	68.0%	68.2%	69.0%	69.0%
Maturity	14.2%	14.3%	15.2%	15.5%	14.5%
Decline	6.0%	6.1%	6.3%	6.9%	7.5%

Similarly, as the project duration increases, the average percentage of site overheads during each of the 4 main construction phases also increase.

4.2.5. Effect of Client Type

Another governing factor is the client type. In this research, the client type was categorized into the 4 major types of clients available in Egypt: national private sector, international private sector, public sector (army) and lastly public sector (ministries/governments). As shown in Figure 19, the dataset of this research consisted of 45% private sector (national), 20% public sector (ministry/government), 18% private sector (international) and 16% public sector (army).

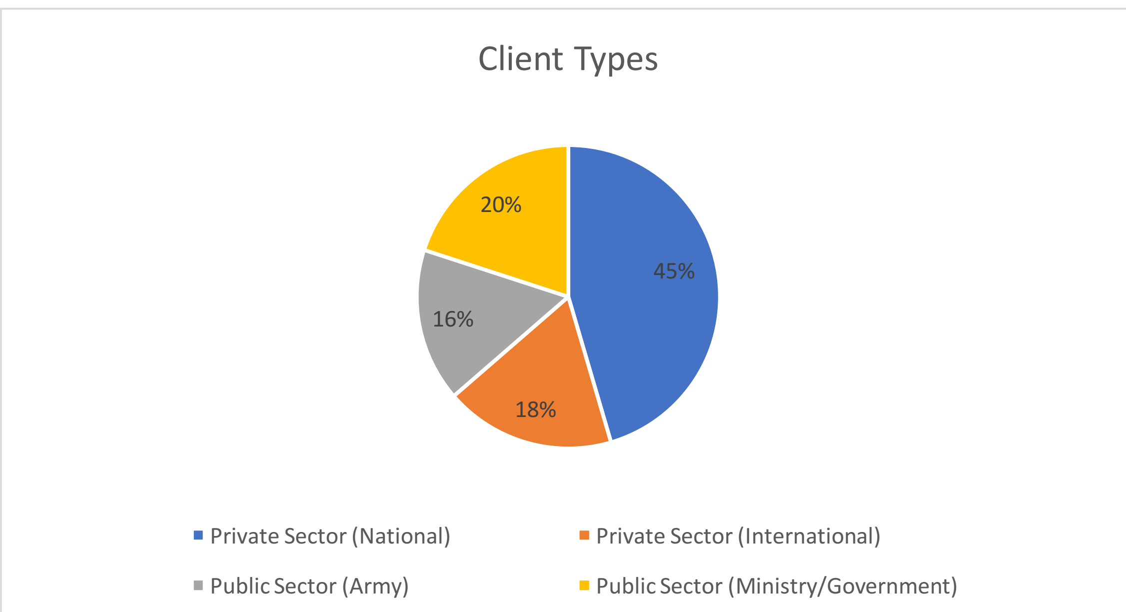

Figure 19: Classification of collected projects according to client type

Effect of Client Type on Total Site Overheads Percentage

Table 14 and Figure 18 below illustrate the minimum value, maximum value and average value of site overheads percentage for each client type.

Table 14: Data Analysis of Client Type

Client type	Min. Value	Max. Value	Average Value
Private Sector (National)	7%	33%	17%
Private Sector (International)	9%	31%	24%
Public Sector (Army)	7%	32%	15%
Public Sector (Ministry/Government)	7%	18%	11%

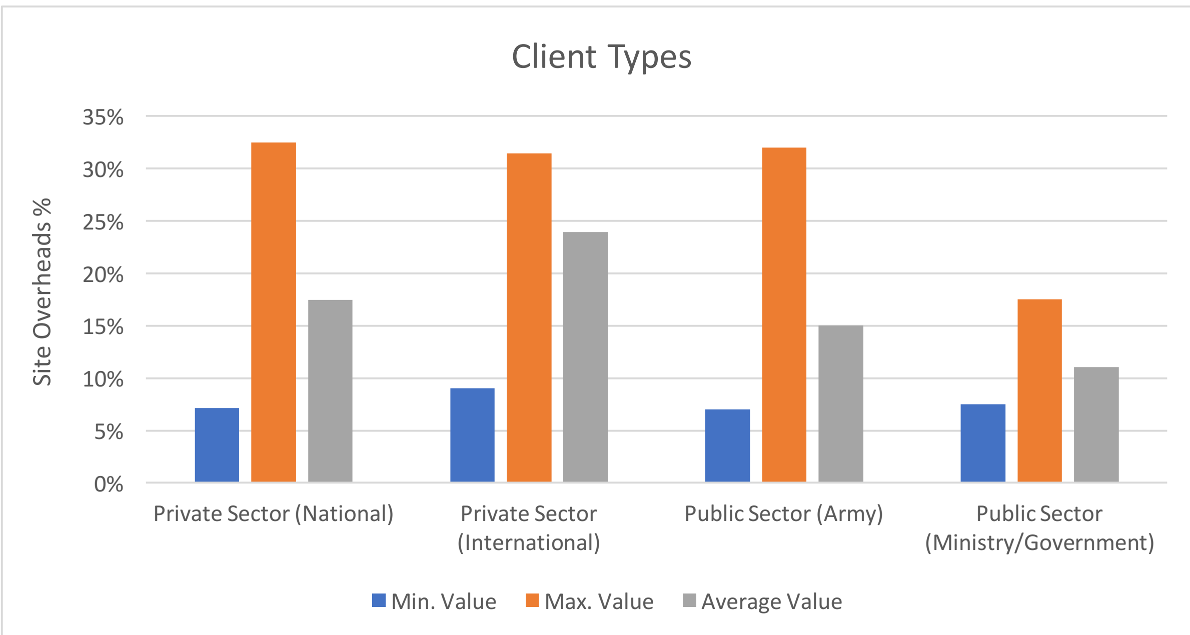

Figure 20: Site Overheads Percentage vs. Client Type

As shown in Figure 20, the client type with the highest average site overheads percentage is the international private sector (24%) while the client type that has the lowest average site overheads percentage is public sector (ministry/government) (11%). As mentioned by El Sawy et al. (2010), this could mainly be due to the fact the international private clients tend to require more safety and quality control measures, high technical specification requirements and strict project management plans, all of which have a significant effect on the indirect cost of projects.

Effect of Client Type on Site Overheads Subcategories

Table 15 shows the average percentage of each category of site overheads for each client type.

Table 15: Average Percentage of Site Overheads Allocated to Each Subcategory for Each Client Type

	Private Sector (National)	**Private Sector (International)**	**Public Sector (Army)**	**Public Sector (Ministry/Government)**
Salaries & Wages	48.4%	50.7%	40.2%	47.8%
Site Facilities	6.2%	6.9%	6.8%	3.7%
Accommodation	1.1%	2.5%	4.7%	4.1%
Mobilization & Demobilization	4.5%	7.0%	3.3%	3.4%
Communication & IT	2.5%	2.0%	1.5%	0.5%
Site Equipment	25.0%	25.7%	23.8%	22.0%
Personnel & Material Transportation	8.0%	6.3%	10.3%	6.9%
Quality and Safety	2.2%	2.4%	2.0%	1.4%
Engineering Expenses	2.7%	2.9%	3.1%	3.5%
Client/Consultant Expenses	1.8%	1.8%	1.0%	1.1%

As shown in Table 15, projects involving the international private sector tend to have the highest average percentage of salaries and wages, site facilities, mobilization and demobilization, quality and safety and client/consultant expenses. As previously discussed, this could be due to the fact the international private clients tend to require more safety and quality control measures, high technical requirements and strict project management plans. They also tend to sometimes provide better mobilization and site facilities to their employees. Furthermore, projects involving the public sector (army) tend to have higher average percentage of accommodation and transportation expenses. This could be due to the fact that public sector projects are often executed for the benefit of the general public, and thus are being located in remote areas with limited existing infrastructure. This often requires providing worker camps and housing labor for extended periods as well as more transportation costs.

Effect of Client Type on Project's Construction Phases

Table 16 shows the average percentage of site overheads incurred during each construction phase for each client type (national private sector, international private sector, public sector-army, public sector-ministry/government).

Table 16: Average Percentage of Site Overheads During Each Construction Phase for Each Client Type

	Client Type			
	Private Sector (National)	**Private Sector (International)**	**Public Sector (Army)**	**Public Sector (Ministry/Government)**
Initiation	10.9%	12.0%	10.4%	10.5%
Growth	68.6%	69.3%	66.1%	67.2%
Maturity	14.0%	16.0%	13.5%	15.1%
Decline	6.9%	6.9%	5.9%	6.8%

It can be concluded from Table 16 that projects involving the international private sector tend to have the highest percentage of site overheads during each of the 4 phases. This may be mainly due to better mobilization and site facilities along with potentially stricter safety regulations compared to national public companies which can elevate costs during each phase, in addition to all the reasons previously explained.

4.2.6. Effect of Contract Type

Generally, there are many types of contracts used in construction projects worldwide. However, this research will only focus on the contract types used the most in Egypt, which are lump sum, re-measured and cost plus contracts. As shown in Figure 21, 69% of the collected projects had lump sum contracts, 25% had re-measured contracts and 5% had cost plus contracts.

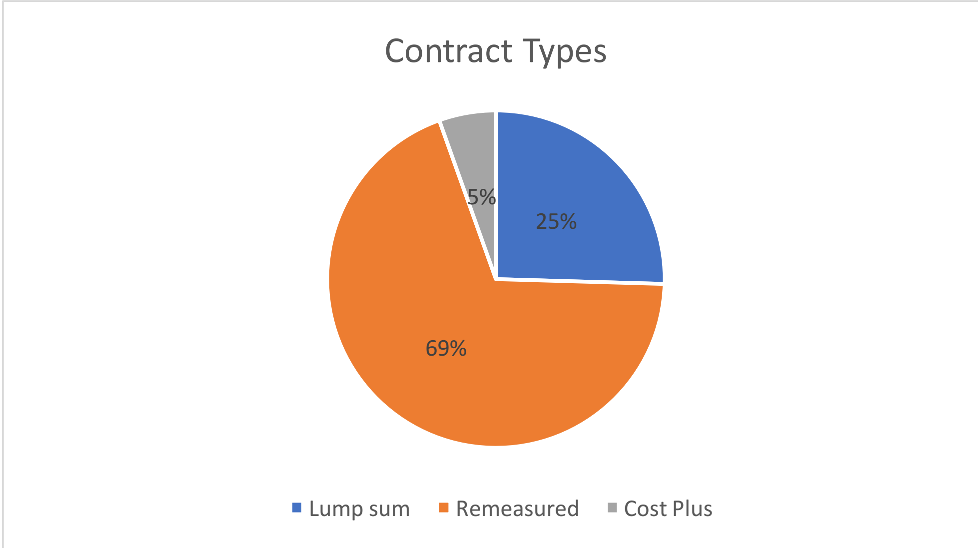

Figure 21: Classification of collected projects according to contract type

Effect of Contract Type on Total Site Overheads Percentage

Table 17 and Figure 20 below illustrate the minimum value, maximum value and average value of site overheads percentage for each contract type.

Table 17: Data Analysis of Contract Type

Contract Type	Min. Value	Max. Value	Average Value
Lump sum	7%	31%	18%
Re-measured	7%	33%	16%
Cost Plus	7%	32%	19%

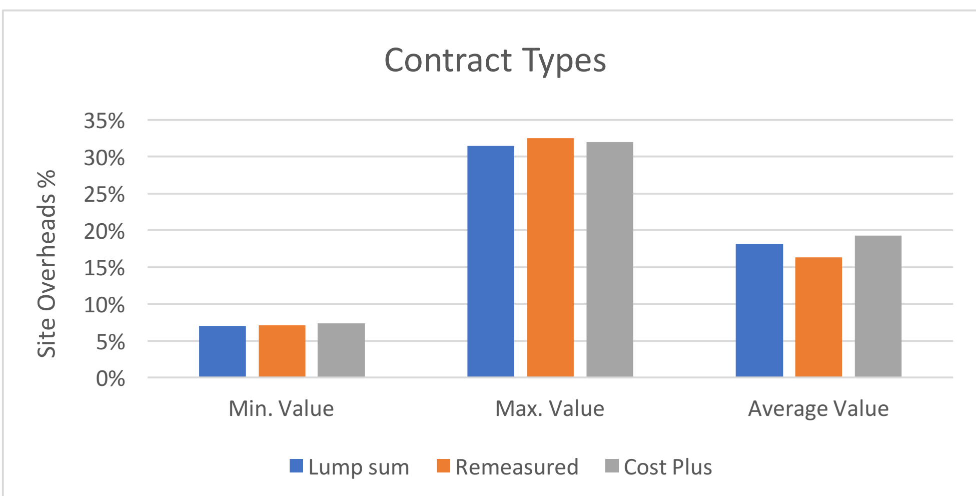

Figure 22: Site Overheads Percentage vs. Contract Type

As shown in Figure 22, the type of contract that tends to have the highest average site overheads percentage is cost plus contracts (19%) while the type that has the lowest average site overheads percentage is lump sum contracts (16%). This may be due to contractors being more motivated to reduce their costs, especially overheads, in lump sum contracts to avoid any risks of having to pay the overheads from the project's profit or from their own pockets.

Effect of Contract Type on Site Overheads Subcategories

Table 18 below shows the average percentage of each category of site overheads for each contract type.

Table 18: Average Percentage of Site Overheads Allocated to Each Subcategory for Each Contract Type

	Contract Type		
	Lump sum	Re-measured	Cost Plus
Salaries & Wages	45.0%	46.8%	49.2%
Site Facilities	5.7%	5.8%	8.4%
Accommodation	4.0%	1.9%	4.9%
Mobilization & Demobilization	3.9%	4.0%	7.3%
Communication & IT	1.3%	1.0%	1.5%
Site Equipment	22.8%	19.4%	25.4%
Personnel & Material Transportation	6.3%	8.6%	5.7%
Quality and Safety	1.8%	2.1%	2.2%
Engineering Expenses	3.2%	2.7%	2.8%
Client/Consultant Expenses	2.0%	1.2%	2.8%

From Table 18, it can be concluded that projects involving cost plus contracts tend to have higher percentages across various categories like salaries, wages, site facilities, site equipment, mobilization and accommodation due to a lack of cost control incentive for the contractor. With reimbursement guaranteed for all expenses, contractors might not prioritize cost-effective solutions for labor, materials, or temporary facilities, potentially leading to inflated costs. Projects including re-measured contracts also have relatively high percentages across all categories and generally highest personnel and material transportation as well as quality and safety expenses than all other contract types as re-measured contracts offer shared cost responsibility which incentivizes higher actual personnel and material transportation and quality and safety expenses. On the other side, lump sum contracts projects tend have the lowest average percentage of site categories across most categories because they involve more cost control to avoid any risks of contractors paying from their profit or their own money.

Effect of Contract Type on Project's Construction Phases

Table 19 below shows the average percentage of site overheads incurred during each construction phase for different project durations.

Table 19: Average Percentage of Site Overheads during each Construction Phase for Each Contract Type

	Contract Type		
	Lump sum	**Re-measured**	**Cost Plus**
Initiation	10.5%	11.3%	11.7%
Growth	67.0%	68.2%	70.0%
Maturity	13.0%	14.4%	15.0%
Decline	5.7%	6.9%	6.3%

As shown in Table 19, projects with cost plus contracts tend to have the highest percentage of site overheads spent during initiation, growth and maturity phases. In addition, projects with re-measured contracts tend to have the highest percentage of site overheads spent during decline phase. As previously discussed, this is the case with cost plus contracts due to reimbursement guaranteed for all expenses and with re-measured contracts due to shared responsibility between client and contractor. Lastly, projects with lump sum contracts tend to incur the lowest average percentage of site overheads during all phases due to efficient control on resources and equipment to avoid any risks.

4.2.7. Effect of Project Size (Total Contract Amount)

Another significant factor that surely has an effect on the site overheads percentage is the project size which can be quantified through the project's total contract amount or the total direct cost. The total direct cost of the collected projects ranged from around 9 million EGP to 2.5 billion EGP. Thus, the projects were categorized into 7 groups. The number of projects that fall under each group is shown in Figure 23 below.

Figure 23: Classification of collected projects according to total contract amount

Effect of Project Size on Total Site Overheads Percentage

Table 20 and Figure 24 below illustrate the minimum value, maximum value and average value of site overheads percentage for each category of total contract amount.

Table 20: Data Analysis of Total Contract Amount

Total Contract Value (Egyptian Pounds)	Min. Value	Max. Value	Average Value
≤ 50 Million	7.0%	32.5%	17.9%
≤ 100 Million	7.1%	32.2%	17.5%
≤ 200 Million	7.7%	28.1%	17.2%
≤ 400 Million	7.2%	27.8%	17.7%
≤ 800 Million	7.3%	31.4%	18.4%
≤ 1 Billion	14.3%	32.0%	22.6%
> 1 Billion	7.5%	17.2%	10.1%

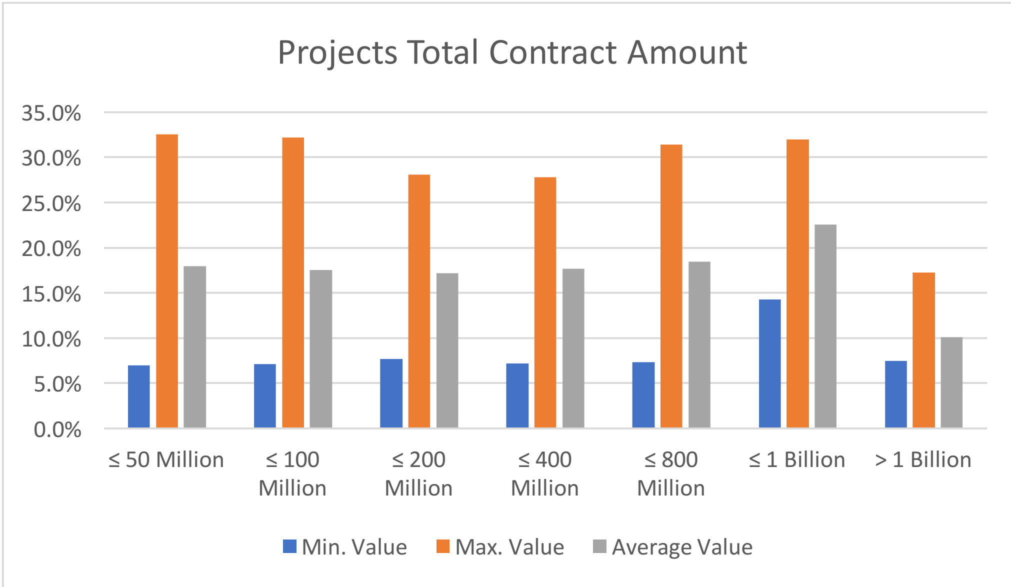

Figure 24: Site Overheads Percentage vs. Projects Total Contract Amount

As illustrated in Figure 24, similar to the project duration, the average total contract amount has a linear relationship with the site overheads percentage for projects less than or equal to 1 billion EGP. For projects below this point, the site overheads percentage increases as the total contract amount increases. After this point, the average site overheads percentage relatively decreases due to the fact that distributing fixed expenses over a large project value could result in a reduced total overhead percentage.

