ADDIS ABABA UNIVERSITY

ADDIS ABABA INSTITUTE OF TECHNOLOGY

SCHOOL OF CIVIL AND ENVIRONMENTAL ENGINEERING

Development of an Approximate Construction Time Prediction Models During the Project Planning Phase. A Case Study of Selected Private Mixed Use and Public Office Building Projects in Addis Ababa.

By; Miliyon Fikre

Advisor: Abebe Dinku, Prof (Dr.-Ing.)

Co-Advisor: Selam Yazew (MSc)

A Thesis submitted to the School of Graduate Studies of the Addis Ababa University in partial fulfillment of the requirements for the degree of Master of Science in Civil Engineering(Stream: Construction Technology and Management)

September ,2021

Addis Ababa, Ethiopia

The undersigned have examined the thesis entitled '**Development of an Approximate Construction Time Prediction Models During the Project Planning Phase. A Case Study of Selected Private Mixed Use and Public Office Building Projects in Addis Ababa**' presented by **Miliyon Fikre**, a candidate for the degree of **Master of Science** and hereby certify that it is worthy of acceptance.

1. Abebe Dinku, Prof (Dr.-Ing.)__ ____________________ ____________
   Advisor Signature Date

2. Dr. Abraham Assefa_________ ______________________ _____________
   Internal Examiner Signature Date

3. Dr. Asregidew Kassa__________ ______________________ ____________
   External Examiner Signature Date

4. ___________________________ ______________________ ____________
   Chairperson Signature Date

DECLARATION

I confirm that research work titled "**Development of an Approximate Construction Time Prediction Models During the Project Planning Phase. A Case Study of Selected Private Mixed Use and Public Office Building Projects in Addis Ababa**" is my own work. The work has not been presented elsewhere. Where material has been used from other sources it has been properly acknowledged.

Miliyon Fikre (GSR/8988/11) ………………………. ……..……………

Signature Date

ABSTRACT

Time, Cost and quality are key performance indicators of construction projects. However, failure to rely on standard duration prediction model lead towards delay of construction projects, which is currently an observed situation in the construction sector. Therefore, the Key objective of this study is to examine and validate the Bromillow's Time-cost (BTC) model and the Love et al.'s Time-Floor (LTF) model to estimate contract durations for private mixed use and public office building projects in Addis Ababa. The study also propose an alternative duration prediction model by considering potentially influential project scope factors identified from literature by checking their appropriateness and comparing their prediction performances. The LTF model formulates the project duration in terms of gross floor area and floor numbers, while the BTC model formulates in terms of cost. Research data were collected from grade one consultants for private mixed-use buildings and from Addis Ababa City Administration Construction Bureau (AACACB) and Federal Government Buildings Construction Project Office (FGBCPO) for public office buildings. IBM SPSS statistics 26 and WEKA tool 3.8.5 were used to develop and validate the developed construction duration prediction models. Applying linear regression method, the study developed Bromillow's Time-cost (BTC), Love et al.'s Time-Floor (LTF) and proposed 'best-fit model' indicating that the gross floor area is the sole predictor of duration. The model validation result also shows that BTC is superior model over linear regression models. Further, a multilayer perceptron neural network (MLP - NN) predictive model to the same data was applied to develop construction duration prediction models. As a result, (MLP - NN) model result shows significant improvement of the accuracy of the construction time prediction over linear regression models with mean absolute percentage error of 22.2% in private mixed use and 16.3 % in public office building projects. The practical implications of this study can help stakeholders participating in the construction industry to get multiple benefits such as client satisfaction, efficient and effective use of resources, minimizing occurrence of claims and healthy relationship among stakeholders in the construction business sector.

Key words: *Bromillow Time-Cost, Building projects, Linear regression, Love Time Floor, Multilayer neural network, Time prediction model*

ACKNOWLEDGMENTS

My deepest appreciation goes to the Almighty God who has made me who I am today. He is truly the Alpha and Omega and this thesis is a testimony of this attribute of God.

I wish to express my profound gratitude to my supervisors; Abebe Dinku, Prof. (Dr.-Ing) and Selam Yazew (MSc) who did not only team up to provide the guidance and constructive criticism required for the work, but also encouraged me at every stage of this research. Moreover, I would like to express my sincere thanks to Dr. Abraham Assefa for his support by giving direction and heart bit of this research.

I would like to thank staff of grade one consultants, involved in responding to project data collection format. Their participations and honesty to show direction for me was wonderful. Those who gave their time and willingness for information, I would like to thank them once again. Further, I would like to thank all those who were participated directly or indirectly for the success of data collection.

Finally, I must express my very profound gratitude to my parents for providing me with unfailing support and continuous encouragement throughout my years of study. My friends your help and encouragement through the process of researching and writing this thesis was great. This accomplishment would not have been possible without them.

Miliyon Fikre

September, 2021

DEDICATION

This work is dedicated to the glory of God and for the benefit of mankind.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AC	Asphalt Concrete Surfaced Road Project
AACACB	Addis Ababa City Administration Construction Bureau
BaTCoDA	Building and Transport Construction and Design Authority
BTC	Bromillow Time-Cost model
CPM	Critical Path Method
CPM	Critical Path Method
DB	Design and Build
DBST	Double Surface Asphalt Treatment Road Project
EPRDF	Ethiopian People Revolutionary Democratic Front
ETB	Ethiopian Birr
ETBRC	Ethiopian building and road construction
FDRE	Federal Democratic Republic of Ethiopia
FGBCPO	Federal Government Buildings Construction Project Office
GDP	Growth Domestic Product
GTP	Growth and Transformation Plan
KPI	Key Performance Indicators
LTF	Love et al.'s Time-floor
MLP-NN	Multilayer Perceptron Neural Network
MoE	Ministry of Education
MoWUD	Ministry of Works and Urban Development
NEC	National Engineers and Contractors
PMBOK	Project Management Body of Knowledge
PMI	Project Management Institute
PTM	Project Time Management
TCT	Time-Cost Trade-off
TCT	Time-Cost Trade-off
TCTP	Time-Cost Trade-off Problem
TCTP	Time-Cost Trade-off Problem

WEKA Waikato Environment for Knowledge Analysis

UK United Kingdom

CHAPTER 1 INTRODUCTION

1.1 Back ground of the study

Construction industry plays an essential role in any developing country. This is mainly because developing countries are considerably dependent on the growth and development of their physical infrastructures and because the linkage of the construction industry to both economic and social sectors is very significant. A wide range of buildings and construction facilities are required by a modern society, including residential and commercial property, manufacturing facilities, schools, hospitals, complex transport infrastructure. Thus, the construction industry is a major sector of the economy of a country[1].

The construction industry in Ethiopia, as in most developing countries, has made a significant contribution to the growth of the economy through infrastructure development and job creation. Public construction projects are parts of the country's development initiative; whereby they share considerable amount of the country's scarce financial resources. In line with this, the building sector has seen a double digit growth, expanding by 37% annually, and is ushering in a new phase of development for the country [2]. However, this study focuses on Addis Ababa since the city is one of the diplomatic center of Africa and capital city of Ethiopia where majority of public office and private mixed use building construction projects are implemented to bring economic development in the country.

In Ethiopia in general and in Addis Ababa Administration in particular, the construction industry is the highest recipient of government budget in terms of government development programs. Consequently, public construction projects consume an average annual rate of nearly 60% of the government's capital budget as reported by Ministry of Works and Urban Development [3]. Moreover, the quick increase of population in Addis Ababa results extraordinary influence on the existing housing policy and on the infrastructure like electrical power, water, road and supply drainage. For instance, The Ethiopian Federal Democratic Government took considerable effort to enhance the housing conditions in Addis Ababa city by maintaining different housing plans like- 10/90, 20/80 and 40/60 schemes respectively. Currently, the same situation such as urbanization and proper usage of land leads the government and private developers to increase private mixed use and public office building projects to satisfy society needs.

A construction project is often acknowledged as effective and efficient when it is finished on schedule, within budget and within the defined quality in compliance to the written specifications and to stakeholders' satisfaction [4]. Functionality, profitability to contractors, non-existence of claims and court proceeding and "fitness for purpose" for occupiers have also been used as determinant of project success. In the construction projects, time and cost have a very close relationship and are key performance indicators of construction project. This relationship can be illustrated in a linear fashion, which means for the same type of project, the greater the volume of work, the greater the cost and time are required in order to finish the whole project [5]. Barrie and Paulson [6]also described that in construction, the prediction of a reasonably accurate time and cost relationship is necessary for a successful planning at all phases of the projects.

A prediction of the appropriate duration of construction at the beginning of a project is vital in ensuring its success eventually [7]. Construction time is usually estimated in the planning phase when an analysis of the feasibility of the project is made [8].Kim and Yeom [9] described construction time is used as a standard to assess the viability of a given business contract and execution of the project and serves as a basis to make necessary business decisions. However, the estimation of construction time in the planning phase is more challenging than when it is done in the design phase, due to the limited amount of information available to project planners and frequent changes that occur during the consultation between the client and designer [8].

To resolve the problem Bromillow's Time-cost (BTC) is the first empirical model developed to predict construction time in the early planning phase of the project using cost as a predictor [10]. Afterwards, Love et al.'s Time-floor (LTF) model checked poor predictive capacity of BTC and suggested that gross floor area and floor numbers as a predictors for both residential and office building projects [11].

In Ethiopia based on observations, the way to predict the span of time that is needed to complete a construction project had been a problem due to subjectivity in duration prediction. Thus, recognizing the problem and importance, many researchers have conducted research on developing time prediction models based on performance data. However, the aim of this study is to develop Bromillow's Time-cost (BTC), Love et al.'s Time-Floor (LTF) and an alternative model. Then, this study recommends best-fit model for private mixed use and public office building construction projects in Addis Ababa.

1.2 Statement of the problem

Several attempts have been made to predict construction duration as it represents a problem of continual concern and interest to both researchers and construction industry practitioners. In the construction industry, contractors usually use previous experiences to predict project duration and cost of a new project [12,13]. The competitive nature of the industry places pressure on contractors to keep project costs as low as possible. At the same time, project durations are also kept to a minimum [14,15,16]. Despite, many contractors simply take contract time set by the client as realistic and prepare their bids accordingly. However, time overrun is one of the common problems often experienced on construction project sites [17,18, 19]. According to Werku and Jha [20] most of government building construction projects in Ethiopia is subjected to time overrun, where only 8.25% projects were completed on planned finish dates, while 91.75% are delayed with 352% of its original contract time.

Delay in building projects increase contractors' costs thereby decrease the contractor's profit margin and reputation; they may also cause clients to incur additional holding charges, professional fees and income lost through late occupancy [21]. As a result, the Government of Ethiopia and other stakeholders for private mixed use and public office building construction projects are increasingly becoming uncomfortable at seeing projects being completed beyond its expected duration.

1.3 Objectives of the Study

1.3.1 General Objectives

The aim of this research is to develop scientific construction time estimation model for use in private mixed use and public office building projects in Addis Ababa.

1.3.2 Specific Objectives

This study specifically tries;

- ✓ Validate applicability of Bromillow's Time–Cost and Love et al.'s Time-Floor models for private mixed use and public office building construction projects in Addis Ababa.
- ✓ Develop an alternative duration prediction model by considering project scope factors identified in literature using linear regression and Artificial neural network.

1.4 Research Questions

I. Can Bromillow's Time –Cost (BTC) model be used to predict the contract duration for public office and private mixed use building construction projects?

II. Can Love et al.'s Time-floor (LTF) model be used to predict the contract duration for public office and private mixed use building projects?

III. Is there any another model than linear regression that can be used to predict the contract duration for private mixed use and public office building projects?

1.5 Hypotheses of the research

It is hypothesized that the total construction time for private mixed use and public office building projects in Addis Ababa is positively correlated with total construction cost of a project. It is also correlated with gross floor area and floor numbers. Hence, the research paper has two hypotheses;

- To verify the validity and applicability of Bromillow's Time-Cost model and Love et al.'s Time-Floor models for use in private mixed use and public office building construction projects in Addis Ababa.
- To verify existence of the proposed model other than Bromillow's Time-Cost and Love et al.'s Time-Floor models for private mixed use and public office building construction projects in Addis Ababa.

The corresponding null hypotheses that would be the subject of acceptance and rejection based on the study findings are:

H01 = "There is no significant relationship between construction time and construction cost" and also no significant relationship between construction time and both gross floor area and floor numbers.

H02 = "There is no model other than BTC and LTF to predict project duration."

1.6 Scope and Limitation of the Study

The scope of this study covered completed above G+3 and below G+20 floors of private mixed use and public office building construction projects in Addis Ababa. The project delivery method and contract type used for these projects were DBB and Admeasurement contract respectively. The data sets were collected only from (grade one consultants) for private mixed use buildings, Addis Ababa City Administration Construction Bureau (AACACB) and Federal

Government Buildings Construction Project Office (FGBCPO) for public office building projects.

It is evident from the literature review that the completion of construction projects is affected by numerous hierarchy of factors that can contribute to the duration of construction projects, some of which include location, productivity, type of contract and weather. However, this study limited to project scope factors to determine project duration for private mixed use and public office building projects in Addis Ababa. It does not incorporate the implications of these likely factors that can influence the total time required for the completion of private mixed use and public office building construction projects. The limitation of this study also includes inaccessibility to necessary data such as drawing, claim case of data sets, limited project scope factors and construction cost price index. In addition, there is a challenge to identify whether each project with in data sets were completed either in normal duration or optimum duration or crash duration. Before model development, data reduction was carried out using Filter methods of feature selection only using WEKA 3.8.5.

1.7 Motivation of the study

Construction industry has been implementing many projects to meet society needs in the future. Moreover, urbanization is increasing gradually which leads towards the need for expansion of private mixed use and public office buildings construction in our country. However, time overrun is one of the factors, which leads stakeholders in construction industry towards construction disputes and additional cost. Consequently, liquidated damages take construction contractors and clients out of market competition and makes the industry unattractive. Therefore, this problem inspired the researcher to develop construction time prediction model as a prerequisite to predict contract time duration for construction projects to minimize the time-overrun problem.

1.8 Significance of the study

This study is important for all professionals involved in the construction industry to estimate project duration during early planning phase as part of bid preparation without any detail activity scheduling. Further, this research is also important for academic and further study related construction time prediction model. In general, the study has practical relevance as to reduce subjectivity in duration prediction during planning phase of a project in the construction sector.

1.9 The research Contents

This research paper has five broad categories. Each category has been described below

Chapter I: Introduction/The research background

Chapter II: Literature Review

Chapter III: The Research Design and Methodology

Chapter IV: The research Data analysis and Discussions

Chapter V: The research Conclusions and Recommendations

Each of the above chapters has the following contents;

Chapter I: This chapter describes the research overview, background, hypothesis, Context, objectives and hypothesis. It also indicates the contents of the research.

Chapter II: It covers the basic concepts of the research and literature review part of the thesis. It discusses project management, time- cost trade off analysis, performance indicators, identifying factors influencing project duration and past developed models.

Chapter III: Covers the research methodology. The methodological approach consists of the research definition, process, Design validity, reliability, and writing of the research paper.

Chapter IV: It contains the data analysis and research discussion. It also contains time cost, cost-area, time-cost-area relationships, hypothesis testing, and model development and validation. And finally

Chapter V: In this part the research conclusion and recommendations are presented

CHAPTER 2 LITERATURE REVIEW

2.1 Basic Conceptual Literature

2.1.1 Project

A project is a temporary endeavor carried out to make a unique product, service, or result. It clearly describe more a temporary nature of projects as a project features a definite starting and finishing point [23]. A major setback in temporary organizations is that project workers know that their contribution is only needed for a limited period of time.

A project is designed up of a collection of interrelated work activities defined by a specific scope, budget, and schedule to deliver tangible asset required to achieve the strategic targets of an agency [24].

A project is a temporary venture that exists to provide a defined outcome and it will have agreed, and specific achievements as its project plan, budget, schedule, deliverables and tasks. It can also involve people from different professions within a firm who are brought together to fulfilling a defined goal [25]**.**

Despite most of the definitions given by different scholars are more or less similar, this research takes into account the definition given by PMBOK.

2.1.2 Project Management

Project management is the application of knowledge, skills, tools, and techniques to project activities to meet the project requirements. It is accomplished through the appropriate application and integration of project management processes at each project phases, which are categorized into five Process Groups discussed as follows [23]:

(1) **Initiating** is the processes conducted to design a new project or a new phase of an existing project by getting authorization to commence the project or phase.

(2) **Planning** is the processes needed to develop the work scope of the project refine the objectives and specify the series of work required to achieve the deliverables that the project was implemented to achieve.

(3) **Executing** is done to finish the task specified in the project management plan to attain the project requirements.

(4) **Monitoring and controlling** is necessary to track, evaluate, and orchestrate the actual and planned progress of the project; pick out any parts in which revisions to the plan are needed; and commencing the corresponding changes.

(5) **Closing** is done to formally finish or close the project, life cycle, or contract.

Managing a project typically covers, but is not restricted to:

- ✓ Identifying requirements;
- ✓ Defining the several needs, concerns, and expectations of the stakeholders in planning and implementing the project;
- ✓ Setting up, maintaining, and carrying out communications among actors in construction that are active, effective, and collaborative in nature;
- ✓ Managing actors in construction projects towards meeting project objectives and creating project deliverables;
- ✓ Balancing the competing project constraints, which involves, but are not restricted to; scope, quality, time, cost, resources, and risks

Project management is a necessary practice that implement knowledge of process, skills, tools, deliverables, and techniques to project activities to make sure a solid path to project achievement by meeting targets and requirements [26]. Furthermore, Project management is often defined because the discipline of applying specific processes and principles to initiate, plan, implement, and manage the way that new initiatives or changes are implemented within a firm, however, it is different to the management of the business as usual activity, which is a continual process because it involves creating new work packages to attain agreed ends or goals [25].

Besides the Process Groups, Knowledge Areas also categorize processes. A Knowledge Area is an identified area of project management defined by its knowledge requirements and described in terms of its component processes, practices, inputs, outputs, tools, and techniques [23].

The ten Knowledge Areas described in PMBOK [23] were described in this paper;

Project Integration Management; Includes the processes and activities to identify, define, combine, unify, and coordinate the various processes and project management activities within the Project Management Process Groups.

Project Scope Management; Includes the processes required to ensure the project includes all the work required, and only the work required, to complete the project successfully.

Project Schedule Management; Includes the processes required to manage the timely completion of the project.

Project Cost Management; Includes the processes involved in planning, estimating, budgeting, financing, funding, managing, and controlling costs so the project can be completed within the approved budget.

Project Quality Management; Includes the processes for incorporating the organization's quality policy regarding planning, managing, and controlling project and product quality requirements, in order to meet stakeholders' expectations.

Project Resource Management; Includes the processes to identify, acquire, and manage the resources needed for the successful completion of the project.

Project Communications Management; Includes the processes required to ensure timely and appropriate planning, collection, creation, distribution, storage, retrieval, management, control, monitoring, and ultimate disposition of project information.

Project Risk Management; Includes the processes of conducting risk management planning, identification, analysis, response planning, response implementation, and monitoring risk on a project.

Project Procurement Management; Includes the processes necessary to purchase or acquire products, services, or results needed from outside the project team.

Project Stakeholder Management; Includes the processes required to identify the people, groups, or organizations that could impact or be impacted by the project, to analyze stakeholder expectations and their impact on the project, and to develop appropriate management strategies for effectively engaging stakeholders in project decisions and execution.

Table 2.1 Shows forty-nine project management processes grouped under project management process group and applied with in knowledge areas of project management. Proper application of these project management processes are basis to achieve key performance indicators of the project and to satisfy stakeholders expectation.

Table 2. 1:Project Management Process Group and Knowledge Area Mapping

(PMBOK®Edition, Guide –Sixth).

Knowledge Areas	Project Management Process Groups				
	Initiating Process Group	Planning Process Group	Executing Process Group	Monitoring and Controlling Process Group	Closing Process Group
4. Project Integration Management	4.1 Develop Project Charter	4.2 Develop Project Management Plan	4.3 Direct and Manage Project Work 4.4 Manage Project Knowledge	4.5 Monitor and Control Project Work 4.6 Perform Integrated Change Control	4.7 Close Project or Phase
5. Project Scope Management		5.1 Plan Scope Management 5.2 Collect Requirements 5.3 Define Scope 5.4 Create WBS		5.5 Validate Scope 5.6 Control Scope	
6. Project Schedule Management		6.1 Plan Schedule Management 6.2 Define Activities 6.3 Sequence Activities 6.4 Estimate Activity Durations 6.5 Develop Schedule		6.6 Control Schedule	
7. Project Cost Management		7.1 Plan Cost Management 7.2 Estimate Costs 7.3 Determine Budget		7.4 Control Costs	
8. Project Quality Management		8.1 Plan Quality Management	8.2 Manage Quality	8.3 Control Quality	
9. Project Resource Management		9.1 Plan Resource Management 9.2 Estimate Activity Resources	9.3 Acquire Resources 9.4 Develop Team 9.5 Manage Team	9.6 Control Resources	
10. Project Communications Management		10.1 Plan Communications Management	10.2 Manage Communications	10.3 Monitor Communications	
11. Project Risk Management		11.1 Plan Risk Management 11.2 Identify Risks 11.3 Perform Qualitative Risk Analysis 11.4 Perform Quantitative Risk Analysis 11.5 Plan Risk Responses	11.6 Implement Risk Responses	11.7 Monitor Risks	
12. Project Procurement Management		12.1 Plan Procurement Management	12.2 Conduct Procurements	12.3 Control Procurements	
13. Project Stakeholder Management	13.1 Identify Stakeholders	13.2 Plan Stakeholder Engagement	13.3 Manage Stakeholder Engagement	13.4 Monitor Stakeholder Engagement	

2.1.3 Projects Life Cycle

A project life cycle is that the series of stages that a project passes through from its commencing to its closure [23]PMBOK.

Projects follow a predictable pattern or life cycle. A project life cycle consists of several stages during which deliverables are created and end with approval of the deliverables. The easy way to envision this is that a project must somehow begin - therefore, there is an initiating phase that starts with the conceive of an idea for a project and culminates in a decision to perform the project (or at least a decision to plan it in more detail and then make the decision whether to implement the project. In the vast middle time on most projects there is a combination of planning and executing of project work.

The most deliberative approach would have all of the planning precede starting of any project execution. The last stage in a project life cycle, closing, when the project's customers officially accept the project deliverables and ends when all the books are closed, documentation is complete, resources are reassigned, etc [27].

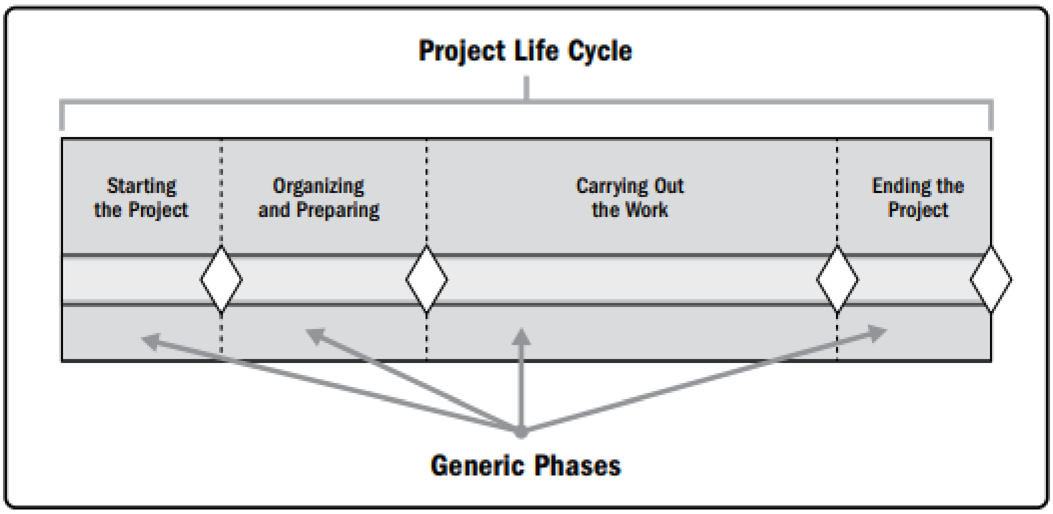


Figure 2. 1: Generic Depiction of a Project Life Cycle (PMBOK® Edition, Guide Sixth).

First, two phases are also called as the feasibility and engineering phases or when the project is built on paper, while in the third and fourth phases, preparation of detailed drawings including specification and physical implementation takes place. Commissioning and completion approvals are required at the end of the last two phases.

To summarize prediction of time and cost have vital importance along the project life cycle starting from feasibility studies under economic and scheduling feasibility to assess viability of project. Moreover, by considering the most feasible options design of the project carried out and ready for construction purpose. However, construction time is essential element of a contract incorporated in contract documents and cost estimation follows for engineering estimate and bid quotation offered by contractors to award the contract for responsible bidders during planning phase.

2.1.4 Project Stakeholders

The stakeholder is an actor who may impact, be impacted by, or realize itself to be impacted by a decision, activity, or deliverables of a project. They may be consciously engaged in the project or have interests that may be positively or negatively impacted by the progress or finish of the project. Stakeholders may have different expectations that might create disputes within the project. They may also exert pressure over the project, its deliverables, and the project team in order to accomplish a set of outcomes that comply strategic business aims or other requirement [23].

Key Stake holders in every Project described in PMBOK [23] includes;

- ✓ **Sponsor:** The person or group who bring resources, assist for the project, and is responsible for ensuring success.
- ✓ **Customers and users:** The individuals or organizations who will confirm and manage the project's product, service, or result. Users are the individuals or organizations who will utilize the project's product, service, or result.
- ✓ **Sellers:** They are, also named vendors, suppliers, or contractors, are external enterprise that make the contractual agreement to delivery parts or services crucial for the project**.**
- ✓ **Business partners**: They are external and independent organizations that have a special interaction with the enterprise, sometimes attained through a certification process. As a result, they bring specialized expertise or execute a specified role such as installation, customization, training, or support.
- ✓ **Organizational groups**: Organizational groups are internal actors who are impacted by the activities of the project team. Examples of several business elements of an organization that may be impacted by the project include human resources, manufacturing, marketing and sales, finance, legal, operations, and customer service.
- ✓ **Functional managers:** They are key individuals who do a management functions within an administrative or functional area of the business, such as human resources, finance, accounting, or procurement. They are assigned their own

permanent workers to implement the day to day activities, and they have a straightforward directive to administer all tasks within their functional area of responsibility**.**

- ✓ **Other stakeholders:** extra stakeholders, such as procurement body, financial institutions, regulatory body, subject matter experts, engineers, and others, may have a financial interest in the project, contribute inputs to the project, or have an interest in the ends of the project.

2.1.5 Project Planning and Scheduling

Several definitions for project planning and scheduling in construction sector can be found. Project planning as a process that is more than support that is vital in order to finish a project effectively [28]. According to Mubarak [29], project planning is foundation and involve a lot of linked functions such time planning, control of schedule and cost planning. Ideally, project planning is meant to accomplish several common factors, which covers criteria of design, project requirements, safety and health as well as to satisfy stakeholders' expectations.

Scheduling on the other hand, were defined as precedence relationship of activities, which include its durations. PMBOK [23]**,** was used to identify the PTM processes that need to be in place to complete construction projects on time. According to the aim of the PTM processes they can be related to the planning (i.e., Defining activity, Sequencing Activity, Estimating Activity Resource, Estimating Activity Duration, and Schedule Development) or to controlling (i.e., Schedule Monitoring and Controlling of Schedule) of the project.

Project scheduling is the development of a baseline that comprises of logical sequence, activity commence and completion dates as well as types of various resources needed with its amount for each time [29]. In addition, Baldwin and Bordoli [28] defined scheduling as the use of mathematical calculations and logical thinking to estimate the place and date for job to be undertaken efficiently and effectively. Although several definitions of project scheduling, its key objective is to collaborate construction activities precedence and timing systematically.

According to Baldwin and Bordoli [28]**,** planning and scheduling tasks cannot be carried out simultaneously. Scheduling process can only be done after planning and it is usually were handled by different people. stated that project planning answers the questions of what should be done? How? Whom? When should we do it? Where to do it? While project scheduling only focuses on 'when'.

Figure 2.2 describe that project scheduling answers when it is to be done while project planning answers what is to be done, what are activities to be done, what is needed to do it, who is to do, how it is to be done and where is to be done.

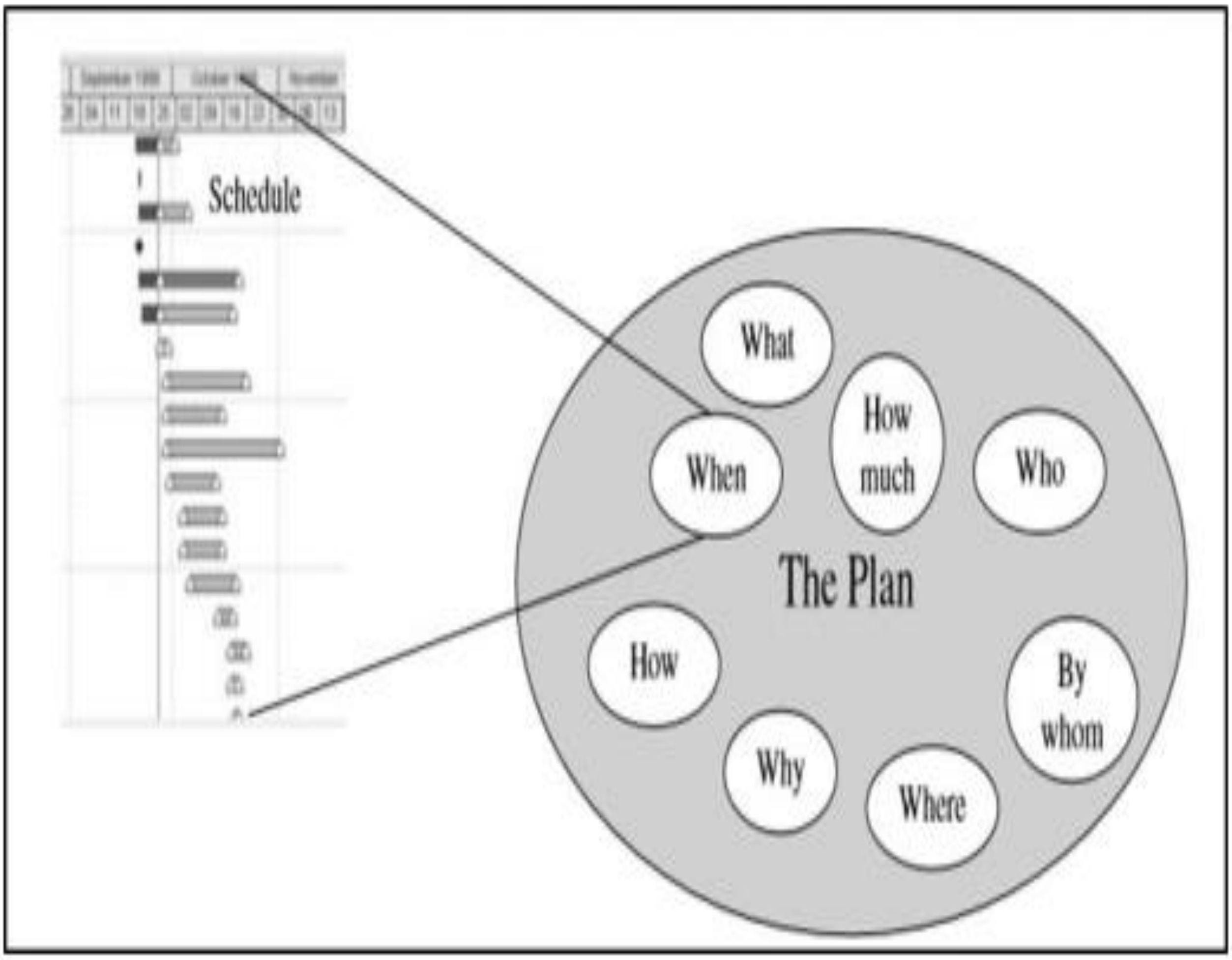


Figure 2. 2: Planning and scheduling (Mubarak, 2010).