Effect of Project Size on Site Overheads Subcategories

Table 21 below shows the average percentage of each category of site overheads for each contract type. As previously noted, the average total contract amount has a linear relationship with the average site overheads percentage related to each category for projects less than 1 billion EGP. After this point, the average site overheads percentage relatively decreases due to the fact that distributing fixed expenses over a large project value could result in a reduced total overhead percentage.

Table 21: Average Percentage of Site Overheads Allocated to Each Subcategory for Different Total Contract Amounts

	Project Total Contract Amount						
	≤ 50 Million	**≤ 100 Million**	**≤ 200 Million**	**≤ 400 Million**	**≤ 800 Million**	**≤ 1 Billion**	**> 1 Billion**
Salaries & Wages	36.6%	42.1%	48.2%	50.0%	55.4%	52.3%	46.6%
Site Facilities	5.9%	4.1%	6.8%	7.3%	8.7%	5.0%	3.3%
Accommodation	0.7%	1.1%	1.5%	3.4%	3.6%	4.9%	4.6%
Mobilization & Demobilization	2.3%	3.3%	4.0%	4.1%	8.0%	4.5%	3.9%
Communication & IT	0.6%	1.0%	1.2%	1.3%	1.9%	0.9%	0.7%
Site Equipment	17.1%	24.1%	25.3%	26.0%	28.8%	24.6%	26.4%
Personnel & Material Transportation	7.1%	7.3%	7.6%	8.7%	11.3%	4.0%	7.2%
Quality and Safety	2.3%	2.2%	2.2%	1.6%	2.3%	2.2%	1.4%
Engineering Expenses	3.8%	2.2%	4.1%	1.3%	1.5%	1.1%	4.7%
Client/Consultant Expenses	0.8%	1.2%	1.2%	2.2%	2.6%	0.6%	1.1%

Effect of Project Size on Project's Construction Phases

Table 22 below shows the average percentage of site overheads incurred during each construction phase for different total contract amounts.

Table 22: Average Percentage of Site Overheads During Each Construction Phase for Different Contract Amounts

	Project Total Contract Amount						
	≤ 50 Million	**≤ 100 Million**	**≤ 200 Million**	**≤ 400 Million**	**≤ 800 Million**	**≤ 1 Billion**	**> 1 Billion**
Initiation	9.2%	10.7%	10.8%	11.3%	12.1%	11.5%	11.3%
Growth	66.7%	67.0%	67.4%	67.8%	68.7%	68.5%	68.2%
Maturity	14.2%	14.3%	14.4%	15.0%	15.2%	15.0%	14.9%
Decline	6.4%	6.6%	6.9%	7.0%	7.2%	7.1%	7.0%

Similarly, the percentage of site overheads incurred during each of the phases increases as the total contract amount increases for projects less than 1 billion. Thus, the projects that have the highest average percentage of site overheads during initiation, growth, maturity and decline are ones with total contract amount less than or equal 800 million.

4.2.8. Effect of Macroeconomic Indicators

The Egyptian construction industry has seen a sharp increase in inflation rate, interest rate and USD to EGP exchange rate over the last ten years, which has had a substantial impact on existing projects. For every collected project, an average for each of these 3 rates during the entire duration of the project was obtained. Additionally, in order to focus on the major economic changes that took place, each economic variable was categorized into different groups. For the interest rate, the analysis was done for projects with average interest rates less than 8%, average interest rates between 8% and 10%, average interest rates between 10% and 12% and average interest rates more than 12%. For the inflation rate, the analysis was done for projects with average inflation rates less than 5%, average inflation rates between 5% and 15%, average inflation rates between 15% and 30% and average inflation rates more than 30%. For USD to EGP exchange rate, the analysis was done for projects with average exchange rates less than 16, average exchange rates between 16 and 18, average exchange rates between 18 and 30 and average exchange rates more than 30%. Figures 24, 25 and 26 classify the collected projects according to their average rates.

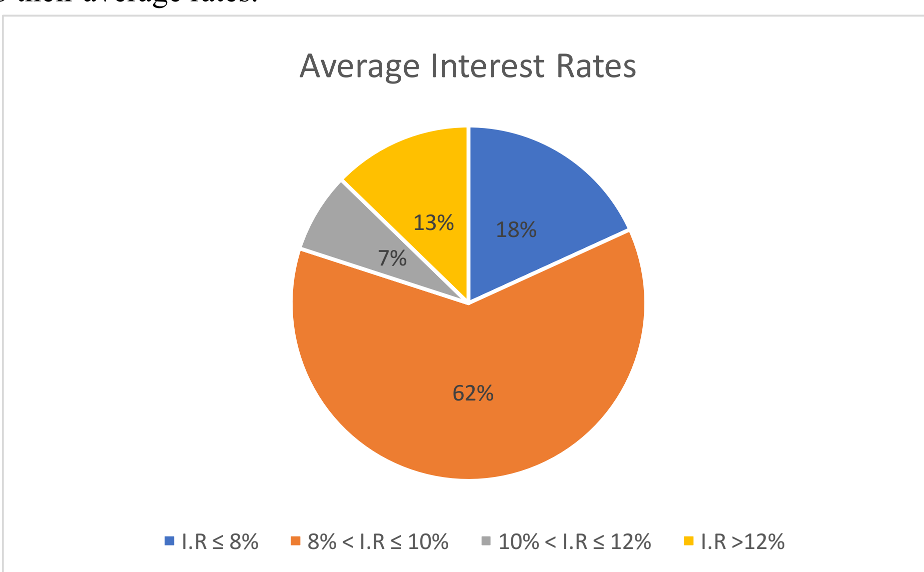

Figure 25: Classification of projects according to average interest rates

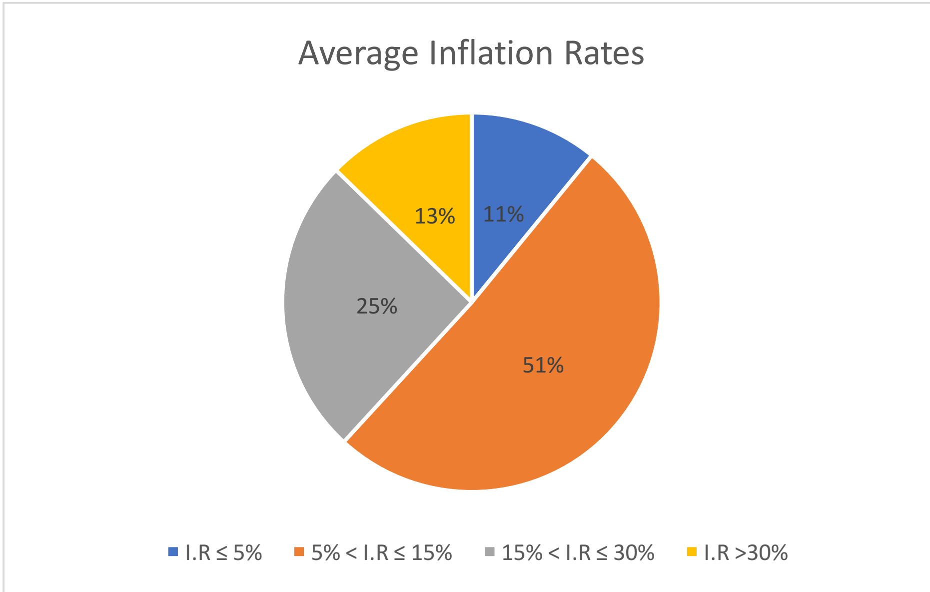

Figure 26: Classification of projects according to average inflation rate

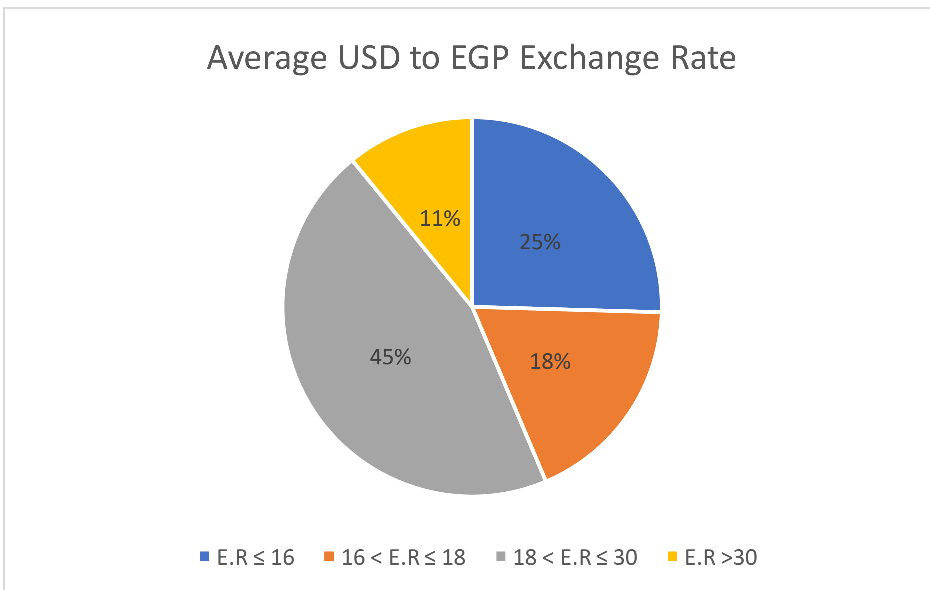

Figure 27: Classification of projects according to average exchange rate

Effect of Macroeconomic Indicators on Site Overheads Percentage

Figures 28, 29 and 30 below illustrate the minimum value, maximum value and average value of site overheads percentage for each category of interest, inflation and exchange rate.

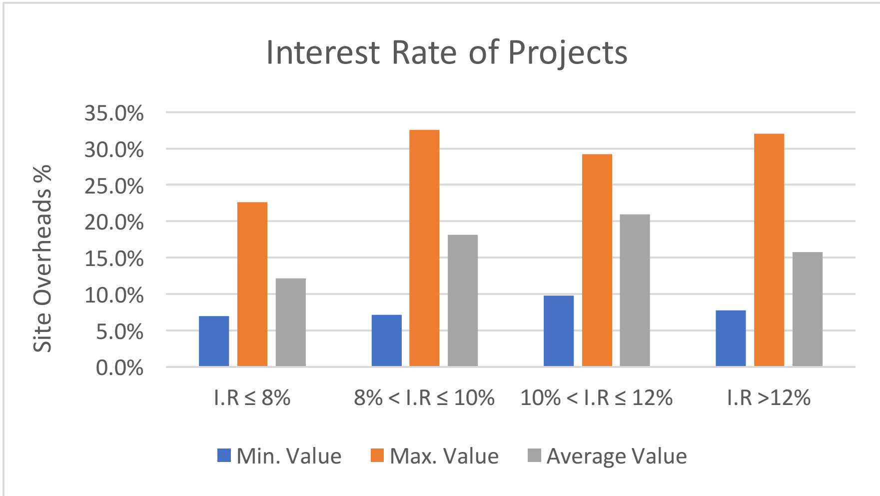

Figure 28: Site Overheads Percentage vs. Interest Rate

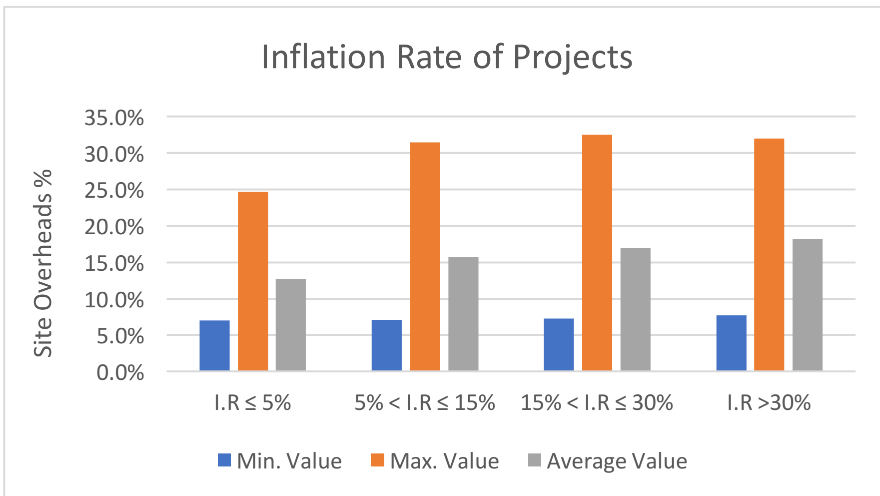

Figure 29: Site Overheads Percentage vs. Inflation Rate

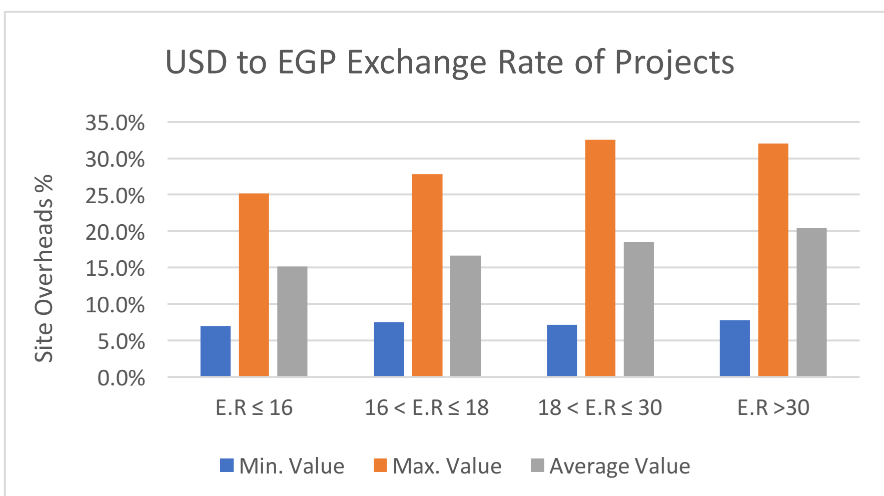

Figure 30: Site Overheads Percentage vs. Exchange Rate

As shown in Figures 28, 29 and 30, the average site overheads percentage increases as the inflation rate and USD to EGP exchange rate increases. This can be attributed to a number of things, such as new laws, shifting market dynamics, and rising fuel costs. Thus, this further confirms that these specific macroeconomic indicators have a considerable effect on the indirect cost of projects. Regarding the interest rates, the average site overheads percentage increases as the interest rate increases till the interest rate reaches 12% and then it slightly decreases. This means that there may be a relationship between site overheads and interest rate, which will be further examined in the coming sections.

Effect of Macroeconomic Indicators on Site Overheads Subcategories

Tables 23, 24 and 25 below show the average percentage of each category of site overheads for different rates of macroeconomic indicators.

Table 23: Average Percentage of Site Overheads Allocated to Each Subcategory for Different Interest Rates

	Average Interest Rate			
	I.R ≤ 8%	8% < I.R ≤ 10%	10% < I.R ≤ 12%	I.R >12%
Salaries & Wages	47.2%	47.5%	58.8%	40.1%
Site Facilities	7.5%	5.7%	6.4%	4.8%
Accommodation	2.1%	2.2%	1.8%	5.4%
Mobilization & Demobilization	4.7%	4.0%	2.1%	5.4%
Communication & IT	0.9%	1.3%	0.8%	0.7%
Site Equipment	23.0%	25.0%	16.0%	28.2%
Personnel & Material Transportation	5.6%	7.9%	8.5%	10.6%
Quality and Safety	2.7%	1.9%	1.5%	1.9%
Engineering Expenses	4.0%	2.9%	0.9%	2.1%
Client/Consultant Expenses	1.4%	1.4%	3.2%	0.8%

Table 24: Average Percentage of Site Overheads Allocated to Each Subcategory for Different Inflation Rates

	Average Inflation Rate			
	I.R ≤ 5%	5% < I.R ≤ 15%	15% < I.R ≤ 30%	I.R >30%
Salaries & Wages	31.1%	49.2%	54.2%	56.2%
Site Facilities	4.7%	4.8%	6.6%	7.1%
Accommodation	1.6%	2.5%	4.1%	5.4%
Mobilization & Demobilization	2.9%	3.4%	5.4%	9.2%
Communication & IT	0.7%	0.9%	1.2%	1.6%
Site Equipment	20.4%	24.3%	28.2%	30.1%
Personnel & Material Transportation	6.6%	7.2%	10.6%	11.7%
Quality and Safety	1.4%	1.8%	1.9%	2.3%
Engineering Expenses	3.1%	2.7%	3.3%	2.1%
Client/Consultant Expenses	0.8%	1.0%	1.4%	2.2%

Table 25: Average Percentage of Site Overheads Allocated to Each Subcategory for Different Exchange Rates

	Average USD to EGP Exchange Rate			
	E.R ≤ 16	16 < E.R ≤ 18	18 < E.R ≤ 30	E.R >30
Salaries & Wages	38.6%	51.6%	52.5%	39.1%
Site Facilities	4.5%	5.1%	6.3%	7.8%
Accommodation	2.1%	2.3%	2.9%	4.4%
Mobilization & Demobilization	2.5%	3.4%	5.7%	7.1%
Communication & IT	0.5%	0.8%	1.1%	1.7%
Site Equipment	22.7%	23.1%	26.6%	28.8%
Personnel & Material Transportation	5.2%	7.0%	9.6%	11.7%
Quality and Safety	1.8%	1.9%	2.0%	2.4%
Engineering Expenses	3.7%	3.6%	2.2%	2.2%
Client/Consultant Expenses	0.8%	1.0%	1.0%	2.1%

Effect of Macroeconomic Indicators on Project's Construction Phases

Tables 26, 27, and 28 below show the average percentage of each category of site overheads for different rates of macroeconomic indicators.

Table 26: Average Percentage of Site Overheads During Each Construction Phase for Different Interest Rates

	Average Interest Rate			
	I.R ≤ 8%	**8% < I.R ≤ 10%**	**10% < I.R ≤ 12%**	**I.R >12%**
Initiation	9.2%	11.2%	10.7%	11.5%
Growth	68.6%	67.7%	70.3%	66.9%
Maturity	14.9%	14.4%	13.1%	15.0%
Decline	6.0%	6.6%	6.0%	6.6%

Table 27: Average Percentage of Site Overheads During Each Construction Phase for Different Inflation Rates

	Average Inflation Rate			
	I.R ≤ 5%	**5% < I.R ≤ 15%**	**15% < I.R ≤ 30%**	**I.R >30%**
Initiation	9.7%	11.1%	11.4%	11.5%
Growth	66.9%	67.2%	68.5%	69.7%
Maturity	13.4%	13.6%	15.0%	15.1%
Decline	6.5%	6.6%	6.6%	7.1%

Table 28: Average Percentage of Site Overheads During Each Construction Phase for Different Exchange Rates

	Average USD to EGP Exchange Rate			
	E.R ≤ 16	**16 < E.R ≤ 18**	**18 < E.R ≤ 30**	**E.R >30**
Initiation	10.3%	10.3%	10.9%	11.0%
Growth	67.2%	68.0%	68.1%	68.2%
Maturity	14.2%	14.3%	14.4%	15.3%
Decline	6.5%	6.6%	6.9%	7.1%

As shown in Tables 23 and 26, the varying interest effect do not seem to have a direct effect on the site overheads subcategories nor the different construction phases. However, the relationship between both will be further examined using the correlation tests in the coming section to assess the extent to which they correlate. Regarding the inflation rate and USD to EGP exchange rate, they both have a direct relationship with the site overheads subcategories and the construction phases of the project. This is because, as shown in Tables 24, 25, 27 and 28, when the rates increase, the average percentage of site overheads for each subcategory and for each phase also increases. Thus, it can be concluded from the data analysis done that the inflation and exchange rate have a direct effect on the project's site overheads while the inflation rate does not have a significant effect on it.