In the perspective of project management in construction, project planning and scheduling are interrelated inputs that are planned with the intention to comply with the assigned objectives. These objectives must be defined carefully so that successful project performance can be achieved.

Figure 2.3 shows summary of knowledge requirements for planning and scheduling process to develop adequate project management plan which can use as baseline to measure progress during project implementation phase.

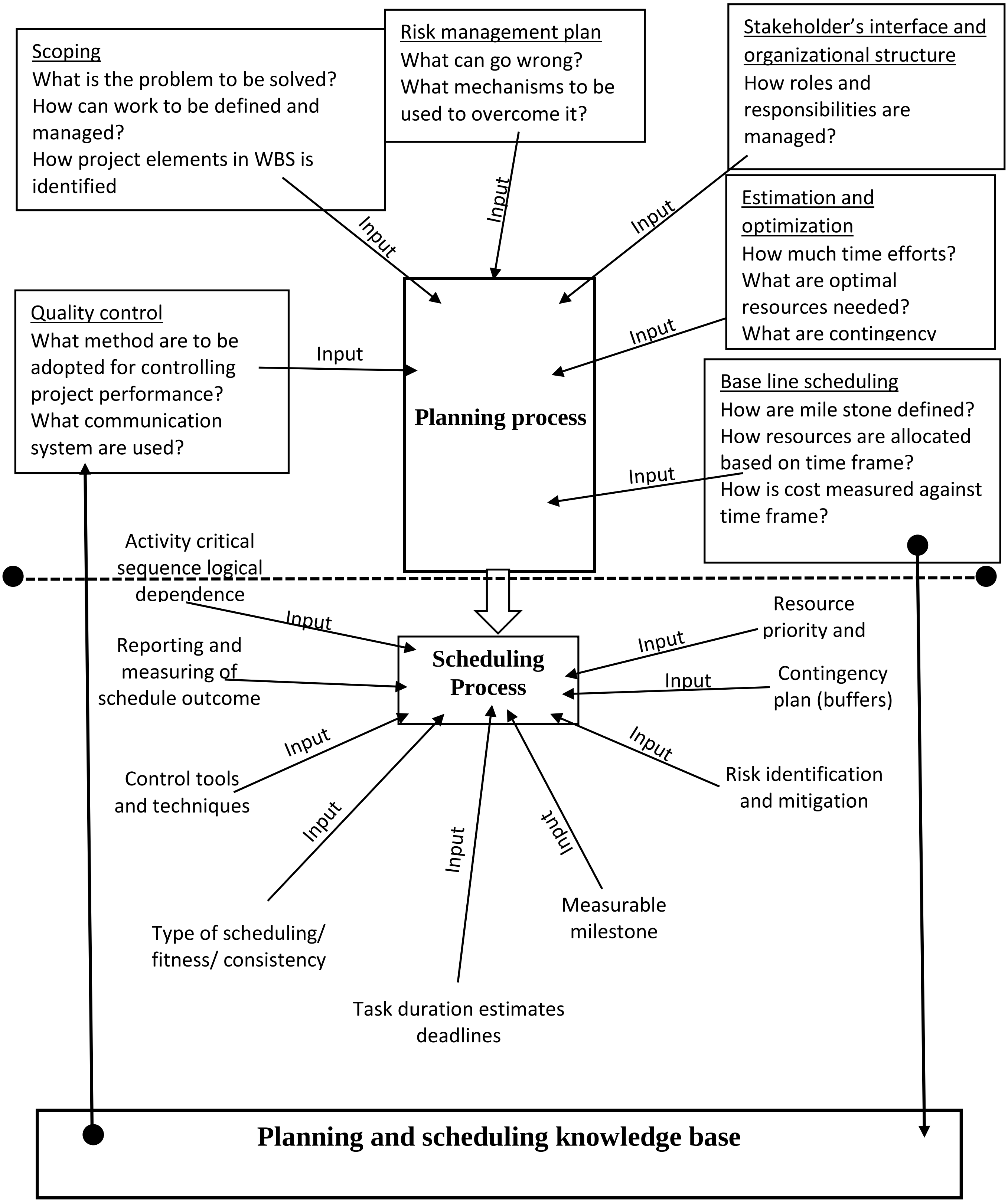


Figure 2. 3: Overview of project planning and scheduling Source (Kerzner (2013) and PMI (2008)).

2.1.6 Cost Estimation

Estimation is the scientific way of determining out the possible cost of an engineering project before implementation of the work [30].

Cost Estimate is defined as an approximation or anticipated cost for stated work scope of a work, project, or operation that is the process of forecasting the cost of a facility through quantitative determination of the work required by the design documents to evaluate a single total value and may have identifiable component values. However, underestimation of resources and costs is one of the most common contributors to project failure [31]. Consequently, realistic cost prediction is important for responsible management at every phase of the project [32].

The process of estimating cost serves three key importance in the construction industry:

- An estimate of the probably cost of construction is needed in the early before construction in order to determine whether or not a project is financially feasible. This prediction is produced from a minimum detail since no developed designs are available and the project is still in infancy.
- Estimates are required in cost control programs to facilitate the control of the expenditure of funds on a project. Cost management during the design phases of a project include considerations of alternative designs that help direct the decision making process better. Also, once construction is underway, this estimate provides a benchmark for the contractor to identify deficiencies and take corrective actions so as to maintain profit margins.
- This results in the competitive bidding process by which most construction contracts are awarded. Providing accurate forecast is basis to the contractor winning a bid.

2.1.6.1 Key Components of a Cost Estimate

A cost estimate is an addition of all the costs used in successfully completing a project, from initiation to commissioning (project contract time). These project costs can be classified in a number of ways and phases of detail, but the easiest category divides costs into two key categories: direct costs and indirect costs [33].

- ✓ **Direct costs** are mostly categorized as those directly related with a unique area (such as a department or a project). In management of project, direct costs are money spent only to a specific project. They can involve project team salary, the costs of key resources to create physical facilities, fuel for machinery, and money spent to analyze any project- specific risks.

- ✓ **Indirect costs,** on the other ways, cannot be related with a specific cost center and are instead spent by a number of projects simultaneously, sometimes in varying amounts. In management of project, quality management, utilities, and security costs are often classified as indirect costs since they are interchangeable used with in a concurrent projects and are not directly spent to any one project.

2.1.6.2 Costs Associated with Constructed Facilities

The costs of a constructed asset to the client involves both the initial working capital cost and the following daily operation and maintenance costs. Each of these main cost categories comprise of a number of cost components [30].

The construction project capital cost involves the expenses associated to the initial development of the facility:

- Land possession, including assembly, holding and improvement
- Planning and viability studies
- Architectural and engineering design
- Construction, including materials, machinery and manpower
- Field inspection of construction
- Construction financing
- Taxes and Insurance during construction
- Owner's head and site office overheads
- Equipment and furnishings not consisted in construction
- Inspection and testing

The operation and maintenance cost in subsequent years over the project life cycle includes the following expenses:

- Land rent, if applicable
- Labor and material for maintenance and repairs
- Periodic renovations
- Insurance and taxes

- Financing costs
- Utilities
- Owner's other expenses

In majority of construction budgets, there is an allowance for contingencies or unexpected costs arising during construction. This contingency amount may be added within each cost of activity or be included as defined contingency percentage of total cost of a single building. The percentage of contingency considered is based on past experience and the anticipated uncertainty of a specific construction project. For instance, one construction organization undertakes prediction of the expected cost in five different areas[30].

- Design development changes,
- Schedule adjustments
- General administration changes (such as wage rates),
- Differing site conditions for those expected, and
- Third party requirements imposed during construction, such as new permits.

2.1.6.3 Types of Construction Cost Estimates

Construction cost comprises only a fraction, though a considerable fraction, of the overall project cost. Although, it is the part of the cost under the management of the construction project manager. The needed levels of accuracy of construction cost estimation vary at different phases of project development, starting from the ballpark amount in the early stage to fairly reliable amount for budget control early before construction. Since design decisions made at the planning stage of a project life cycle are more tentative than those made at a later stage, the cost estimates made at the earlier stage are expected to be less accurate. Generally, the accuracy of a price estimate will reflect the knowledge obtained at the time of estimation.

Construction cost estimates could also be viewed from a distinct perspective due to different institutional requirements. In spite of the various sorts of cost estimates used at different stages of a project, cost estimates can best be classified into three major categories in line with their functions. A construction cost estimate serves one among the three basic functions: design, bid, and control. For establishing the financing of a project, either a design estimate or a bid estimate is employed [34].

1. Design Estimates. For the employer or its assigned design professionals, the kinds of cost estimates undertaken run parallel with the planning and design as follows:

- Screening estimates (or order of magnitude estimates)
- Preliminary estimates (or conceptual estimates)
- Detailed estimates (or definitive estimates)
- Engineer's estimates supported plans and specifications

2. Bid Estimates. For the contractor, a bid estimate submitted to the owner either for competitive bidding or negotiation consists of direct construction cost including field supervision, plus a markup to cover general overhead and profits. The direct cost of construction for bid estimates is usually derived from a combination of the following approaches.

- Subcontractor quotations
- Quantity takeoffs
- Construction procedures.

3. Control Estimates. For monitoring the project during construction, a control estimate is derived from available information to establish:

- Budget estimate for financing
- Budgeted cost after contracting but prior to construction
- Estimated cost to completion during the progress of construction.

2.1.6.4 Cost estimation Techniques

The most common cost estimating techniques in construction projects [35]**.**

1. Analogous Estimating - Also called top-down estimating, analogous estimating looks at historical project data to develop estimates for new projects. The idea of this system is that the reliance on an archive of past and similar projects.

The main advantage of this technique is that it is very quick, and there are not alot of inputs required. The only estimation that has to be done is how much the cost should be scaled based on the matching criteria.

Conversely, if no elements are found to be within the threshold for comparison, then the model is completely ineffective and unusable. Clearly, this poses a largeconcern if the database from which the user is pooling is limited. In extreme cases, if no element is found within the threshold, there will be no estimate that is generated.

2. Bottom-Up Estimating - Also called analytical estimating, bottom-up estimating divides projects deliverables into a series of work packages, which are made up from a series of tasks. The idea of this system is to estimate the entire cost of completing each task and add all of them up to calculate the entire costs. This system is understood to be one among the foremost accurate, but also one among the foremost time-consuming.

The greatest advantage of Bottom –up estimation method is that an accurate model is developed for each process. Because the accuracy of these models can be very good, and every process is accounted for, this method is a very powerful way to estimate cost.

The main drawback of process-based cost estimation is that an expert in the area of manufacture must estimate the amount of raw material that is used by each process model. Because this process is very tedious, it does not afford a quick way to estimate cost. The design must be very detailed as well, in order for the input data to be precise.

3.Parametric Estimating - parametric cost estimation is a way of estimating costs using Cost Estimating Relationships (CERs), along with mathematical algorithms and other logic to establish a cost estimate. Moreover, the parametric method seeks to evaluate the costs of a product from the parameters characterizing the product but without describing it completely. Simply put, if the original cost is known of the item being manufactured, users can take simple inputs and generate a model that will produce a cost that is reasonably accurate.

The main advantage of parametric cost estimation is its ease of use. Once the model has been defined, application of the method is straightforward. Data is easy to implement, and any user can look at a design and find the input data needed. By taking general, broad attributes, a fairly accurate estimate can be generated. This allows an estimate to be made without the presence of a detailed conceptual design.

Some disadvantages of parametric cost estimation are that the model itself can be very difficult to develop. To begin, one needs to first develop appropriate parameters,and this person usually has to be an expert on the topic. Next, historical data must be found, as well as relationships that can take vague inputs and produce a cost that is as accurate as possible. On top of being difficult to develop, accuracy is another great drawback of this technique. Other methods, while more complicated for the user, are generally more accurate.

4. Expert Judgment - This technique can only be conducted by experts and specialists who have an enormous background knowledge with similar projects in the past. The specialist must compare a past project with the present project to make and adjust the cost estimate. This technique can't give an exact number, so it most frequently used when determining if a project viable or not.

The main advantages of cost estimation using expert judgment are quick to produce, requires little resource in terms of time and cost and can be as accurate as other more expensive methods.

The major drawbacks are related to its practicality. The first is related to the time needed to obtain the group opinion, and the second is related to the number of experts required to produce worthwhile results. Estimators make many qualitative judgements as they generate estimates and are often under time constraints and working with limited amounts of information.

5.Three-Point Estimating -The three-point estimating technique comes from the Program Analysis and Review Technique, which analyzes activity, project costs, and durations by determining optimistic, pessimistic, and presumably estimates. This statistical procedure uses various formulas to calculate expected costs. this technique can work well, but only if the initial estimates are accurate.

The three-point estimating technique offers several advantages such as more accurate estimates, minimizing risk, analyzing and prioritizing risks.

The cons of three-point estimating techniques are subjectivity of activity sequence weakens accuracy, bias of participants submitting estimates reduces precision of data and resources of people and time needed to collect estimates can be cost prohibitive.

To conclude, cost estimation techniques are applied within project phases. Analogous, parametric, empirical, expert judgement and Three point based estimation technique most of the time suitable during viability analysis and design development phase. However, Bottom-up estimation technique is fundamental during bid offer preparation by professional contractors in Ethiopia.

2.1.7 Time-cost Tradeoff Analysis

Time and cost analysis are the most prominent factors for planning and monitoring of construction projects. Construction planners carry out the selection of different sources and technologies, such as materials, labor force, equipment, and methods, for the realization of an activity. Uncertainties and risks in construction applications according to the region where the project is implemented are taken into account in this selection process. In this respect, project duration and project cost have been determined consistent with the technique chosen within the context of regional conditions.

The critical path method (CPM) is employed so as to monitor time and costs in construction project planning also on monitor resources. The trade-off between quality, time, and cost is vital in project planning [36]. Therefore, most vital goal of the construction management engineers in project planning is determining the optimum point between time and cost by using trade-off rates.

According to Liao [37], Time-Cost Trade-off (TCT) is employed in meeting project deadline especially when considering such projects whose deadline is fixed or a project that is already behind schedule. Furthermore, Siemens [38] also described that TCT analysis involves replacing some activities found on the critical path with shorter methods of construction for the purpose of saving time, even though this might increase the cost of such critical activities, even if non-critical activities were left undisturbed to save cost. Therefore, the purpose of the time-cost trade-off problem (TCTP) is to compress the first project time acquired from the critical path analysis, to fulfill a specific due date with the minimum direct and indirect cost of the project.

2.1.8.1 Activity Time- Cost Relation Ship

In general, there is a trade-off between the time and the direct cost to complete an activity; the less expensive the resources, the larger duration they take to complete an activity. compressing the duration on an activity will normally increase its direct cost which includes the cost of manpower, machinery, and material [39].

A simple way of representing the possible relationship between the activity time and its direct costs appears in Figure 2. 4 Considering only this activity in isolation and without reference to the project completion due date, a manager would select a duration, which took minimum direct cost, called the normal duration. At the other ways, a manager might choose to complete the activity in the minimum possible time, called crashed duration, but at a maximum cost.

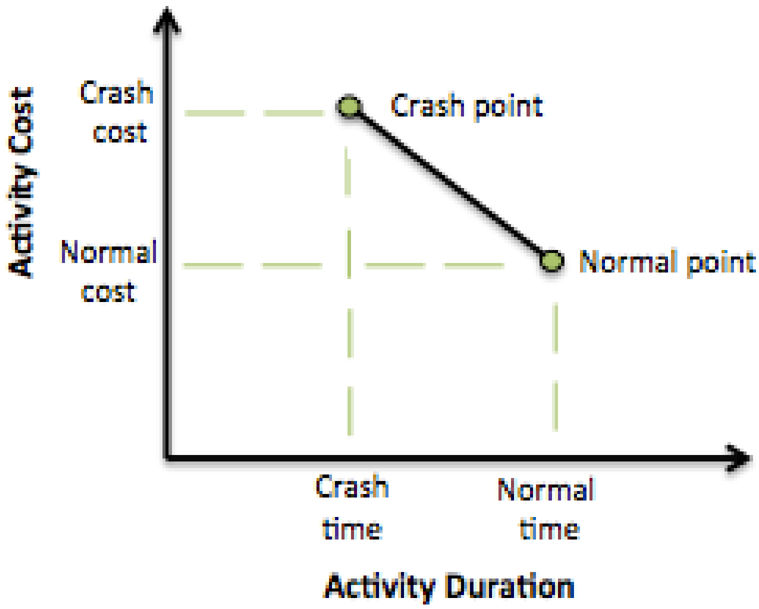


Figure 2. 4: Activity Cost and Time Relationship (Deterministic Decision Models. (n.d.). Course material in Network Optimization at DTU autumn 2015)

The linear relationship shown within the Figure 2.4 between these two points depicts that any intermediate duration could even be chosen. It is possible that some intermediate value may reflect the ideal or optimal trade-off between time and cost for this activity. The slope of the line linking the normal point (lower point) and the crash point (upper point) is termed the cost slope of the activity. The slope of this line are often determined mathematically by knowing the coordinates of the normal and crash points.

Cost slope = crash cost – normal cost / normal duration – crash duration

As shown in Figure 2. 4, the smallest amount direct cost needed to finish an activity is called the normal cost (minimum cost), and also the corresponding duration is called the normal duration. The minimum possible duration needed for finishing the activity is called the crash duration, and also the corresponding cost is called the crash cost. Normally, a planner commences his/her estimation and scheduling process by assuming the smallest amount costly option.

2.1.8.2 Project Time –cost relation ship

Total project costs include both direct costs and indirect costs of executing the activities of the project. Direct costs for the project include the cost of materials, manpower, machinery, and subcontractors. Indirect costs, on the other hand, are the required costs of doing work which can't be associated with a specific activity, and in some cases can't be associated with a specific [39].

If each activity was scheduled for the duration that resulted within the minimum direct cost during this way, the time to finish the whole project could be too long and substantial penalties related to the late project completion might be incurred. Thus, planners carry out what is called time-cost trade-off analysis to reduce the project duration. This could be done by choosing some activities on the critical path to minimize their duration.

As the direct cost for the project equals the sum of the direct costs of its activities, afterwards the project direct cost will increase by decreasing its duration. On the other hand, the indirect cost will have reduced by reducing the project duration, because the indirect cost are almost a linear function with the project duration. Figure 8.5 demonstrates the direct and indirect cost relationships with the project duration.

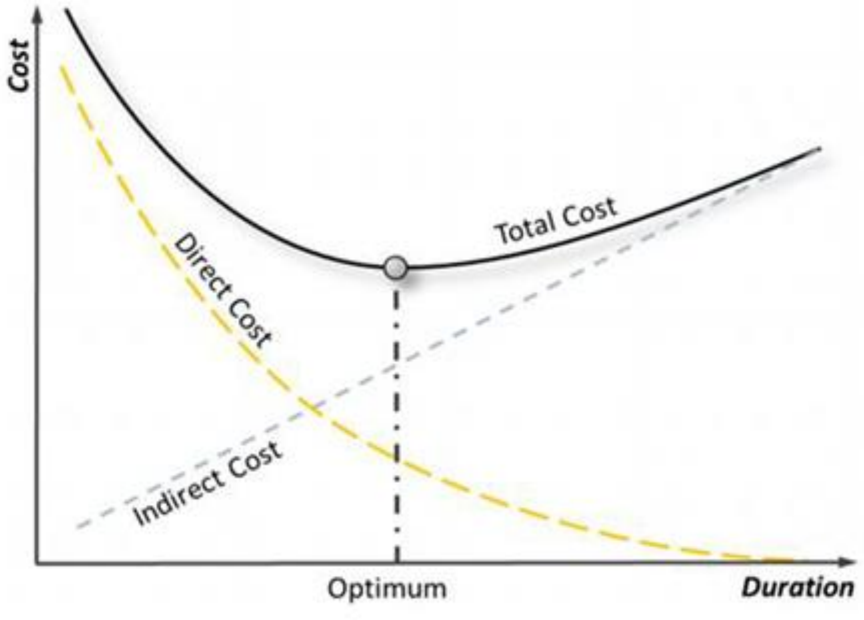


Figure 2. 5: Project time-cost relationship (Hegazy T.,1999)

The project total time-cost relationship can be determined by adding up the direct cost and indirect cost values together as shown in Figure 2.5. The optimum project duration can be calculated as the project duration that results in the minimum project total cost.

After having utilized the CPM, the procedure for project crashing covers the following steps [39, 40, 41].

1. Compute the crash cost per period for all activities in the project network
2. Find critical path(s) and critical activities in the project network
3. In the case of only one critical path, identify a critical activity with the least crash cost per period that can still be crashed. Otherwise, identify critical activity/activities from each critical path with the least crash cost per period, which can still be crashed.
4. Reduce the duration of the critical activity/activities identified in step 3 by one time period.
5. Reconstruct or adjust the project network based on changes made in step 4. Calculate the total costs = directs + indirect costs. Stop the procedure if the desired completion deadline is reached, otherwise return to step 3. Continue until the crash point has been reached, that is, no further shortening is possible.

2.2 Related Literature Review

2.2.1 Modelling Time

Time, cost and quality are one of the traditional key performance indicators for successful project management. In addition, the realistic schedule is a critical success factor that appears in many publications which focused on research about the success factors. Although time in the construction projects is a factor that is highly depend on the experience to arrange the time, technology, financial, and the other factors, there are some attempts to create a construction duration model based on costs and project scope factors that have been realized.

In practice, there are two common methods of estimating project completion time: (1) according to the client's time constraints eg. occupancy need, or (2) through a detailed analysis of work to be done and resources available, using estimates of the time requirements for each specific activity [21]**.** Method (2) is known to be very tedious and is often impractical in view of the time limitations imposed on contractors at the tendering stage or uncertainty regarding the specific tasks and materials required of the project. Detailed estimating of construction activities also relies on the estimators' experience and judgement to correctly interpret project and site information and make the best possible decisions according to cited by [21]**.**

2.2.2 Previous studies on Time - Cost relationship based on Bromillow's Model

Several research is carried out to develop Time- Cost model for estimating construction duration since 1969, using Bromillow Time – Cost as a baseline model. The first empirical study was carried out and published for the Australian building industry by Bromillow using a total of 329 building projects with a total value exceeding 270 million Australian dollars (A$) were analyzed [10]. These projects were conducted in Australia during the period between June 1964 and June 1967. First, the expected duration of each project was compared with the actual. The projects were classified according to the type of building and the area in which they were to be built. The results showed that only one contract in eight was completed on or before the date originally expected and the overall average extra time taken exceeded 40% of the original. This was far worse than expected.

The second part of the work was the build-up of the relationship between the actual construction cost of the building and the time taken. The equation describing the average construction duration as a function of project value was found to be:

$T = KC^{B}$ ....[Eq.2.1]

Where

T = duration of the project period from the commencement of construction to practical completion, in working days.

C = final cost of building in millions of dollars, adjusted to constant labor and material prices.

K = a constant describing the general level of time performance for a $1 million project; and

B = a constant delineating how the time performance is affected by project size as measured by cost.

The constants K and B may vary depending on the construction conditions of the country, but also depending on the construction sector. Thus, He finally summarized his model as $T = 350C^{0.3}$ and argued that the variation in construction duration can be explained by project scope that is measured as project cost. Furthermore, Bromillow's surveyed 370 building projects in Australia and developed empirical model $T = 313C^{0.3}$.In 1980, Bromillow's further analyzed 419 projects between the periods 1970-1976 to determine whether the model still holds. After using the Data programme to perform the necessary analysis, he concluded that the model $T = KC^{B}$ still holds for the Australian situation [42].

After development of the BTC model other researchers around the world also carried out similar studies based on this model.

A similar research was undertaken by Ireland [43] to predict the construction time of high-rise commercial properties in Australia. From analysis of 25 high–rise buildings, he concluded that the best predictor of average construction time of high rise commercial buildings based on cost in millions which have been indexed to June 1979 was $T = 219C^{0.47}$ for $R^2 = 0.58$ where R^2 is the coefficient of determination. Furthermore, buildings of a particular area tend to be more quickly constructed if they are of fewer stories.

Kaka and Price [44] also conducted a similar research on buildings and road projects in the United Kingdom. They considered two different samples. Sample 1 consisted of 661 building projects with total value exceeding £695 million while sample 2 was made up of 140 road projects with a total value exceeding £120 million. They produced a similar relationship after adjusting the contract values to 1988 prices using the adjustment formula and tender price index available in the building cost information service manual. They obtained the following relationships for the UK situation: For Government building projects, T = $486C^{0.205}$ for R = 0.68.For private building projects, T = $491.2C^{0.082}$ for R = 0.61. For road construction projects, T = 436.3 $C^{0.437}$for R = 0.97 where R is the coefficient of correlation. Kaka and Price (1991) also tested the influence of the type of client (public or private), type of tender, and the form of tender. Through their study, they arrived at the following conclusions: (i) the type of bid competition did not affect the reliability of the BTC model; (ii) the type of client (public or private) does influence the time-cost relationship with public building works generally taking longer than private works; (iii) the type of project affected the relationship considerably with civil engineering works taking less time to complete than buildings of the same value; and (iv) the form of the contract significantly influenced the time-cost relationship with adjusted price contracts being the largest (in dollars) and longest (in working days).

Kumaraswamy and Chan [45] surveyed a combination of building and civil engineering projects and confirmed that the time – cost relationship for both types of project can be modelled in the form of T = KC^{B} . In a survey of 111 projects in Hong Kong, they obtained the K and B values for building and civil engineering projects, which are shown in Tables 2.2 and 2.3 respectively. The correlation coefficients showing a relationship between the two variables are also shown.

Table 2. 2:Time- cost performance of Government building projects in Hong Kong (Kumaraswamy and Chan 1995).

Type of building	Estimated			Actual		
	K	B	R	K	B	R
Total building projects	182.3	0.277	0.81	216.3	0.253	0.79
Public (government) housing	188.8	0.262	0.77	178.8	0.279	0.70
Public (government) buildings	166.4	0.294	0.78	207.1	0.266	0.76

Table 2. 3: Time- cost performance of civil engineering projects in Hong (Kumaraswamy and Chan 1995).

Type of building	Estimated			Actual		
	K	B	R	K	B	R
Total civil projects	252.5	0.213	0.80	291.4	0.205	0.78
Road works	233.1	0.248	0.89	301.4	0.215	0.80
Other civil projects	270.6	0.190	0.71	272.3	0.211	0.77

In addition, they suggested that the inclusion of other project characteristic macro variables such as construction cost, gross floor area, number of storeys and micro factors affecting productivity as well as other significant factors that may influence project duration.

Love,Tse and Edwards [11] investigated Project time– cost relationships using project scope factors as predictors for 161 construction projects that were completed in various Australian States using the multiple regression technique of weighted least squares. Project scope factors included project size, project type, facility type, contract value, project duration, procurement method, the tender type, Gross floor area (GFA) and number of floors (i.e., project height), and their influence on the time and cost performance was examined. An alternative model to that proposed by Bromillow more than three decades ago for forecasting time and cost performance was proposed. Logarithmic regression was performed between project duration (i.e., time), project type, procurement method, tender type, gross floor area (GFA), and number of stories. The result of the analysed data shows that GFA and the number of stories in a building are key determinants of time performance in projects. However, BTC results indicate that cost is a poor predictor of project time. As we do not know cost (ex post) before a project is completed, there is no practical use for the time–cost relationship to predict (ex ante) project time. According to Yousef and Baccarini cited by [11] found unreliable predictions using BTC, since a number of projects were completed rather differently from that estimated by client’s project managers at tender award (rather than based on actual cost). Thus, they suggested that it is more sensible to predict project time based on GFA and the number of floor levels, rather than cost.

Therefore, the suggested project time (T) can be predicted using the following equation model:

$$\text{Log}(T) = \beta 0 + ((1+\beta 1) \times \text{Log}(GFA)) + \beta 2 \times \text{Log}(F) \qquad \text{...............[Eq.2.2]}$$

Thus, established LTF model was converted to the following form by replacing the found coefficients (βi) from regression output:

$$\text{Log (T)} = 3.178 + 0.274 \times \text{Log (GFA)} + 0.142 \times \text{Log (F)} \qquad \text{...............[Eq.2.3]}$$

Furthermore, Love et al.'s Time-Floor suggested that project cost essentially comprises of two elements: Labor and material. Consequently, they hypothesized that C=(T,GFA), where T is used as a proxy for the amount of labor input. If we transform this into an empirical model, the following regression is derived:

$$\text{Log(C)} = b0 + b1 \log(T) + b2 \log(GFA) \qquad \text{...............[Eq.2.4]}$$

Thus after performance data was used the following model obtained in Equation 5,

$$\text{Log(C)} = 4.461 + 0.715 \log(T) + 0.837 \log(GFA) \qquad \text{...............[Eq.2.5]}$$

The result of the model shows that, an adjusted R^2=0.86 and standard error=0.63 are derived from the analysis. The results indicate that construction costs increase by 0.71 and 0.82%, respectively, when project time and area increase by 1%. This implies that the cost per unit of area decreases with GFA—possibly as a result of economies of scale.