4.3. Pearson and Spearman Correlation Analysis & Checking for Multi-Collinearity

In order to assess the correlation between the factors and the site overheads percentage and to determine which factors have the highest impact and which factors do not have a major impact, correlations tests like Pearson and Spearman correlation tests. Pearson correlation coefficient measures the strength as well as the direction of the linear relationship present between two different quantitative variables. If Pearson coefficient is equal to 0, this supports the null hypothesis that no correlation exists between the tested variables. On the other side, if the coefficient is not equal to zero, it conveys that there exists a correlation between the variables (Xiao et al., 2016). The strength of the correlation is determined through the value of the coefficient. Table 29 further explains how the strength of the relationship can be determined.

Table 29: Strength of Pearson Correlation (Xiao et al., 2016)

Pearson Correlation Coefficient Value	Strength of Relationship between Variables
-1 to -0.5 / +0.5 to +1	Strong
-0.5 to -0.3 / 0.3 to 0.5	Moderate
-0.3 to -0.1 / 0.1 to 0.3	Weak
-0.1 to 0 / 0 to 0.1	Non-existing relationship
0	None

If Pearson correlation coefficient conveys a non-existing relationship, this can only mean that there is no linear relationship between the variables but does not mean that there is no any relation at all as it could be a monotonic relationship instead. For this purpose, Spearman correlation test should also be done for all the quantitative factors. Spearman correlation coefficient is a form of Pearson correlation coefficient which represents a non-parametric measure of the strength and direction of a monotonic relationship between two different ranked variables or two different variables, one of which one is ranked and one is not. A ranked variable is mainly a new variable created by assigning a rank order to each data point in the original variable (Xiao et al., 2016). As previously noted, both correlation tests can only be done to numerical/quantitative factors only. For the categorical factors, they should first be converted to numerical factors by giving each category a code. Along with coding the categorical factors, normalization using min-max scaler was done to all numerical inputs before performing the correlation test. The coding and normalization steps are explained in more details in the next section of this report.

For this research, Pearson and Spearman correlation test was done on Microsoft Excel using excel add-in called StatTools to assess the correlation between the factors (inputs) and themselves and between the factors (inputs) and the site overheads percentage, the percentage of site overheads for each subcategory and lastly the percentage of site overheads incurred during the main project construction phases (outputs). The results are shown below in Tables 30 and 31.

4.3.1. Checking for Multi-Collinearity

As shown in Tables 30 and 31, the factors that tend to have high multicollinearity are mainly the interest rate, inflation rate, and exchange rate. These variables are highly correlated, meaning they might overlap in the information they provide. To address this multicollinearity and assess their impact on model accuracy, several runs will be performed on the model. Each model will exclude one of these three variables to determine if removing a specific factor improves performance. Additionally, a model will be run including all 3 variables. The model with the highest accuracy will be selected.

Table 30: Pearson Correlation Coefficients Between Independent Variables

Linear Correlation Table	**Project Type Data Set #1**	**Project Location Data Set #1**	**Project Duration (months) Data Set #1**	**Client Type Data Set #1**	**Contract Type Data Set #1**	**Avg. Interest Rate Data Set #1**	**Avg. Inflation Rate Data Set #1**	**Avg. USD to EGP Exchange Rate Data Set #1**	**Total Direct Cost Data Set #1**
Project Type	1.000	0.349	-0.332	0.265	0.127	0.218	0.355	0.432	0.223
Project Location	0.349	1	0.162	0.002	0.028	0.261	0.366	0.328	0.307
Project Duration (months)	-0.332	0.162	1	-0.09	-0.121	0.09	-0.019	-0.069	0.295
Client Type	0.265	0.002	-0.09	1	-0.231	0.281	0.213	0.287	0.609
Contract Type	0.127	0.028	-0.121	-0.231	1	-0.178	-0.091	-0.023	-0.146
Avg. Interest Rate	0.218	0.261	0.09	0.281	-0.178	1	0.844	0.759	0.282
Avg. Inflation Rate	0.355	0.366	-0.019	0.213	-0.091	0.844	1	0.915	0.299
Avg. USD to EGP Exchange Rate	0.432	0.328	-0.069	0.287	-0.023	0.759	0.915	1	0.341
Total Direct Cost	0.223	0.307	0.295	0.609	-0.146	0.282	0.299	0.341	1

Table 31: Spearman Correlation Coefficients Between Independent Variables

Rank-Order Correlation Table	Project Type Data Set #1	Project Location Data Set #1	Project Duration (months) Data Set #1	Client Type Data Set #1	Contract Type Data Set #1	Avg. Interest Rate Data Set #1	Avg. Inflation Rate Data Set #1	Avg. USD to EGP Exchange Rate Data Set #1	Total Direct Cost Data Set #1
Project Type	1	0.312	-0.347	0.318	0.123	0.278	0.349	0.409	0.028
Project Location	0.312	1	0.138	-0.011	0.037	0.159	0.393	0.371	0.236
Project Duration (months)	-0.347	0.138	1	-0.071	-0.129	0.204	0.083	0.053	0.385
Client Type	0.318	-0.011	-0.071	1	-0.292	0.314	0.126	0.191	0.491
Contract Type	0.123	0.037	-0.129	-0.292	1	-0.232	-0.107	-0.059	-0.224
Avg. Interest Rate	0.278	0.159	0.204	0.314	-0.232	1	0.653	0.626	0.431
Avg. Inflation Rate	0.349	0.393	0.083	0.126	-0.107	0.653	1	0.938	0.434
Avg. USD to EGP Exchange Rate	0.409	0.371	0.053	0.191	-0.059	0.626	0.938	1	0.47
Total Direct Cost	0.028	0.236	0.385	0.491	-0.224	0.431	0.434	0.47	1

4.3.2. Correlation Analysis Between Inputs and Site Overheads Percentage

As shown in Tables 32 and 33, the factor that has the highest impact on site overheads percentage is project duration, followed by total direct cost. In fact, the project duration has a positive linear relationship with the site overheads percentage while the total direct cost has a negative linear relationship. The rest of the factors also have a considerable impact on the site overheads percentage but not as strong as the first two factors.

Table 32: Pearson Correlation Test Between Inputs and Site Overheads Percentage

Linear Correlation Table	Site Overheads % Data Set #1
Project Type	-0.101
Project Location	0.114
Project Duration (months)	0.402
Client Type	-0.311
Contract Type	-0.145
Avg. Interest Rate	0.107
Avg. Inflation Rate	-0.127
Avg. USD to EGP Exchange Rate	-0.274
Total Direct Cost	-0.315
Site Overheads %	1.000

Table 33: Spearman Correlation Test Between Inputs and Site Overheads Percentage

Rank-Order Correlation Table	Site Overheads % Data Set #1
Project Type	-0.107
Project Location	0.114
Project Duration (months)	0.401
Client Type	-0.262
Contract Type	-0.132
Avg. Interest Rate	0.148
Avg. Inflation Rate	-0.107
Avg. USD to EGP Exchange Rate	-0.189
Total Direct Cost	-0.262
Site Overheads %	1.000

4.3.3. Correlation Analysis Between Inputs and Site Overheads Percentage Allocated to Each Subcategory

Tables 34 and 35 show the results of the Pearson and Spearman correlations tests between inputs and percentages allocated to each subcategory.

Table 34: Pearson Correlation Test for Inputs and Site Overheads Percentage Allocated to Each Subcategory

Linear Correlation Table	Salaries & Wages % Data Set #1	Site Facilities % Data Set #1	Accommodation % Data Set #1	Mobilization & Demobilization % Data Set #1	Communication & IT % Data Set #1	Site Equipment % Data Set #1	Personnel & Material Transportation % Data Set #1	Quality and Safety % Data Set #1	Engineering % Data Set #1	Client/Consultant % Data Set #1
Project Type	-0.08	-0.28	0.02	0.25	0.16	-0.07	0.26	-0.15	0.21	-0.09
Project Location	-0.08	-0.15	0.19	-0.01	-0.11	0.12	0.10	-0.04	-0.06	0.02
Project Duration (months)	0.29	0.05	0.19	-0.27	0.03	-0.12	-0.24	-0.02	-0.16	-0.03
Client Type	-0.09	-0.26	0.47	0.24	-0.08	0.01	-0.01	-0.16	0.06	-0.11
Contract Type	-0.09	0.13	-0.16	0.14	-0.04	0.03	0.11	0.07	-0.06	-0.05
Avg. Interest Rate	-0.03	-0.22	0.33	-0.01	-0.13	0.00	0.22	-0.10	-0.11	-0.04
Avg. Inflation Rate	0.10	-0.32	0.25	-0.19	-0.22	-0.05	0.09	-0.01	-0.12	0.07
Avg. USD to EGP Exchange Rate	0.08	-0.37	0.15	-0.15	-0.18	-0.03	0.15	-0.05	-0.11	0.07
Total Direct Cost	0.15	-0.37	0.42	-0.11	-0.20	-0.01	-0.27	-0.16	0.04	0.02
Site Overheads %	0.07	0.02	-0.11	-0.22	0.14	0.21	-0.07	-0.11	-0.21	-0.24

Table 35: Spearman Correlation Test for Inputs and Site Overheads Percentage Allocated to Each Subcategory

Rank-Order Correlation Table	**Salaries & Wages % Data Set #1**	**Site Facilities % Data Set #1**	**Accommodation % Data Set #1**	**Mobilization & Demobilization % Data Set #1**	**Communication & IT % Data Set #1**	**Site Equipment % Data Set #1**	**Personnel & Material Transportation % Data Set #1**	**Quality and Safety % Data Set #1**	**Engineering % Data Set #1**	**Client/Consultant % Data Set #1**
Project Type	-0.11	-0.31	0.23	0.31	0.22	-0.07	0.26	-0.12	0.24	-0.19
Project Location	-0.12	-0.11	0.17	0.03	-0.17	0.17	0.12	-0.07	0.02	-0.08
Project Duration (months)	0.26	0.15	0.07	-0.38	-0.03	-0.11	-0.15	0.09	-0.23	0.02
Client Type	-0.07	-0.17	0.58	0.29	-0.03	0.01	-0.10	-0.12	-0.15	-0.16
Contract Type	-0.09	0.02	-0.25	0.13	-0.12	0.05	0.13	-0.04	0.07	-0.11
Avg. Interest Rate	0.09	-0.19	0.36	-0.05	-0.02	-0.07	0.10	-0.01	-0.17	-0.10
Avg. Inflation Rate	0.20	-0.39	0.35	-0.25	-0.13	-0.09	-0.10	0.05	-0.26	0.03
Avg. USD to EGP Exchange Rate	0.22	-0.42	0.32	-0.27	-0.06	-0.08	-0.06	0.00	-0.24	0.04
Total Direct Cost	0.33	-0.28	0.45	-0.21	-0.29	-0.11	-0.36	-0.07	-0.24	0.10
Site Overheads %	0.11	0.05	-0.14	-0.23	0.11	0.16	0.12	-0.08	-0.14	-0.18

As shown in Tables 34 and 35, while most correlations are weak to moderate, some trends offer valuable insights. Project duration emerges as a key driver across categories. Longer durations are associated with higher salaries and wages as well as accommodation costs. Client type also seems to influence overall spending as it affects mainly the site facilities and mobilization and demobilization costs. Furthermore, project type tends to have an impact on the mobilization and demobilization as well as the site facilities percentage. Also, the project location affects mainly the site facilities, accommodation as well as personnel and material transportation. The total direct cost shows a positive correlation with most subcategories, like salaries & wages, site facilities, and potentially mobilization & demobilization. The impact of contract type on the different subcategories appears to be less prominent based on the weak correlations observed. Regarding the macroeconomic indicators, they seem to have a weak to moderate correlation with some categories like salaries, site facilities and accommodation. This suggests that higher rates might lead to increased costs in these areas due to rising material or service costs.

4.3.4. Correlation Analysis Between Inputs and Site Overheads Percentage Allocated to Each Project Construction Phase

Tables 36 and 37 show the results of the Pearson and Spearman correlations tests between inputs and percentages allocated to construction phase.

Table 36: Pearson Correlation Test Between Inputs and Site Overheads Percentage Allocated to Each Phase

Linear Correlation Table	Initiation Phase %	Growth Phase %	Maturity Phase %	Decline Phase %
Project Type	-0.328	0.321	-0.249	0.118
Project Location	0.024	-0.037	-0.086	0.256
Project Duration (months)	0.208	-0.12	0.051	-0.092
Client Type	0.03	-0.065	0.081	-0.03
Contract Type	-0.155	0.148	-0.12	0.07
Avg. Interest Rate	0.104	0.002	-0.014	-0.146
Avg. Inflation Rate	0.002	-0.008	0.002	0.017
Avg. USD to EGP Exchange Rate	-0.038	0.064	-0.079	0.042
Total Direct Cost	0.02	-0.123	0.148	0.014
Site Overheads %	0.089	0.045	-0.03	-0.212

Table 37: Spearman Correlation Test Between Inputs and Site Overheads Percentage Allocated to Each Phase

Rank-Order Correlation Table	Initiation Phase %	Growth Phase %	Maturity Phase %	Decline Phase %
Project Type	-0.312	0.31	-0.202	0.161
Project Location	0.018	-0.045	-0.055	0.309
Project Duration (months)	0.213	-0.149	0.037	-0.094
Client Type	0.026	-0.092	0.138	-0.056
Contract Type	-0.159	0.139	-0.104	0.177
Avg. Interest Rate	0.116	0.064	-0.035	-0.204
Avg. Inflation Rate	-0.015	-0.021	0.017	0.029
Avg. USD to EGP Exchange Rate	-0.023	0.012	-0.032	0.068
Total Direct Cost	0.145	-0.211	0.105	0.011
Site Overheads %	0.067	0.137	-0.048	-0.207

As shown in Tables 36 and 37, project duration exhibits a weak positive correlation with initiation phase overheads and a weak negative correlation with decline phase overheads, suggesting a potential for slightly higher initial setup costs and lower overhead needs towards the project's end. Client type and total direct cost also have weak correlations, with total direct cost showing a weak positive association with initiation phase overheads but a weak negative correlation with growth phase overheads. This might indicate that projects with higher overall costs have slightly higher initial overhead requirements but potentially lower overhead needs during the growth phase. Finally, macroeconomic indicators like interest rate, inflation rate, and exchange rate show negligible correlations with site overhead allocation across project phases.

5. Developing the ANN Model

5.1. Introduction

Several different software and tools can be used to generate prediction models, all of which have different strengths and weaknesses. One of the most powerful alternatives that could be used is Python programming language. Python presents itself as a powerful programming language that is easy to learn and considered an ideal language for scripting and rapid application development in numerous domains across various platforms (Jawahar, 2023). Python offers an extensive selection of libraries and tools for predictive data analytics, making it a robust programming language. Keras, NumPy, Tensor-flow, pandas and scikit-learn are among the widely used libraries in Python for predictive analytics. They offer diverse functionalities for manipulating data, preprocessing, modeling, and evaluating (Jawahar, 2023).

Python codes can be implemented on various softwares and platforms, one of which is Google Colaboratory. It is provided by Google as a free cloud service to promote research in Machine Learning and Artificial Intelligence. This cloud-based platform eliminates the need for any software installation by executing the Python code directly within the web browser. Operating entirely in the cloud, it offers a user-friendly platform that is highly configurable. Moreover, it provides access to Google's robust computing resources, enhancing the potential of Python and enabling users to handle demanding tasks without concerns about hardware limitations (Naik, 2022).

For this research, three different models were created using Python:

- The first ANN model estimates the total site overhead percentage based on project characteristics.
- The second ANN model then utilizes both the predicted total overhead and existing project data to forecast the breakdown of site overheads across its different subcategories.
- The third model forecasts the breakdown of the site overheads across the different construction phases (initiation, growth, maturity and decline).

In contrast to a single, complicated model, the three-stage ANN approach provides a multitude of advantages. Initially, it enhances accuracy as each distinct model concentrates on a specific task.

The first model handles the overall site overhead percentage, potentially resulting in an improved precision for this critical value. The second and third model then enhance the prediction by combining the site overheads obtained from the first model and existing project data to provide a more detailed breakdown across subcategories and phases. Furthermore, this method addresses the problem of model complexity. A single model predicting a large number of outputs all at once can become overly complex, leading to overfitting, causing the model to perform well on training data but poorly on unseen data. By separating the tasks, the models will remain focused and can be individually optimized for their specific goals.

5.2. Model 1 Code Development

Step 1: Data Sheet Preparation

On Microsoft Excel, a data set consisting of 55 different projects executed during the period from 2018 to 2024 in Egypt was prepared. The data set contained 10 variables (9 inputs and 1 output). The inputs were project type, project location, project duration, client type, contract type, average interest rate, average inflation rate, average USD to EGP exchange rate and total direct cost. It is important to note that several trials were conducted that included removing each of the 3 economic indicators. The most optimum scenario was the one that included the 3 economic indicators in the model. The output was site overheads percentage. The categorical data was coded as explained in the previous section, while the numerical data was left without normalization as the model will then perform this function itself. The excel file was then saved as CSV file to be used in the model. A sample of the python model data file is shown in Figure 31.

Project Type	Project Location	Project Duration	Client Type	Contract Type	Avg Interest Rate	Avg Inflation Rate	Avg USD to EGP Exchange Rate	Total Direct Cost	Site Overheads Percentage
2	1	28	2	1	0.0817	0.1158	18.7197	92982503.46	0.28
2	1	34	2	1	0.0924	0.0624	15.9109	411389278.6	0.24
1	1	46	2	1	0.0904	0.1347	19.4692	542555635.1	0.29
1	1	9	1	2	0.0796	0.1373	18.8586	81583232.93	0.07
1	2	37	2	1	0.0882	0.1484	24.2455	339596363.3	0.31
2	2	57	2	1	0.0959	0.1373	19.4722	1266327223	0.17
4	2	23	3	1	0.1046	0.2611	26.2181	410088807.8	0.1
3	1	30	1	3	0.0945	0.1973	23.0192	491061838.8	0.07
3	2	38	1	2	0.0885	0.1694	21.3048	403753952	0.25
2	1	41	4	2	0.0865	0.115	18.814	323828300.6	0.18
5	1	14	3	2	0.1045	0.2766	27.4782	211342369.8	0.17
5	1	20	3	2	0.0985	0.2435	25.0367	536838861.2	0.1
2	2	12	1	2	0.0762	0.0958	16.9446	115548112	0.14
1	1	30	1	2	0.0969	0.2103	23.2308	217603165	0.17
1	1	27	1	2	0.0885	0.0504	15.766	127697803	0.16
1	1	40	1	2	0.0857	0.0769	16.6041	124444448	0.17
1	2	24	1	2	0.0791	0.0678	16.0864	151055536.3	0.18
3	1	5	1	2	0.0752	0.0852	15.9024	18955408.39	0.11
1	2	25	1	2	0.08	0.0998	17.3644	113005839.6	0.19
1	2	26	1	2	0.0817	0.1158	18.2597	92714840	0.2
1	2	37	1	2	0.0919	0.0636	15.9602	155724197.4	0.2
1	1	12	1	2	0.076	0.0991	16.9499	116528525.4	0.08
1	1	38	2	1	0.1075	0.0878	16.5553	159071597	0.28
3	1	12	1	2	0.0918	0.0521	15.7803	90190443.66	0.25

Figure 31: Model 1 CSV Data File

Step 2: Importing Libraries

In order to perform the required functions, certain libraries need to be imported as shown in the Figure 32. For example, pandas library was required for data manipulation and analysis, MinMaxScaler was needed for data normalization and TensorFlow.keras was needed for building the artificial neural network model.

```
# Import libraries
import pandas as pd
import numpy as np
from tensorflow import keras
from sklearn.preprocessing import MinMaxScaler
from sklearn.model_selection import train_test_split
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense
from tensorflow.keras.layers import Activation
```

Figure 32: Code for Importing Libraries for Model 1

Step 3: Loading Data

In order to import and load the data into the model, the code shown in Figure 33 was included.

```
# Loading data

data = pd.read_csv("/content/ANN Model's Database.csv")
```

Figure 33: Code for Loading Data of Model 1

Step 4: Defining the Model's Inputs and Outputs

As shown in the Figure 34, the next step was defining the model's 9 inputs and its targeted output.

```
# Defining features (inputs) and target (output)

features = [
    "Project Type",
    "Project Location",
    "Project Duration",
    "Client Type",
    "Contract Type",
    "Avg Interest Rate",
    "Avg Inflation Rate",
    "Avg USD to EGP Exchange Rate",
    "Total Direct Cost",
]
target = 'Site Overheads Percentage'
```

Figure 34: Code for Defining Model 1 Variables

Step 5: Data Preprocessing

The first step was defining the categorical and numerical features available in the loaded data set. The categorical features were already coded prior to importing them; thus, no further preprocessing was needed for them. Regarding the numerical features, they were normalized using MinMaxScaler. This technique scales the data to a range of 0 to 1, ensuring all features contribute equally during model training. The code needed for this function is shown in Figure 35.

```
# Data Preprocessing

# Identify categorical features (assuming these are already coded)
categorical_features = ['Project Type', 'Project Location', 'Client Type', 'Contract Type']

# Numerical Data Normalization

scaler = MinMaxScaler(feature_range=(0, 1))
numerical_features = ['Project Duration', 'Avg Interest Rate',
                      'Avg Inflation Rate', 'Avg USD to EGP Exchange Rate',
                      'Total Direct Cost']
data[numerical_features] = scaler.fit_transform(data[numerical_features])
```

Figure 35: Code for Data Preprocessing of Model 1

Step 6: Splitting Data into Training and Testing Sets

Then, the data was split into training and testing tests as follows: 80% for training and 20% for testing. The code performing this step is shown in Figure 36.

```
#Splitting the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(data[features], data[target], test_size=0.2, random_state=0)
```

Figure 36: Code for Splitting Data into Training and Testing Sets of Model 1

Step 7: Defining the Neural Network Model

This section of the code is responsible for defining the architecture of the artificial neural network using Keras. The model has 2 hidden layers. The first hidden layer contains 64 neurons and ReLU (Rectified Linear Unit) activation function. The second hidden layer contains 32 neurons and ReLU activation function. The number of hidden layers, neurons within each hidden layer and their activation functions were selected after conducting several trials, each having different number of hidden layers, neurons and activation functions and reaching the most optimum neural network. The output layer contains only one neuron representing the site overheads percentage.

The activation function used was sigmoid activation function, which is suitable for predicting percentages as it outputs values between 0 and 1. The code used for defining the model is shown in Figure 37.

```
# Defining the neural network model
model = Sequential()
model.add(Dense(64, activation='relu', input_shape=(data[features].shape[1],))) # First hidden layer with 64 neurons and ReLU activation
model.add(Dense(32, activation='relu'))  # Second hidden layer with 32 neurons and ReLU activation
model.add(Dense(1, activation='sigmoid'))  # Output layer with 1 neuron and sigmoid activation (for percentage)
```

Figure 37: Code for Defining Model 1

Step 8: Compiling the Model

In this step, the model is compiled, specifying the loss function, optimizer, and the metrics used for evaluation. Regarding the optimizer, Adam optimizer, which is an efficient optimization algorithm for training ANNs that is well-suited for large datasets and parameters, was used to update the model's weights during training to minimize the loss function (Kingma, 2014). For the loss function, mean squared error is used to measure the difference between the model's predictions and the actual target values. Lastly, the mean absolute error was used as an additional metric to evaluate the model's performance. The code performing this step is shown in the Figure 38.

```
# Compiling the model
model.compile(loss='mse', optimizer='adam', metrics=['mae'])  # Mean squared error loss, mean absolute error metric
```

Figure 38: Code for Compiling Model 1

Step 9: Training the Model

The next step was training the model on the training data, as shown in the Figure 39. The training process iterated through the training data for 100 iterations in order to learn the patterns.

```
#Fitting the ANN to the training set
model.fit(X_train, y_train, batch_size=32, epochs=100)
```

Figure 39: Code for Model 1 Training

Step 10: Evaluating the Model

After that, the model was evaluated by calculating the MAE and MSE for the entire data set, as shown in Figure 40.

```
# Evaluating the model
loss, mae = model.evaluate(data[features], data[target])
print(f"Loss: {loss}, Mean Absolute Error: {mae}")
print("Mean Squared Error", loss)
```

Figure 40: Code for Model 1 Evaluation

Step 11: Validating the Model

After that, it was necessary to evaluate the model's performance on new, unseen data. MSE and MAE were used to assess the validation process. The code used for Model 1 validation is shown in Figure 41.

```
# Model Validation using New Project Data

# Loading new project data from a separate CSV file
new_data = pd.read_csv("/content/Validation Data File.csv")

# Preprocessing new data
new_data[numerical_features] = scaler.transform(new_data[numerical_features])

# Predicting site overheads percentage for new projects
new_predictions = model.predict(new_data[features])

# Print the predicted site overhead percentages
print("Predicted Site Overhead Percentage for New Projects:")
print(new_predictions)
```

Figure 41: Code for Model 1 Validation

5.3. Model 2 Code Development

Step 1: Data Sheet Preparation

The data set contained 20 variables (10 inputs and 10 outputs). The inputs were project type, project location, project duration, client type, contract type, average interest rate, average inflation rate, average USD to EGP exchange rate, total direct cost and site overheads percentage. The outputs were the percentages of site overheads allocated to salaries and wages, site facilities, accommodation, mobilization and demobilization, communication and IT, site equipment, material and personnel transportation expenses, engineering fees, owner and consultant expenses. The categorical data was coded as explained in the previous section, while the numerical data was left without normalization as the model will then perform this function itself. The excel file was then saved as CSV file to be used in the model. A sample of the python Model 2 data file is shown in the Figure 42.

Project Type	Project Location	Project Duration	Client Type	Contract Type	Avg Interest Rate	Avg Inflation Rate	Avg USD to EGP Exchange Rate	Total Direct Cost	Site Overheads Percentage	Salaries & Wages %	Site Facilities %	Accommodation %	Mobilization & Demobilization %	Communication & IT %	Site Equipment %	Personnel & Material Transportation %	Quality and Safety %	Engineering %	Client/Consultant %
2	1	28	2	1	0.082	0.116	18.72	92982503.46	0.28	0.57	0.06	0.01	0	0.03	0.26	0.02	0.01	0	0.04
2	1	34	2	1	0.092	0.062	15.91	411389278.6	0.24	0.56	0.07	0.02	0.01	0.02	0.23	0.01	0.02	0.04	0.01
1	1	46	2	1	0.09	0.135	19.47	542555635.1	0.29	0.62	0.07	0.02	0	0.02	0.2	0.03	0.01	0.01	0.01
1	1	9	1	2	0.08	0.137	18.86	81583232.93	0.07	0.62	0.05	0.02	0	0	0.11	0.03	0.02	0.02	0.04
1	2	37	2	1	0.088	0.148	24.25	339596363.3	0.31	0.52	0.07	0.02	0.01	0.01	0.23	0.08	0.04	0.02	0
2	2	57	2	1	0.096	0.137	19.47	1266327223	0.17	0.51	0.03	0.03	0.01	0.02	0.27	0.09	0.03	0.01	0.01
4	2	23	3	1	0.105	0.261	26.22	410088807.8	0.1	0.51	0.06	0.03	0.01	0.01	0.14	0.12	0.01	0	0.11
3	1	30	1	3	0.094	0.197	23.02	491061838.8	0.07	0.74	0.02	0.01	0	0.01	0.11	0	0.01	0.02	0.07
3	2	38	1	2	0.089	0.169	21.3	403753952	0.25	0.61	0.07	0.03	0.01	0.02	0.14	0.09	0.04	0	0
2	1	41	4	2	0.086	0.115	18.81	323828300.6	0.18	0.62	0.04	0.03	0.01	0.01	0.1	0.08	0.01	0	0
5	1	14	3	2	0.105	0.277	27.48	211342369.8	0.17	0.64	0.05	0.03	0.03	0.01	0.13	0.08	0.03	0	0
5	1	20	3	2	0.098	0.243	25.04	536838861.2	0.1	0.54	0.07	0.03	0.04	0.01	0.18	0.08	0.04	0	0
2	2	12	1	2	0.076	0.096	16.94	115548112	0.14	0.27	0.08	0	0.03	0.01	0.39	0.07	0.07	0.07	0.01
1	1	30	1	2	0.097	0.21	23.23	217603165	0.17	0.49	0.08	0.01	0.01	0.01	0.27	0.05	0.02	0.02	0.03
1	1	27	1	2	0.088	0.05	15.77	127697803	0.16	0.55	0.03	0	0.02	0	0.26	0.07	0.01	0.04	0.02
1	1	40	1	2	0.086	0.077	16.6	124444448	0.17	0.58	0.06	0	0.02	0	0.19	0.08	0.01	0.04	0.01
1	2	24	1	2	0.079	0.068	16.09	151055536.3	0.18	0.59	0.07	0.02	0.01	0	0.18	0.06	0.02	0.04	0.01
3	1	5	1	2	0.075	0.085	15.9	18955408.39	0.11	0.37	0.14	0	0.04	0.02	0.29	0.06	0.02	0.06	0

Figure 42: Model 2 CSV Data File

Step 2: Importing Libraries

In order to perform the required functions, certain libraries need to be imported, as shown in the Figure 43.

```
# Import libraries
import pandas as pd
import numpy as np
from tensorflow import keras
from sklearn.preprocessing import MinMaxScaler
from sklearn.model_selection import train_test_split
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense
from tensorflow.keras.layers import Activation
```

Figure 43: Code for Importing Libraries for Model 2

Step 3: Loading Data

In order to import and load the data into the model, the code shown in Figure 44 was included.

```
# Loading data

data = pd.read_csv("/content/ANN Model 2 Database.csv")
```

Figure 44: Code for Loading Data of Model 2

Step 4: Defining the Model's Inputs and Outputs

As shown in the Figure 45, the next step was defining the model's 10 inputs and its targeted outputs.

```
features = [
    "Project Type",
    "Project Location",
    "Project Duration",
    "Client Type",
    "Contract Type",
    "Avg Interest Rate",
    "Avg Inflation Rate",
    "Avg USD to EGP Exchange Rate",
    "Total Direct Cost",
    "Site Overheads Percentage",
]
target = [
    "Salaries & Wages %",
    "Site Facilities %",
    "Accommodation %",
    "Mobilization & Demobilization %",
    "Communication & IT %",
    "Site Equipment %",
    "Personnel & Material Transportation %",
    "Quality and Safety %",
    "Engineering %",
    "Client/Consultant %",
]
```

Figure 45: Code for Defining Model 2 Variables

Step 5: Data Preprocessing

The first step was defining the categorical and numerical features available in the loaded data set. The categorical features were already coded prior to importing them, thus no further preprocessing was needed for them. Regarding the numerical features, they were normalized using MinMaxScaler. The code needed for this function is shown in Figure 46.

```
# Data Preprocessing

# Identify categorical features (assuming these are already coded)
categorical_features = ['Project Type', 'Project Location', 'Client Type', 'Contract Type']

# Numerical Data Normalization

scaler = MinMaxScaler(feature_range=(0, 1))
numerical_features = ['Project Duration', 'Avg Interest Rate',
                      'Avg Inflation Rate', 'Avg USD to EGP Exchange Rate',
                      'Total Direct Cost','Site Overheads Percentage']
data[numerical_features] = scaler.fit_transform(data[numerical_features])
```

Figure 46: Code for Data Preprocessing of Model 2

Step 6: Splitting Data into Training and Testing Sets

Then, the data was split into training and testing tests as follows: 80% for training and 20% for testing. The code used for performing this step is shown in Figure 47.

```
#Splitting the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(data[features], data[target], test_size=0.2, random_state=0)
```

Figure 47: Code for Splitting Data into Training and Testing Sets of Model 2

Step 7: Defining the Neural Network Model

This section of the code is responsible for defining the architecture of the artificial neural network using Keras. The model has 2 hidden layers. The first hidden layer contains 128 neurons and ReLU (Rectified Linear Unit) activation function. The second hidden layer contains 64 neurons and ReLU activation function. The number of hidden layers, neurons within each hidden layer and their activation functions were selected after conducting several trials, each having different number of hidden layers, neurons and activation functions and reaching the most optimum neural network. The output layer contains 10 neurons, each neuron representing a site overheads subcategory. The activation function used was softmax, which is suitable for making sure that the outputs sum to 1. The code used for defining Model 2 is shown in Figure 48.

```
# Defining the neural network model
model = Sequential()
model.add(Dense(128, activation='relu', input_shape=(data[features].shape[1],))) # First hidden layer with 128 neurons and ReLU activatic
model.add(Dense(64, activation='relu'))  # Second hidden layer with 64 neurons and ReLU activation
model.add(Dense(len(target), activation='softmax'))  # Output layer with 10 neurons and softmax activation
```

Figure 48: Code for Defining Model 2

Step 8: Compiling the Model

In this step, the model is compiled, specifying the loss function, optimizer, and the metrics used for evaluation. For the loss function, mean squared error is used. Additionally, the Adam optimizer was used to update the model's weights during training to minimize the loss function. Lastly, the mean absolute error was used as an additional metric to evaluate the model's performance. The code performing this step is shown in Figure 49.

```
# Compiling the model
model.compile(loss='mse', optimizer='adam', metrics=['mae'])  # Mean squared error loss, mean absolute error metric
```

Figure 49: Code for Compiling Model 2

Step 9: Training the Model

The next step was training the model on the training data, as shown in Figure 50. The training process iterated through the training data for 100 iterations in order to learn the patterns.

```
#Fitting the ANN to the training set
model.fit(X_train, y_train, batch_size=32, epochs=100)
```

Figure 50: Code for Model 2 Training

Step 10: Evaluating the Model

After that, the model was evaluated by calculating the MAE and MSE for the entire data set, as shown in Figure 51.

```
# Evaluating the model
loss, mae = model.evaluate(data[features], data[target])
print(f"Loss: {loss}, Mean Absolute Error: {mae}")
print("Mean Squared Error", loss)
```

Figure 51: Code for Model 2 Evaluation

Step 11: Validating the Model

After that, it was necessary to evaluate the model's performance on new, unseen data. MSE and MAE were used to assess the validation process. The code used for Model 2 validation is shown in Figure 52.

```
# Model Validation using New Project Data

# Loading new project data from a separate CSV file
new_data = pd.read_csv("/content//Model 2 Validation Data File.csv")

# Preprocessing new data
new_data[numerical_features] = scaler.transform(new_data[numerical_features])

# Predicting the percentage of each category
new_predictions = model.predict(new_data[features])

# Print the predicted percentage of each category
print("Predicted Percentage of Each Category:")
print(new_predictions)
```

Figure 52: Code for Model 2 Validation

5.4. Model 3 Code Development

Step 1: Data Sheet Preparation

The data set contained 14 variables (10 inputs and 4 outputs). The inputs were project type, project location, project duration, client type, contract type, average interest rate, average inflation rate, average USD to EGP exchange rate, total direct cost and site overheads percentage. The outputs were the percentages of site overheads allocated to the initiation phase, growth phase, maturity phase and decline phase. The categorical data was coded as explained in the previous section, while the numerical data was left without normalization as the model will then perform this function itself. The excel file was then saved as CSV file to be used in the model. A sample of the python Model 3 data file is shown in the Figure 53.