Love,Tse and Edwards [46] concluded their study; cost and time were both different at the end of the project than expected at the beginning and stated that change orders and rework are the primary reasons for such deviations. Hence, they suggested that a significant amount of rework, and too some extent change orders, can be eliminated with effective design and project management.

Hoffman et al [47] also described BTC based on the data collected for 856 facility projects completed between the period of 1988 and 2004. The data were analyzed using the BTC model and multiple linear regression. Applying the BTC model, they came up with $T = 26.8C^{0.202}$ .The result indicated that majority of the variability was not explained by the model since $R^2 = 0.337$.And also, study therefore attempted to explain the larger portions of this variability by considering other factors.

The first study carried out in Africa on the BTC model was by Ojo cited by [48] in Nigeria. The study carried out a survey in southwestern part of Nigeria and arrived at a similar relationship, $T = 27C^{0.125}$. The coefficients of correlation (R) and determination (R^2) for this model are 0.431 and 0.186 respectively. She attributed the low value of R^2 and for that matter poor performance of the BTC model in Nigeria to the unstable economic climate. Ogunsemi cited by [48] also attributed the poor performance of the BTC model in Nigeria to price fluctuation, which he identified as the most prominent cause of cost overrun of construction projects in South-Western Nigeria. Because the model formulated by Ojo cited by [48] had low predictive ability (R^2= 0.186),Ogunsemi cited by[48] , Ogunsemi and Jagboro [48] also carried out studies in the same part of Nigeria as Ojo.

After a survey of 87 building projects, their models based on the BTC are as follows: For private projects, $T = 55C^{0.312}$ for R = 0.567 and R^2 =0.322. For public projects, $T = 69C^{0.255}$ for R = 0.443 and R^2 = 0.196. For all projects, $T = 63C^{0.262}$ for R = 0.453 and R^2 = 0.205 .The above models were obtained after the cost data were adjusted to 2000 prices. Based on the coefficient of determination, Ogunsemi and Jagboro [48],indicated that for all the projects, 20.5% of the variance in construction duration is explained by the project scope expressed in terms of the estimated final cost of construction. This according to them means that 79.5% of the variance in construction is explained by other variables that are not included in the model. They therefore indicated that the BTC model is not applicable to the Nigerian situation. They however developed another model using the piecewise linear model with breakeven point. The following models for the three categories were obtained as: For private projects, T = 168.895 + 0.491C ($C \leq 557$) or 709.66 + 0.884C ($C>557$). For public projects, T = 98.01 + 0.357C ($C \leq 353$) or 567.967 + 0.283C ($C>353$). For all projects, T = 118.563 - 0.401C ($C \leq 408$) or 603.427 + 0.61C ($C>408$).The coefficient of determination (R^2) for the private, public and all projects were 77.62%, 83.06% and 76.56% respectively. They therefore indicated that these models have high predictive abilities than the BTC model for the Nigerian situation.

Furthermore, Ameyaw, Mensah and Arthur [49] tested applicability of bromillow's time – cost model on building projects in Ghana. They analyzed 62 building projects consist of 13 offices, 27 residentials and 22 class rooms. The analysis results , R^2 values of the various samples show a very low figure with the exception of office buildings which recorded an R^2 = 0.684 which was considered to be moderately good in terms of the goodness of fit of the models derived from the data. The R^2 obtained with respect to classrooms, residential and the combined sample ranging from 0.463, 0.399 and 0.378 was observed to be low and for that matter weak to predict the project time in Ghana. Based on R^2 values , they concluded that the original Bromillow's time cost model is not applicable in the Ghanaian construction industry.

Mačková and Bašková [19] checked applicability of Bromillow's Time- Cost model for residential projects in Slovakia using 28 residential projects completed within the period 2010–2013 and the output of the final linear regression model, where dependent variable is natural logarithm of time and independent variable is natural logarithm of cost, has the high coefficient of determination and adjusted coefficient of determination values of 0.776 and 0.768, respectively. It is shown that that over 76 percent of the variance in natural logarithm of time estimation can be explained by natural logarithm of cost. As result, He confirmed Bromillow's Time-Cost model is suitable for duration estimation for construction projects.

Jarkas [16],also conducted comparable research to develop a scientific model for predicting construction duration using 113 residential and 74 office building completed projects with in a period of 2004-2010 as a sample data for analysis. The data were analyzed using BTC model and multiple regression model. Applying BTC model, they came up with $T = 28.79C^{0.192}$ for residential and, $T = 7.92C^{0.277}$ for office building projects. The study results showed variability in construction duration was explained by the model since $R^2 = 0.803$ for residential $R^2 = 0.849$ public office-building projects. Even if, He affirmed that Bromillow's time –cost model is applicable to Kuwait construction industry. Multiple regression model results; T =338.70 +0.00574CA +13.95NFAG + 27.58NFBG, $R^2 = 0.928$ for residential, T =453.97 + 0.0103CA +1.92NFAG + 58.54NFBG, $R^2 = 0.965$ for office building projects. Finally, He compared between the outcomes of both models tested demonstrates that, on average, the multifactor model proposed is associated with an enhanced prediction accuracy of 16 and 14% for residential and office buildings, respectively, and therefore can be deemed more suitable for forecasting local construction duration for the building types investigated.

Additional study by, Otali and Adewuyi [50] reported on duration prediction using time- cost relationship for public building projects in north central, Nigeria for both projects types that were executed under traditional; and design and build procurement arrangements based on time-cost relationships. Forty projects were found valid and application for this study comprising of twenty (20) executed under traditional and twenty (20) executed under design and build procurement arrangements. The model was built using regression statistical tool. As a result, they obtained, CT= 0.46 +0.63* CC, with R^2 = 0.38 which shows that construction cost accounts for 38% of variation in construction time of building projects executed with Traditional procurement method and CT= 0.36+ 0.59 * CC, R^2 = 0.26 which shows that construction cost accounts for 26% of variation in construction time of building projects executed with Design and Build (DB) procurement method. The remaining 62% under Traditional procurement method and 74% under Design and Build (DB) procurement method may be attributed to other variables and factors affecting construction the accurate prediction and reliability of construction cost. These include the project location, project type, contractor selection method, corrupt practices in public projects procurement and inflation. Others include urgency of completion, incomplete drawing before the commencement of projects, lack of collaboration between the Design and Build contractors and other professionals at the planning stage, the correctness of the prepared project cost, lowest bid cost syndrome and the nature and type of project risks that were not identified nor budgeted for at the point of estimating the construction cost. All these will affect the reliability of construction cost as a single criterion for predicting project duration. The pearson correlation coefficient of 0.62 and 0.51 under Traditional procurement method and Design and Build (DB) procurement method respectively shows that the construction cost correlates with time of projects under study.

A more recent study Bayram [7] conducted on duration Prediction Models for construction projects: In Terms of Cost or Physical Characteristics using 'Bromillow's Time-cost (BTC) model' and the 'Love et al.'s for re-validation. The study used a total of 530 Turkish public building projects which were completed in seven geographic regions by performing an archive search work 80% of the total data (i.e. from 424 projects) were used to establish the models while the remaining 20% (106 projects) were used for validation. Eight project characteristics of; total construction area, building height, gross floor area, floor numbers, contract sum, actual cost, contract duration and the actual duration were collected. The data collected were analyzed using BTC and Multiple linear regression model. The result of analysis shows $T = 192C^{0.467}$, $R^2 = 0.419$ for contract duration and $T = 209C^{0.353}$, $R^2 = 0.255$ for Actual duration. While LTF model result shows $T = 14.72 \times GFA^{0.332} \times F^{0.401}$, $R^2 = 0.358$ for contract duration and $T = 32.51 \times GFA^{0.246} \times F^{0.318}$, $R^2 = 0.247$ for Actual duration. Consequently, He suggested that the BTC model is superior to the LTF model, which confirms that 'cost' is a more significant predictor of the 'duration' than'gross floor area' & 'floor numbers.

2.2.3 Previous Time - Cost relationship study in Ethiopia

Although the first study in duration modelling using BTC in African was undertaken in Nigeria, further research carried out in other African countries. The first construction duration prediction model was carried out in Ethiopia cited by Abraham [51] The Building and

Transport Construction Design Authority (BaTCoDA), the former government building regulatory body, in 1987 formulated completion time based on cost by using the data on completed projects." The formula is:

$T = 7 \text{ to } 12\ C^{1/3}$ ....[Eq.2.6]

Where:

7 - Stand for small sized buildings and 12- taken for most complex building projects.

T - Actual completion time in months

C - Actual project cost Birr in millions (Abraham, 2008)

Furthermore, In Ethiopia Andinet et al.'s [52] developed construction time bench marking BTC in 2008**.** The study collected data for 29 public educational building projects in Ethiopia and concluded that the time-cost relationship for the sample education sector construction projects in Ethiopia can be expressed using Bromillow's model as; **T = 7.06***$C^{0.47}$ where,T=total time required to complete the project in months and C= actual completion cost in million of ETB. He obtained output on relationship between time and cost using regression tools on figure;

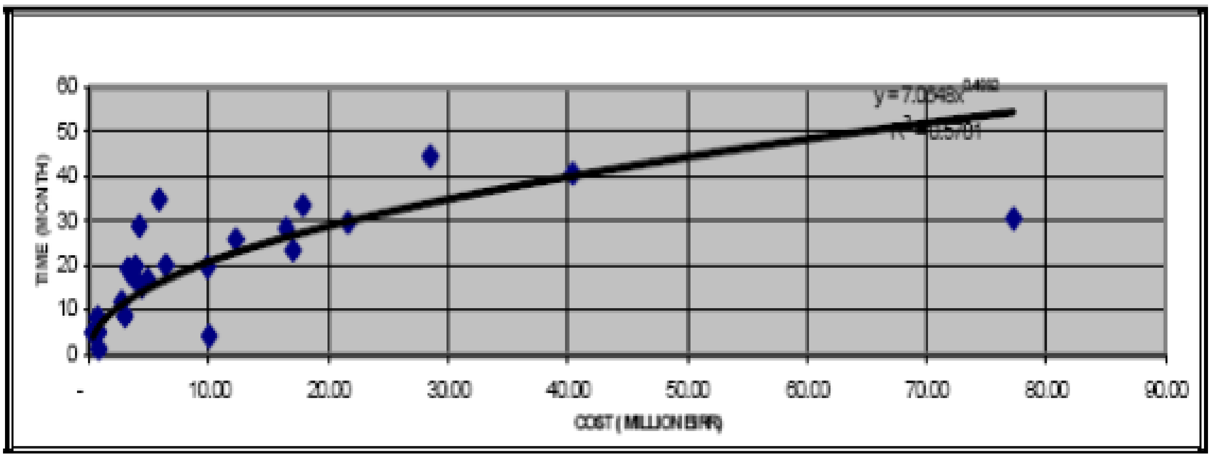


Figure 2. 6: Scatter plot for the relationship between Time and Cost.

Abraham [51] also conducted additional study on road projects. He collected data for a total of 33 road projects contracted by International and Domestic contractors in Ethiopia. He carried out separate analysis for different categories of road projects. From a total of 33 projects, 21 of them were constructed by international contractors and 12 were constructed by domestic contractors. Out of the 21 implemented by international contractors 14 are AC Surface projects and 7 are DBST surface projects. All the 12 projects implemented by domestic contractors were Gravel Surface projects. He concluded his study by developing time and cost models for each category of road projects as follows:

Time prediction model using Bromillow's principle has been found valid for IC projects carried out under International Competitive Bidding Contracts for both AC and DBST road projects. The study established formulas Log (T) =-47.058+8.7680Log(C)-0.0394$(\text{Log(C)})^3$ for AC projects and Log (T) =32.795+6.1661Log(C) -0.0268$(\text{Log(C)})^3$for both AC and DBST road project , where, T – Time in calendar days, C – cost in Ethiopian Birr, IC -projects contracted by international contractors ,AC-asphalt concrete and DBST-double surface treatment. Furthermore, his study revealed that Bromillow's Time- cost relationship is invalid for DC projects carried out under National Competitive Bidding Contracts. Though, the relationship with final Time and cost was weak for DC projects the relationship between contract Time and length has been found to be strong enough to predict is modelled as ; (Tc)=15.06515+20.224(L)-0.2631(L) 2+0.0013$(\text{L})^3$ Where, T – Time in calendar days, L – project length in km, DC-projects contracted by domestic contractors and GS-gravel surface. Cost prediction model was investigated based on project length and as a result, the analysis of the data came up with the model Log(C) =6.9174+0.6332Log (L) + 0.0683$(\text{Log (L)})^2$for IC - AC Road Projects and Log(C) =6.2154+1.3132Log (L)-0.0464 $(\text{Log (L)})^3$ for IC - DBST Road Projects where C – cost in Ethiopian Birr and L project length in km.

Same research on educational buildings Muluken [53] investigated additional study on Time, Cost and floor area relationship for Educational buildings in Ethiopia. He collected data from 40 educational building construction projects for analysis which 19 are UCBP Lot 1A projects located at different regional states of the country and the remaining 21 projects are under the classification of other school building projects which are collected from Ministry of Education (MoE), Construction design share Company (CD.S.C) and sub city offices of works and urban development of Addis Ababa. He analyzed time overrun for projects, presented empirically mean average overrun as 62.63% for UCBP, 195.91% for other school buildings and 129.27% for all projects. Moreover, He analyzed cost overrun for projects obtained 21.18% for UCBP,34.18% for other school buildings and 27.68% all projects. After the analysis of overrun for sampled construction projects, his study came up with the cost prediction models; Log (C) = 4.498*$e^{(0.123(\text{LogA})}$ for UCBP and Log(C) =4.194*$(1.133)^{(\mathbf{Log(A)})}$ for school building projects. As a result , he concluded that Bromillow's Time-cost models has not been found valid for educational building projects in Ethiopia classified under UCBP and school building projects, Cost-Total floor area relationship is found valid for both project categories and hence the cost prediction models was developed which can to be used as cost prediction tool for educational building projects in Ethiopia and time performance measured by Cost and Floor Area is not applicable for Educational building projects in Ethiopia.

Recently, in Ethiopia Dawit [22] examined study on model for estimating construction duration of public building projects in Addis Ababa. The study collected secondary data from Addis Ababa City Administration Construction Bureau and Federal Government Buildings Construction Project Office for 34 projects completed between the period of 2010 to 2019.After data collection, he carried out adjustment for cost due to effects of inflation using base year of 2018.Then, the study developed BTC model for construction time prediction. Moreover, alternative model with a better statistical significance was developed with gross floor area as a sole predictor variable. Finally, He concluded that the adjusted R^2 for the BTC was 0.397 while that of the developed model was 0.425 which implies the developed model was a better fit for the data sets.

Table 2. 4: Summary of previous studies based on BTC to estimate construction time.

Authors	Country	Year	Time- Cost model	Constants		Types construction
				K	B	
Bromilow	Australia	1969	$T=350C^{0.3}$	350	0.3	Overall Public projects
Bromilow	Australia	1974	$T=313C^{0.3}$	313	0.3	370 overall building projects
Ireland	Australia	1985	$T=219C^{0.47}$	219	0.47	25 high rise commercial building projects
Kaka and Price	UK	1991	$T=216.3C^{0.253}$	216.3	0.25	37 Total Public projects
			$T=291.4C^{0.205}$	291.4	0.21	38 Total Civil engineering projects
Ogunsemi and Jagboro	Nigeria	2006	$T=63C^{0.262}$	63	0.26	32 Private and 55 public building projects
Hoffman et al		2007	$T=26.8\ C^{0.202}$	26.8	0.2	856 Facility projects
Andinet	Ethiopia	2008	$T=7.06C^{0.47}$	7.06	0.47	29 public educational building projects
Ameyaw and Mensah	Ghana	2012	$T=344.586C^{0.684}$	344.59	0.68	13 Offices projects
			$T=512.2C^{0.463}$	512.2	0.46	27 Residential projects
			$T=2.807C^{0.399}$	2.807	0.4	22 Class room projects
			$T=3.17C^{0.378}$	3.17	0.38	62 Combined overall projects
Mackova and Baskova	Slovakia	2014	$T=386C^{0.295}$	386	0.3	28 Residential projects
Jarkas	State of Kuwait	2015	$T=28.79C^{0.192}$	28.79	0.19	113 Residential building projects
			$T=7.92C^{0.277}$	7.92	0.28	74 Office building projects
Bayram	Turkey	2016	$T=209C^{0.353}$	209	0.35	530 Public building projects
Dawit	Ethiopia	2020	$T=10.48C^{0.252}$	10.48	0.252	34 Public building projects

2.2.4 Limitation of BTC model

The BTC model is the most well noted pioneer prediction model that has been cited frequently in the last few decades with regard to predicting the construction time at the initial stage of a project. Several studies, however, pointed out its shortcomings which are summarized below:

1.The BTC model is a simple regression model with only one independent variable, namely, construction cost, and does not consider other factors such as gross floor area, number of stories above the ground or below the ground, when establishing the construction time [54].

2. The model is limited to being used in the planning phase of the project when the information available is very basic, such as the use of the building. Therefore, it is not possible to accurately predict the construction cost [11] and the resulting error adversely affects the predictive ability of the BTC model.

3. The use of cost as a predictor has a limitation in terms of time value of money (the change in cost for the same quantum of work over a period of time) [55].

2.2.5 Previous studies on multiple regression models

Before the report of previously multiple regression model in this section the concept behind multiple regression highlighted.

The multiple linear regression equation consists of independent variables (explanatory variables) and a dependent variable (response variable). The relationship between the dependent and independent variables is expressed by [Eq.2.7] as below [56]:

$$Y = f(X1; X2; ... ;X\rho) + \varepsilon \qquad \text{...............[Eq.2.7]}$$

[Eq.2.7] can be re-written as;

$$Y = \beta 0 + \beta 1\ X1 + \beta 2\ X2 + ... + \beta\rho\ X\rho + \varepsilon \qquad \text{...............[Eq.2.8]}$$

In the above [Eq.2.7], Y represents the dependent variable, X1, X2, . . ., Xp represents the set of independent variables, p is the number of independent variables, and ε refers to the probable error, which shows the deviation of the data from the approximation. The probable error denotes the degree to which a model does not fit the given data. The function f (X1, X2, . . ., Xp) describes the relationship between Y and X1, X2, . . ., Xp. And, also in the above [Eq.2.8], β0, β1, . . ., βp are called regression parameters or coefficients, representing the known constants that can be estimated from the given data set.

Realizing the importance of multiple regression modelling several Authors have been developed models for predicting time duration of construction project using project scope factors as independent variables. Thus, it is stated in the above section Love [11] suggested that GFA and F are one of the determinant variable to predict project duration.

Yeom,Seo,Kim, CHO and Kim [57] carried out further research to develop duration prediction for general office building project .They collected data comprised of 47 cases in total that are made up of 22 cases of offices and 25 cases of efficiency apartments, and except for one case (began construction in 1995), all were constructions that began and were completed after 2000 (between 2002–2013).The collected data was based on literature identified influential factor impacting time duration. As a result, the collected data were analyzed through a multiple linear regression model using SPSS 18.0. They came up with the result that independent variables that are strongly correlated with the construction duration were derived to be total floor area, number of floors above the ground, and number of basement floors with which there is a strong linear relationship. Using multiple linear regression and ANOVA analyses, an approximate construction duration prediction model was derived with high explanatory power (R^2= 0.905, adjusted R^2 = 0.882) and statistical significance. The developed model was a multiple linear regression model with excellent applicability to the end user's prediction of an approximate construction duration. Independent variables, such as the total floor area, building area, number of floors above the ground, number of basement floors, RC, SRC, and winter season were incorporated.

A recent survey Kim and Yeom [9] conducted empirical duration prediction model in Korea. They collected a performance data of 1969 mixed-use buildings. They developed a models through step wise selection method based on literature reviewed influential factors. Consequently, the gross floor area, number of stories below ground, number of stories above ground, building area, multi-family housing, and the Chungcheong region were selected as the independent variables for Prediction Model I. When six identified outliers (0.51% of the total data) were deleted, the model showed an R^2 of 0.605, Ra^2 of 0.603 with the Durbin-Watson value being 1.931, and it was observed that there was no autocorrelation of the residuals. and then they transformed the quantitative independent variables to obtain the highest values of r with regard to the dependent variable and found that ln (building area), ln (gross floor area), (number of stories above ground)2, and number of stories below ground had the strongest relationship with the construction duration to develop model II. Through a stepwise selection method, they chose the ln (gross floor area), number of stories below ground, (number of stories above ground)2, ln (building area), the Chungcheon region, and multi-family housing as the independent variables for the Prediction Model II. Between the two variables that caused multi-collinearity, the ln (building area) was omitted as it had a lower correlation with the dependent variable. When 23 identified outliers were removed (20 first and then three), the Prediction Model II showed an R^2 of 0.637, Ra^2 of 0.635, and no evidence of any violation of the assumptions of multiple linear regression analysis. As a result, they concluded that Prediction Model II was superior as it had higher values of R^2 and Ra^2. Furthermore, they used a holdout sample of 787 project cases, to examine the accuracy and validity of the final prediction model. The results showed that the final model had lower values of RMSE and MAPE than those of Prediction Model I with its average prediction accuracy being 88.51%.

Table 2. 5: Summary of previous studies on multiple regression models to estimate construction.

Authors	Country	Year	Multiple regression model	Types construction
Love	Australia	2005	Log (T) = 3.178 + 0.274 × Log (GFA) + 0.142 × Log (F)	161 Building projects
Jarkas	State of Kuwait	2015	T =338.70 +0.00574CA +13.95NFAG + 27.58NFBG	113 Residential building projects
			T =453.97 + 0.0103CA +1.92NFAG + 58.54NFBG	74 Office building projects
Bayram	Turkey	2016	$T = 32.51 \times GFA^{0.246} \times F^{0.318}$	530 Public building projects
Yeom et al	Korea	2018	T =897.272+0.004 x Total Floor Area 0.003x Building Area + 9.896 x Number of Above Ground +40.646 x Floors Number of Basement Floors – 437.319 x SRC - 578.234x RC -110 *110.37 x Winter	47 Office building projects
Kim	China	2019	T= -571.777 + 105.34 x In gross floor area + 0.319 x No of stories above ground + 38.969 x No of stories below ground 28.042 x Chucheong Region + 15.539 multi-family Housing. Values of Chucheong Region and multi-family Housing are 1 if it is constructed there otherwise it is 0.	1,969 Mixed-use building projects

2.2.6 Previous studies on construction time modelling using Machine Learning

Algorithm

Traditionally, analysts generally employ statistical techniques to discover patterns or knowledge from collected data. Nevertheless, the large sets of data in the construction industry are far from fully utilized by applying statistical methods. However, the recent developments in the field of science and technology has led to many data driven software and other innovations in the construction field [58]. Every aspect of the construction is getting adapted to new software, whether it might be scheduling, estimation, safety analysis or mitigating risks [58]. Use of machine learning helps to analyze the data of the construction works, which would further provide effective insights and guides the engineers/project managers during decision making phases. Time, cost, quality, safety, operations and maintenance, and many other aspects of project management have adopted machine learning (ML) algorithms [59].

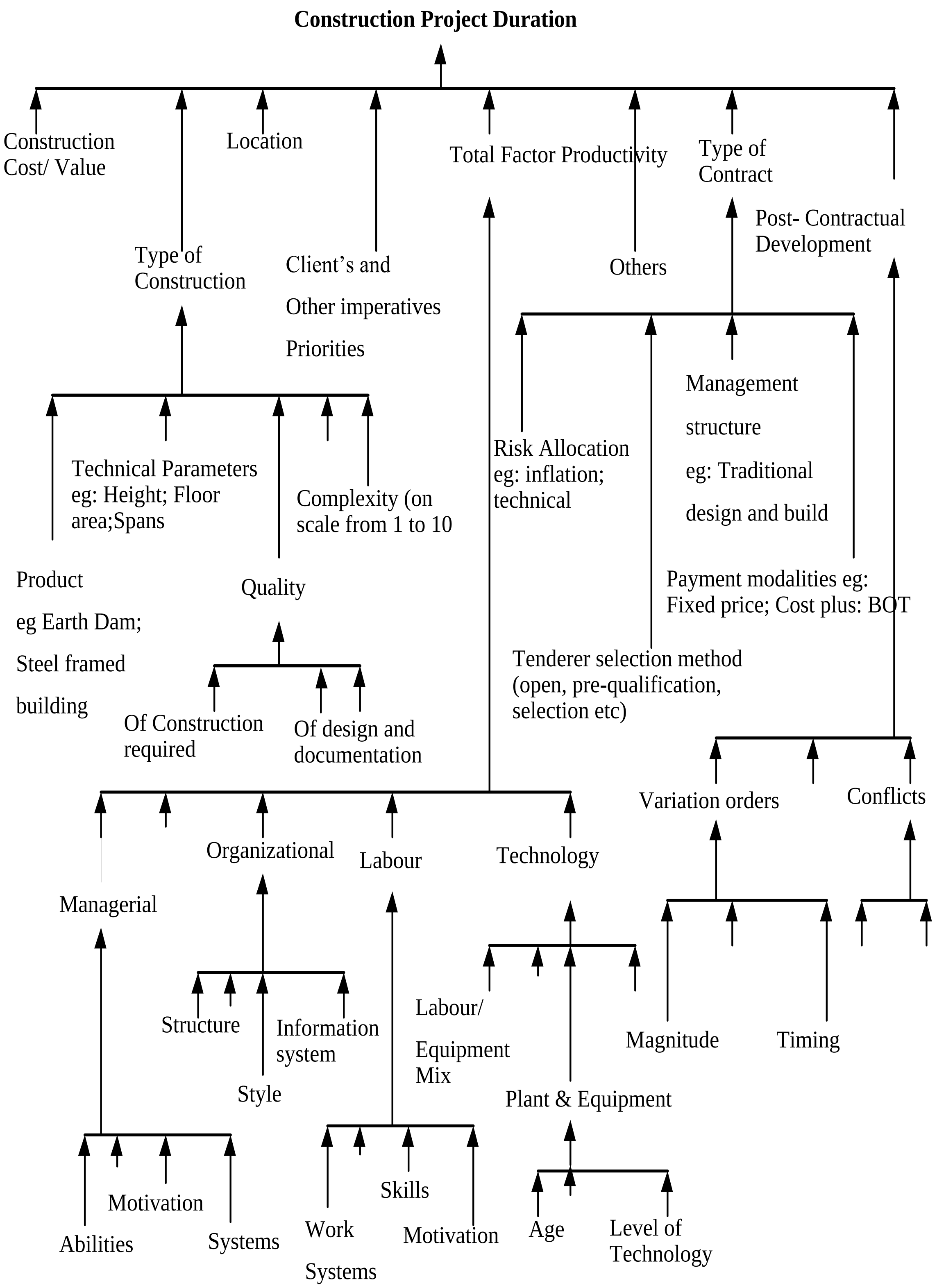
There are four different types of algorithms that can be used to project management areas for planning and estimation; they are classification, regression, association, and clustering[60]. Both classification and regression are used as predictive models where classification has categorical outputs and regression has continuous variable outputs. Association is used for finding the relationships between the variables and clustering is a segmentation process where the similar variables are grouped [60].

In works related to construction, a lot of attempts has been made to adopt machine learning techniques. For instance, in 1997,[61] implemented machine learning algorithms (neural networks and regression models) for performing cost estimation to low rise buildings. This was one of the initiatives taken towards the emergence of automation in the field of construction. Hammad et.al [62] in their paper tried to build a predictive model for determining the duration of a steel fabrication using ML. The results turned out to be more accurate than the existing traditional practice. Furthermore, Petruseva, Zujo & Pancovska [63] reported neural network prediction model of 75 buildings constructed in the Federation of Bosnia and Herzegovina shows significant improvement prediction accuracy over linear regression for construction project duration. They found out the value of coefficient determination (R^2) = 0.96989, and mean absolute percentage error (MAPE) = 2.4984055, using multilayer perceptron neural network (MLP - NN).

To summarize, it can be said that poor estimation leads to unrealistic duration, improper budget usage and sub-standard quality in the overall works and badly affects the performance of the project in terms of cost and time. As a result, it leads stakeholders towards disputes, unnecessary cost and eventually to bankruptcy. In order to provide valuable solutions, this paper aims to develop a mathematical optimization model for construction time predictions in future use. In addition, this approach could be implemented in the planning stages of any private mixed use and public office building projects, so that the major stakeholders get a standardized method to decide construction time required for private mixed use and public office building projects.

2.2.7 Others Factors influencing Contract duration

There are other factors affecting project duration apart from cost and there have been studies on some of these factors worldwide. According to Chan [45] a range of significant factors influencing the duration of a construction project are postulated hierarchically as illustrated in Figure 2.8, based on the general international literature, observed common construction practice and the survey results. These factors include both qualitative and quantitative contributors. Construction time can be considered to be a function of all such primary, secondary and tertiary factors in the hierarchy of determinants of construction project duration, so construction time = f (factors from Figure 2.7. hierarchy). Primary factors include construction cost, type of construction, location, stakeholder's priorities, productivity, type of contract and post contractual developments. Looking at for example post contractual developments at the secondary level, factors like variation, and conflicts will be noticed. Detailing the variation factor even further will result in the tertiary factors like magnitude, interference level and timing of the variation. However, they did not identify a conclusive detail, productivity and factors affecting it have been determined as key factors that can affect duration of construction projects.



Note : Unlabeled arrows indicate other factors also contribute

- Cost/ value is in turn affected by all others factors listed: whereas some others also interact to varying degrees.

Figure 2. 7: Some factors affecting construction project duration(Chan et al,1995).

Additional study by, Wijaya [64] identified factors that were potential to influence construction time performance from different literatures .The factors were grouped into seven main categories. They were labors, materials, equipment's, site characteristics, managerial, financial, and other factors as illustrated on Figure 2.8.