Project Type	Project Location	Project Duration	Client Type	Contract Type	Avg Interest Rate	Avg Inflation Rate	Avg USD to EGP Exchange Rate	Total Direct Cost	Site Overheads Percentage	Initiation Phase %	Growth Phase %	Maturity Phase %	Decline Phase %
2	1	28	2	1	0.082	0.116	18.72	92982503.5	0.28	0.12	0.6	0.2	0.08
2	1	34	2	1	0.092	0.062	15.91	411389279	0.24	0.15	0.6	0.2	0.05
1	1	46	2	1	0.09	0.135	19.47	542555635	0.29	0.15	0.69	0.11	0.05
1	1	9	1	2	0.08	0.137	18.86	81583232.9	0.07	0.1	0.65	0.2	0.05
1	2	37	2	1	0.088	0.148	24.25	339596363	0.31	0.15	0.6	0.2	0.05
2	2	57	2	1	0.096	0.137	19.47	1266327223	0.17	0.1	0.7	0.15	0.05
4	2	23	3	1	0.105	0.261	26.22	410088808	0.1	0.12	0.7	0.1	0.08
3	1	30	1	3	0.094	0.197	23.02	491061839	0.07	0.14	0.7	0.11	0.05
3	2	38	1	2	0.089	0.169	21.3	403753952	0.25	0.1	0.7	0.1	0.1
2	1	41	4	2	0.086	0.115	18.81	323828301	0.18	0.12	0.72	0.11	0.05
5	1	14	3	2	0.105	0.277	27.48	211342370	0.17	0.12	0.66	0.16	0.06
5	1	20	3	2	0.098	0.243	25.04	536838861	0.1	0.05	0.68	0.19	0.07
2	2	12	1	2	0.076	0.096	16.94	115548112	0.14	0.12	0.66	0.15	0.07
1	1	30	1	2	0.097	0.21	23.23	217603165	0.17	0.09	0.61	0.2	0.1
1	1	27	1	2	0.088	0.05	15.77	127697803	0.16	0.1	0.7	0.12	0.08
1	1	40	1	2	0.086	0.077	16.6	124444448	0.17	0.06	0.78	0.11	0.06
1	2	24	1	2	0.079	0.068	16.09	151055536	0.18	0.09	0.72	0.12	0.07
3	1	5	1	2	0.075	0.085	15.9	18955408.4	0.11	0.05	0.69	0.2	0.06
1	2	25	1	2	0.08	0.1	17.36	113005840	0.19	0.11	0.61	0.18	0.1
1	2	26	1	2	0.082	0.116	18.26	92714840	0.2	0.14	0.7	0.11	0.06

Figure 53: Model 3 CSV Data File

Step 2: Importing Libraries

In order to perform the required functions, certain libraries need to be imported as shown in Figure 54.

```python
# Import libraries
import pandas as pd
import numpy as np
from tensorflow import keras
from sklearn.preprocessing import MinMaxScaler
from sklearn.model_selection import train_test_split
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense
from tensorflow.keras.layers import Activation
```

Figure 54: Code for Importing Libraries for Model 3

Step 3: Loading Data

In order to import and load the data into the model, the code shown in Figure 55 was included.

```
# Loading data

data = pd.read_csv("/content/ANN Model 3 Database.csv")
```

Figure 55: Code for Loading Data of Model 3

Step 4: Defining the Model's Inputs and Outputs

As shown in Figure 56, the next step was defining the model's 10 inputs and its targeted outputs.

```
# Defining features (inputs) and target (output)

features = [
    "Project Type",
    "Project Location",
    "Project Duration",
    "Client Type",
    "Contract Type",
    "Avg Interest Rate",
    "Avg Inflation Rate",
    "Avg USD to EGP Exchange Rate",
    "Total Direct Cost",
    "Site Overheads Percentage",
]
target = [
    "Initiation Phase %",
    "Growth Phase %",
    "Maturity Phase %",
    "Decline Phase %",
]
```

Figure 56: Code for Defining Model 3 Variables

Step 5: Data Preprocessing

The first step was defining the categorical and numerical features available in the loaded data set. The categorical features were already coded prior to importing them, thus no further preprocessing was needed for them. Regarding the numerical features, they were normalized using MinMaxScaler. The code needed for this function is shown in Figure 57.

```
# Data Preprocessing

# Identify categorical features (assuming these are already coded)
categorical_features = ['Project Type', 'Project Location', 'Client Type', 'Contract Type']

# Numerical Data Normalization

scaler = MinMaxScaler(feature_range=(0, 1))
numerical_features = ['Project Duration', 'Avg Interest Rate',
                      'Avg Inflation Rate', 'Avg USD to EGP Exchange Rate',
                      'Total Direct Cost','Site Overheads Percentage']
data[numerical_features] = scaler.fit_transform(data[numerical_features])
```

Figure 57: Code for Data Preprocessing of Model 3

Step 6: Splitting Data into Training and Testing Sets

Then, the data was split into training and testing tests as follows: 80% for training and 20% for testing. The code performing this step is shown in Figure 58.

```
#Splitting the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(data[features], data[target], test_size=0.2, random_state=0)
```

Figure 58: Code for Splitting Data into Training and Testing Sets of Model 3

Step 7: Defining the Neural Network Model

This section of the code is responsible for defining the architecture of the artificial neural network using Keras. The model has 2 hidden layers. The first hidden layer contains 128 neurons and ReLU (Rectified Linear Unit) activation function. The second hidden layer contains 64 neurons and ReLU activation function. The number of hidden layers, neurons within each hidden layer and their activation functions were selected after conducting several trials, each having different number of hidden layers, neurons and activation functions and reaching the most optimum neural network. The output layer contains 4 neurons, each neuron representing a project construction phase. The activation function used was softmax, which is suitable for making sure that the outputs sum to 1. The code used for defining Model 3 is shown in Figure 59.

```
# Defining the neural network model
model = Sequential()
model.add(Dense(128, activation='relu', input_shape=(data[features].shape[1],))) # First hidden layer with 128 neurons and ReLU activatic
model.add(Dense(64, activation='relu'))  # Second hidden layer with 64 neurons and ReLU activation
model.add(Dense(len(target), activation='softmax'))  # Output layer with 4 neurons and softmax activation
```

Figure 59: Code for Defining Model 3

Step 8: Compiling the Model

In this step, the model is compiled, specifying the loss function, optimizer, and the metrics used for evaluation. For the loss function, mean squared error is used. Additionally, the Adam optimizer was used to update the model's weights during training to minimize the loss function. Lastly, the mean absolute error was used as an additional metric to evaluate the model's performance. The code performing this step is shown in Figure 60.

```
# Compiling the model
model.compile(loss='mse', optimizer='adam', metrics=['mae'])  # Mean squared error loss, mean absolute error metric
```

Figure 60: Code for Compiling Model 3

Step 9: Training the Model

The next step was training the model on the training data, as shown in Figure 61. The training process iterated through the training data for 100 iterations in order to learn the patterns.

```
#Fitting the ANN to the training set
model.fit(X_train, y_train, batch_size=32, epochs=100)
```

Figure 61: Code for Model 3 Training

Step 10: Evaluating the Model

After that, the model was evaluated by calculating the MAE and MSE for the entire data set, as shown in Figure 62.

```
# Evaluating the model
loss, mae = model.evaluate(data[features], data[target])
print(f"Loss: {loss}, Mean Absolute Error: {mae}")
print("Mean Squared Error", loss)
```

Figure 62: Code for Model 3 Evaluation

Step 11: Validating the Model

After that, it was necessary to evaluate the model's performance on new, unseen data. MSE and MAE were used to assess the validation process. The code used for Model 3 validation is shown in Figure 63.

```
# Model Validation using New Project Data

# Loading new project data from a separate CSV file
new_data = pd.read_csv("/content//Model 3 Validation Data File.csv")

# Preprocessing new data
new_data[numerical_features] = scaler.transform(new_data[numerical_features])

# Predicting the percentage of each category
new_predictions = model.predict(new_data[features])

# Print the predicted percentage of each category
print("Predicted Percentage of Each Category:")
print(new_predictions)
```

Figure 63: Code for Model 3 Validation

6. Results and Discussion

6.1. Python Model 1 Results

6.1.1. Training Phase

As it was previously stated, 80% of the project data was used for training and 20% for testing. The accuracy of the training and testing processes was evaluated using the two parameters: mean absolute error (MAE) and mean squared error (MSE). The model experienced a training process that involved 100 iterations, during which it showed a constant improvement in its accuracy. The MSE function decreased from 0.0135 during the first iteration to 0.0013 in the last one. Regarding the MAE, it decreased from 8.79% to 2.75%. This implies that the model has effectively learned from the training data and has progressively improved its predictions.

6.1.2. Testing Phase

For the testing data set, the MSE was 0.0031 and the MAE was 3.9%. While the model performed well on the testing set, it displayed a slight increase in both the MSE and MAE compared to its performance on the training set. This minor difference indicates a potential overfitting problem. Given the relatively small difference in performance metrics between the training and testing sets, it was determined that the observed overfitting is considered acceptable for the current model iteration as it will not have a significant effect on the prediction accuracy for new, unseen data. This will be further assessed using the validation data set.

6.1.3. Validation Phase

After the model's training and testing, it was important to further evaluate the model using a dataset that it has not seen before to assess its ability to predict using new data. For the validation phase, the data set in Table 38 was used as input.

Table 38: Model 1 Validation Set

Project No.	Project Type	Project Location	Project Duration (months)	Client Type	Contract Type	Avg. Interest Rate	Avg. Inflation Rate	Avg. USD to EGP Exchange Rate	Total Direct Cost (EGP)
1	Marine	outside the city/rural	27	Public Sector (Ministry/Government)	Remeasured	12.5%	21.1%	17.75	97,900,000.00
2	Administrative	inside the city/urban	30	Private Sector (International)	Lump Sum	8.0%	11.2%	18.65	102,280,753.80
3	Residential	inside the city/urban	45	Private Sector (International)	Lump Sum	8.5%	13.2%	19.55	569,683,416.83
4	Commercial	outside the city/rural	40	Private Sector (National)	Remeasured	8.9%	16.1%	21.56	423,941,649.60
5	Infrastructure	outside the city/rural	14	Public Sector (Ministry/Government)	Lump Sum	12.2%	35.1%	30.60	1,201,857,757.21

The results are shown in Table 39.

Table 39: Model 1 Validation Results

Project No.	Model's Predicted Value	Actual Value	Absolute Percentage Error
1	11.80%	12.69%	7.01%
2	24.46%	26.10%	6.28%
3	28.69%	28.54%	0.53%
4	26.30%	24.54%	7.17%
5	9.03%	8.74%	3.32%

From Table 39, it can be noted that the absolute percentage error of each of the 5 projects was within 10%, with project 2 having the lowest error (0.53%). The mean absolute percentage error of the validation set is 4.86%.

Figure 64 shows a graph of the predicted versus actual values for the site overheads percentage.

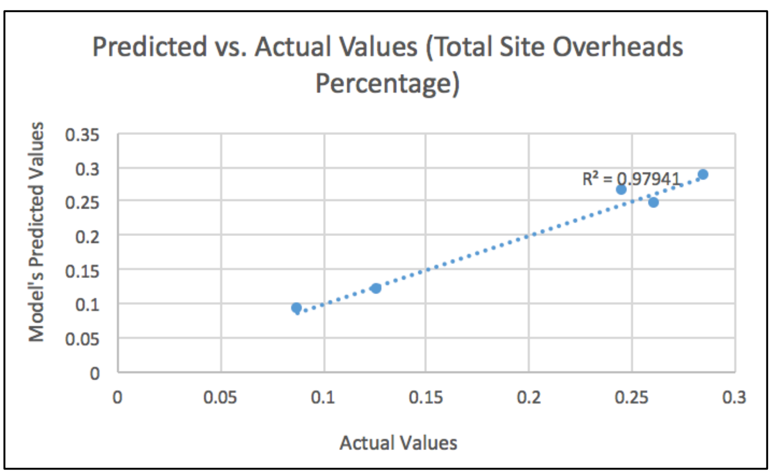

Figure 64: Predicted vs. Actual Values (Total Site Overheads Percentage)

6.2. Python Model 2 Results

6.2.1. Training Phase

Similar to Model 1, 80% of the project data was used for training and 20% for testing. The accuracy of the training and testing processes was evaluated using same two parameters: mean absolute error (MAE) and mean squared error (MSE). The model experienced a training process that involved 100 iterations, during which it showed a constant improvement in its accuracy. The MSE decreased from 0.0250 during the first iteration to 0.0017 in the last one. Regarding the MAE, it decreased from 11.02% to 2.62%. Overall, the results suggest the model has learned from the training data and may perform well on unseen data.

6.2.2. Testing Phase

For the testing data set, the MSE was 0.0024 and the MAE was 2.83%. The MSE and MAE of the testing set are slightly higher than the training MSE and MAE, but still within a reasonable range, indicating that this difference will not have a significant effect on the prediction accuracy for new, unseen data. This will be further assessed using the validation data set.

6.2.3. Validation Phase

After the model's training and testing, it was important to further evaluate the model using a dataset that it has not seen before to assess its ability to predict using new data. For the validation phase, the data set in Table 40 was used as input.

Table 40: Model 2 Validation Set

Project No.	Project Type	Project Location	Project Duration (months)	Client Type	Contract Type	Avg. Interest Rate	Avg. Inflation Rate	Avg. USD to EGP Exchange Rate	Total Direct Cost (EGP)	Site Overheads Percentage
1	Marine	outside the city/rural	27	Public Sector (Ministry/Government)	Remeasured	12.5%	21.1%	17.75	97,900,000.00	12.69%
2	Administrative	inside the city/urban	30	Private Sector (International)	Lump Sum	8.0%	11.2%	18.65	102,280,753.80	26.10%
3	Residential	inside the city/urban	45	Private Sector (International)	Lump Sum	8.5%	13.2%	19.55	569,683,416.83	28.54%
4	Commercial	outside the city/rural	40	Private Sector (National)	Remeasured	8.9%	16.1%	21.56	423,941,649.60	24.54%
5	Infrastructure	outside the city/rural	14	Public Sector (Ministry/Government)	Lump Sum	12.2%	35.1%	30.60	1,201,857,757.21	8.74%

The results for each category are shown in Tables 41 to 50.

1. **Salaries and Wages Percentage**

Table 41: Model 2 Validation Results (Salaries and Wages)

Project No.	Model's Predicted Value	Actual Value	Absolute Percentage Error
1	69.60%	66.40%	4.82%
2	52.20%	54.58%	4.36%
3	58.65%	60.97%	3.81%
4	51.34%	50.09%	2.50%
5	47.96%	48.85%	1.82%

For the salaries and wages subcategory, the mean absolute percentage error of the validation set is 3.46%.

Figure 65 shows a graph of the predicted versus actual values for the salaries and wages percentage.

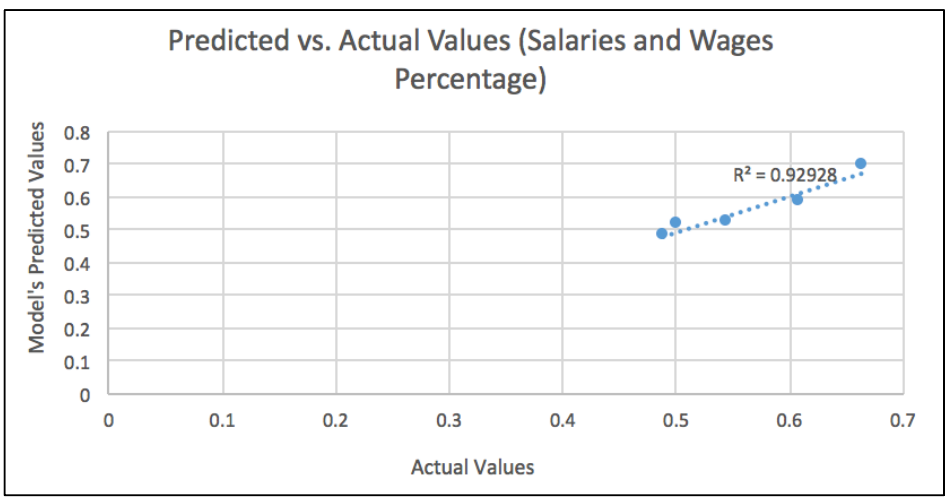

Figure 65: Predicted vs. Actual Values (Salaries and Wages Percentage)

2. Site Facilities Percentage

Table 42: Model 2 Validation Results (Site Facilities)

Project No.	Model's Predicted Value	Actual Value	Absolute Percentage Error
1	2.67%	3.25%	17.85%
2	6.19%	5.80%	6.72%
3	5.03%	4.55%	10.55%
4	5.51%	6.57%	16.11%
5	4.90%	4.23%	15.84%

For the site facilities subcategory, the mean absolute percentage error of the validation set is 13.41%.

Figure 66 shows a graph of the predicted versus actual values for the site facilities percentage.

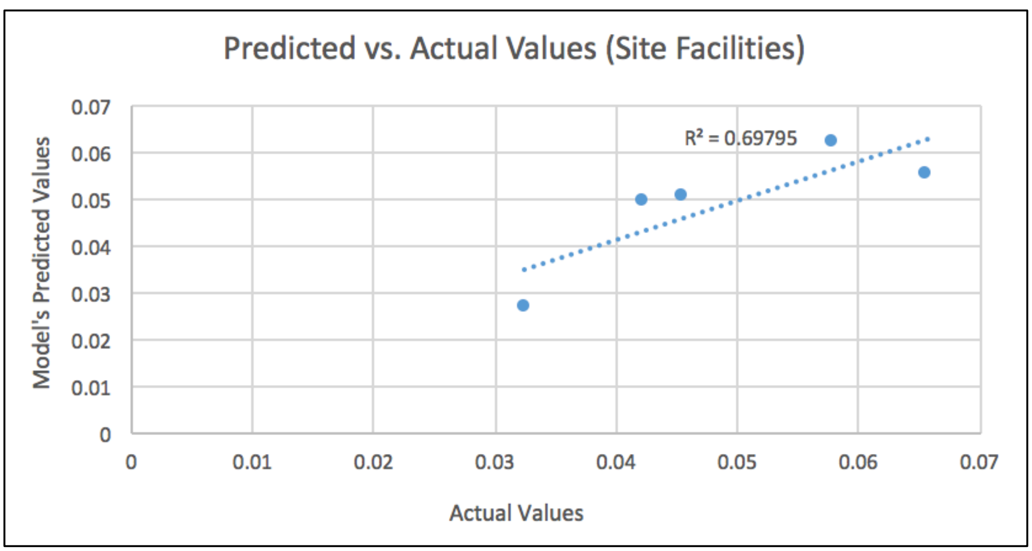

Figure 66: Predicted vs. Actual Values (Site Facilities Percentage)

3. Accommodation Percentage

Table 43: Model 2 Validation Results (Accommodation)

Project No.	Model's Predicted Value	Actual Value	Absolute Percentage Error
1	0.73%	1.01%	27.72%
2	3.79%	3.46%	9.66%
3	3.64%	3.14%	15.92%
4	1.78%	2.23%	20.18%
5	3.51%	2.69%	30.48%

For the site facilities subcategory, the mean absolute percentage error of the validation set is 20.79%.

Figure 67 shows a graph of the predicted versus actual values for the accommodation percentage.

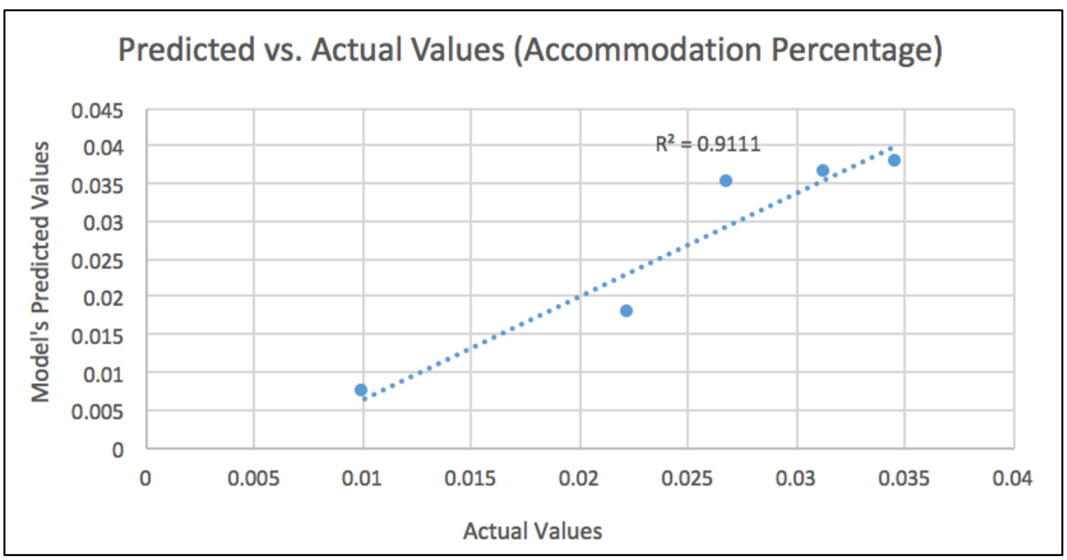

Figure 67: Predicted vs. Actual Values (Accommodation Percentage)

4. Mobilization and Demobilization Percentage

Table 44: Model 2 Validation Results (Mobilization and Demobilization)

Project No.	Model's Predicted Value	Actual Value	Absolute Percentage Error
1	1.58%	1.98%	20.20%
2	3.87%	3.22%	20.37%
3	2.91%	2.41%	20.75%
4	2.79%	3.46%	19.36%
5	4.13%	6.58%	37.23%

For the mobilization and demobilization subcategory, the mean absolute percentage error of the validation set is 23.58%.

Figure 68 shows a graph of the predicted versus actual values for the mobilization and demobilization percentage.

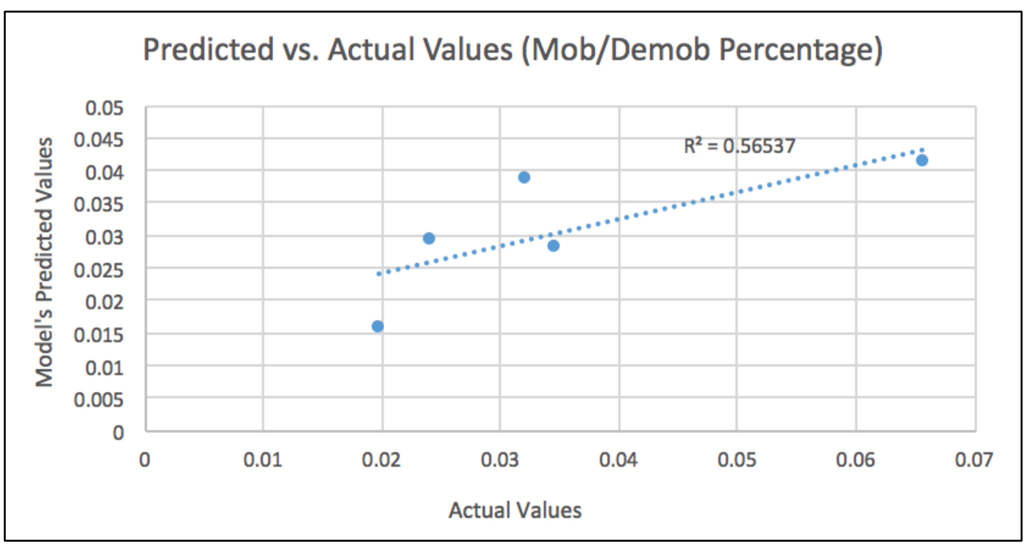

Figure 68: Predicted vs. Actual Values (Mob/Demob Percentage)

5. Communication and IT Percentage

Table 45: Model 2 Validation Results (Communication and IT)

Project No.	Model's Predicted Value	Actual Value	Absolute Percentage Error
1	0.17%	0.25%	32.00%
2	1.47%	2.27%	35.24%
3	1.14%	0.94%	21.28%
4	1.06%	1.63%	34.97%
5	1.10%	0.89%	24.01%

For the communication and IT expenses subcategory, the mean absolute percentage error of the validation set is 29.50%.

Figure 69 shows a graph of the predicted versus actual values for the communication and IT percentage.

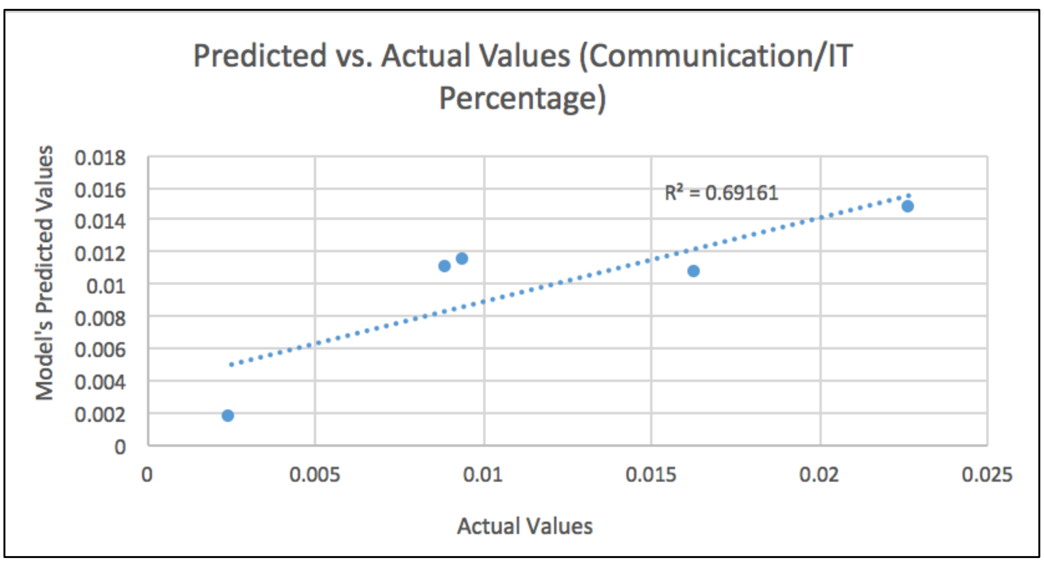

Figure 69: Predicted vs. Actual Values (Communication/IT Percentage)

6. Site Equipment Percentage

Table 46: Model 2 Validation Results (Site Equipment)

Project No.	Model's Predicted Value	Actual Value	Absolute Percentage Error
1	18.06%	20.37%	11.34%
2	18.79%	18.24%	3.02%
3	18.04%	18.92%	4.63%
4	25.71%	23.41%	9.82%
5	21.50%	22.42%	4.10%

For the site equipment subcategory, the mean absolute percentage error of the validation set is 6.58%.

Figure 70 shows a graph of the predicted versus actual values for the site equipment percentage.

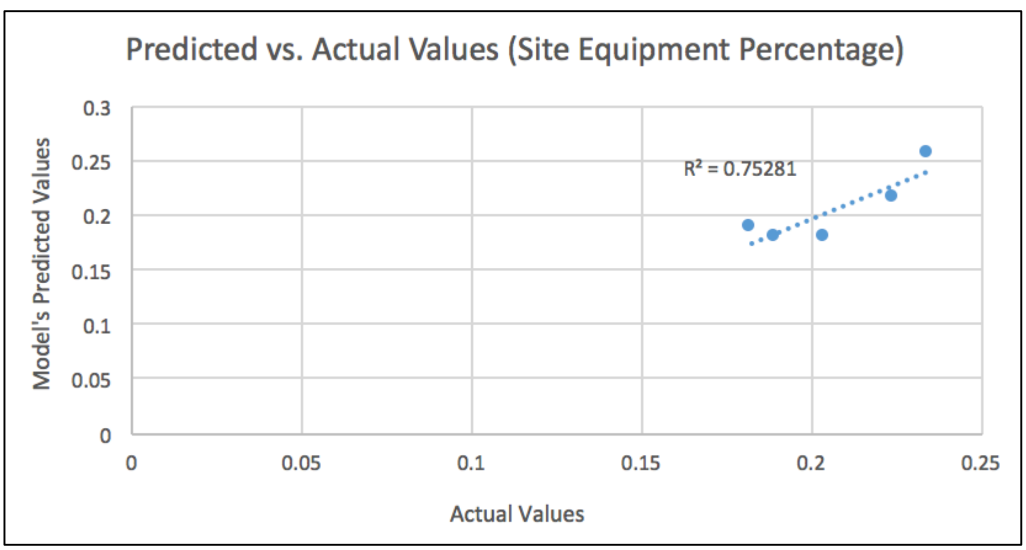

Figure 70: Predicted vs. Actual Values (Site Equipment Percentage)

7. Personnel and Material Transportation Percentage

Table 47: Model 2 Validation Results (Personnel and Material Transportation)

Project No.	Model's Predicted Value	Actual Value	Absolute Percentage Error
1	5.96%	4.45%	33.93%
2	6.80%	5.10%	33.33%
3	4.75%	3.70%	28.38%
4	7.49%	9.05%	17.24%
5	10.13%	8.43%	20.17%

For the personnel and material transportation subcategory, the mean absolute percentage error of the validation set is 26.61%.

Figure 71 shows a graph of the predicted versus actual values for the personnel and material transportation percentage.

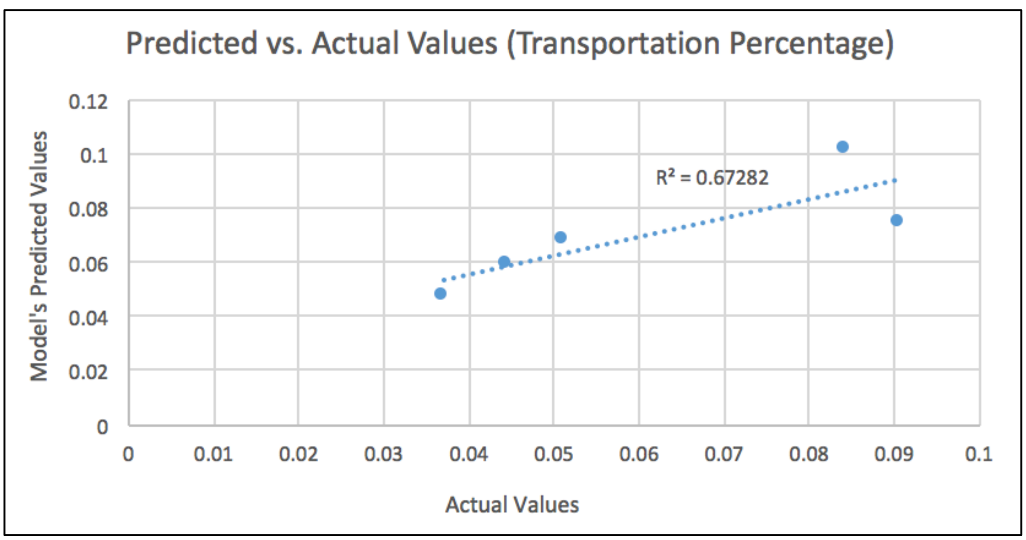

Figure 71: Predicted vs. Actual Values (Transportation Percentage)

8. Quality and Safety Percentage

Table 48: Model 2 Validation Results (Quality and Safety)

Project No.	Model's Predicted Value	Actual Value	Absolute Percentage Error
1	0.19%	0.30%	36.67%
2	2.36%	1.99%	18.59%
3	2.26%	2.46%	8.13%
4	1.10%	1.48%	25.68%
5	1.69%	1.23%	37.40%

For the quality and safety subcategory, the mean absolute percentage error of the validation set is 25.29%.

Figure 72 shows a graph of the predicted versus actual values for the quality and safety expenses percentage.

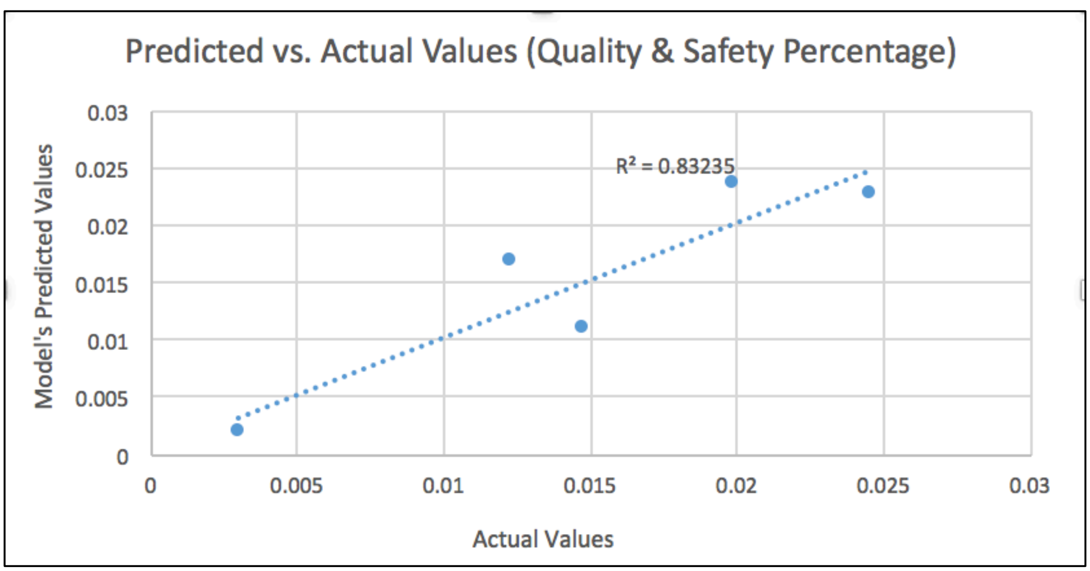

Figure 72: Predicted vs. Actual Values (Quality and Safety Percentage)

9. Engineering Percentage

Table 49: Model 2 Validation Results (Engineering Expenses)

Project No.	Model's Predicted Value	Actual Value	Absolute Percentage Error
1	0.89%	1.21%	26.45%
2	3.10%	2.52%	23.02%
3	2.37%	1.83%	29.44%
4	2.41%	1.97%	22.34%
5	3.56%	4.41%	19.27%

For the engineering expenses subcategory, the mean absolute percentage error of the validation set is 24.10%.

Figure 73 shows a graph of the predicted versus actual values for the engineering expenses percentage.

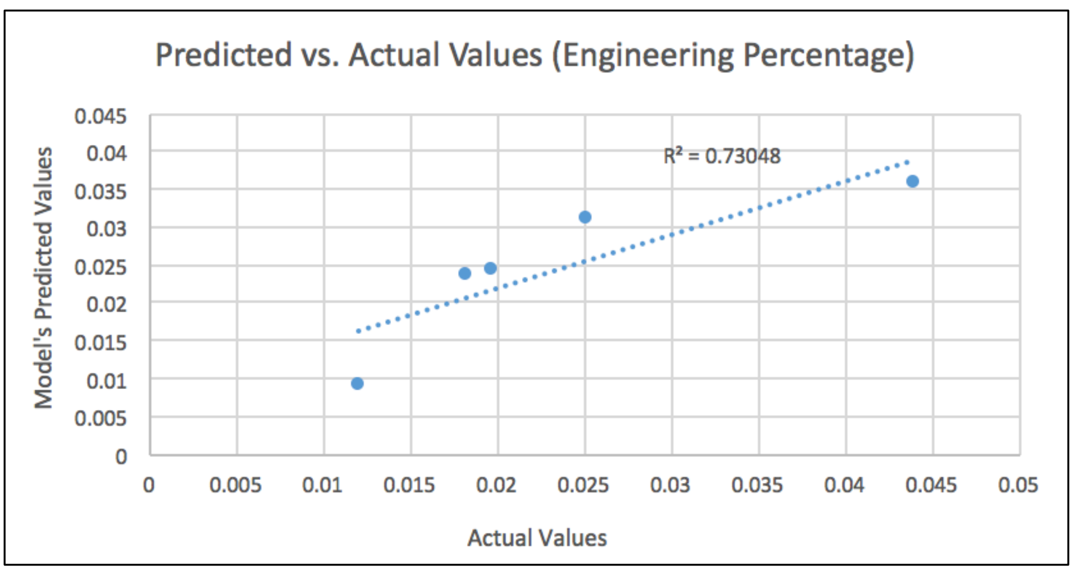

Figure 73: Predicted vs. Actual Values (Engineering Percentage)

10. Client/Consultant Percentage

Table 50: Model 2 Validation Results (Client/Consultant)

Project No.	Model's Predicted Value	Actual Value	Absolute Percentage Error
1	0.15%	0.28%	45.45%
2	1.43%	3.00%	52.33%
3	1.22%	1.08%	12.96%
4	0.82%	0.63%	30.16%
5	1.51%	1.09%	38.53%

For the client/consultant expenses subcategory, the mean absolute percentage error of the validation set is 35.89%.

Figure 74 shows a graph of the predicted versus actual values for the client/consultant expenses percentage.

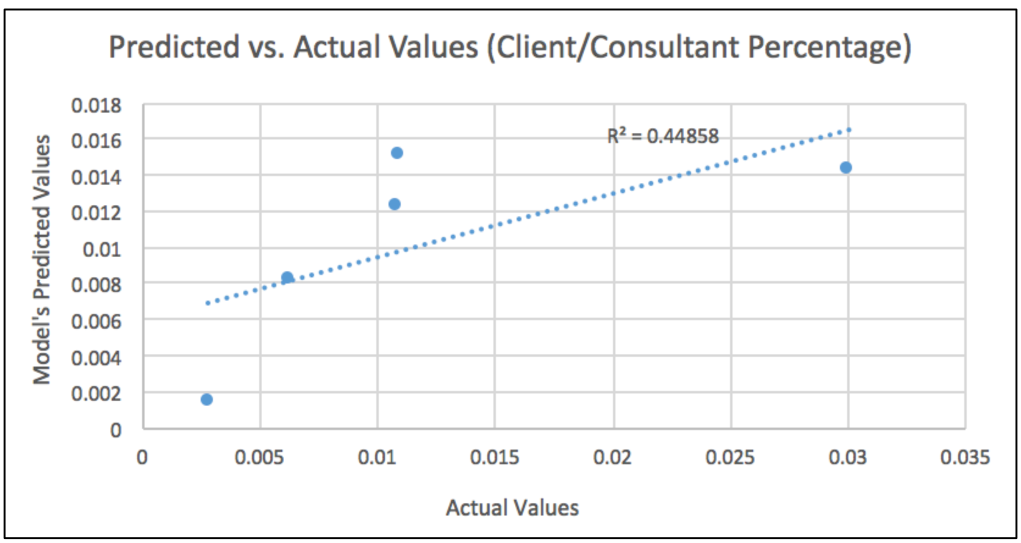

Figure 74: Predicted vs. Actual Values (Client/Consultant Percentage)

As shown in Tables 41 to 50 and Figures 65 to 74, Model 2 appears to perform well for predicting the percentages of site overheads allocated to the major subcategories, with absolute percentage errors typically below 10% for salaries and wages and site equipment. In fact, the MAPE for these 2 categories are 3.46% and 6.58% respectively, indicating an acceptable accuracy for the two major subcategories that account for most of the site overheads percentage (around 60-80% of any projects total site overheads). For the other categories, the model shows higher absolute percentage errors, often exceeding 15%. For these subcategories, it is important to note that the larger discrepancies observed in subcategories like site facilities, accommodation and mobilization and demobilization might be partially attributed to the inherently small percentages associated with these categories. In such cases, even a slight difference between predicted and actual values can translate into a high absolute percentage error. In addition to this, the larger discrepancies could also be attribute to the fact that these categories, unlike salaries and site equipment, tend to be present in some projects and absent in others, depending on the different conditions of each project.