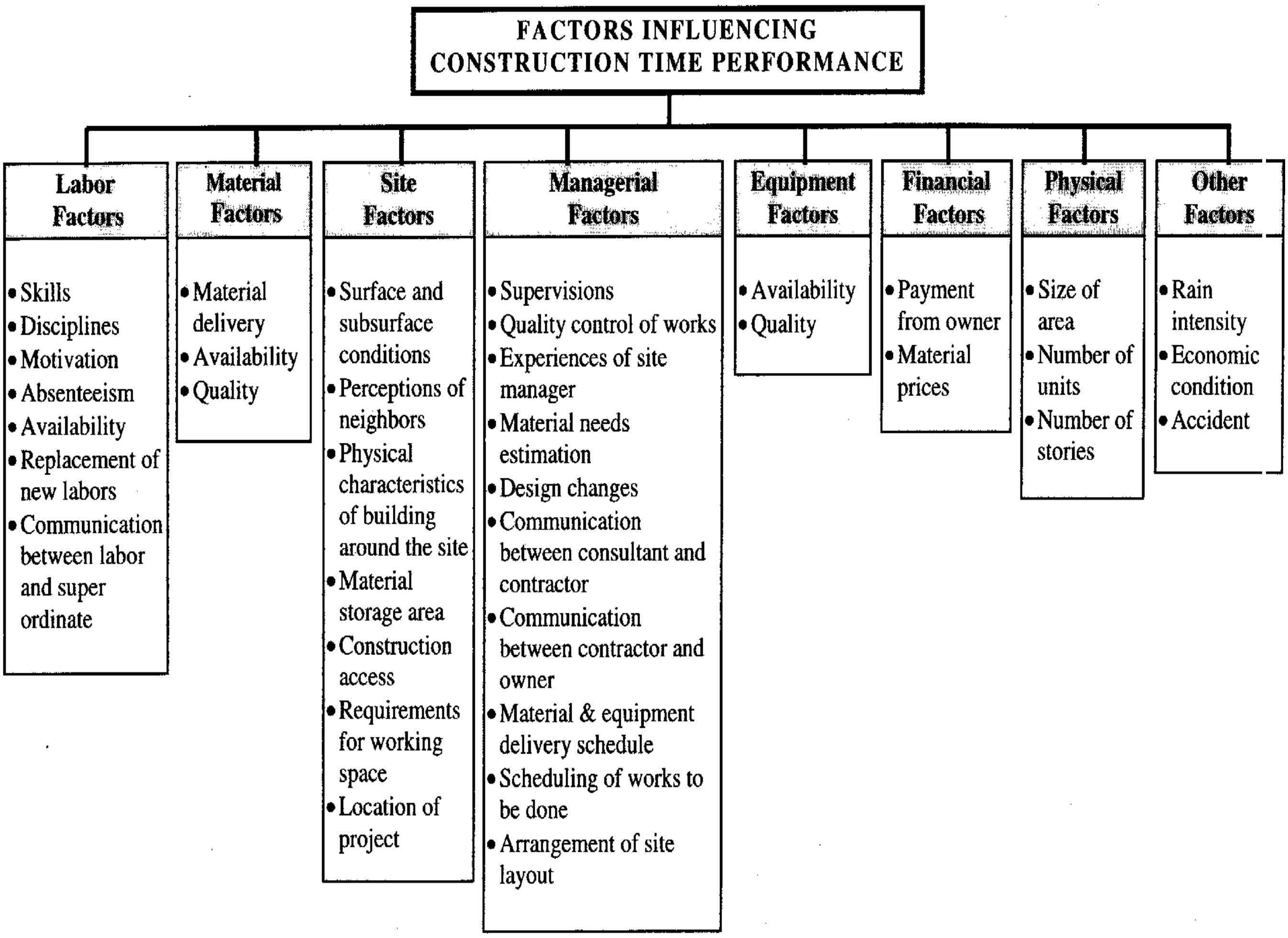


Figure 2. 8: Potential Factors Influencing Time Performance of Shop-Houses Constructions (Wijaya, 2003)

In Ethiopia, several studies were conducted on factors influencing construction time. One of the prominent research by Sinasilassie,Shahid and Jha [65], on Critical factors affecting schedule performance: a case of Ethiopian public construction projects – engineers' perspective and found out 'conflict among project participants', 'poor human resource management' and 'project manager's ignorance and lack of knowledge' are detrimental to schedule performance of Ethiopian public construction project .

Further study, by Abdurezak and Neway [66] on the Causes of Delay in Public Building Construction Projects: A Case of Addis Ababa Administration, Ethiopia was undertaken using questionnaire survey. The study find revealed that the top ten factors that cause construction delays in the public building construction projects in Addis Ababa are: the top ten factors that cause construction delays in the public building construction projects in Addis Ababa are: (1)Difficulty in project financing (poor financial system); (2) Poor Project management system;(3) Delay in issuance of designs and working drawings; (4) Shortage of availability of imported construction materials and goods on market; (5) Design errors and complexity of designs;(6) Delay in progress payments for completed works; (7) Late start & resource mobilization to site; (8) Financing problems; (9) Inaccurate Site investigation Report; and (10) Price Inflation.

Although, several factors affect duration of construction projects, for this study only project scope is used to prove or disprove previously developed model, try to develop an alternatives model to suggest as determinant of project duration in the planning phase.

2.2.8 Effect of time on cost data of projects.

This research relies on secondary cost data collected for private mixed use completed under grade one consultants and for public office building projects completed under management of Addis Ababa City Administration Construction Bureau (AACACB) and Federal Government Buildings Construction Project Office (FGBCPO). Besides, since the time data values range from 2010 to 2020, necessary price adjustments are required to be carried out.

Estimates for construction work are determined at a specific point in time and the prices used therein are (unless other parameters are specifically set) relevant only for that date. This is because prices for items supplied and work undertaken are continually subject to market forces. These forces arise from two main directions: Inflation (and potentially and alternatively, ‘deflation’) and the ever-changing relationship between supply and demand for construction in the market place.

Inflation in construction acts differently than consumer inflation. When there is more work available, inflation increases. When work is scarce, inflation declines. The level of construction activity has a direct influence on labor and material demand and margins and therefore on construction inflation. In active markets overhead and profit margins increase in response to increased demand [67].

2.2.5.1 Adjustment for inflation

Construction costs are volatile, and the prices of materials and other costs are continuously fluctuating, creating volatility in economic growth. Not only economic growth but also the labour market and consumer price index are affected by inflation [68].

Windapo and Cattell [69] presented a conceptual model of the factors influencing the price of building materials as shown in Fig.2.9. They reported that there are many factors involved in deviating the material costs mainly; supply and demand, market conditions, transport and energy costs, raw materials, labour costs, crude oil prices, exchange rates, import duties and inflation that influence construction costs as well as the output of the construction industry. From these it is clear that cost of construction project is mainly depend on the price of materials with in the implementation period. Therefore, inflation is one of important techniques of comparing cost of construction projects completed at the different period of time.

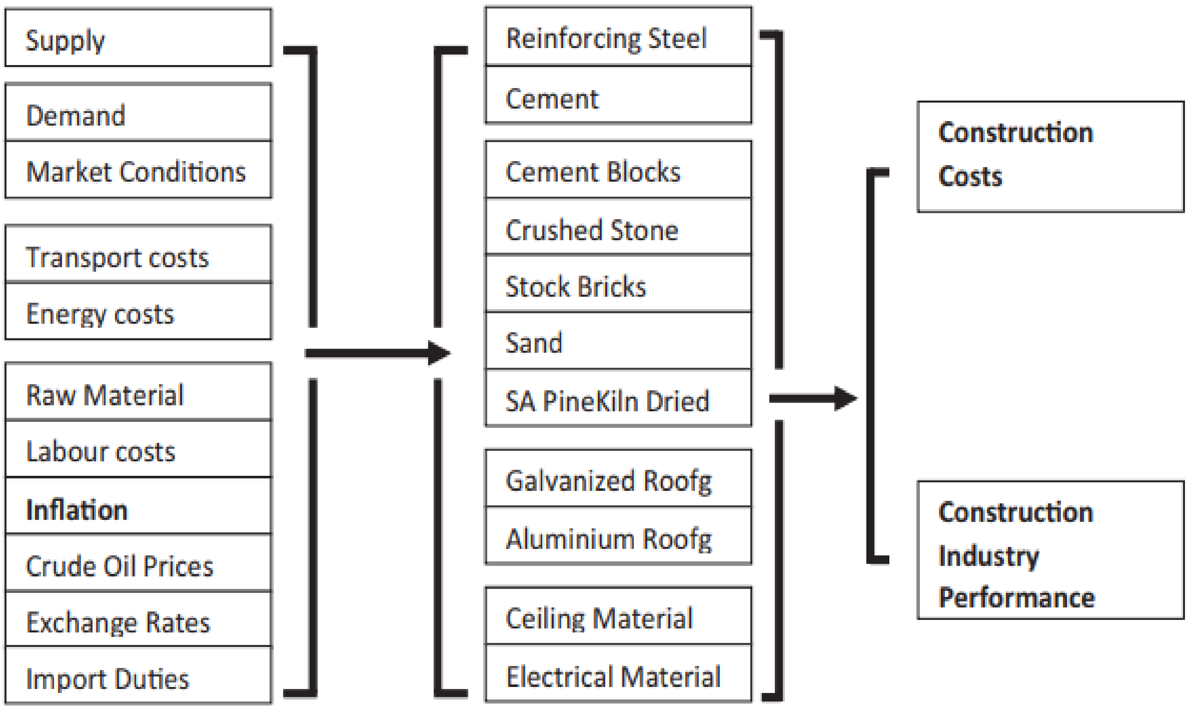


Figure 2. 9:Conceptual Model of the Factors Affecting Building Material Prices(Windapo and Cattell,2012)

In Ethiopia, raw inflation figures are reported monthly using the Consumer Price Index (CPI) by the Central Statistical Agency. The CPI is an estimation of the price changes for a typical basket of goods. In other words, the prices of everyday goods such as housing, food, education, clothing, etc., are compared from one month to the next and the difference represents the CPI. The CPI published by CSA composed of the weighted average of two sub-indexes that reflect the development of prices of goods production in certain sectors of economy, namely, food and nonfood prices. Therefore, for this research purpose CPI of non-food item is used to bring the cost of construction projects completed at the different period of time to specific period of time.

Table 2. 6: Ethiopian Inflation Data (Central Statistical Agency of Ethiopia).

Year	Inflation (%)
2010	67.5
2011	87
2012	111.7
2013	105.9
2014	83.2
2015	92.7
2016	98.9
2017	109.1
2018	124.2
2019	143.9
2020	173.1

2.2.9 Construction industry development in Ethiopia

According to a report of Ethiopian Economic association (2007), the progress of contemporary construction industry in Ethiopia is a current phenomenon and can generally be concluded into four different periods. The first period covers the period prior to the year 1968 when most civil works (including roads) were carried out by foreign contractors through international competitive bids. Relevant skilled manpower was also largely employed from abroad.

The second period in the growth of the construction sector in Ethiopia covers between 1968 - 1982 when some small local contractors started to emerge. In order to build capacity and improve their competitiveness, the government took initiatives to assist contractors to involve in the construction of feeder road projects. Following initiatives, three local contractors: National Engineers and Contractors (NEC), Ethiopian building and road construction (ETBRC) and the BERTA Construction Company [70].

The third era in the growth of the industry was the period of the Derg regime which had created the then evolving local private construction organization under government control in 1982. In addition, state-owned construction companies were established. It was regarded as the lost opportunity for the launching of a competitive construction industry in the country [70].

The fourth period begins from the time the EPRDF-led transitional government of Ethiopia took power in May 1991. The fourth era begins from the period the EPRDF-directed transitional government of Ethiopia took authority in May 1991. Economic administration has changed from command to a free market system and several reform action focused at encouraging the private sector including private construction companies have been introduced. As a result, the role of private contractors in the industry has commenced expansion while that of public companies weakened since 1991[70].

Furthermore, the evidence shows that the construction industry has been overgrowing over the past decade, especially after the country's growth and transformation plans have been launched. In the first phase of the Growth and Transformation Plan (GTPI) from 2010/11 to 2014/2015, the government announced and implemented a fundamental construction industry structure to enhance the industry's capabilities and role. This reform includes a policy framework for the construction industry, the establishment of the Ethiopian Construction Project Management Institute and the preparation and implementation of a strategic reform plan for the construction industry [71].

The construction industry in Ethiopia is characterized by a large number of micro-entrepreneurs, most of them operating in the country's informal economy. Although a large number of local contractors appeared after the proposed reforms were implemented, and foreign private construction companies also joined the sector. It is well-appreciated that the launch of the national growth and transformation plan (GTP) boosting the construction sector (i.e. 4.5% (2009/2010) share in GDP to 8.5% in 2014/15) [71]. The entire government agencies, construction companies, various professional associations, consulting companies, etc. that are committed to improving the living standards of the society by expanding the physical and social infrastructure are the main stakeholders of the Ethiopian construction industry and GTP. The municipalities, Ethiopian Road Authority, Ministry of Energy and Water Resources, Engineering and Urban Development Authority are specially included. In this study, Addis Ababa construction bureau and private or corporate developers are the stake in construction of public office and private mixed use building projects respectively to expand physical infrastructures for significant economic development.

According to Mengistu et al., [72] report the major challenges in developing the Ethiopian construction industry were identified as: (i) delay in construction industry development (CID) policy implementation and corruption; (ii) weak capacity of contractors and consultants; (iii) lack of collaboration and professionalism; and (iv) lack of benchmarking CID practice from role of government, resource related variables, nature of the industry and industry's vision for its own development, respectively. Therefore, this report is strong input in order to ensure sustainable growth of construction industry by government through collaboration with the rest stakeholders of construction industry.

To summarize Ethiopian Economic Association report showed that the number of local contractors involved in the industry across the periods increased which shows a positive change and enables to meet the society need. And also, inviting foreign contractors to operate in our country helps our professional to obtain their experience in terms of discipline, usage of technology, planning techniques, scheduling techniques and overall management principles. Further, this report announced that diminish in state owned construction companies during the fourth period results reduce competition and open the space of winning the bid among the competent professional contractors and create rich individuals. The change in Economic administration from command to free market system loosen the control of government on private construction organization. Although various reform action was carried out during the fourth period of construction sector development construction time overrun is not avoided. Therefore, this report failed to identify major impediments to complete project on schedule, within budget and specified quality which is fundamental to forward scientific project management principles and policy to meet project requirements and stakeholder's expectation.

2.2.10 Critical Appraisal of Literature Review

This chapter has reviewed studies on time – cost models, time-floor models for the construction industry worldwide as well as some factors affecting the duration of projects. After the development of the BTC model researchers were confirmed applicability of this model. Whereas, validated the poor predictive ability of Bromillow's time-cost.

Construction delay is a worldwide concern in construction industry. Therefore, appropriate construction project duration estimation is a basis for good plan and also a key for successful management during project execution phases. In Ethiopia Andinet et al.'s [52] reported construction time prediction model for educational building projects based on BTC as a base line. Moreover, Muluken [53] presented additional study to check applicability Bromillow's time -cost and tried to check whether gross floor area predict duration of educational building construction projects. Both study carried out by Andinet et al.'s and Muluken were limited to educational building projects which overlook another projects. In addition, the developed model took long time without updated which fails to consider current knowledge of construction project management and technologically advancement in the industry.

Recently in 2020 dawit reported BTC model applicability for public building construction projects using traditional statistically analysis. However, dawit study didn't cover applicability of LTF model for public office building projects and construction duration estimation for private mixed use building projects. Despite several authors developed construction time using machine learning algorithm internationally there is no any attempt by local researchers. Therefore, this study tries to develop construction time using machine learning algorithm.

Nevertheless, the recorded success of applying the time-cost model to estimate project durations in specific geographical settings, many studies reported weak prediction ability of the model in others. Furthermore, due to the differences in socio-economic and political conditions, construction business environment, weather and geophysical conditions, and technological advancements between one country and another, it becomes vital to validate the model before it can be successfully applied within a particular environment [73].

Proper utilization of this construction time prediction model is basis for all stakeholders engaged in the project to complete project on time, with in approved budget and specified quality by minimizing claims, risks, conflict, cancellation and termination. The benefit of successfully accomplishing projects create stakeholders competent, liquid and reputable in the construction sector. Therefore, this study come up with construction time best prediction plat form in construction industry.

CHAPTER 3 RESEARCH METHODOLOGY

3.1. **Introduction**

The research methods include all techniques/methods that are adopted for conducting research. Thus research techniques or methods are the methods the researchers adopt for conducting the research operations. On the other hand, research methodology is the way of systematically solving the research problem. It is a science of studying how research is conducted scientifically. Under it, the researcher acquaints himself/here self with the various steps adopted to study a research problem, along with the underlying logic behind them. Hence, it is not only important for the researcher to know the research method/techniques, but also scientific approach called methodology.

3.2 Definition of Research

Research refers to a careful, well-defined (or redefined), objective, and systematic method of search for knowledge, or formulation of a theory that is driven by inquisitiveness for that which is unknown and useful on a particular aspect so as to make an original contribution to expand the existing knowledge base. Research involves formulation of hypothesis or proposition of solutions, data analysis, and deductions; and ascertaining whether the conclusions fit the hypothesis. Research is a process of creating, or formulating knowledge that does not yet exist [74]. Although, research have several definitions above stated definition is suitable for this study because the outcome of this research is based on proving or disproving hypothesis.

3.3 Types of Research

Despite, types of research are written by several authors their concept are almost nearly the same. Hence, According to Deb,Dey and Balas [74] the different types of research are;

I. Descriptive versus Analytical**:** Descriptive research includes comparative and correlational methods, and fact-finding inquiries, to effectively describe the present state of art. The researcher holds no control over the variables; rather only reports as it is. Descriptive research also includes attempts to determine causes even though the variables cannot be controlled. On the contrary, in analytical research, already available facts for analysis and critical evaluation are utilized. Some research studies can be both descriptive and analytical.

II. Applied versus Fundamental**:** Research can either be applied research or fundamental (basic or pure) research. Applied research seeks to solve an immediate problem facing the organization, whereas fundamental research is concerned with generalizations and formulation of a theory.

III. Quantitative versus Qualitative: Quantitative research uses statistical observations of a sufficiently large number of representative cases to draw any conclusions, while qualitative researchers rely on a few non representative cases or verbal narrative in behavioral studies such as clustering effect in intersections in Transportation engineering to make a proposition.

This research is descriptive, correlation, applied and quantitative research. Thus, it clearly describe and prove or dis-prove the relationship between time and cost, time and gross floor area, number of floors. The outcome of this research is applied in the respective study areas after developed model validated using statistical analysis.

3.4 Research process

The Research process consists of series of actions or steps necessary to effectively carry out research and the desired sequencing of these steps [75].

The research process is therefore concerned with collecting data and processing it into information. People can use the information thus created to add to their knowledge, perhaps even developing wisdom.

The Figure 3.1 shows that the research process having a number of closely related actions, as shown from step I to VII. But these activities should be following in a strictly prescribed sequence otherwise researcher may face the problem in completion of the research. In the research process, each step is specific and they are separate and distinct from each other.

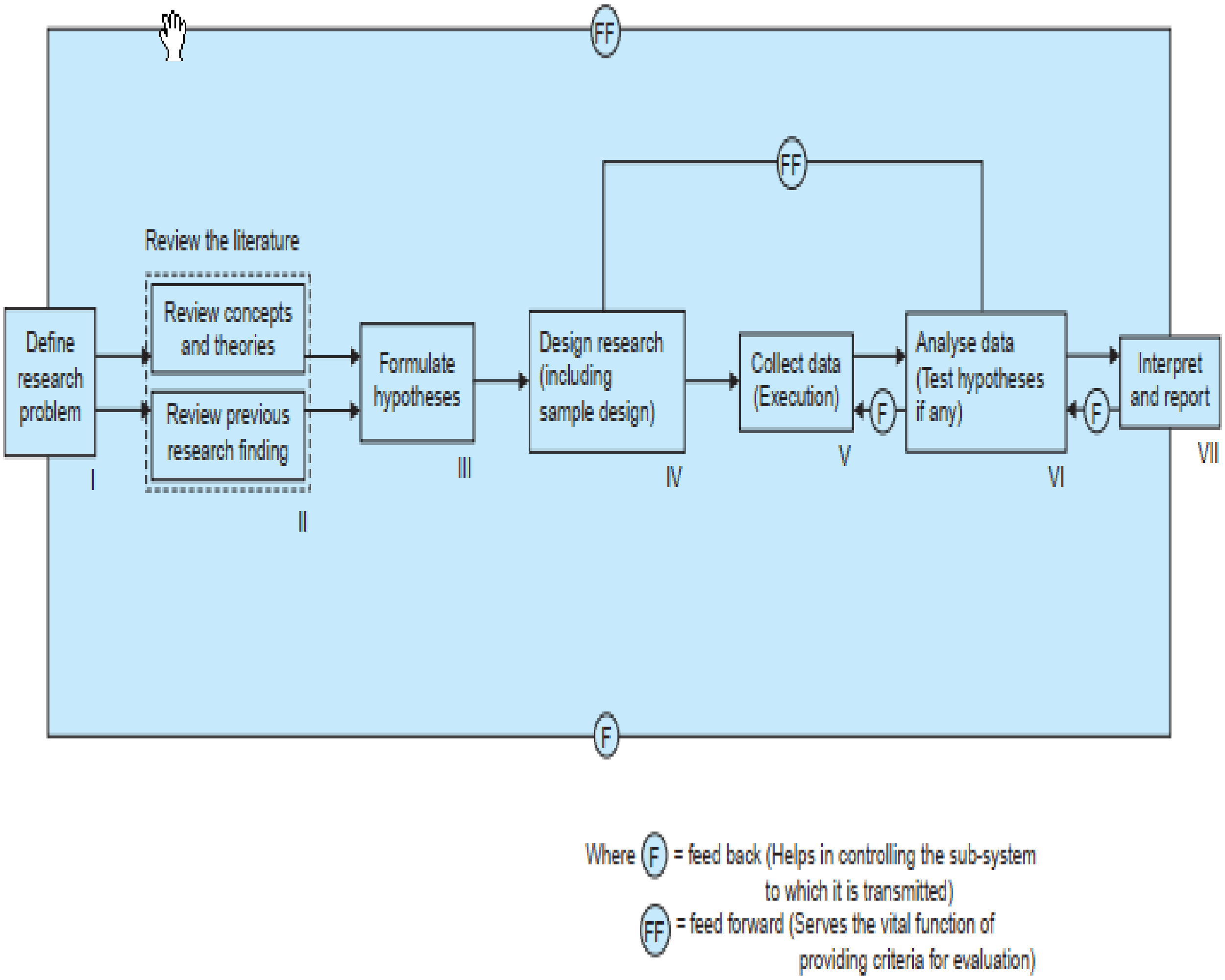


Figure 3. 1:Research Process (Kothari C.R.,1990)

As shown in the Figure 3.2 the research process used in this study are as follows: Based on investigation of previous studies potential variables were identified to determine construction duration estimation. Before model development outliers were eliminated from data sets to minimize effect of outliers on prediction ability of developed models. Afterwards, both statistical and machine learning tool were used to develop the models and their process were discussed under the method of data analysis section.

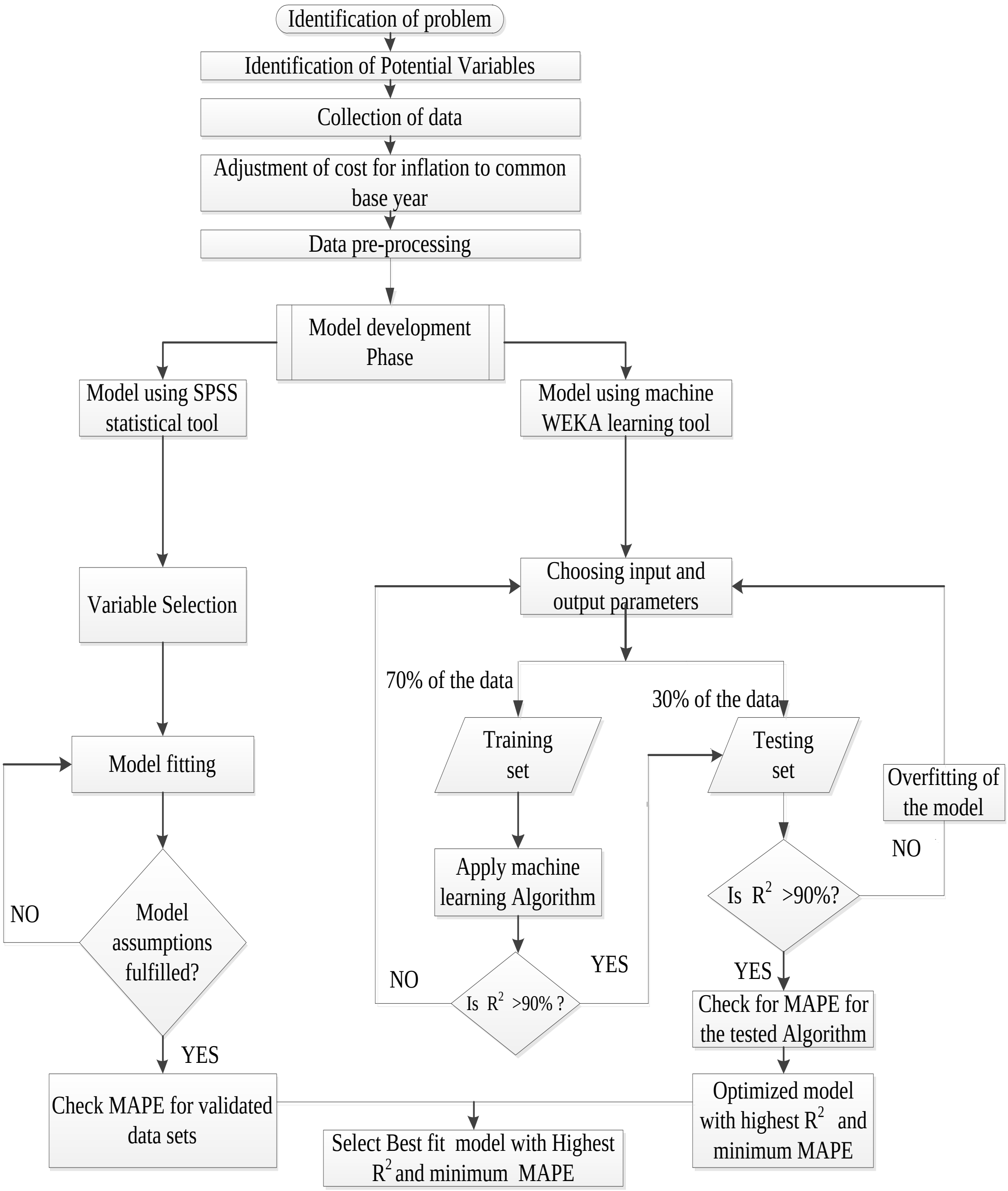


Figure 3. 2:Research process adopted for this Study

3.5 Research Hypothesis

Hypothesis may be defined as a proposition or a set of propositions set forth as an explanation for the occurrence of some specified group of phenomenon either asserted merely as a provisional conjecture to guide some investigation or accepted as highly probable in the fact of established facts [75].

Further research hypothesis is a proposed explanation of an observable phenomenon that can be tested via research. Alternatively stated, a hypothesis is a declarative statement wherein the researcher-typically in quantitative research studies or the quantitative phase(s) of mixed research studies makes a prediction or judgment about the relationship that exists among the variables of interest. As stated by Johnson and Christensen [76], “the stated hypothesis typically emerges from the literature review or from theory”.

Hence, a prediction or a hypothesized relationship is tested by adopting scientific methods. It is research hypothesis. The research hypothesis is a predictive statement, which relates to dependent variable and independent variable. Generally, a research hypothesis must consist of at least one dependent variable and one independent variable. Whereas the relationships that are assumed but not be tested are predictive, statements that are not objectively verified are not classified as research hypothesis. Therefore, the hypotheses of this research are:

1. To verify the validity and applicability of Bromillow’s Time-Cost model and Love et al.’s Time-Floor for private mixed use and public office building projects in Addis Ababa.
2. To develop an alternative construction time prediction model based on literature identified independent variables for private mixed use and public office building projects in Addis Ababa.

3.6 The Research design

The most important problem after defining the research problem is preparing the design of the research project. Which is properly known as the research design. A research design helps to decide up on issues like what, when and where, by what much, by what means, etc with regard to an enquiry or a research study research design is the arrangement of conditions for collection and analysis of data in a manner that aims to combine relevance to the research purpose with economy in procedure.

In fact, the research design is the conceptual structure within which research is conducted; it constitutes the blueprint for the collection, measurement and analysis of data. As such the design includes an outline of what the researcher will do from writing the hypothesis and its operational implications to the final analysis of data.

According to Kothari [75] the research design decisions answers the below raised questions :

(i) What is the study about?

(ii) Why is the study being made?

(iii) Where will the study be carried out?

(iv) What type of data is required?

(v) Where can the required data be found?

(vi) What periods of time will the study include?

(vii) What will be the sample design?

(viii) What techniques of data collection will be used?

(ix) How will the data be analyzed?

(x) In what style will the report be prepared?

3.6.1 Selection of Research Method

A good research methodology will be prepared if a research problem should be stated clearly. In other words, the purpose of research methodology is refers as general procedure that you choose to combine the various components of the study in a consistent and logical way. It comprises the outline for the collection, measurement, and analysis of data. A flexible research methodology which offers the opportunity for allowing the different aspects of a problem is considered suitable if the purpose of the research study is to be clear. There are several research designs, such as, Descriptive (e.g., case-study, naturalistic observation, survey), Correlational (e.g., case-control study, observational study), Semi-experimental (e.g., field experiment, quasi-experiment), Experimental (experiment with random assignment), Review (literature review, systematic review) and Meta-analytic (meta-analysis).

This research paper has used a descriptive (case study) and correlational (observational study) to test the hypothesis based on 40 private mixed use and 31 public office building projects completed in the past ten years (2010-2020). This is mainly because the research is focused on data analysis to validate the applicability of Bromillow's Time-Cost model and Love et al.'s Time-Floor for private mixed use and public office building projects in Addis Ababa.

3.6.2 Sources and Collection of data

Kumar[77] has classified data sources into two Primary and Secondary. Primary data is collected from Primary sources like Observation, Interviewing and Questionnaire and Secondary data is collected from Secondary sources like documents: Government publications, earlier researches, and census and personal records. The sources of data for this thesis paper are contract documents, payment certificates on provisional acceptance, floor plans and final payment certificates. Therefore, data used for this research is considered as quantitative secondary data since data existed as a formal and binding agreement among stakeholders involved in each project.

The source of data for this study are Grade one consultants, Addis Ababa City Administration Construction Bureau (AACACB) and Federal Government Buildings Construction Project Office (FGBCPO), together with contractors who had actively acted as stake in execution of the project of each project used for this study. The data collected from stated source was checked whether it matches with contractors for some selected projects for reliability of the data before data analysis.