6.3. Python Model 3 Results

6.3.1. Training Phase

Similar to Model 1, 80% of the project data was used for training and 20% for testing. The accuracy of the training and testing processes was evaluated using same two parameters: mean absolute error (MAE) and mean squared error (MSE). The model experienced a training process that involved 100 iterations, during which it showed a constant improvement in its accuracy. The MSE decreased from 0.0305 during the first epoch to 0.00070 in the last one. Regarding the MAE, it decreased from 14.39% to 2.12%. Overall, the results suggest the model has learned from the training data and may perform well on unseen data.

6.3.2. Testing Phase

For the testing data set, the MSE was 0.000846 and the MAE was 2.31%. The MSE and MAE of the testing set are slightly higher than the training MSE and MAE, but still within a reasonable range, indicating that this difference will not have a significant effect on the prediction accuracy for new, unseen data. This will be further assessed using the validation data set.

6.3.3. Validation Phase

For the validation phase, the data set in Table 51 was used as input.

Table 51: Model 3 Validation Set

Project No.	Project Type	Project Location	Project Duration (months)	Client Type	Contract Type	Avg. Interest Rate	Avg. Inflation Rate	Avg. USD to EGP Exchange Rate	Total Direct Cost (EGP)	Site Overheads Percentage
1	Marine	outside the city/rural	27	Public Sector (Ministry/Government)	Remeasured	12.5%	21.1%	17.75	97,900,000.00	12.69%
2	Administrative	inside the city/urban	30	Private Sector (International)	Lump Sum	8.0%	11.2%	18.65	102,280,753.80	26.10%
3	Residential	inside the city/urban	45	Private Sector (International)	Lump Sum	8.5%	13.2%	19.55	569,683,416.83	28.54%
4	Commercial	outside the city/rural	40	Private Sector (National)	Remeasured	8.9%	16.1%	21.56	423,941,649.60	24.54%
5	Infrastructure	outside the city/rural	14	Public Sector (Ministry/Government)	Lump Sum	12.2%	35.1%	30.60	1,201,857,757.21	8.74%

The results for each phase are shown in Tables 52 to 55.

1. Initiation Phase Percentage

Table 52: Model 3 Validation Results (Initiation Phase)

Project No.	Model's Predicted Value	Actual Value	Absolute Percentage Error
1	5.00%	5.52%	9.42%
2	11.92%	13.00%	8.31%
3	11.05%	12.23%	9.65%
4	10.19%	10.00%	1.90%
5	13.60%	12.45%	9.24%

For the initiation phase, the mean absolute percentage error of the validation set is 7.70%.

Figure 75 shows a graph of the predicted versus actual values for the initiation phase percentage.

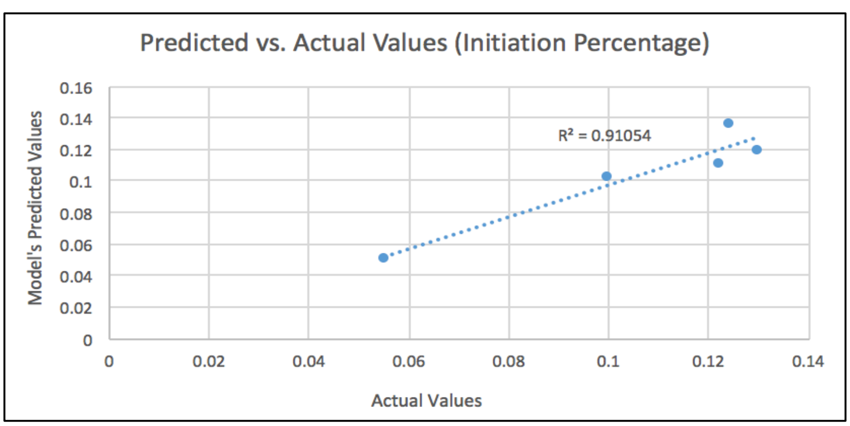

Figure 75: Predicted vs. Actual Values (Initiation Percentage)

2. Growth Phase Percentage

Table 53: Model 3 Validation Results (Growth Phase)

Project No.	Model's Predicted Value	Actual Value	Absolute Percentage Error
1	84.32%	82.21%	2.57%
2	63.14%	60.00%	5.23%
3	64.14%	66.04%	2.88%
4	69.68%	70.00%	0.46%
5	66.09%	68.00%	2.81%

For the growth phase, the mean absolute percentage error of the validation set is 2.79%.

Figure 76 shows a graph of the predicted versus actual values for the growth phase percentage.

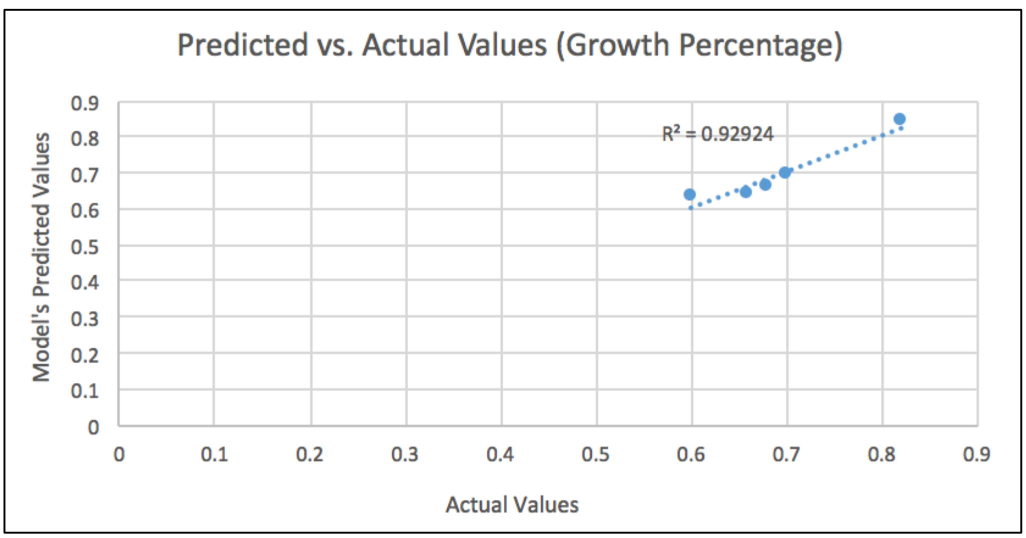

Figure 76: Predicted vs. Actual Values (Growth Percentage)

3. Maturity Phase Percentage

Table 54: Model 3 Validation Results (Maturity Phase)

Project No.	Model's Predicted Value	Actual Value	Absolute Percentage Error
1	8.64%	9.54%	9.43%
2	16.92%	18.44%	8.24%
3	17.75%	16.01%	10.87%
4	13.65%	12.35%	10.53%
5	13.73%	14.40%	4.65%

For the maturity phase, the mean absolute percentage error of the validation set is 8.74%.

Figure 77 shows a graph of the predicted versus actual values for the maturity phase percentage.

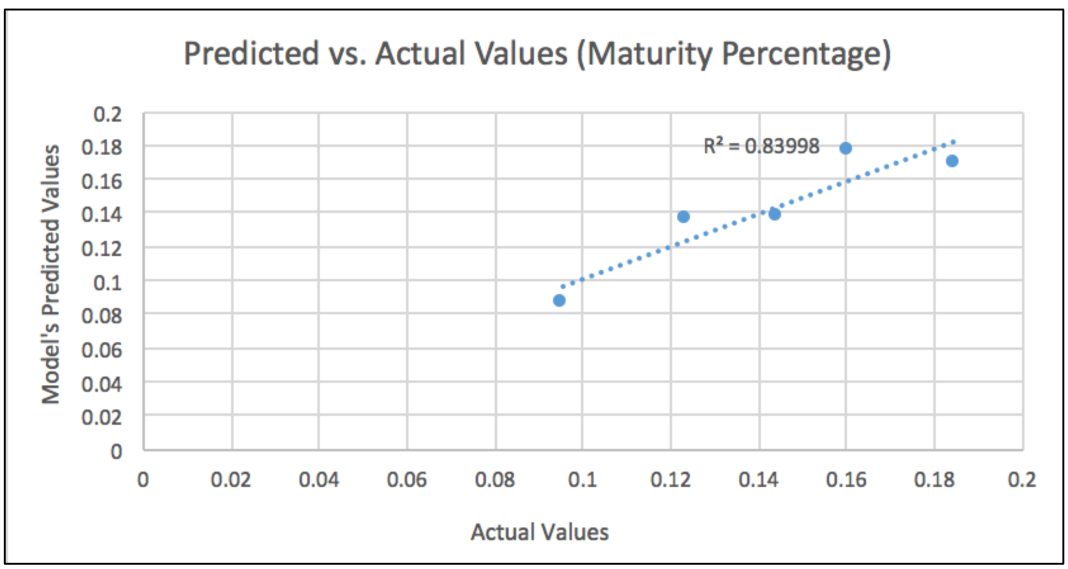

Figure 77: Predicted vs. Actual Values (Maturity Percentage)

4. Decline Phase Percentage

Table 55: Model 3 Validation Results (Decline Phase)

Project No.	Model's Predicted Value	Actual Value	Absolute Percentage Error
1	2.05%	2.72%	24.63%
2	8.02%	8.00%	0.25%
3	7.06%	6.55%	7.79%
4	6.47%	7.11%	9.00%
5	6.57%	6.00%	9.50%

For the decline phase, the mean absolute percentage error of the validation set is 10.23%.

Figure 78 shows a graph of the predicted versus actual values for the decline phase percentage.

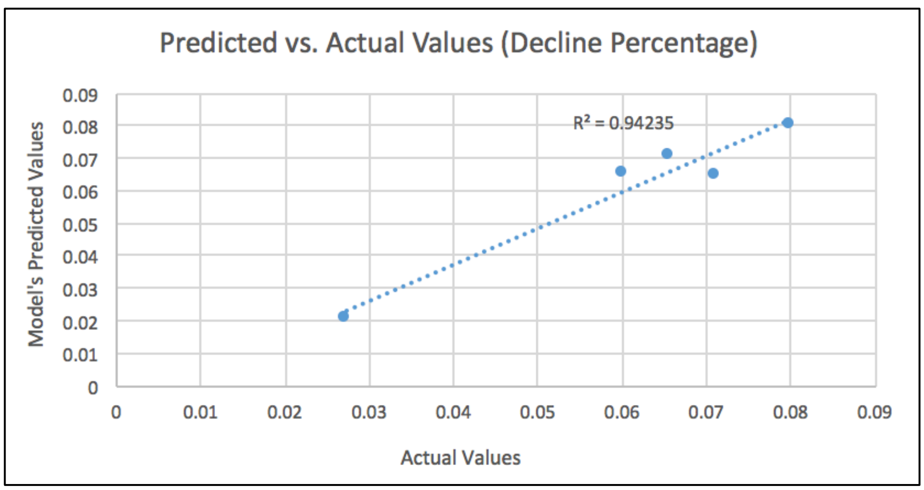

Figure 78: Predicted vs. Actual Values (Decline Percentage)

As noted in Tables 52 to 55 and Figures 75 to 78, the model overall performs well with absolute percentage errors being relatively low for most projects in all phases (often below 10%). In addition, the MAPE in initiation, growth, maturity and decline phases is 7.70%, 2.79%, 8.74% and 10.23% respectively. This further indicates that model performs well in all phases, specifically in the growth phases which has the lowest MAPE and which usually accounts for more than 50% of the total site overheads in any project. Thus, this model can be considered a reliable source for predicting the percentages allocated to each of the 4 phases and assisting in plotting an s-curve for the site overheads of construction projects.

7. Conclusion and Recommendations

7.1. Conclusion

In conclusion, the main objective of this research was to enhance the contractor's ability to accurately predict the percentage of site overheads in construction projects in Egypt through the identifying and analyzing the key factors influencing site overheads in the Egyptian construction industry, developing a robust dataset containing historical cost data of projects executed in the past 10 years and lastly developing ANN models capable of accurately predicting the total percentage of site overheads for construction projects in Egypt and subsequently allocating these costs across different site overheads subcategories and across the different project construction phases, incorporating both economic and non-economic variables to enhance the predictive accuracy

An extensive literature review was done to identify the major factors affecting indirect cost in Egypt and worldwide in order to be later used as inputs for the ANN model. These factors included:

- project type
- project location
- project duration
- contract type
- project direct cost
- client type
- class of contracting company
- macroeconomic indicators (inflation rate, interest rate and currency exchange rate).

A data collection form was prepared to be sent to experts in the field in order to collect data related to the previously identified factors from real-life projects executed and completed in the past 10 years. Cost data for 55 projects constructed in Egypt executed during the period from 2018 to 2024 was collected. This data mainly served as a database for the neural network's learning process. This data was then used as a database for the learning process of the ANN model. In addition, data related to Egypt's macroeconomic indicators was collected from the website of the Central Bank of Egypt (CBE).

The gathered project data was then analyzed and examined in order to determine how each factor affected the overall site overhead percentage. A comparative analysis of each factor and how it affects the site overhead percentage was done along with Pearson and Spearman Correlation tests which mainly conveyed that the factor that has the highest impact on site overheads percentage is project duration, followed by total direct cost.

Three models were created using Python on Google Colab.

- The first ANN model estimates the total site overhead percentage based on project characteristics.
- The second ANN model then utilizes both the predicted total overhead and existing project data to forecast the breakdown of site overheads across its different subcategories.
- Third model forecasts the breakdown of the site overheads across the different construction phases.

Each of the developed models had different architectures, and they were as follows:

- Model 1 architecture consists of 9 input neurons and 1 output layer representing the percentage of site overhead from the total direct cost of a construction project. The model has 2 hidden layers. The first hidden layer contains 64 neurons and ReLU (Rectified Linear Unit) activation function. The second hidden layer contains 32 neurons and ReLU activation function.
- Model 2 architecture consists of 10 input neurons and 1 output layer contains 10 neurons, each neuron representing one of the site overheads subcategories. The model has 2 hidden layers. The first hidden layer contains 128 neurons and ReLU (Rectified Linear Unit) activation function. The second hidden layer contains 64 neurons and ReLU activation function.
- Model 3 architecture consists 10 input neurons and 1 output layer containing 4 neurons, each neuron representing one of the project's construction phases. The model has 2 hidden layers. The first hidden layer contains 128 neurons and ReLU (Rectified Linear Unit) activation function. The second hidden layer contains 64 neurons and ReLU activation function.

Each model was then evaluated using mean absolute error (MAE) of its predictions.

- Model 1 had 2.75% MAE for training set and 3.90% MAE for testing set.
- Model 2 has 2.62% MAE for training and 2.83% MAE for testing set.
- Model 3 had 2.12% MAE for training and 2.31% MAE for testing data set.

The models were also evaluated using a dataset that it has not seen before to assess its ability to predict using new data.

- For Model 1, the mean absolute percentage error of the validation set was 4.86%.
- For Model 2, it appears to perform well for predicting the percentages of site overheads allocated to the different subcategories, with absolute percentage errors typically below 10% for salaries and wages and site equipment. In fact, the MAPE for these 2 categories was 3.46% and 6.58% respectively, indicating an acceptable accuracy for the two major subcategories that account for most of the site overheads percentage. For the other categories, the model showed higher absolute percentage errors, often exceeding 15%. The larger discrepancies observed in these subcategories might be partially attributed to the inherently small percentages associated with these categories and the fact that these categories tend to be present in some projects and absent in others, depending on the different conditions of each project.
- For Model 3, it performs well with absolute percentage errors being relatively low for most projects in all phases (often below 10%). The MAPE in initiation, growth, maturity and decline phases was 7.70%, 2.79%, 8.74% and 10.23% respectively.

Overall, the models performed well and can be considered a useful tool for the predicting the percentage of site overheads.

7.2. Research Limitations

1. The study's primary limitation lies in the reliance on expert-provided data, which may introduce subjectivity.
2. The focus on the Egyptian construction industry limits the generalizability of findings to other regions.
3. The model's accuracy might be influenced by the availability and quality of data, as well as the complexity of the construction projects included in the dataset.
4. This research dealt with category A companies only and thus cannot be generalized to projects relating to other categories.
5. The dataset contained projects completed between 2018 and 2024, potentially hindering the model's ability to capture long-term economic patterns and trends.

7.3. Recommendations for Future Work

1. Developing a user-friendly interface (web-based application or software tool) that allows users to input project data and receive predictions without requiring technical expertise.
2. Exploring the incorporation of additional factors such as project complexity and schedule constraints (including fast-tracking) into the ANN models.
3. Expanding the dataset to include projects from different regions to enhance the model's generalizability.
4. Expanding the dataset to include more projects in order to improve the accuracy of the outputs.
5. Investigating the integration of other machine learning techniques is recommended to assess if there are other techniques that lead to higher accuracy.
6. Enhancing the model's ability to detect economic patterns over time, it recommended to extend the study period and develop a dataset encompassing projects from older time periods, ideally spanning several economic cycles. This would enable the model to identify and learn from historical economic fluctuations and trends.

By addressing these limitations and exploring the suggested avenues for future research, the understanding of site overheads in the construction industry can be further enhanced.

References

AACE (2013) "AACE recommended practice and standard cost engineering terminology", no 10s-90, AACE, inc.

AACE International Recommended Practice (2017), Cost Engineering Terminology, TCM Framework: No. 10S-90, the Association of the Advancement of Cost Engineering International (AACE International), available at: https://web.aacei.org/resources/publications.

Aneja, N. (2011), "Neural networks approach v/s Algorithmic approach: a study through pattern recognition", Advanced Computing: An International Journal, Vol. 2 No. 6, pp. 183-187.

Al-Tawal, Dareen & Arafah, Mazen & Sweis, Ghaleb. (2020). A model utilizing the artificial neural network in cost estimation of construction projects in Jordan. Engineering Construction & Architectural Management. 10.1108/ECAM-06-2020-0402.

Arafa, M and Alqedra, M. (2011). Early Stage Cost Estimation of Buildings Construction Projects Using ANN. Journal of Artificial Intelligence, vol. 4, no. 1, pp. 63-75.

Awad, Rayya Muhammad Ahmad. (2017). *Artificial neural network (ANN) for estimating of overhead cost for school construction projects Gaza Strip.* (Master's theses Theses and Dissertations Master). Islamic University, Palestine (Gaza Strip) https://search.emarefa.net/detail/BIM-904767

Avinash R, Swamy R. (2018). Factors affecting cost and inflation of a project. Int Res J Eng Technol. 5(2):1713.

Bakr, A., Ragab, A., Yehia, N. (2018). Estimating of Site Overheads for Residential Projects in Egypt Using Neural Network. Proceedings of the 2nd International Conference of Sustainable Construction and Project Management.

Chan, C. T. (2012). The principal factors affecting construction project overhead expenses: an exploratory factor analysis approach. Construction Management and Economics, 30(10), 903-914.

Estimation of Contractor's Project Overhead Rate as Research on Building Cost. Proceedings from International Conference on Building Education and Research. https://www.semanticscholar.org/paper/Contractors-Perceptions-of-Effects-of-Project-Costs-Ujene-Idoro/7c84cdcdcbc3337a8f1e10269fe7eea63fb1cb5d.