3.6.3 Sample size determination

Sampling may be defined as the selection of some part of an aggregate or totality on the basis of which a judgement or inference about the aggregate or totality is made. In other words, it is the process of obtaining information about an entire population by examining only a part of it. In most of the research work and surveys, the usual approach happens to be to make generalizations or to draw inferences based on samples about the parameters of population from which the samples are taken[75]**.** Identifying number of private mixed use building projects completed between the period of 2010 to 2020 is difficulty due to improper data handling by construction stakeholders. Moreover, recently completed projects were suitable for modelling due to their similarity in construction project management practice and uniformity in terms of technological utilization. Therefore, this study used purposive sampling to reach on representative sample size for modelling construction time of private mixed use buildings. In the case of public office building projects this study used all completed projects between the period of 2010 to 2020 due to limited number of projects completed within the period.

3.7 Method of data analysis

Appropriate method of data analysis was very necessary to be able to accurately process the data that was collected. For this research, both quantitative and graphical solutions are required, thus both graphical and quantitative data analysis was used. Thus, this research has used the quantitative hypothesis testing by using regression techniques and also the scatter and residual plot techniques. The analysis of this research has been carried out by using statistical software, IBM SPSS (Statistical Package for Social Studies), version 26 and WEKA 3.8.5 as data mining and machine learning tool.

3.7.1 Data Preprocessing

According to the data analysis process, the next step after data collection is data pre-processing. Generally, the data collected from the large databases will have errors. These errors might be in different forms - (i) incomplete data which has certain missing attributes values, contains aggregate data. (ii) Noisy data which has errors and outlier's values that deviates from the larger pattern of data. (iii) Inconsistent data which has discrepancies in the codes, negative or incorrect values. These kinds of errors are quite common in real world databases due to various reasons such as manual entries, multiple people handling the data, computer errors at time, technical limitations while transferring or merging data, incorrect format of entering the parameters and so on [78]. The data which has to be analyzed using data mining techniques should be in a clean and consolidated state. The processing of data involves multiple steps such as data cleaning, data integration, data transformation and data reduction. Data cleaning involves filling up the missing values, identifying the outlier's and removing them using statistical methods. Data integration is a process of merging data from multiple databases and data transforming is a process of normalising the data to have the uniformity in the distribution of values. Further, data reduction is removing unrelated parameters (by correlation and other feature selection methods) which does not add value to end result of data mining process [78]. In this research, the initial step of data pre-processing, that is data cleaning is done by removing the uncertainties such as missing/incorrect values, duplicated data entries, and removing the potential outliers with the help of statistical methods.

The data points in the dataset which do not lie in the general behaviour of the model are referred as outliers. They disturb the distribution pattern of any data set [78]. There are various methods of handling the outliers such as least square fitting (regression), standard deviation, interquartile range (IQR), and cook's distance and so on. In addition, there are visual methods such as scatter plots, boxplots and histograms help to analyze the behaviour of data and identify outliers. However, there is no standard rule that is followed to remove the outlier's. In this research, an iterative method IQR (Inter Quartile Range) is adopted to identify potential outliers.

Figure 3.3 shows the general representation of IQR ranges through the boxplot. To illustrate IQR, the box plot defines lower quartile as the 25th percentile, median as the middle point and upper quartile as the 75th percentile. The difference between the upper quartile (Q3) and the lower quartile (Q1), which spreads over half of the data is referred as IQR.

$$IQR = Q3 - Q1 \quad \text{...............[Eq.3.1]}$$

A point beyond the inner fence on either side is considered a mild outlier's. A point beyond the outer fence is considered an extreme outlier. A potential outlier is a data point that is 1.5 times the IQ range from the edge of the box [78]. IQR method is similar to the standard deviation; However, IQR range is advantageous because they are not affected by the extreme values of data points [79].

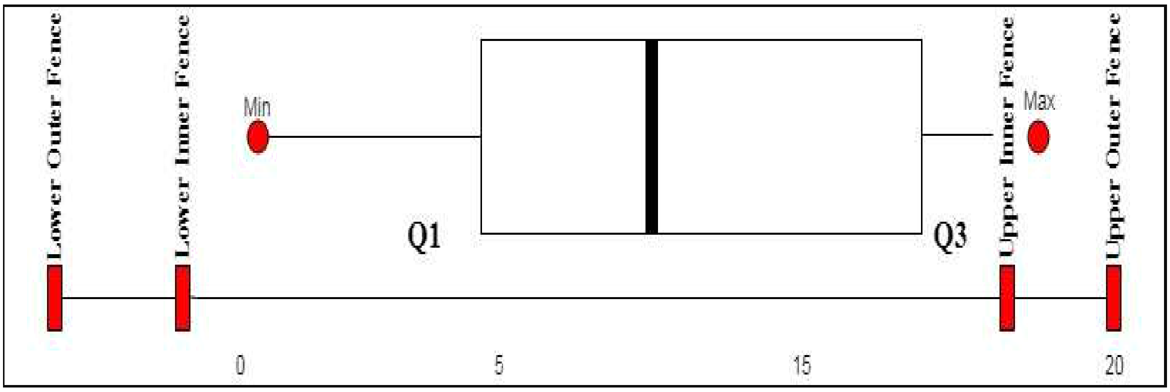


Figure 3. 3: Boxplot showing IQR ranges (Siddappa,2019)

Feature Selection: Feature selection can be described has a process of identifying and removing irrelevant and redundant information from any data set [80]. There are three different methods of identifying the variables impacting the output in predictive model building- filter methods, wrapper methods, and embedded methods. In order to reduce the data efficiently, a feature selection method can be used in the data pre-processing stage. This process helps in identifying accurate data models.

a) Filter methods – It uses statistical tests and measures, then assign a scoring for each feature or variable. The variables are ranked and based on the score they are either selected or removed from the dataset. Examples of filter based methods are Chi squared test, information gain and correlation coefficient scores [81]. The process is represented in Figure 3.4 and the advantage and disadvantages are shown in Table 3.1.

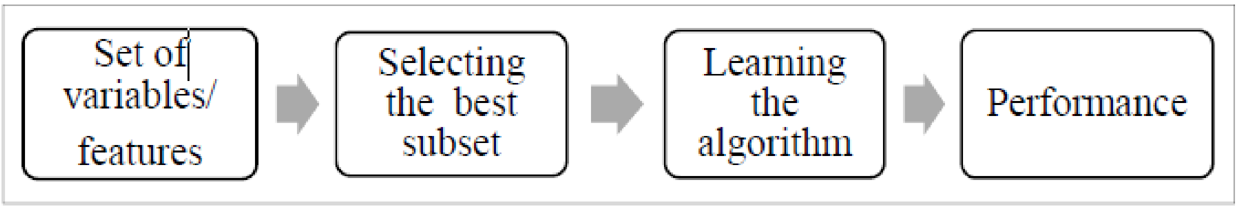


Figure 3. 4: General process of filter method feature processing (Siddappa,2019)

Table 3. 1: Advantages and disadvantages of filters methods (Aziz et al., 2016).

Model search	Advantages	Disadvantages	Examples
Filter	**Univariate**		
	Fast, Scalable Independent of the classifier	Ignores feature dependencies Some features which as a group have strong discriminatory power but are weak as individual features will be ignored Features are considered independently	χ^2 Euclidean distance t-test Information gain Gain ratio
	Multivariate		
	The models feature dependencies Independent of the classifier Better computational complexity than wrapper methods	Slower than univariate techniques Less scalable than univariate techniques Ignores interaction with the classifier Redundant features may be included	Correlation-based feature selection (CFS) Markov blanket filter (MBF) Fast correlation-based feature selection (FCBF)

b) Wrapper methods – A wrapper method will choose a particular feature subset based on how effectively a modelling algorithm performs, this is considered as a black box evaluator. Thus, a classifier performance is chosen to evaluate a subset for classification tasks and similarly a cluster algorithm's performance is considered to evaluate a subset for clustering. Examples of wrapper methods are random hill-climbing algorithm, Boruta algorithm, forward and backward propagation. The process is represented in Figure 3.5 and the advantage and disadvantage of both the methods are presented in Table 3.2 with example.

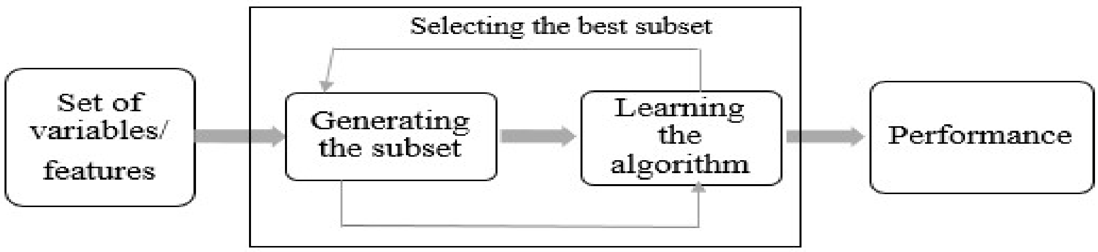


Figure 3. 5: General process of wrapper method feature processing (Siddappa,2019)

Table 3. 2: Advantages and disadvantages of wrapper methods (Aziz et al., 2016).

Model search	Advantages	Disadvantages	Examples
Wrapper	**Sequential selection algorithms**		
	Simple Interacts with the classifier Small overfitting risk Less computationally Prone to local optima Consider the dependence among features	Risk of over fitting More prone than randomized Algorithms for getting stuck in a local optimum (greedy search) Classifier dependent methods The solution is not optimal	Sequential forward selection (SFS) Sequential backward elimination (SBE) Plus q take away r Beam search
	Evolutionary selection algorithms		
	Less prone to local optima Interacts with the classifier Models feature dependencies Higher performance accuracy than filter	Computationally intensive Discriminative power Lower shorter training times Classifier dependent selection Higher risk of over-fitting than deterministic algorithms	Simulated annealing Randomized hill climbing Genetic algorithms Ant Colony Optimization Rough set methods Particle Swarm Optimization Artificial Bee Colony (ABC)

c) Embedded methods– These methods as the name goes are embedded in the algorithm either as its normal or extended functionality and perform the feature selection processes during modelling algorithm's execution [81]. Examples of embedded methods are stepwise regression, regularized trees. The process is represented in Figure 3.6 and the advantages and disadvantages are listed in Table 3.3.

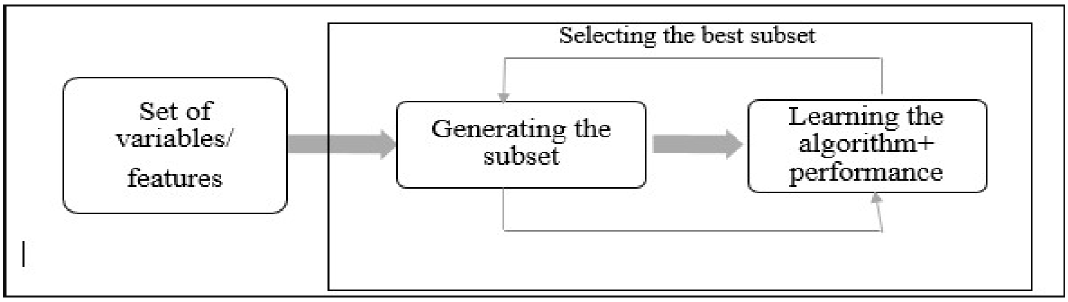


Figure 3. 6: General process of embedded method feature processing (Siddappa,2019)

Table 3. 3: Advantages and disadvantages of embedded methods (Aziz et al., 2016).

Model search	**Advantages**	**Disadvantages**	**Examples**
Embedded	Interacts with the classifier The models feature dependencies better computational complexity than the wrapper Higher performance, accuracy than filter Less prone to over-fitting than wrapper Preserving data characteristics for interpretability	Classifier dependent selection Consider the dependence among features	Decision trees Weighted naive Bayes Feature selection using the weight vector of SVM Random forests Least absolute shrinkage and selection operator (LASSO)

3.7.2 Linear Regression models

Regression analysis is a technique that finds a formula or mathematical model which best describes a set of data collected [56]. Moreover, Linear regression is an approach to model the relationship between a scalar dependent variable Y and one or more explanatory variables denoted x. The case of one explanatory variable is called simple linear regression. For more than one explanatory variable, it is called multiple linear regression.

In linear regression data are modeled using linear predictor functions and unknown model parameters are estimated from the data.

(a) Simple linear regression

This is represented by the mathematical formula

$Y = a + bX + e$ ....[Eq.3.2]

Where Y is the dependent variable

X is the independent variable, a and b are constants called regression parameters and

e is the error term.

3.5.1.1 Assessment of models

To choose the best out of the models generated in this study, the following assessment criteria were used.

(a). Correlation Coefficient (R)

The coefficient of correlation, r, identifies the degree and nature of the relationship between the two variables, from a perfect positive relationship (+1) to a perfect negative relationship (−1), that is $-1 \leq r \leq +1$. The coefficient implies, r =+1 means that an increase in variable x is matched by an equiproportional increase in y. If r = 0, there is no relationship; changes in the variables are quite independent of each other; they are random [82].

b) Coefficient of Determination (R^2)

The coefficient of determination (R^2),is statistic required to measure how much of the change in the values of a dependent variable is caused, given the logic of the relationship, by a change in the values of the independent variable [82].

The formula for calculating R^2 is

R^2= Regression sum of squares

$$R^2 = \frac{\text{Regression sum of squares}}{\text{Total sum of squares}} \qquad \text{.............[Eq.3.3]}$$

The values also range between zero and one ($0 \leq R^2 \leq 1$). A high value of R^2 indicates a good model. R^2 gives an indication of the predictive capability of the regression model using new observations [12,48]. It is usually expressed as a percentage by multiplying by 100.

(c) Significance of Regression (F- ratio)

This was used to determine the significance of a regression equation by testing whether the developed regression model is significant or not [83].

To ascertain the significant level,

$$\text{F- calculated} = \frac{\text{Regression mean square}}{\text{Regression mean square error}} \qquad \text{............[Eq.3.4]}$$

is compared with the critical value of F (F-tabulated) at 5% level of significance. Where F-calculated is greater than F-tabulated, the developed regression model is significant and suitable for prediction.

3.7.3 Machine Learning model (Artificial Neural Network (ANN))

Gurney [84] , a profound data scientist defines neural network as

"An interconnected assembly of simple processing elements, units or nodes, whose functionality is loosely based on animal neuron. The processing ability of the network is stored on interunit connections strengths or weights, obtained by the process of adaptation to, or learning from a set of training patterns."

In the world of growing machine learning algorithms, the artificial neural network is another such technique that adopts the work structure of a neuron in a human body. The neurons work based on the external stimuli (input) received and then by performing its activities (function) passes an end signal (output). The neural network is inspired by the functionality of neuron cells. Artificial neural networks is a machine learning algorithm, in which the system learns to perform some task by analyzing the information or the data received and performs a mathematical function to give desired output [85].The examples of artificial neural network includes image recognition, where the system finds images through visualization and check whether it correlates with particular labels generated. ANN is a set of connected neurons organized in three layers such as input layer, hidden layer and output layer. Input layer is the one where the initial data fed to the system to process the subsequent layers of artificial neurons. The hidden layer is the layer between input and output layers where artificial neurons take weighted inputs and activation function is set up. Then, the output layer is the one which provides the desired output. Unlike linear regression, neural networks don't assume any relationship (linear /nonlinear) between input and output parameters. Every layer has a random amount of nodes associated with it. For example, the number of nodes in the input layer is equivalent to the dimension of the input data features. When it comes for the number of nodes in the output layer, it depends if it is a regression problem (where the number of nodes is 1) or a classification problem (where the number of nodes is equal to a number of classes or groups that are always greater than 1). Certain matrix operations that include multiplying input data with weights associated in the network layer are executed, post that bias is added to each layer at every node. Afterward, the activation functions such as sigmoid, tan hyperbolic, linear functions is then added to those layers before transferring them further. The activation function is a mathematical formula applied to the inputs along with the weights to provide desired outputs [gurney]. A bias is an additional node in the hidden layer used to enhance the output and give a better fit. An example of the artificial neural network built is represented in Figure 3.7.

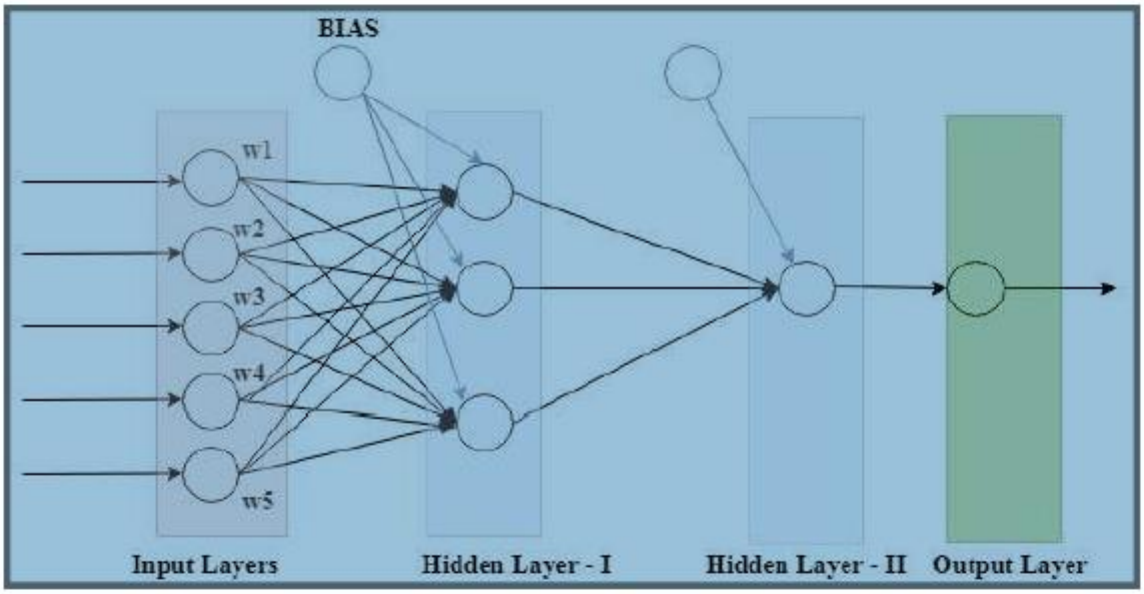


Figure 3. 7:Example of an artificial neural network (Siddappa,2019)

While training the artificial neural networks it is advised to maintain a normal distribution for the values of the input parameters to avoid data redundancy. If the values of each parameter in the data sets are in different ranges, certain algorithms does not perform well. The machine learning algorithms perform well when there is uniformity in the range of data. The rescaling of data is carried out to get uniform distribution. The rescaling can be done by two methods data standardization and data normalization. Data standardization is a rescaling technique which converts the entire data into Gaussian distribution having a mean of 0 and standard deviation ranging [86]. The general formula for data standardization is Eq.(3.5) z_i corresponds to the ith standardized value and x_i represents all values.

$$z = \frac{Xi - mean(x)}{std(x)} \qquad \text{...............[Eq.3.5]}$$

Data Normalization is to scale the values of all the parameters of different range in the dataset to the range of minimum 0 and maximum of 1[86]. The general formula for data normalization is Eq. (3.6)

$$z = \frac{Xi - min(x)}{\max(x) - \min(x)} \qquad \text{...............[Eq.3.6]}$$

The most widely used types of neural networks: Multilayer Perceptron Networks (MLP), Probabilistic Neural Networks (PNN) and General Regression Neural Networks (GRNN), Radial Basic Function (RBF) networks, Polynomial Neural Networks (GMDH), and Cascade Correlation networks. In this paper, Multilayer Perceptron Neural network is used as a predictive model.

A multilayer perceptron (MLP) is a feed forward artificial neural network that maps sets of input data onto a set of appropriate outputs. An MLP consists of multiple layers of nodes in a directed graph, with each layer fully connected to the next one. Except for the input nodes, each node is a neuron (or processing element) with a nonlinear activation function. For training the network MLP utilizes a supervised learning technique called back propagation [87].

What makes a multilayer perceptron different is that each neuron uses a nonlinear activation function which was developed to model the frequency of action potentials. This function is modeled in several ways, but always must be normalizable and differentiable.

There are two main activation functions used in current applications which are both sigmoid:

$$\phi(v_i) = \tanh(v_i) \quad \text{and} \quad \phi(v_i) = \left(1 + e^{-v_i}\right)^{-1}$$

where: the former function is **a hyperbolic tangent** which ranges from -1 to 1, and the latter, the logistic function, is similar in shape but ranges from 0 to 1.

One of the most important characteristics of a multilayer perceptron network is the number of neurons in the hidden layer(s). If an inadequate number of neurons are used, the network will be unable to model complex data, and the resulting fit will be poor. If too many neurons are used, the training time may become excessively long, and, worse, the network may over fit the data. When over fitting occurs, the network will begin to model random noise in the data. The result is that the model fits the training data extremely well, but it generalizes poorly to new, unseen data.

Since linear regression is restricted to fitting linear (straight line/plane) functions to data, it rarely works as well on real-world data as more general techniques such as regression machine learning algorithm which can model non-linear functions. Therefore, this study used WEKA 3.8.5 to examine whether the non-linear model could improve construction time duration prediction further.

3.7.4 Model Evaluation

The purpose of model evaluation is to ascertain the performance of the model. According to Liou and Borcherding [88], model validation can be achieved by using the following techniques:

(i) Analysis of model coefficients and predicted values in comparison with any outside knowledge;

(ii) Using fresh data to test the predictive models; and

(iii) Splitting the original data into two i.e. one set for model calibration while the other is used for validation.

The data splitting method was adopted for this study. Data splitting refers to the process by which the data population is divided into estimation and prediction data sets for regression analysis and validation [83]. The data pool used in building the regression model will be divided randomly into 70% training set and 30% testing set when examining the artificial neuron network model. Numerous studies about application of Artificial Neural Network (ANN) models in construction management showed that Multi-Layer Perceptrons (MLP) model is perceived as the most appropriate approach. This paper will focus on this type of ANN model. The samples representing the calibration and validation were selected randomly. Hence, Table 3.1 shows Number of Projects that were used for calibration and validation.

Model validation is an important step in developing regression relationship. The model performance will be measured using Percentage Errors (PE), Mean Absolute Percentage Error (MAPE) and Root Mean Squared Error (RMSE) according to Eq. (3.7), (3.8), and (3.9).

a) MAPE (Mean Absolute Percentage Error) is a measure of accuracy of a method for constructing fitted times series values in statistics. It usually expresses accuracy as a percentage [89]. Two measures of accuracy dealing with percentage error were used to compare the forecasting performance of the model.

$$\text{Percentage Error} = \frac{\text{Predicted duration} - \text{Actual duration}}{\text{Actual duration}} \times 100\% \qquad \text{.............[Eq.3.7]}$$

$$\text{Mean Absolute Percentage Error (MAPE)} = \frac{\Sigma\,[PE]}{N} \qquad \text{..............[Eq.3.8]}$$

(b)Root Mean Squared Error (RMSE)

RMSE is the root square of the average squared distance between the actual and predicted observations. It is one of the crucial metric that is used on the models, especially when it is a regression model. Since the errors are squared, the metric provides high weight to larger residuals (different between the actual and predicted value) than the smaller residuals [90]**.** Eq (8) represents the general RMSE formula.

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(\text{Predicted} - \text{Actual})^2} \qquad \text{..............[Eq.3.9]}$$

Table 3. 4: Number of Projects that were used for calibration and validation

Type of Projects	Total No. of Projects	No. for Calibration	No. for Validation
Public office buildings	31	22	9
Private mixed use buildings	40	26	11

CHAPTER 4 PRESENTATION OF DATA, ANALYSIS AND DISCUSSION

4.1 Presentation of Data

The data used in this study collected using project data collection format for forty private mixed use buildings from grade one consultants in Table 4.1 and thirty-one public office buildings from Addis Ababa City Administration Construction Bureau (AACACB) and Federal Government Buildings Construction Project Office (FGBCPO) in Table 4.2. The projects were constructed in the period from 2010 to 2020.The collected data contain the name of the project, the expected project duration in days, the actual project duration in days, the expected project cost in birr, the actual project cost in birr, the gross floor area in square meter, the total construction Area in square meter, the building height in meter, the floor numbers in figure, the number of floors above ground in figure and number of floors below ground in figure. The final costs were adjusted to base year of 2020 prices using CPI data from central statistically agency for homogeneity reasons so as to have a fair basis for comparisons.

Key data collected are:

- Expected Project Duration (EPD)
- Actual Project Duration (APD)
- Expected Project Cost (EPC)
- Actual Project Cost (APC)
- Gross Floor Area (GFA)
- Total construction Area (TCA)
- Building Height (BH)
- Floor numbers (F)
- Number of Floors above Ground (NFAG)
- Number of Floors below Ground (NFBG)

Table 4. 1: Performance and scope of private mixed use building projects (Grade one consultants).

P/No.	Project Name	APD (Days)	Adjusted Actual cost in M ETB	GFA	TCA	BH	F	NFAG	NFBG
1	Guest entertainment mixed use complex building	537	192.3	12856	725	53	14	12	2
2	Africa Insurance	1142	247.6	6900	507	54.87	16	13	3
3	Lion Insurance	1402	74.4	5628	585	29.4	9	8	1
4	Amhara National Democratic Movement(ANDM)	1257	169.2	7098	1318	22.4	7	5	2
5	Awash bank	1394	206.5	7600	1600	32	10	10	0*
6	Dashan Bank SC, A.A	480	206.5	7500	1625	17.4	5	5	0
7	1B+G+13 Mixed use building	2200	565.6	20678	1920.1	53	14	13	1**
8	B+G+6 Mixed use building (BRD construction plc)	760	208.9	10500	585	22.5	8	7	1
9	Ethiopian orthodox church B+G+5 Multi- purpose building	1394	171.4	9860	450	24.86	7	6	1
10	G+9 building construction	1401	282.2	12000	1600	32.5	10	10	0
11	G+4 Mixed- use building construction	485	96.6	3800	1235	13.5	5	5	0
12	Mekonnen Bitew Mixed use building(2B+G+14)	817	93.5	6239	525	40.3	17	15	2
13	2B+G+10 Mixed Used Building	1110	142.3	7600	1200	35	13	11	2
14	Tigat Mixed use building building complex	2855	238.5	21500	1746	48.66	16	14	2**
15	Oromia Development Association multi-Purpose building	2170	250.9	11382	1850	29	8	7	1*
16	Mixed use building for Military Tera Merchants	1815	396.5	30866	4000	29	9	7	2
17	S.A Bagerisa Mixed use	1760	323.9	11270	1750	42	17	15	2*
18	Ato Gashaw Mixed use building	2100	508.3	21934	2806	36	11	10	1*
19	2B+G+9 Ethiopian Orthodox thewahdo church Mixed Used Building	1279	234.0	11484	1032	37	12	10	2
20	B+G+8 Kimem Mixed use Building	1730	152.8	7500	1071	31.4	10	9	1

21	Tigray Development Association	2098	255.7	14145	1600	37.5	12	10	2*
22	B+G+11 Mixed use building	720	51.7	4450	380	39.6	12	11	1
23	Solomon Kitew B+G+9 Mixed-use Building	755	41.8	3467	450	34.78	11	10	1
24	ETCON B+G+7 Mixed use building	490	26.1	2100	274	28.6	9	8	1
25	AMT 3B+G+15 Mixed use Building	730	78.4	5965	410	57	19	16	3
26	Lentebur Multi market Center SH.Co	1309	140.4	11500	950	48.6	16	14	2
27	Fezak B+G+7 Mixed use Building	936	67.8	5690	850	24.3	8	7	1
28	Giant Eagle Trading PLC	942	76.5	5850	560	47.1	15	15	0*
29	W/RO Huluye Enaeayehu 2B+G+12 Mixed use Building	1308	41.8	3998	305	45.9	15	13	2*
30	3B+G+19 Mixed use building(gong construction)	1379	307.0	19450	1208	67.4	23	20	3
31	2B+G+19 Mixed use building	431.6	702.4	58500	3000	61.75	21	19	2**
32	2B+G+14 Apartment and commercial	1090	110.4	9450	1320	56.4	17	15	2
33	2B+G+10 Awash Bank Balcha Aba Nefso Branch office	1245	384.0	24107	1980	40.6	13	11	2*
34	B+G+M+5 Mixed use Building	1009	172.7	10960	1507	25.7	8	7	1
35	2B+G+5 Shopping Mall building project	400	75.5	6800	1300	28.5	9	7	2
36	B+G+8 Mixed use Building	493	46.2	3280	412	29.7	10	9	1
37	Nib Bank	970	35.0	3011	510	25.2	8	8	0
38	Dashan Bank SC	910	79.7	3058	560	19.5	6	6	0*
39	B+G+9 W/ro Zehara Yesuf Mixed use building	650	40.9	3600	540	31	11	10	1*
40	B+G+9 Wendafrash Bekele Mixed use building	700	60.5	4500	650	32.5	11	10	1*

Cost adjusted to January 2020 prices.

* Used for validation.

** Outliers in data sets

Table 4. 2: Performance and scope of Public office building projects (AACB and Federal building construction projects).