Cheng, M., Tsai, H., & Sudjono, E. (2010). Conceptual Cost Estimates Using Evolutionary Fuzzy Hybrid Neural Network for Projects in Construction Industry. Expert Syst. Appl., 37, 4224-4231.

Dastres, R. & Soori, M. (2021). Artificial Neural Network Systems. International Journal of Imaging and Robotics. 21. 13-25.

Dikmen, S and Sonmez, M. (2011). An Artificial Neural Networks Model for The Estimation of Formwork Labour. J. Civ. Eng. Manag. 17 (3): 340–347.

Dunford, R., Su, Q., & Tamang, E. (2021). The Pareto Principle. The Race. https://www.semanticscholar.org/paper/The-Pareto-Principle-Dunford-Su/ef82dfe7b0ef7a88727636f5ad680a464e33e345.

El-Sawalhi, N.I. and Shehatto, O. (2014). A neural network model for building construction projects cost estimating. Journal of Construction Engineering and Project Management, Vol. 4 No. 4, pp. 9-16.

ElSawy, Ismaail & Hosny, Hossam & Razek, Mohammed. (2010). Factors Affecting Site Overhead Cost for Building Construction Projects. Journal of Al Azhar University Engineering Section, Vol. 5, No. 16.

ElSawy, Ismaail & Hosny, Hossam & Razek, Mohammed. (2011). A Neural Network Model for Construction Projects Site Overhead Cost Estimating in Egypt. International Journal of Computer Science Issues. 8.

Enshassi, A., & Aziz, A. R. A. (2008). Investigating the overhead costs in construction projects in Palestine. Journal of Financial Management of Property and Construction, 13(1), 35-47.

Enshassi A, Mohamed S, Abdel-Hadi M. (2013). Factors affecting the accuracy of pre-tender cost estimates in the Gaza strip. J Constr Dev Countries. 24(1):2–17.

Fan, R. Y. C., Ng, S. T., & Wong, J. M. W. (2010). Reliability of the Box–Jenkins model for forecasting construction demand covering times of economic austerity. Construction Management and Economics, 28(3), 241–254.

Forbes, T., & Riso, T. (2024b, June 20). Guide to S-curve modeling in construction. Procore. https://www.procore.com/library/s-curve-modeling-construction

Georgy, M. and Barsoum, S. (2005). Artificial Neural Networks Model for Parametric Estimating of Construction Project Costs. Journal of Engineering and Applied Science, Vol. 52, No. 6, pp. 1050-1066.

Google Colaboratory. (2024). https://colab.research.google.com/

Gunaydın, M. and Dogan, Z. (2004). A Neural Network Approach for Early Cost Estimation of Structural Systems of Buildings. International Journal of Project Management, vol. 22, pp. 595–602, 2004.

Hassouna, D. H., Khedr, A. E., Idrees, A. M., & ElSeddawy, A. I. (2020). Intelligent Personalized System for Enhancing the Quality of Learning. Journal of Theoretical and Applied Information Technology, 98(13), 2199–2213.

Hatamleh M, Hiyassat M, Sweis G, Sweis R. (2018). Factors affecting the accuracy of cost estimate: case of Jordan. Eng Constr Architect Manage. 25(1):113–653.

Heaton, J. (2017). The number of hidden layers. Heaton Research, 1. http://www.heatonresearch.com/2017/06/ 01/hidden-layers.html.

Hegazy, T. and Amr, A. (1998). Neural network model for parametric cost estimation of highway projects, J. Constr. Eng. Manag. 124 (3): 210–218.

Hesami, S., & Lavasani, S.A. (2014). Identifying and Classifying Effective Factors Affecting Overhead Costs in Constructing Projects in Iran. *International Journal of Construction Engineering and Management, 3*, 24-41.

Idrees, A. M., Alsheref, F. K. and ElSeddawy, A. I. (2019). A Proposed Model for Detecting Facebook News' Credibility. [IJACSA]. International Journal of Advanced Computer Science and Applications, 10(7), 311–316. doi:10.14569/IJACSA.2019.0100743.

Idrees, A. M., El Seddawy, A. I. and Moaaz, E. L. (2019). A Proposed Mining Based Business Continuity Information System for Educational Institutes. *Journal of Computational Science*, *15*(8), 1133–1149. doi:10.3844/ jcssp.2019.1133.1149

Idrees, A. M., ElSeddawy, A. I. and Zeidan, M. O. (2019). Knowledge Discovery based Framework for Enhancing the House of Quality. [IJACSA]. International Journal of Advanced Computer Science and Applications, 10(7), 324–331. doi:10.14569/IJACSA.2019.0100745

Israel, G.D. (1992) Determining Sample Size. University of Florida Cooperative Extension Service, Institute of Food and Agriculture Sciences, EDIS, Florida. https://www.psycholosphere.com/Determining%20sample%20size%20by%20Glen%20Israel.pdf.

Jawahar, S., Kukunuri, G., Devaraju, S., Gokuldev, S., Jayaprakash, S., Anandaram, H., Manivasagan, C. and Thenmozhi, M.. (2023). An Exploration of Python Libraries in Machine Learning Models for Data Science. 10.4018/978-1-6684-8696-2.ch001.

Idrees, A. M., Hassan, A. H., & Elseddawy, A. I. (2023). Neural Network-Based Prediction Model for Sites' Overhead in Commercial Projects. International Journal of e-Collaboration (IJeC), 19(1), 1-24. http://doi.org/10.4018/IJeC.318143

Khedr, A. E., Idrees, A. M., & El Seddawy, A. I. (2016). Enhancing Iterative Dichotomiser 3 algorithm for classification decision tree. WIREs Data Mining Knowledge Discovery, 6(2), 70–79. doi:10.1002/widm.1177

Kingma, D. P. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Kshirsagar, P., & Rathod, N. (2012). Artificial neural network. *International Journal of*

Computer Applications.

Kulkarni, P. S., Londhe, S. N. & Deo, M. C., 2017. Artificial Neural Networks for Construction Management. *Journal of Soft Computing in Civil Engineering,* 1(2), pp. 70-88.

Leśniak, A., Juszczyk, M. (2018). Prediction of site overhead costs with the use of artificial neural network based model. 2018. Archiv.Civ.Mech.Eng 18, 973–982 (2018). https://doi.org/10.1016/j.acme.2018.01.014

Leśniak, A., Juszczyk, M. (2019). Modelling Construction Site Cost Index Based on Neural Network Ensembles. Symmetry, 11(3), 411. MDPI AG. Retrieved from http://dx.doi.org/10.3390/sym11030411

Lorman. (2014). Current Issues to watch for in construction claims, part III: Overhead. https://www.lorman.com/resources/current-issues-to-watch-for-in- construction-claims-part-iii-overhead-claims-15877.

Luu, T. V. & Kim, S. Y. (2009). Neural Network Model for Construction Cost Prediction of Apartment Projects in Vietnam. Korean Journal of Construction Engineering and Management, 10(3), pp. 139-147.

Mackenzie, M.D. and Briggs, R. (2006). Modelling and Simulation - A Profitable Tool for all Phases of the Lifecycle Engineering Asset Management. Springer London, 691-699.

Matel, E., Vahdatikhaki, F., Hosseinyalamdary, S., Evers, T., & Voordijk, H. (2019). An artificial neural network approach for cost estimation of engineering services. International Journal of Construction Management, 22(7), 1274-1287. https://doi.org/10.1080/15623599.2019.1692400.

Mordor Intelligence. (2022). Egypt Construction Market - Growth, Trends, COVID-19 Impact, and Forecasts (2023 - 2028). https://www.mordorintelligence.com/industry-reports/egypt-construction-market.

Musarat, M., Alaloul, W. and Liew, M.S.. (2020). Impact of inflation rate on construction projects budget: A review. Ain Shams Engineering Journal. 12. 10.1016/j.asej.2020.04.009.

Naik, P., Naik, G. and Patil, M.B. (2022). Conceptualizing Python in Google COLAB.

Nielsen, M. (2015) Neural Networks and Deep Learning. Determination Press. https://jingyuexing.github.io/Ebook/Machine_Learning/Neural%20Networks%20and%20Deep%20Learning-eng.pdf.

Othman, S. (2020). Evaluation of Indirect Cost Estimation in the Egyptian Construction Industry. [Master's Thesis, Joint Study Program of Metropolia University of Applied Sciences and HTW Berlin]. Available at: https://urn.fi/URN:NBN:fi:amk-2021090617468.

Patil, S. S. & Bhangale, P. P., 2014. Overhead Cost In Construction Industry. *International Journal of IndustrialEngineering & Technology (IJIET),* 4(2), pp. 1-6.

Potdar, K. & Pardawala, T & Pai, C. (2017). A Comparative Study of Categorical Variable Encoding Techniques for Neural Network Classifiers. International Journal of Computer Applications. 175. 7-9. 10.5120/ijca2017915495.
Project Management Institute (PMBOK). (2013). A guide to the Project Management Body of Knowledge (PMBOK guide) (5th ed.). Project Management Institute.

Puci, Jona & Demi, Albana & Kadiu, Arjana. (2023). Impact of macroeconomic variables on the construction sector. Corporate and Business Strategy Review. 4. 22-30. 10.22495/cbsrv4i1art2.

Rezaian, A. (2011), "Time-cost-quality-risk of construction and development projects or investment", Middle-East Journal of Scientific Research, Vol. 10 No. 2, pp. 218-223.

Roxas, L. & Maximino, J. (2014). An Artificial Neural Network Approach to Structural Cost Estimation of Building Projects in the Philippines. DLSU Research Congress. Manila.

Sharma, S. & Athaiya, A.. (2020). Activation Function In Neural Networks. International Journal of Engineering Applied Sciences and Technology. 04. 310-316.
https://doi.org/10.33564/IJEAST.2020.v04i12.054.

Shiha, A. (2019). *Prediction of construction material prices using macroeconomic indicators: A neural networks model* [Master's Thesis, the American University in Cairo]. AUC Knowledge Fountain. https://fount.aucegypt.edu/etds/1569

Srinivasan, Kathiravan & Cherukuri, Aswani Kumar & Vincent P M, Durai & Garg, Ashish & Chen, Bor-Yann. (2019). An Efficient Implementation of Artificial Neural Networks with K-fold Cross-validation for Process Optimization. Journal of Internet Technology. 20. 1213-1225. 10.3966/160792642019072004020.

Stolz, J. M., 2010. Overhead and Uncertainty in Cost Estimates: A Guide to Their Review, San Francisco, CA: Jacobs Associates.

Warsame, A. (2006). Construction cost-central concepts, categories and determining factor. https://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-47669.

Yadav, R., Vyas, V., Vyas, M. & Agrawal, S., 2016. Cost Estimation Model (Cem) for Residential Building using Artificial Neural Network. International Journal of Engineering Research & Technology (IJERT), 5(1), pp. 430-432.

Zheng, W. X., Chen, D. X. & Yan, L. J., 2010. Application of Neural Network in the Cost Estimation of Highway Engineering. Journal of Computers, 5(11), pp. 1762-1766.

Appendix

Full Code (Model 1):

```
# Import libraries
import pandas as pd
import numpy as np
from tensorflow import keras
from sklearn.preprocessing import MinMaxScaler
from sklearn.model_selection import train_test_split
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense
from tensorflow.keras.layers import Activation

# Loading data

data = pd.read_csv("/content/ANN Model's Database.csv")

# Defining features (inputs) and target (output)

features = [
    "Project Type",
    "Project Location",
    "Project Duration",
    "Client Type",
    "Contract Type",
    "Avg Interest Rate",
    "Avg Inflation Rate",
    "Avg USD to EGP Exchange Rate",
    "Total Direct Cost",
]
```

```
target = 'Site Overheads Percentage'

# Data Preprocessing

# Identify categorical features (assuming these are already coded)
categorical_features = ['Project Type', 'Project Location', 'Client Type', 'Contract Type']

# Numerical Data Normalization

scaler = MinMaxScaler(feature_range=(0, 1))
numerical_features = ['Project Duration', 'Avg Interest Rate',
                      'Avg Inflation Rate', 'Avg USD to EGP Exchange Rate',
                      'Total Direct Cost']
data[numerical_features] = scaler.fit_transform(data[numerical_features])


#Splitting the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(data[features], data[target], test_size=0.2, random_state=0)

# Defining the neural network model
model = Sequential()
model.add(Dense(64, activation='relu', input_shape=(data[features].shape[1],))) # First hidden layer with 64 neurons and ReLU activation
model.add(Dense(32, activation='relu'))  # Second hidden layer with 32 neurons and ReLU activation
model.add(Dense(1, activation='sigmoid'))  # Output layer with 1 neuron and sigmoid activation (for percentage)

# Compiling the model
model.compile(loss='mse', optimizer='adam', metrics=['mae'])  # Mean squared error loss, mean absolute error metric
```

```
#Fitting the ANN to the training set
model.fit(X_train, y_train, batch_size=32, epochs=100)

# Evaluating the model
loss, mae = model.evaluate(data[features], data[target])
print(f"Loss: {loss}, Mean Absolute Error: {mae}")
print("Mean Squared Error", loss)

# Model Validation using New Project Data

# Loading new project data from a separate CSV file
new_data = pd.read_csv("/content/Validation Data File.csv")

# Preprocessing new data
new_data[numerical_features] = scaler.transform(new_data[numerical_features])

# Predicting site overheads percentage for new projects
new_predictions = model.predict(new_data[features])

# Print the predicted site overhead percentages
print("Predicted Site Overhead Percentage for New Projects:")
print(new_predictions)
```

Figure 79: Full Model 1 Python Code

Full Code (Model 2):

```python
 # Import libraries
import pandas as pd
import numpy as np
from tensorflow import keras
from sklearn.preprocessing import MinMaxScaler
from sklearn.model_selection import train_test_split
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense
from tensorflow.keras.layers import Activation

# Loading data

data = pd.read_csv("/content/ANN Model 2 Database.csv")

# Defining features (inputs) and target (output)

features = [
    "Project Type",
    "Project Location",
    "Project Duration",
    "Client Type",
    "Contract Type",
    "Avg Interest Rate",
    "Avg Inflation Rate",
    "Avg USD to EGP Exchange Rate",
    "Total Direct Cost",
    "Site Overheads Percentage",
```

```python
target = [
    "Salaries & Wages %",
    "Site Facilities %",
    "Accommodation %",
    "Mobilization & Demobilization %",
    "Communication & IT %",
    "Site Equipment %",
    "Personnel & Material Transportation %",
    "Quality and Safety %",
    "Engineering %",
    "Client/Consultant %",
]

# Data Preprocessing

# Identify categorical features (assuming these are already coded)
categorical_features = ['Project Type', 'Project Location', 'Client Type', 'Contract Type']

# Numerical Data Normalization

scaler = MinMaxScaler(feature_range=(0, 1))
numerical_features = ['Project Duration', 'Avg Interest Rate',
                      'Avg Inflation Rate', 'Avg USD to EGP Exchange Rate',
                      'Total Direct Cost','Site Overheads Percentage']
data[numerical_features] = scaler.fit_transform(data[numerical_features])
```

```python
#Splitting the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(data[features], data[target], test_size=0.2, random_sta

# Defining the neural network model
model = Sequential()
model.add(Dense(128, activation='relu', input_shape=(data[features].shape[1],))) # First hidden layer with 128 neurons and ReLU activatic
model.add(Dense(64, activation='relu'))  # Second hidden layer with 64 neurons and ReLU activation
model.add(Dense(len(target), activation='softmax'))  # Output layer with 10 neurons and softmax activation

# Compiling the model
model.compile(loss='mse', optimizer='adam', metrics=['mae'])  # Mean squared error loss, mean absolute error metric

#Fitting the ANN to the training set
model.fit(X_train, y_train, batch_size=32, epochs=100)

# Evaluating the model
loss, mae = model.evaluate(data[features], data[target])
print(f"Loss: {loss}, Mean Absolute Error: {mae}")
print("Mean Squared Error", loss)

# Model Validation using New Project Data

# Loading new project data from a separate CSV file
new_data = pd.read_csv("/content//Model 2 Validation Data File.csv")

# Preprocessing new data
new_data[numerical_features] = scaler.transform(new_data[numerical_features])
```

```python
# Predicting the percentage of each category
new_predictions = model.predict(new_data[features])

# Print the predicted percentage of each category
print("Predicted Percentage of Each Category:")
print(new_predictions)
```

Figure 80: Full Model 2 Python Code

Full Code (Model 3):

```python
 # Import libraries
import pandas as pd
import numpy as np
from tensorflow import keras
from sklearn.preprocessing import MinMaxScaler
from sklearn.model_selection import train_test_split
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense
from tensorflow.keras.layers import Activation

# Loading data

data = pd.read_csv("/content/ANN Model 3 Database.csv")

# Defining features (inputs) and target (output)

features = [
    "Project Type",
    "Project Location",
    "Project Duration",
    "Client Type",
    "Contract Type",
    "Avg Interest Rate",
    "Avg Inflation Rate",
    "Avg USD to EGP Exchange Rate",
    "Total Direct Cost",
    "Site Overheads Percentage",
```

```python
target = [
    "Initiation Phase %",
    "Growth Phase %",
    "Maturity Phase %",
    "Decline Phase %",
]

# Data Preprocessing

# Identify categorical features (assuming these are already coded)
categorical_features = ['Project Type', 'Project Location', 'Client Type', 'Contract Type']

# Numerical Data Normalization

scaler = MinMaxScaler(feature_range=(0, 1))
numerical_features = ['Project Duration', 'Avg Interest Rate',
                      'Avg Inflation Rate', 'Avg USD to EGP Exchange Rate',
                      'Total Direct Cost','Site Overheads Percentage']
data[numerical_features] = scaler.fit_transform(data[numerical_features])
```

```python
#Splitting the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(data[features], data[target], test_size=0.2, random_state=0)

# Defining the neural network model
model = Sequential()
model.add(Dense(128, activation='relu', input_shape=(data[features].shape[1],))) # First hidden layer with 128 neurons and ReLU activatic
model.add(Dense(64, activation='relu'))  # Second hidden layer with 64 neurons and ReLU activation
model.add(Dense(len(target), activation='softmax'))  # Output layer with 10 neurons and softmax activation

# Compiling the model
model.compile(loss='mse', optimizer='adam', metrics=['mae'])  # Mean squared error loss, mean absolute error metric

#Fitting the ANN to the training set
model.fit(X_train, y_train, batch_size=32, epochs=100)

# Evaluating the model
loss, mae = model.evaluate(data[features], data[target])
print(f"Loss: {loss}, Mean Absolute Error: {mae}")
print("Mean Squared Error", loss)

# Model Validation using New Project Data

# Loading new project data from a separate CSV file
new_data = pd.read_csv("/content//Model 3 Validation Data File.csv")
```

```python
# Preprocessing new data
new_data[numerical_features] = scaler.transform(new_data[numerical_features])

# Predicting the percentage of each category
new_predictions = model.predict(new_data[features])

# Print the predicted percentage of each category
print("Predicted Percentage of Each Category:")
print(new_predictions)
```

Figure 81: Full Model 3 Python Code