P/No.	Project Name	APD (Days)	Adjusted Actual cost in M ETB	GFA	TCA	BH	F	NFAG	NFBG
1	Bole Sub-city	1760	338.1	20,658	1,589.10	45.7	13	12	1
2	Kirkos Sub-city	1317	403.6	20,658	1,589.10	45.7	13	12	1
3	Gulele Sub-city	1035	284.6	20,658	1,589.10	45.7	13	12	1
4	Addis Ketema Sub-	1245	341.8	20,658	1,589.10	45.7	13	12	1
5	Arada Sub-city	1192	317.5	20,658	1,589.10	45.7	13	12	1*
6	Kolfe Sub-city	1223	334.0	20,658	1,589.10	45.7	13	12	1*
7	Yeka Sub-city	1095	317.3	20,658	1,589.10	45.7	13	12	1
8	Akaki Kality Sub-city	1095	325.3	20,658	1,589.10	45.7	13	12	1
9	Yeka Sub-city Police	763	78.6	2,218	554.6	14.3	4	4	0*
10	Addis Ketema Sub-city Police	560	76.6	2,218	554.6	14.3	4	4	0
11	Polytechnic (Tegbared) Building	1095	54.4	5,249	583.2	34.14	9	8	1*
12	Akaki TVET Building	740	27.5	2,333	583.2	14.3	4	4	0
13	Entoto Technical	410	11.7	2,218	443.7	17.75	5	5	0
14	Kotebe UniversityCollege Building Project 1	977	13.5	2,500	500	16.5	5	5	0
15	Kotebe University College Building Project 2	702	11.9	2,500	500	16.5	5	5	0*
16	Bole Sub-city Fire and Rescue Agency Building	365	34.4	3,166	791.5	14.3	4	4	0
17	Addis Ketema Sub- city Fire and RescueAgency Building	768	16.1	3,166	791.5	14.3	4	4	0
18	Gulele Sub-city Fire and Rescue Agency Building	949	36.8	3,166	791.5	14.3	4	4	0*
19	Yekatit 12 Hospital Office Building	1580	284.6	10,188	1,132.00	33.41	9	8	1
20	Zewditu Hospital Office Building	480	16.6	2,162	432.5	17.75	5	5	0
21	Justice Bureau Building	803	88.4	8,870	633.6	48.54	14	13	1*
22	Environment Protection Office Building	1760	67.3	3,857	964.2	15	4	4	0
23	Addis Ababa Construction Bureau Laboratory Building	730	20.3	2,728	454.6	20.7	6	5	1
24	Education Bureau Building	1085	81.3	8,870	633.6	48.54	14	13	1
25	Ethiopian Sport Commision Office Building	1420	74.3	8,870	633.6	48.54	14	13	1

26	Urban Integrated Land Information System Development Office Building	1,273.00	231.0	13,500	900	46.5	15	13	2
27	Ministry of Science and Technology Office Building	1,385.00	50.7	7,200	800	26.7	9	9	0*
28	Industry Minister Office Building	1,547.00	48.6	8,800	800	34.5	11	11	0
29	Central Statistics Agency Office Building	1,273.00	39.4	5,853	731.6	28.38	8	7	1
30	Ethiopian Roads Authority Office Building	1,742.00	48.8	5,600	800	22.92	7	7	0
31	Ministry of Foreign Affairs Conference Hall Project	1,505.00	274.0	12,400	3,100.00	16.3	4	3	1*

Cost adjusted to January 2020 prices.

* Used for validation.

** Outliers in data sets

All of the above projects were carried out with the traditional Design-Bid-Build (DBB) form of contract delivery and hence the partitioning of data with this respect or the association of this variable in the analysis was disregarded. Before model development the performance of collected construction projects was determined in order to reaffirm occurrence of delay in construction projects which is the major problem and makes this study indispensable.

Percentage change in project cost was computed as follows: (Otali, 2010; Ijigah *et al*., 2012)

$$PCPC = \frac{APC-EPC}{EPC} * 100 \text{ ,} APCPC = PCPC \qquad \text{...............[Eq.4.1]}$$

PCPC is the percentage change in project cost, CPC is the change in project cost, APC is the actual project cost, EPC is the expected project cost and APCPC is the average percentage change in project cost.

Percentage change in project duration was computed as follows:(Otali,2010; Ijigah et al., 2012)

$$PCP\ D = \frac{APD-EPD}{EPD} * 100 \text{ ,} APCPD = PCPD \qquad \text{...............[Eq.4.2]}$$

PCPD is the percentage change in project duration, APD is the actual project duration, EPD is the expected project duration and APCPD is the average percentage change in project duration.

From a total of 40 private mixed buildings, delay of the contracted deadline was registered at 38 of them (95%), overrun of the contracted price was registered at 31 structures (77.5 %), while simultaneously the delay of contracted deadline and the contracted price overrun were registered at 30 structures (75 %). Furthermore, Maximum contracted deadline overrun was 474.8%, and price 45.1% while the average percentage change in cost and time of the projects is 7.7 % and 110.6% respectively. Contracted deadline reduction was registered at 2 structures (5%), and price 7 structures (17.5%) while simultaneously the contracted deadline and the contracted price reduction was registered at 1 structures (2.5%).

For the case of 31 Public office buildings, delay of the contracted deadline was registered at 30 of them (96.8 %), overrun of the contracted price was registered at 15 structures (48.4 %), while simultaneously the delay of contracted deadline and the contracted price overrun were registered at 15 structures (48.4%). Furthermore, Maximum contracted deadline overrun was 326.7 %, and price 41.1 % while the average percentage change in cost and time of the projects is -1.8 % and 151.3% respectively. Contracted deadline reduction was registered at 0 structures (0%), and price 16 structures (51.6%) while simultaneously the contracted deadline and the contracted price reduction was registered at 0 structures (0 %).

From above result, the delay of contracted deadline for public office buildings was 96.8%. This reaffirms the work of Werku and Jha [20] where 91.75% of the public buildings under his study were not finished in the allotted time. Therefore, this study also consider the effects of cost overrun to develop scientific contract duration prediction model.

4.2 Data preprocessing

Most studies have outlined the importance of partitioning the data to reflect difference in context of the projects such as type of project, type of project delivery system, location, type of structure, type of façade and so on. Although, type of project delivery, location, type of structure and façade used for all projects was nearly the same, this study considered partitioning of data based on type of projects as private mixed use and public office building projects.

The removal of outliers from the data sets is controversial. Some have opined that the removal of outliers in regression analysis help linearize the relationship and result in a best-fit model while others argue the removal of outlier's from the data set violates the purpose of the model. However, this study considered outlier's removal to minimize effect of abnormal observations happened due to various reasons. Consequently, the study carried out data cleaning using box plot method, which internally supported by IBM SPSS statistics 26 to remove outlier's in data sets. As a result, (7.5% of the total data from number of private mixed use building projects) and (4.2% of the total data from number of overall building projects), which were observed to be more than 1.5 times the interquartile range (IQR) away from the 75th percentile and correspondingly eliminated as outlier's. In case of public office building all data sets have normal observations.

4.2.1 Data correlation

The first step in developing regression model is analyzing the relationship among variables. This is done for three reasons – (1) to check how the independent variables are related with the dependent variable so as to show which variables can best predict the dependent variable, (2) to see the pattern of the relationship between each independent variable and the dependent variable and, (3) to check how the independent variables are correlated with each other i.e. the problem of multi-collinearity. In this regard, Pearson's correlation coefficient as well as graphical exploration methods such as scatter plots are widely used. Correlation coefficient is used to measure the strength of association between two variables ranges between -1 (perfect negative correlation) to 1 (perfect positive correlation). while the scatter plots graphically show the pattern of the relationship, most importantly whether the relationship is linear or not. The correlation matrix among the variables under this study shown in Table 4.3 and 4.4.

From Table 4.3 correlation matrix result shows that Actual project cost, gross floor area and Total construction area have moderate to strong relationship with Actual project duration. Therefore, those variables are quite imperative in determining construction duration. However, correlation result between actual project and gross floor area show us the existence of multi-collinearity.

Table 4. 3: Correlation Matrix for Private mixed use buildings before Variables transformation

	APD	APC	GFA	TCA	BH	F	NFAG	NFBG
APD	1	0.578	0.584	0.448	0.156	0.148	0.126	0.202
APC	0.578	1	0.867	0.689	0.183	0.140	0.100	0.275
GFA	0.584	0.867	1	0.797	0.247	0.218	0.173	0.348
TCA	0.448	0.689	0.797	1	-0.155	-0.167	-0.204	0.038
BH	0.156	0.183	0.247	-0.155	1	0.961	0.945	0.767
F	0.148	0.140	0.218	-0.167	0.961	1	0.989	0.776
NFAG	0.126	0.100	0.173	-0.204	0.945	0.989	1	0.673
NFBG	0.202	0.275	0.348	0.038	0.767	0.776	0.673	1

From Table 4.4 correlation matrix result shows that all independent variables have moderate to have strong relationship with Actual project duration. Although, all independent variables are capable to predict actual project duration the correlation result shows the existence of multi-collinearity among independent variables.

Table 4. 4: Correlation matrix for Public office buildings before variable transformation

	APD	APC	GFA	TCA	BH	F	NFAG	NFBG
APD	1	0.408	0.535	0.448	0.489	0.494	0.502	0.340
APC	0.408	1	0.944	0.931	0.759	0.725	0.712	0.657
GFA	0.435	0.944	1	0.938	0.866	0.842	0.841	0.673
TCA	0.448	0.931	0.938	1	0.674	0.635	0.636	0.493
BH	0.489	0.759	0.866	0.674	1	0.994	0.989	0.813
F	0.494	0.725	0.842	0.635	0.994	1	0.996	0.815
NFAG	0.502	0.712	0.841	0.636	0.989	0.996	1	0.759
NFBG	0.340	0.657	0.673	0.493	0.813	0.815	0.759	1

The corresponding scatter plot matrix among the variables is shown in Figure 4.1 and 4.2. The pattern of the relationship is between the predictor variables and the response variable is different across the independent variables. The variables that have good correlation with duration seem to have a fairly linear relationship while some of the independent variables are linearly dependent with each other that could pose the problem of multicollinearity.

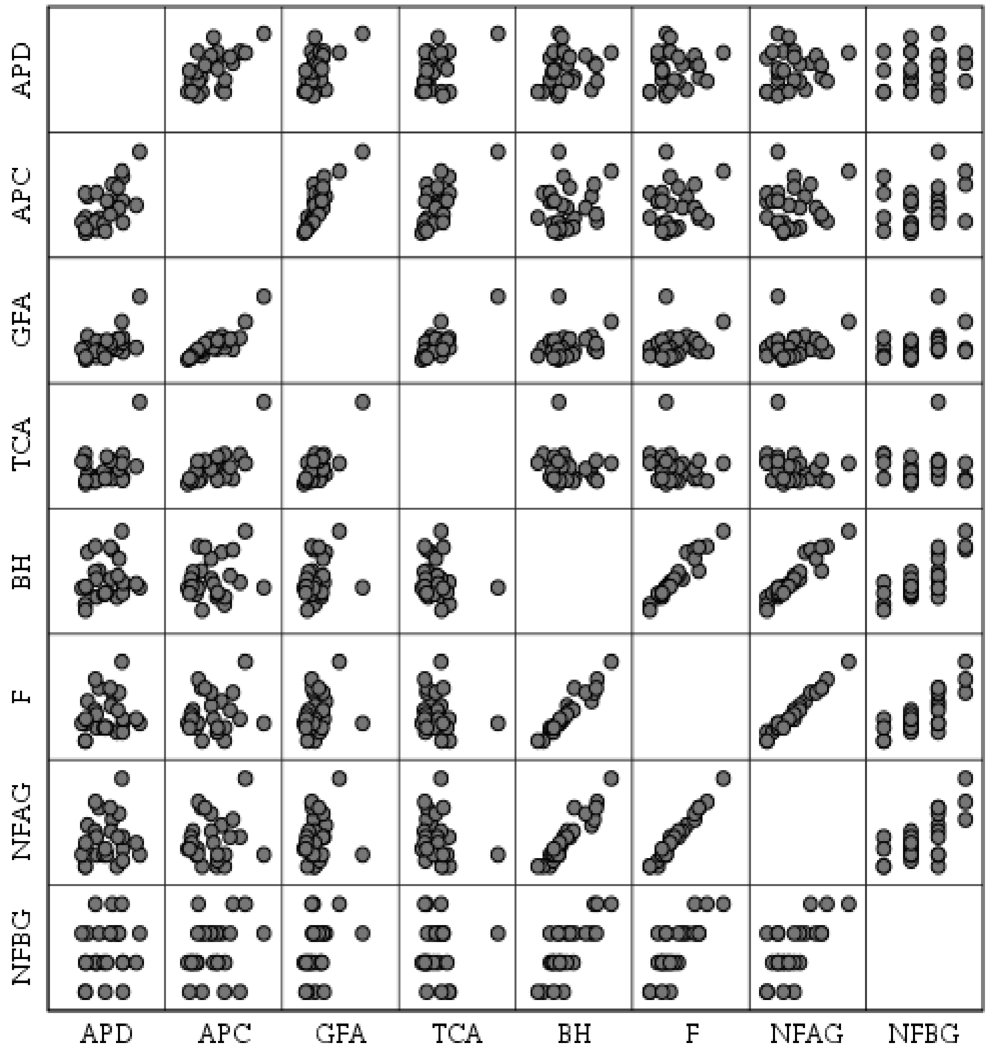


Figure 4. 1: Scatter plot for private building before variable Transformation

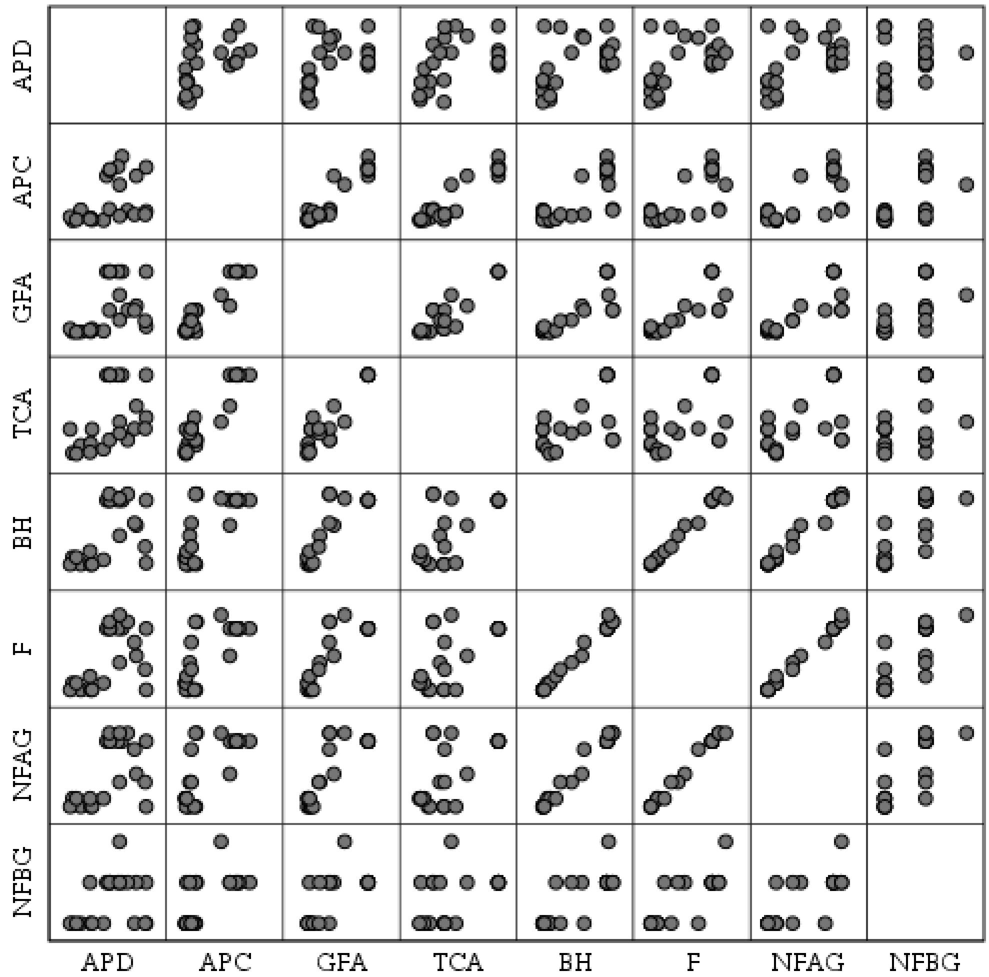


Figure 4. 2: Scatter plot for public office building before variable Transformation

4.2.2 Data Transformation

Data do not always come in a form that is immediately suitable for analysis. Data transformation is often carried out on the variables before carrying out the analysis. Transformations are applied to accomplish certain objectives such as to ensure linearity, to achieve normality, or to stabilize the variance. It often becomes necessary to fit a linear regression model to the transformed rather than the original variables [69].In this regard, logarithmic transformations had proven to be productive in previous studies, as was in the case of the BTC. Accordingly, the data set was transformed into the natural logarithmic form. The correlation of the variables after variable transformation was shown in Table 4.5 and Table 4.6.

Table 4.5 and 4.6 shows transformation of the data into natural logarithm form has resulted in a considerable increase in the coefficient of correlation for variables. Gross floor area was found to be the independent variable with the highest correlation coefficient when compared with other independent variables for all project types.

Table 4. 5: Correlations matrix for private mixed use after variable transformation

	LnAPD	LnAPC	LnGFA	LnTCA	LnBH	LnF	LnNFAG
LnAPD	1	0.516	0.558	0.346	0.259	0.243	0.200
LnAPC	0.516	1	0.899	0.677	0.136	0.099	0.044
LnGFA	0.558	0.899	1	0.696	0.298	0.256	0.188
LnTCA	0.346	0.677	0.696	1	-0.178	-0.185	-0.227
LnBH	0.259	0.136	0.298	-0.178	1	0.97	0.944
LnF	0.243	0.099	0.256	-0.185	0.97	1	0.979
LnNFAG	0.200	0.044	0.188	-0.227	0.944	0.979	1

Table 4. 6: Correlation matrix for Public office buildings after variable transformation

	LnAPD	LnAPC	LnGFA	LnTCA	LnBH	LnF	LnNFAG
LnAPD	1	0.567	0.638	0.548	0.596	0.599	0.599
LnAPC	0.567	1	0.914	0.889	0.792	0.766	0.761
LnGFA	0.638	0.914	1	0.888	0.925	0.912	0.913
LnTCA	0.548	0.889	0.888	1	0.651	0.623	0.629
LnBH	0.596	0.792	0.925	0.651	1	0.996	0.991
LnF	0.599	0.766	0.912	0.623	0.996	1	0.996
LnNFAG	0.599	0.761	0.913	0.629	0.991	0.996	1

The corresponding scatter plot matrix for the transformed data is shown in figure 4.3 and 4.4. It is worth mentioning that the transformation into natural logarithmic form for the data corresponding to number of floors below the ground has not been made in this study, as some of the buildings have no basement floors and this made the log-transformation for this data set inapplicable.

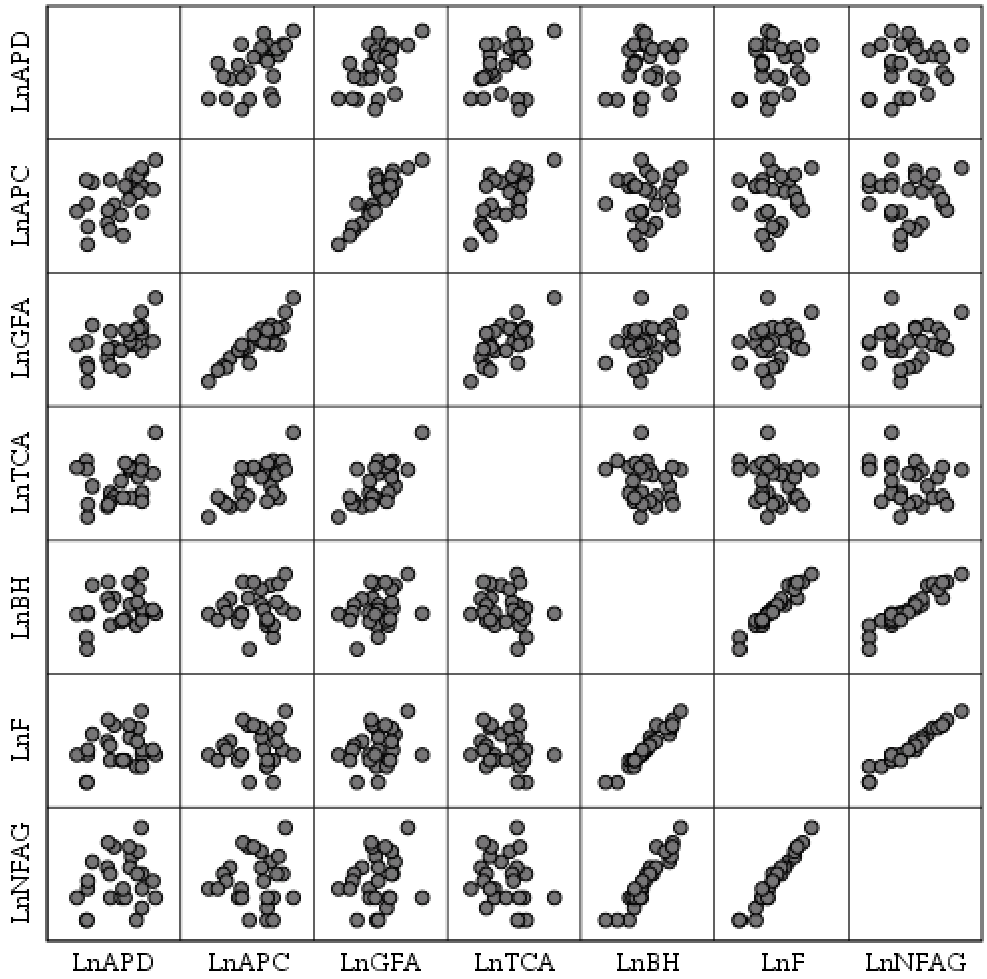


Figure 4. 3: Scatter plot for private mixed use building projects

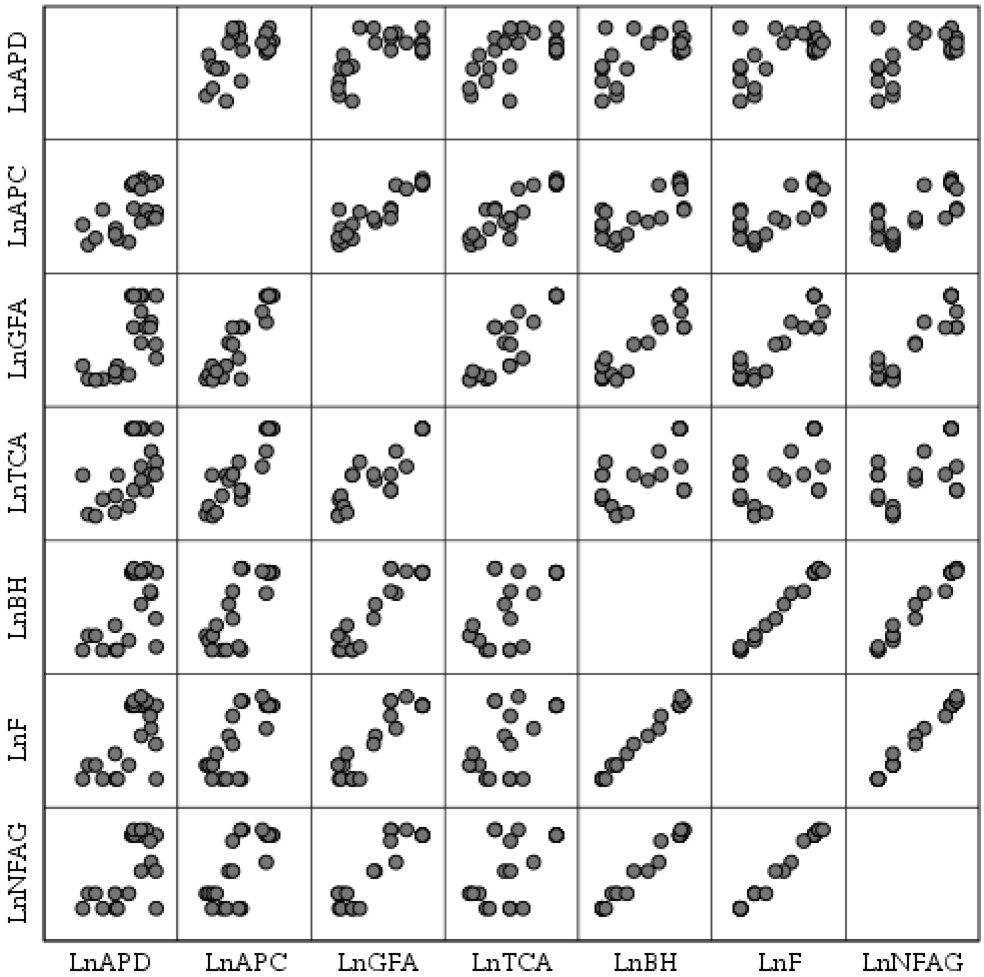


Figure 4. 4: Scatter plot for public office building projects

4.3 Modelling construction time based on BTC model

The study used the data in tables 4.1 and 4.2 to establish the time - cost model developed by Bromillow [10,43] which had been validated in other parts of the world. Hence, for each project type, Bromillow's time-cost relationship was modeled by the least-squares method at 5% significance level, using the IBM SPSS statistics 26.

[Eq.2.1] stated in literature ($T = KC^B$) is non-linear which can be linearized by applying the double – Ln,

i.e. Ln(T) = Ln (KCB) = Ln K + B Ln C.

Ln (T) = Ln K + B Ln C ....[Eq.4.3]

Where

ln(T) is the natural logarithm of project duration;

ln(C) is the natural logarithm of project cost

ln(K) is the natural logarithm of K

[Eq.4.1] can be transformed to the standard simple linear regression form expressed in [Eq.3.2]: Where, Y=Ln (T), a=Ln K, b=B, and X= LnC.

where, for the e value of 2.718, K is quantified by [Eq .4.2]

$$K=e^a \qquad \text{...............[Eq.4.4]}$$

It is evident in figure 4.3 and 4.4, there is a linear dependency between these transformed forms of variables. A simple linear regression technique was carried out to obtain the K and B values in the BTC model using collected data sets. Based on [Eq.3.2], Simple regression model was established and analyzed for each project type. ANOVA was also used and the hypotheses tested at five% significance level. H0: There is no significant relationship between the dependent variable (natural logarithm of construction time) and independent variable (natural logarithm of construction cost), H1: There is a significant relationship between the dependent variable (natural logarithm of construction time) and independent variable (natural logarithm of construction cost).

After data preprocessing, simple linear regression model developed. And, the result of the final simple regression model, where the dependent variable is the natural logarithm of duration (Ln T) and the independent variable is the natural logarithm of cost (Ln C), has the R^2 value of 0.266 for the actual duration. The R^2 values show that there is a positive correlation between duration and cost of actual data for private mixed use building projects.
Assessment of developed model for Private mixed use building projects Tables 4.3, 4.4 and 4.5 show the statistical parameters of the developed model in equation 7.

Table 4. 7:BTC Model summary for Private mixed use building projects.

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.516	0.266	0.236	0.38339	1.567

a. Predictors: (Constant), LnAPC
b. Dependent Variable: LnAPD

An important aspect of a statistical procedure that derives model from empirical data is to indicate or show how well the model predicts results. If there is a perfect relation between the dependent and independent variables, R^2 is 1. Where there is no relationship between the dependent and independent variables R^2 is 0. From table 4.7and 4.10, R^2 are 0.266 and 0.321. The R^2 shows that a percentage of 26.6% and 32.1% of the variance in construction duration of private mixed use and public office building projects respectively in Addis Ababa were explained by the project scope expressed in terms of the final cost of construction indicating weak predictive ability for private mixed use and the moderate predictive ability for public office building projects.

The analysis of variance (ANOVA) in table 4.8 indicates that the developed private mixed use building projects regression model is significant [with $F = 8.713$, $p= 0.007<0.05$]. Moreover, the summarized results for public office building projects was obtained in table 4.10. The result depicts that the developed regression models were significant [with $F = 9.46$, $p= 0.006<0.05$] for public office buildings. This indicate that the BTC model can be extended to private mixed use and public office building projects in Addis Ababa.

Using [Eq.4.1] and Table 4.9 for Actual project duration of private mixed use building projects B= 0.314 and ln K = 5.318⇒ K =203.86. And, also using model parameter in table 4.10 for public office building projects. The BTC model in [Eq.4.1] is formulated as:

$$\textbf{Actual duration}(\textbf{Private mixed use}): \mathbf{T = 203.86C^{0.314}} \quad \text{..............[Eq.4.5]}$$

$$\textbf{Actual duration}(\textbf{Public office}): \mathbf{T = 390.87C^{0.218}} \quad \text{...............[Eq.4.6]}$$

Table 4. 8: ANOVA for private mixed use building projects model

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1.281	1	1.281	8.713	.007
	Residual	3.528	24	0.147		
	Total	4.809	25			

a. Dependent Variable: LnAPD
b. Predictors: (Constant), LnAPC

Table 4. 9: Coefficients for private mixed use building model

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	5.318	0.513		10.361	0.000	4.259	6.377
	LnAPC	0.314	0.106	0.516	2.952	0.007	0.095	0.534

a. Dependent Variable: LnAPD

Table 4. 10: Summary of BTC model for public office building projects.

Type of building	Adjusted Actual cost							
	LnK	K	B	R	R^2	Adjusted R^2	F	Significance (F)
Public office buildings	5.969	390.87	0.218	0.567	0.321	0.287	9.460	0.006

4.3.1 Searching for violation of assumptions

Residual search is conducted to test the validity of the linear regression model. A scatter plot is a good means of judging how well a straight line fits the data Choudhury and Rajan [12] . As can be seen in figure 4.5 and 4.6, a straight line fits the data well as there is a random distribution which is clustered around the horizontal line through the origin. This means that the assumptions of linearity and homogeneity have been achieved.

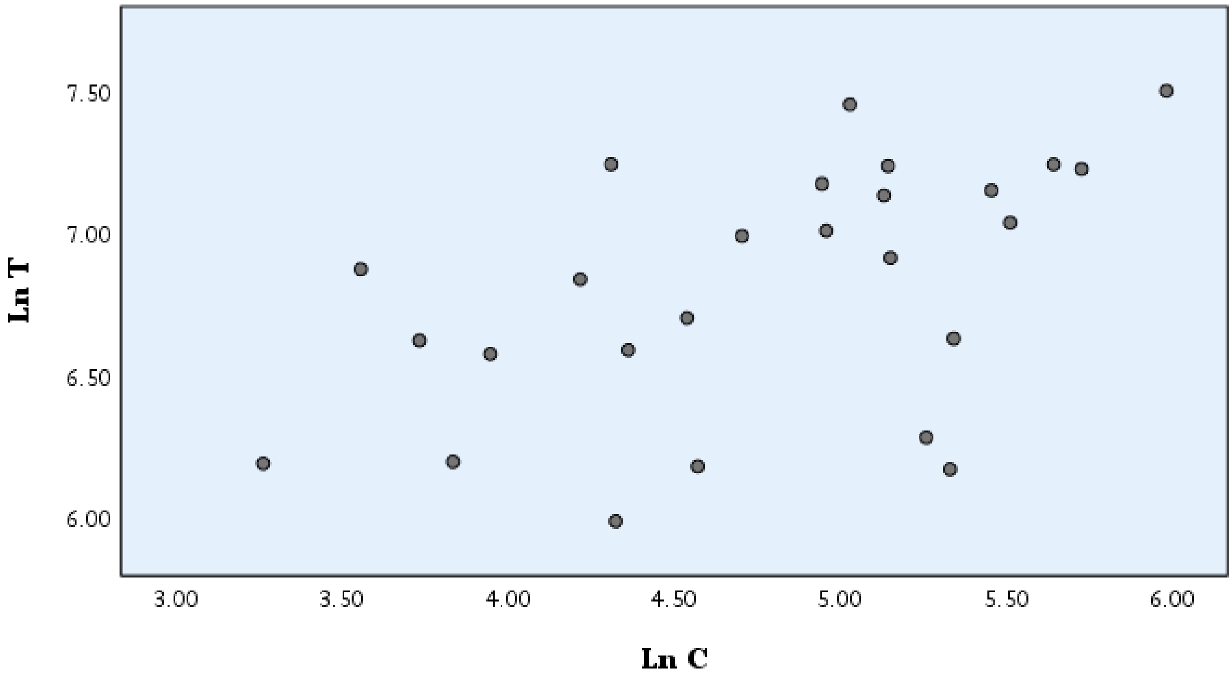


Figure 4. 5: Scatter plot of Ln T against Ln C for private mixed use building projects.

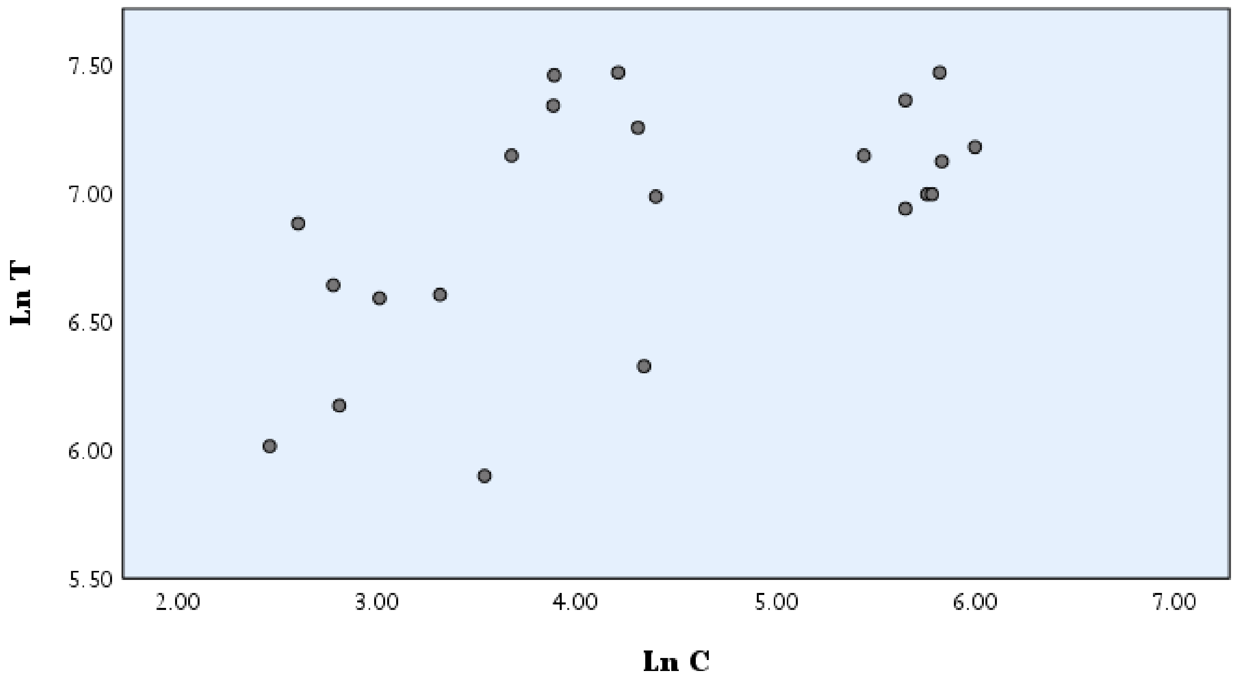


Figure 4. 6: Scatter plot of Ln T against Ln C for Public office building projects

4.4 Modelling construction time based on LTF model

This research also investigated applicability of LTF model for projects under this study. Based on LTF we can define the effect of Gross floor area and number of floors on construction duration in non-linear form:

$T=K \times GFA^{\beta 1} \times F^{\beta 2}$ ....[Eq.4.7]

However, in order to use multiple linear regression analysis, it is necessary to linearize formula into form: This can be done by applying double log on to [Eq.4.7] and give us [Eq.4.8] below.

$LnT = LnK + \beta1 \times LnGFA + \beta2 \times Ln F$ ....[Eq.4.8]

LnT = natural logarithm of construction time,

LnGFA = natural logarithm of gross floor area,

LnF = natural logarithm of floor numbers,

LnK = natural logarithm of constant K,

$\beta1, \beta2$ = coefficients.

[Eq.4.7] can be transformed to the standard multiple linear regression form expressed in [Eq.2.8]

$Y = \beta0 + \beta1\ GFA + \beta0\ F$ ....[Eq.2.8]

Where, Y=Ln (T), $\beta0$=Ln K, X1= LnGFA and X2= LnF

It is evident in figure 4.3 and 4.4, there is a linear dependency between these transformed forms of variables. Regression analysis was used for defining the coefficients K, $\beta1$, $\beta2$ in [Eq.4.4]. Based on [Eq.2.8], multiple regression models were established and analyzed. ANOVA was also used and the hypotheses tested at five% significance level as; H0: There is no significant relationship between the dependent variable (natural logarithm of construction time) and independent variables (natural logarithm of gross floor area and natural logarithm of floor numbers),H1: There is a significant relationship between the dependent variable (natural logarithm of construction time) and independent variables (natural logarithm of gross floor area and natural logarithm of floor numbers).

From table 4.11 and 4.14 , R^2 are 0.322 and 0.409 .R Square value of 32.2 % and 40.9% shows the variance in construction duration of private mixed use and public office building projects respectively in Addis Ababa were explained by the project scope expressed in terms of gross floor area and floor numbers indicating moderate predictive ability for all project types.

The analysis of variance (ANOVA) in table 4.12 indicates that the developed private mixed use building projects regression model is significant [with $F = 5.465$, $p= 0.011<0.05$]. Moreover, the summarized LTF results for public office building projects was obtained in table4.14. The result depicts that the developed regression models was significant [with $F = 6.579$, $p= 0.007<0.05$] for public office buildings. This indicate that the LTF model can be extended to private mixed use, public office building projects in Addis Ababa.

Using [Eq.2.8] and Table 4.13 for Actual project duration of private mixed use buildings $\beta1 = 0.390$, $\beta2 =0.121$ and $\ln K = 3.061 \Rightarrow K =21.34$. And, also using table 4.14 for public office building projects the LTF model using [Eq.4.5] is formulated as:

Actual duration (Private mixed use): $\mathbf{T= 21.34 \times GFA^{0.390} \times F^{0.121}}$ ....[Eq.4.7]

Actual duration (Public office): $\mathbf{T= 64.24 \times GFA^{0.291} \times F^{0.09}}$ ....[Eq.4.8]

Table 4. 11: BTC Model summary for Private mixed use building projects

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	0.568	0.322	0.263	0.37646	1.600

a. Predictors: (Constant), LnF, LnGFA

b. Dependent Variable: LnAPD

Table 4. 12: ANOVA for private mixed use building projects

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1.549	2	0.775	5.465	.011[b]
	Residual	3.260	23	0.142		
	Total	4.809	25			

a. Dependent Variable: LnAPD

b. Predictors: (Constant), LnF, LnGFA

Table 4. 13: Coefficients for private mixed use building projects

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	3.061	1.139		2.687	0.013	0.705	5.417
	LnGFA	0.390	0.131	0.531	2.988	0.007	0.120	0.661
	LnF	0.121	0.200	0.107	0.603	0.552	-0.294	0.535

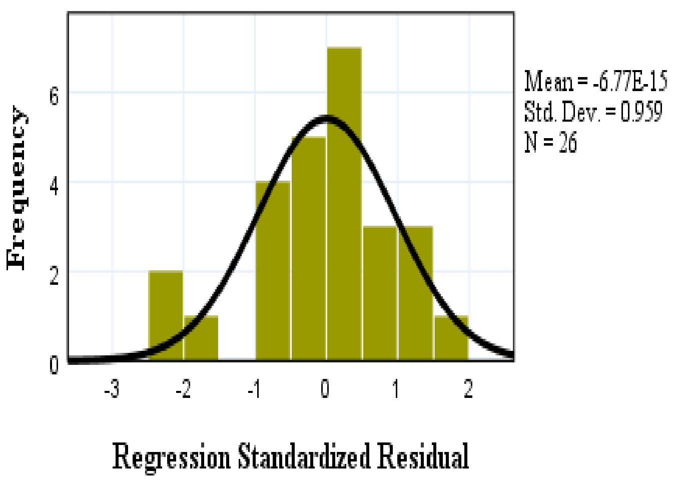
a. Dependent Variable: LnAPD

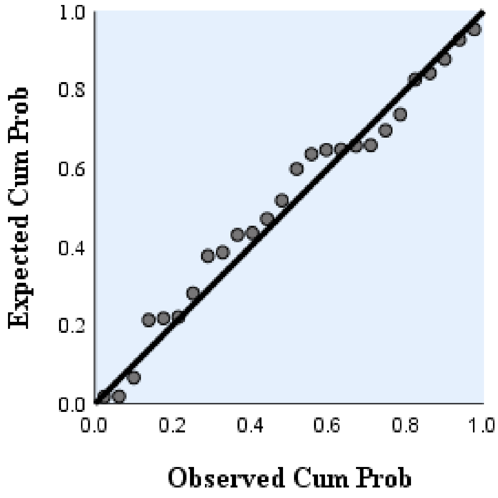
Table 4. 14: Summary of LTF model for public office building projects.

Type of building	Adjusted Actual cost							
	β_0	β_1	β_2	R	R^2	Adjusted R^2	F	Significance (F)
Public office buildings	4.163	0.291	0.090	0.640	0.409	0.347	6.579	0.007

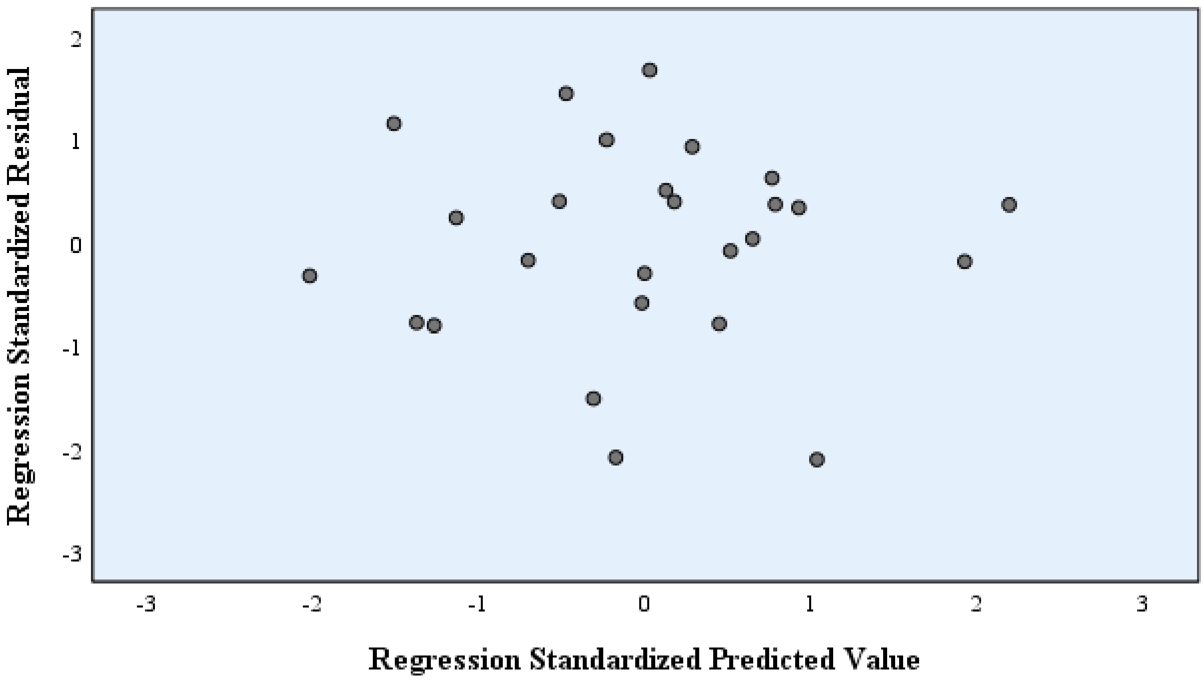
4.4.1 Searching for violations of assumptions

Residual analyses, such as the residual histogram, normal probability plot, etc. were carried out to determine if the residual is distributed normally, as shown in Figures 4.8 and 4.9. According to the result, the residual histogram showed a distribution that is close to a normal distribution, and the normal probability plot also showed points that are lined up linearly, suggesting that the residual is distributed normally. In addition, if there is a clear trend shown by the points of a scatter plot for the dependent variables of a standardized residual, it is interpreted as lacking randomness in the data or having not considered an important independent variable(s). As there was no clear trend in this study, this model is considered to be appropriate.



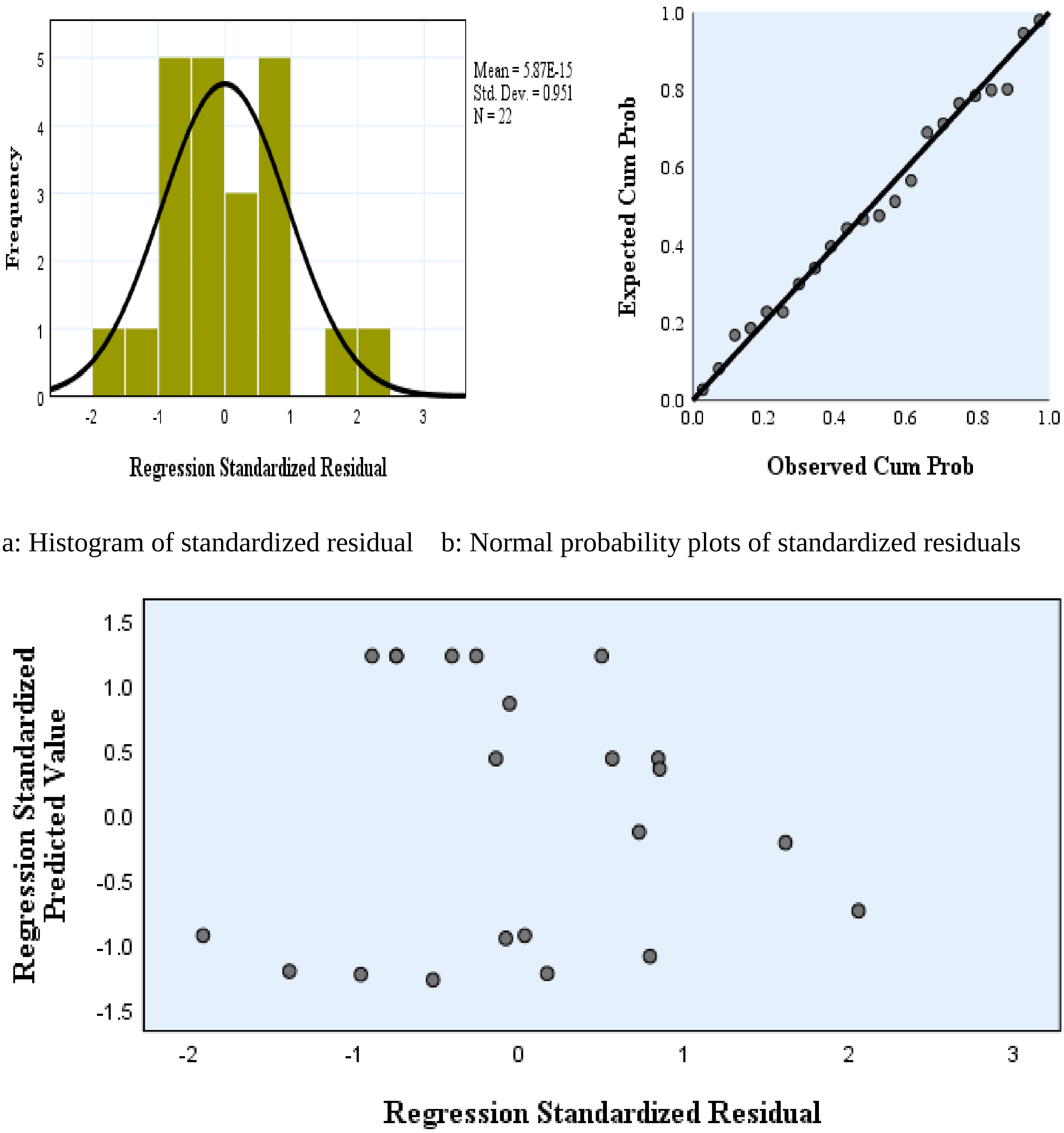


1-Histogram of standardized residual   2-Normal probability plots of standardized residuals



3-Scatter plot of the standardized residuals against the fitted values

Figure 4. 7: Result of residual analysis of LTF for private mixed building projects: 1-Regression standardized residual histogram; 2– normal probability plot of standardized residual; 3- scatter plot for dependent variable of standardized residual.



a: Histogram of standardized residual b: Normal probability plots of standardized residuals

c: Scatter plot of the standardized residuals against the fitted values

Figure 4. 8: Result of residual analysis of LTF for public office building projects: a – Regression standardized residual histogram; b – normal probability plot of standardized residual; c – scatter plot for dependent variable of standardized residual.

4.5.Proposed construction duration prediction model other than BTC and LTF

The impact of cost on construction time is determined by the Bromillow's model and gross floor area and floor numbers are determined by Love time floor model. Further, this study looks for better significant variables determining construction duration by considering all literature identified independent variables including variables identified by BTC and LTF.

Multiple linear regression in [Eq.2.8] can be re-written as;

$T= \beta_0 + \beta_1 PC + \beta_2 GFA + \beta_3 TCA + \beta_4 BH + \beta_5 F + \beta_6 NFAG$ ....[Eq.4.9]

in which T is the Actual duration (in days); PC is the project cost (in ETB); GFA is Gross floor area (in m^2); TCA is total construction area (in m^2); BH is the building height (in m) ; F is the floor numbers ; NFAG is the Number of floors above ground ; NFBG is the Number of floors above ground; and $\beta_0, \beta_1, \ldots \beta_7$ are the regression coefficients.

Through a stepwise selection method, multiple linear regression for private mixed use buildings was applied. ANOVA was also used and the hypotheses tested at five% significance level similar to the previous models, the H0 hypothesis was rejected and the H1 hypothesis was accepted, which implies validity for all the models. The results of the model were shown in table 4.15 to 4.19. And, also summarized results of stepwise multiple linear regression model output were obtained for public office in table 4.20.

The resulting model has only one predictor variable which is gross floor area for all project types as the inclusion of other variables in the model didn't improve the significance of the model.

The overall regression model was significant with [$F =10.854$, $\rho = 0.003 < 0.05$ & $R_2 = 0.311$] for private mixed use and with [$F =13.756$, $\rho = 0.001 < 0.05$ & $R_2 = 0.408$] for public office building projects for their fitted data sets.

The resulting regression equation of private mixed use buildings based on the model coefficients shown in Table 4.17, the typical linear regression equation expressed in [Eq. 4.2] would be:

$Ln(T) = \mathbf{3}.165 + \mathbf{0}.411 \ln(GFA)$ ....[Eq.4.10]

The non-linear form of the proposed model, based on [Eq.4.2], is expressed in Eq. 4.8.

Actual duration (Private mixed use); $\mathbf{T= 23.68 \times GFA^{0.411}}$ ....[Eq.4.11]

Actual duration (public office); $\mathbf{T= 50.63 \times GFA^{0.339}}$ ....[Eq.4.12]

Table 4. 15: Excluded variables entered / removed - Proposed model

Variables Entered/Removed

Model	Variables Entered	Variables Removed	Method
1	LnGFA		Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).

a. Dependent Variable: LnAPD

Table 4. 16: Summary of proposed model

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.558[a]	0.311	0.283	0.37143	0.311	10.854	1	24	0.003

a. Predictors: (Constant), LnGFA

b. Dependent Variable: LnAPD

Table 4. 17:ANOVA –Proposed model

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1.497	1	1.497	10.854	.003[b]
	Residual	3.311	24	0.138		
	Total	4.809	25			

a. Dependent Variable: LnAPD

b. Predictors: (Constant), LnGFA

Table 4. 18: Coefficients-Proposed model

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Correlations			Collinearity Statistics	
		B	Std. Error	Beta			Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	3.165	1.111		2.849	0.009					
	LnGFA	0.411	0.125	0.558	3.295	0.003	0.558	0.558	0.558	1.000	1.000

a. Dependent Variable: LnAPD

Table 4. 19: Excluded Variables-Proposed model

Model		Beta In	T	Sig.	Partial Correlation	Collinearity Statistics		
						Tolerance	VIF	Minimum Tolerance
1	LnAPC	.075[b]	0.191	0.851	0.040	0.192	5.212	0.192
	LnTCA	-.083[b]	-0.345	0.733	-0.072	0.515	1.941	0.515
	LnBH	.102[b]	0.568	0.575	0.118	0.911	1.097	0.911
	LnF	.107[b]	0.603	0.552	0.125	0.934	1.070	0.934
	LnNFAG	.099[b]	0.565	0.577	0.117	0.965	1.037	0.965

a. Dependent Variable: LnAPD

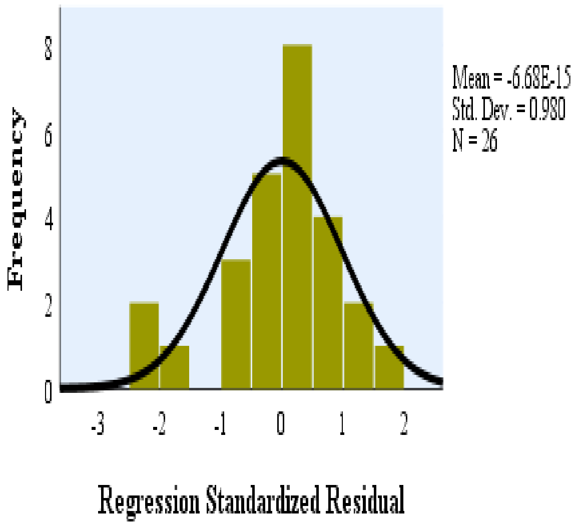
b. Predictors in the Model: (Constant), LnGFA

Table 4. 20:Model parameter for Proposed model.

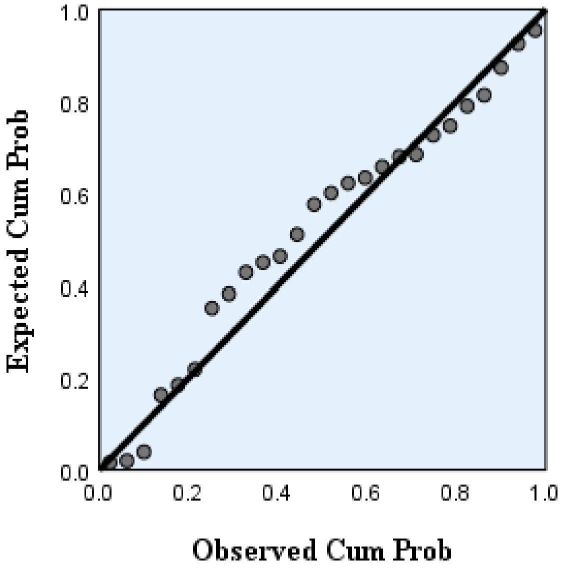
Type of building	Summary of step wise regression model						
	β_0	β_2	R	R^2	Adjusted R^2	F	Significance (F)
Public office buildings	3.925	0.339	0.638	0.408	0.378	13.756	0.001

4.5.1 Searching for violations of assumptions

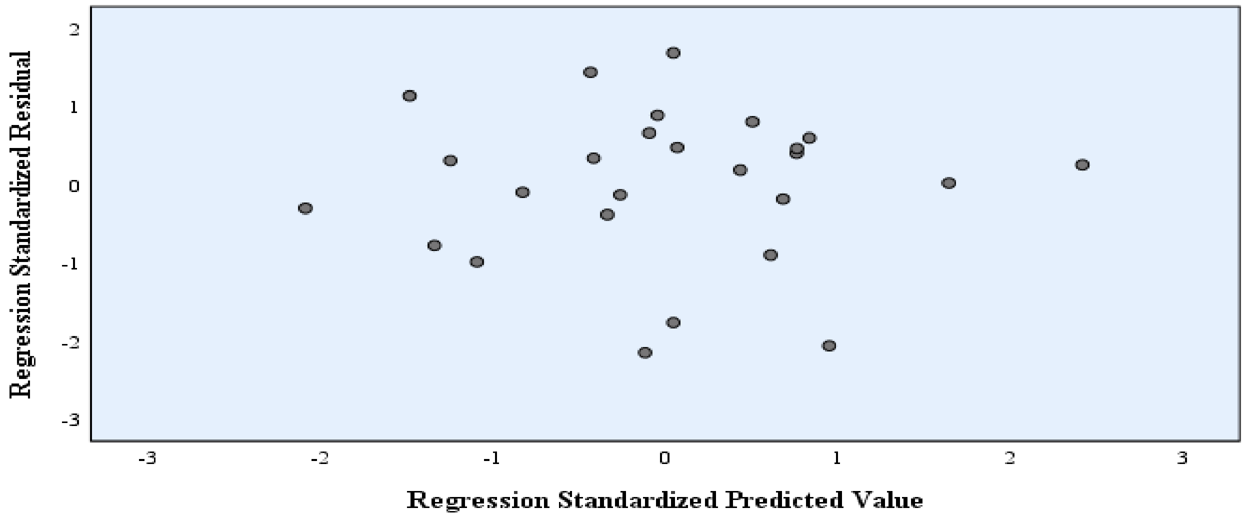
residual analyses, such as the residual histogram, normal probability plot, etc. were carried out to determine if the residual is distributed normally, as shown in Figures 4.10 and 4.11. According to the result, the residual histogram showed a distribution that is close to a normal distribution, and the normal probability plot also showed points that are lined up linearly, suggesting that the residual is distributed normally. In addition, if there is a clear trend shown by the points of a scatter plot for the dependent variables of a standardized residual, it is interpreted as lacking randomness in the data or having not considered an important independent variable(s). As there was no clear trend in this study, this model is considered to be appropriate.



1- regression standardized residual histogram

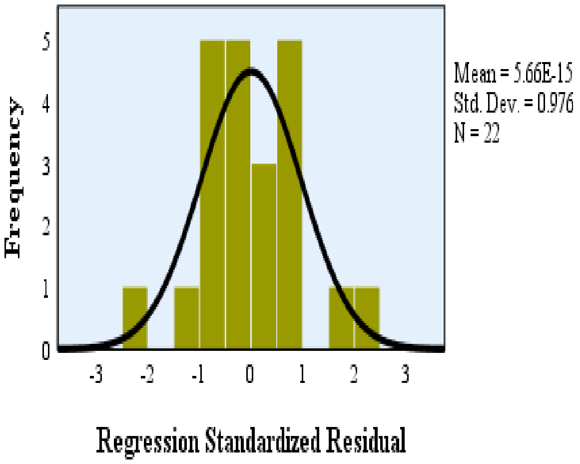


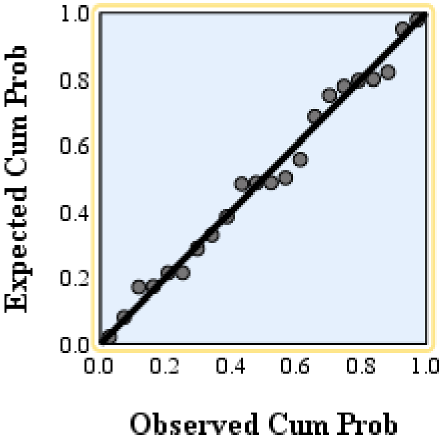
2- normal probability plot of standardized residual



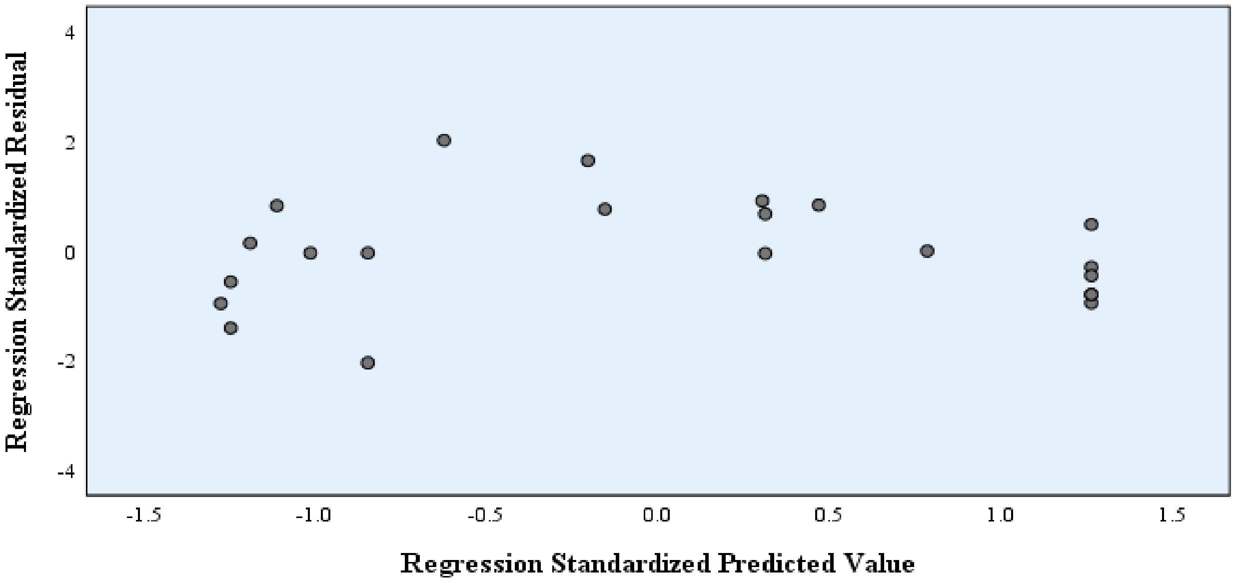
3- scatter plot for dependent variable of standardized residual.

Figure 4. 9: Result of residual analysis of proposed model for private mixed building projects: 1- regression standardized residual histogram; 2– normal probability plot of standardized residual;





a: Histogram of standardized residual b: Normal probability plots of standardized residuals



c: Scatter plot of the standardized residuals against the fitted values

Figure 4. 10: Result of residual analysis of proposed model for public office building projects: a – Regression standardized residual histogram; b – normal probability plot of standardized residual;

4.6 Model Validation

The validity of the model is usually assessed in terms of predictive accuracy. That is, the predicted values obtained from the developed model are compared with the actual observed values to verify the predictive efficacy. To further confirm the predictive ability of the model, the eleven (11) projects for private projects and the nine (9) projects (marked as * in tables 4.1

and 4.2) which were not used in the formulation of the BTC, LTF, Proposed and BATCODA model were used for validation.

The MAPE of all models for private mixed use were determined and the following results were obtained in Table 4.21 to 4.24. The result of model validation implies standard error measurement(MAPE) is high in all models. However, the BTC model reflect low error compared with another models. BaTCoDA model shows high error due lack of continual update for changing technology and project management perceptions.

Table 4. 21: Comparison of actual values and predicted values for private mixed use buildings construction durations using the BTC model.

Project no.	Actual duration (days)	Predicted duration (days)	Difference		
			Error (days)	Error %	Absolute %
5	1394	1082.6	311.4	22.3	22.3
15	2170	1155.5	1014.5	46.8	46.8
17	1760	1252.0	508.0	28.9	28.9
18	2100	1442.2	657.8	31.3	31.3
21	2098	1162.4	935.6	44.6	44.6
28	942	795.8	146.2	15.5	15.5
29	1308	658.24	649.8	49.7	49.7
33	1245	1320.72	-75.7	-6.1	6.1
38	910	806.1	103.9	11.4	11.4
39	650	653.75	-3.8	-0.6	0.6
40	700	739.27	-39.3	-5.6	5.6
Mean	**1388.8**	**1006.2**			**23.9**

*See tables 4.1 and 4.2

Mean Absolute Percentage Error (MAPE)=23.9%

Table 4. 22: Comparison of actual values and predicted values for private mixed use buildings construction durations using the LTF model

Project no.	Actual duration (days)	Predicted duration (days)	Difference		
			Error (days)	Error %	Absolute %
5	1394	919.8	474.2	34.0	34.0
15	2170	1048.1	1121.9	51.7	51.7
17	1760	1143.74	616.3	35.0	35.0
18	2100	1406.8	693.2	33.0	33.0
21	2098	1198.1	899.9	42.9	42.9
28	942	872.4	69.6	7.4	7.4
29	1308	752	556.0	42.5	42.5
33	1245	1489.4	-244.4	-19.6	19.6
38	910	606.3	303.7	33.4	33.4
39	650	695.3	-45.3	-7.0	7.0
40	700	758.5	-58.5	-8.4	8.4
	1388.8	**990.0**			**28.6**

*See tables 4.1 and 4.2

Mean Absolute Percentage Error (MAPE)=28.6%

Table 4. 23: Comparison of actual values and predicted values for private mixed use buildings construction durations using the Proposed model.

Project no.	Actual duration (days)	Predicted duration (days)	Difference		
			Error (days)	Error %	Absolute %
5	1394	932.0	462.0	33.1	33.1
15	2170	1100.2	1069.8	49.3	49.3
17	1760	1095.8	664.2	37.7	37.7
18	2100	1440.7	659.3	31.4	31.4
21	2098	1203	895.0	42.7	42.7
28	942	836.9	105.1	11.2	11.2
29	1308	715.7	592.3	45.3	45.3
33	1245	1497.8	-252.8	-20.3	20.3
38	910	641.1	268.9	29.5	29.5
39	650	685.5	-35.5	-5.5	5.5
40	700	751.4	-51.4	-7.3	7.3
	1388.8	**990.9**			**28.5**

*See tables 4.1 and 4.2

Mean Absolute Percentage Error (MAPE)=28.5 %

Table 4. 24: Comparison of actual values and predicted values for private mixed use buildings construction durations using the BATCODA

Project no.	Actual duration (days)	Predicted duration (days)	Difference		
			Error (days)	Error %	Absolute %
5	1394	1814.4	-420.4	-30.2	30.2
15	2170	1936	234.0	10.8	10.8
17	1760	2108.07	-348.1	-19.8	19.8
18	2100	2449.8	-349.8	-16.7	16.7
21	2098	1948.3	149.7	7.1	7.1
28	942	1303.1	-361.1	-38.3	38.3
29	1308	1059.2	248.8	19.0	19.0
33	1245	2231.1	-986.1	-79.2	79.2
38	910	1322.1	-412.1	-45.3	45.3
39	650	1057.6	-407.6	-62.7	62.7
40	700	1205	-505.0	-72.1	72.1
	1388.8	**1675.9**			**36.5**

*See tables 4.1 and 4.2

Mean Absolute Percentage Error (MAPE)=36.5 %

Table 4. 25: Predictive ability for public office building models

Following the same procedure for private mixed use buildings the summarized model validation result for public office building projects were presented in Table 4.25.

Measure	BTC	LTF	Proposed	BATCODA
Formula	$T=390.87C^{0.218}$	$T= 64.24\ GFA^{0.291} \times F^{0.09}$	$T= 23.68\ GFA^{0.411}$	$T = 7 \text{ to } 12^* C^{1/3}$
MAPE(%)	18.3	20.1	18.9	39.7

Based on Table 4.25, Mean Absolute Percentage Error for all models of public office building projects were determined. The result of standard error measurement(MAPE) shows BTC model with hold minimum error compared with another models.

4.7 Modelling construction time based on Machine Learning Algorithm

To build the predictive models for construction actual duration, the processed data is trained by multiple machine learning algorithms and tested for future predictions. Also, as discussed earlier the data is divided into two sets, namely training set and test set. Generally, the split will be 70 percent of the data for training set and 30 percent of the data for set. After the training set data is trained by different algorithms, it has to be tested by the new inputs. Hence, the test data is applied to the trained models built by different algorithms to crosscheck their performance. In this section, the training and building process of the model is executed with the help of WEKA 3.8.5. WEKA supports several standard data mining tasks, more specifically, data preprocessing, clustering, classification, regression, visualization, and feature selection.

Before applying the algorithms to the data, the data has to be in its processed state. The processed or the cleaned form of the data includes eliminating the parameters not affecting the output, clearing all the unwanted or empty data points (outlier's removal) and having uniform distribution of parameters. In this section, correlation based feature selection is adopted to select the relevant and potential variables in Table 4.26.

Table 4. 26: Selected variables for modelling after correlation based feature selection

Types of project	Selected variables after feature selection
Private mixed use buildings	APC, GFA,TCA
Public office building projects	APC,GFA,TCA,BH,F,NFAG,NFBG,APD

There are certain times wherein few algorithms work well for training data, but then they fail to give good results for new data or test data. This situation where the model is unable to perform well on the newer data as against its performance on the training data is called overfitting. A good model will learn the pattern from the training data and then generalize it on new data (from a similar distribution). If the test error is decreasing with new set of inputs, the model is still right. On the other hand, an increase in the test error indicates that the algorithm is probably overfitting. In such situations, it's always recommended to reconsider the data set, which means sub setting of data in different approaches. There is no solution other than doing trial and error. Linear regression with a reasonable number of the variable will never over fit the data, since the model is simple and restricted to linear relationships between the variables. After the data is clustered and is in its processed state, the work flow of building predictive model is defined, and the final stage is to apply different machine learning algorithms and evaluate their accuracy as well as error percentage. The machine learning algorithms used in this research are multiple linear regression and artificial neural networks (ANN).

a) Multiple linear regression

Multiple linear regression was applied to the prepared data for modelling. The results of the model in table 4.27 shows that coefficient of determination and Mean absolute percentage error [67.24% &25.92%] and for public office [65.09% & 27.52%].

Table 4. 27: Summary of multiple linear regression results

Types of project	Multiple linear regression	
	R^2	MAPE%
Private mixed use buildings	0.6724	25.92
Public office buildings	0.6509	27.52

b) Artificial neural network (ANN)

To develop the model, trial and error practices were carried out to conclude the best structure of the model. As mentioned before the total number of collected data for private mixed use and public office were 40 and 31 respectively. The model development trials and errors were applied on 26 projects for private mixed use and 22 projects for public office building projects after outlier’s removal. While the other projects 11 from private mixed use and 9 from public office building were chosen randomly for testing the network.

There are several ways to determine a good number of hidden neurons. One solution is to train several networks with varying numbers of hidden neurons and select the one that tests best. A second solution is to begin with a small number of hidden neurons and add more while training if the network is not learning. A third solution is starting to get the right number of hidden neurons. This study developed the model with varying numbers of hidden neurons and select the one that tests best.

Before model development data visualization was carried out for representing data through graphs and plots with the aim to understand data clearly as shown in figure 4.11and 4.12.

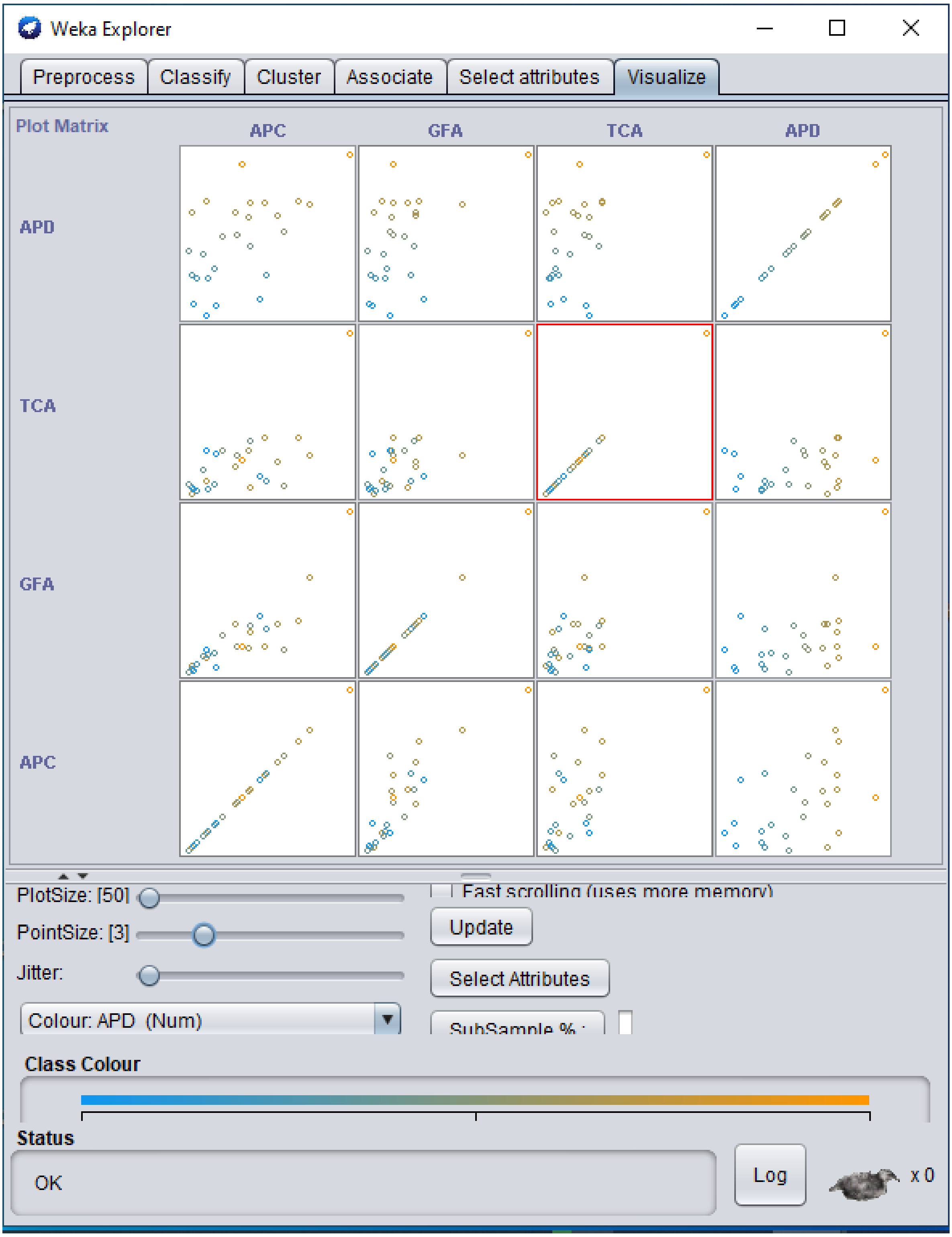


Figure 4. 11: Weka Scatter Plot Matrix for private mixed use building projects.

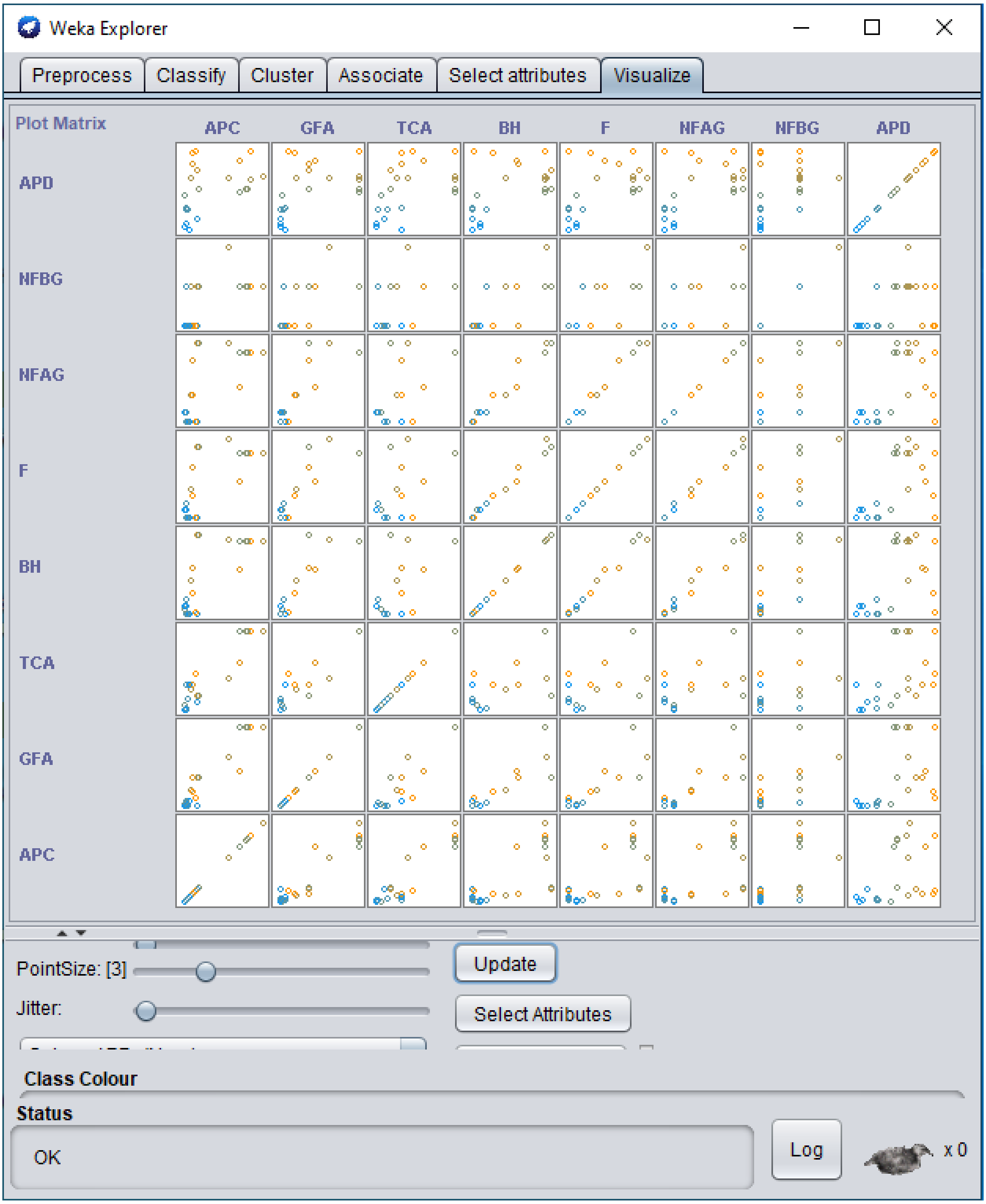


Figure 4. 12: Weka Scatter Plot Matrix for public office building projects.

After training and testing of data sets by trial and error approach the result of the trained and tested shows neural network structure of private mixed use and public office building projects in figure 14, and figure 15 respectively. And, also the result of neural network structure was summarized in table 4.28.

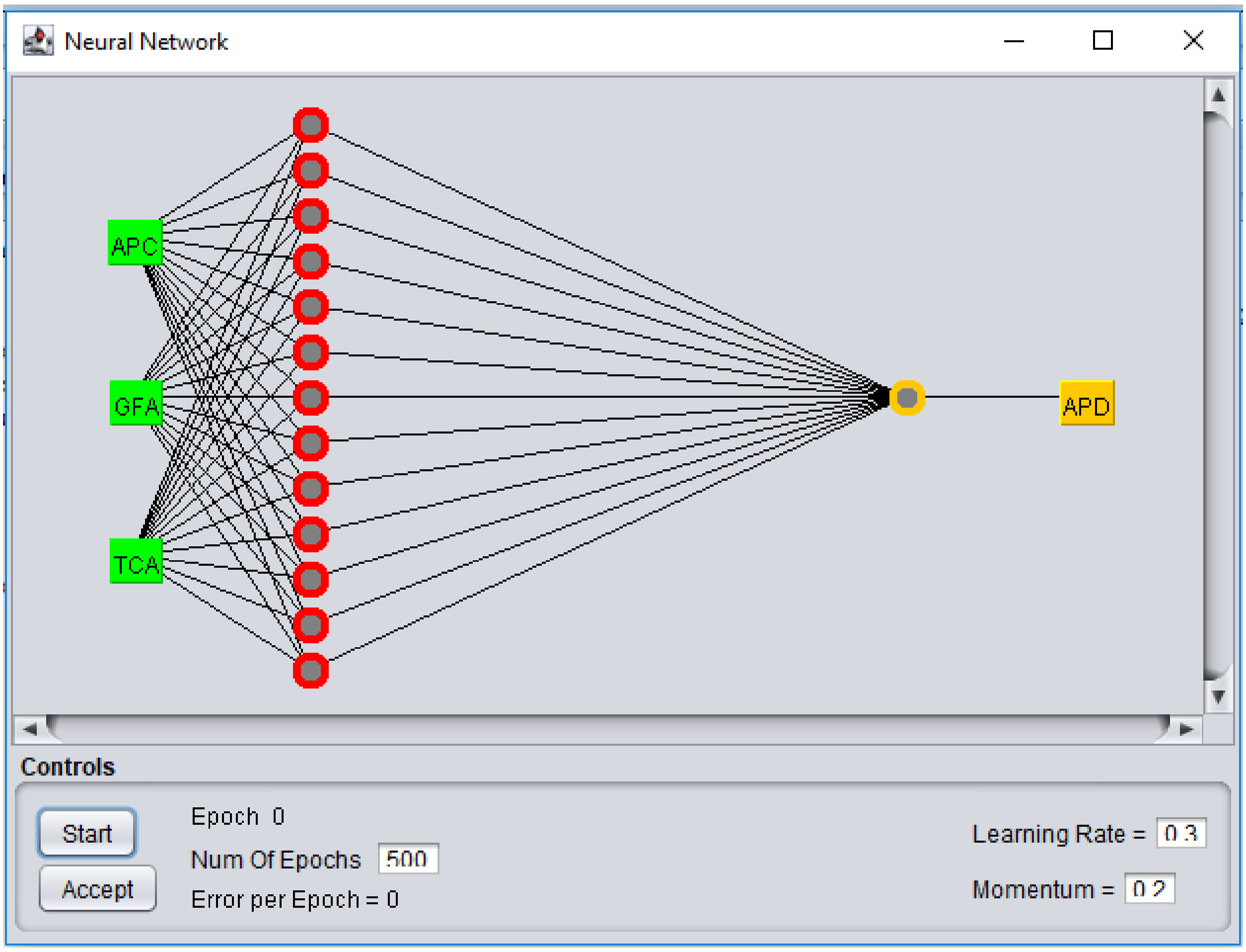


Figure 4. 13:Weka neural network structure for private mixed use buildings

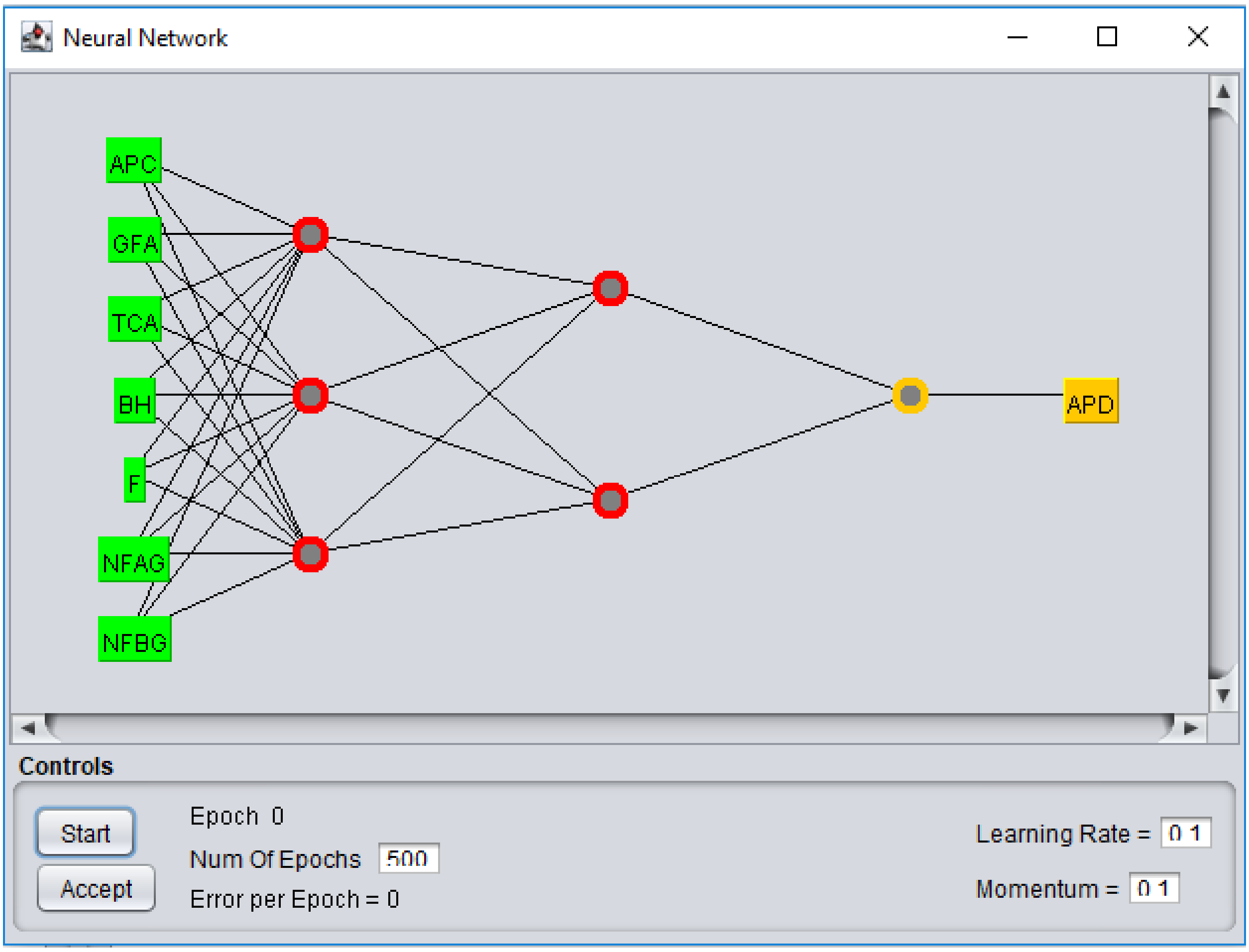


Figure 4. 14: Weka neural network structure for public office buildings

The result of neural network structure was achieved after numerous trial and error of changing number of neurons in hidden layers from one to fourteen, changing hidden layers from one to two. The ANN models for private mixed use buildings fit data sets with in one hidden layer and four number of neurons in hidden layers for training data sets, one hidden layer and thirteen number of neurons in hidden layers for testing data sets, for the case of public building projects the model fit with one hidden layer and three number of neurons in hidden layers for training data sets, two hidden layer with three and two number of hidden neurons respectively for testing data sets.

Table 4. 28: Summary of (MLP-NN) results

Types of project	Multilayer perceptron neural network	
	R^2	MAPE
Private mixed use buildings	0.7247	22.19
Public office building projects	0.7322	16.33

This table shows that coefficient of determination for private mixed use and public office building projects were 72.47% and 73.22% respectively. And, also the mean absolute percentage error for the projects were 22.19% and 16.33% respectively.

CHAPTER 5 CONCLUSIONS AND RECCOMENDATIONS

5.1 Conclusions

This section summarises the findings of the study. Based on the analysis carried out in the previous chapter using forty (40) private mixed use and thirty-one (31) public office building projects, the results of the study are as follows:

5.1.1 Review of objective one

The main aim is to develop appropriate construction time models for private mixed and public office building projects in Addis Ababa. In this regard, the study had put forth the first hypotheses to realize its objectives.

The first objective of this study is to validate applicability of Bromillow's time–cost and Love Time Floor models for public office and private mixed use building construction projects in Addis Ababa.

Based on this for private mixed use building projects from Addis Ababa, a time – cost , time-gross floor area and floor numbers models were derived as $\mathbf{T = 203.86C^{0.314}}$ and $\mathbf{T= 21.34 \times GFA^{0.390} \times F^{0.121}}$ respectively. And, also this study formulated BTC and LTF for public office building projects as $\mathbf{T = 390.87C^{0.218}}$ and $\mathbf{T= 64.24 \times GFA^{0.291} \times F^{0.09}}$ respectively. Although, bromillow's time-cost and love time floor were valid to determine construction time the developed models prediction is not very strong enough to forecast contract duration of projects.

5.1.2 Review of objective two

The second objective of this study is to develop an alternative duration prediction model by considering project scope factors identified in literature.

To achieve this objective, models were developed based on Multiple linear regression and artificial neural network. The results of the study for private mixed use and public office building projects using step wise linear regression shows that GFA is the significant variable which determine construction time for both projects. The formulated model using gross floor area for private mixed and public office building projects were $\mathbf{T= 23.68 \times GFA^{0.411}}$ and $\mathbf{T= 50.63 \times GFA^{0.339}}$ respectively. Furthermore, construction time prediction models were developed through machine learning algorithm using WEKA 3.8.5. Before model development correlation based feature selection was carried out to identify relevant variables. After wards, the duration prediction models were developed using Multiple linear regression and Artificial neural. The Multiple linear regression model result shows the coefficient of determination 67.24% and mean absolute percentage error 25.92% for private mixed use buildings. And, also it shows the coefficient of determination 65.09% and mean absolute percentage error 27.52% for public office building projects. Moreover, using the same data, the Artificial neural network models result shows the coefficient of determination 72.47% and mean absolute percentage error 22.2% for private mixed use buildings respectively. And, also it shows the coefficient of determination 73.2% and mean absolute percentage error 16.3 % for public office building projects.

5.1.3 Review of objective three

The third objective of this study is to identify which of the formulae derived could be used to predict the duration of public office and private mixed use building construction projects in Addis Ababa by validating the models that have been derived.

To accomplish this objective, this study evaluated the developed model for their predictive accuracy using standard error measurement for unseen data. The measurement used in this study was Mean Absolute Percentage Error. The results of Mean Absolute Percentage Error of private mixed use buildings for BTC, LTF, proposed, BaTCoDA and WEKA based multiple linear regression for testing data sets were 23.9 %, 28.6%,28.5% ,36.5% and 25.92% respectively. And also, for public office building projects the result of MAPE for BTC, LTF, proposed, BaTCoDA and WEKA based multiple linear regression were 18.3% 20.1%,18.9% 39.7% and27.52 % respectively. However, for the case of ANN model the MAPE of private mixed use and office building projects were 22.2% and 16.3 % respectively. Therefore, ANN result shows that error encountered using Artificial neural network models were comparatively low to linear regression models.

5.1.4 Summary for all objectives comparing with related papers

Although, Dawit reported that GFA is significant variable which better determine the duration of public office building projects this study validated that BTC is superior model comparing with other linear regression models for private mixed use and public office building projects under this study. However, the study results of MLP-NN shows significant improvement of prediction over linear regression models formulated by Dawit **[**22**]**.

Finally, this study concludes that ANN model for construction duration prediction better determine duration of private mixed and public office building projects.

5.2 Recommendations

This study recommends, the stakeholders of construction projects;

Project owners

- The project managers or those who carry out feasibility study on the behalf of the client shall consider developed models for estimating duration of private mixed use and public office building projects at the feasibility study stage.

Regulatory body

- The regulatory body should have to continually update duration estimation model for construction projects to minimize and resolve claims happened in construction industry. As a result, it makes construction industry attractive for professional and brings customer satisfaction.

Contractors

- Contractors engaged in the building construction should have to use developed models of the buildings to check estimated time by the client is realistic reasonably.
- Contractor should have continual update duration estimation model in order to include effects of technological change, project management perceptions and increase in number of contractors.

Consultants

- Consultants should have to use developed model during preparation of contract documents for private mixed use and public office building projects.
- Consultants shall strongly urge project owners to use developed models at feasibility stages of the project.

5.3 Further study

The areas for further investigation are being proposed as follows:

- Construction time is dependent on so many other factors apart from project scope factors. For future research work, it may be necessary to incorporate other factors weather conditions and management attributes among others to improve the model. This could be done by using the relative importance index to identify the most tactical options affecting the duration of private mixed use and public office building projects in Addis Ababa. The sensitivities of these significant factors to the duration of the projects can then be tested and their coefficients could be incorporated as appropriate.

- This study covered duration estimation model of design bid build and admeasurement contract projects of private mixed use and public office building projects. Subsequently, further research should have to be carried for the rest of project delivery system, another building types and infrastructure projects.
- This study covered filter methods of feature selection to identify relevant variables which determine construction time using ANN. Therefore, subsequent work on construction modelling will consider other feature selection techniques.
- Most of the study were carried using project scope factors and this study also covers construction time estimation using project scope. Therefore, the next study should have to carry out construction duration estimation using Bill of quantity.

BIBLIOGRAPHY

[1] Habenom, G.(2017).Research in Ethiopian Construction Industry: Review of Past Studies and Future Need Assessment, Addis Ababa University School of Graduate Studies.

[2] ITE Build & Interiors (2015).Ethiopia's Construction Industry: Transforming a Nation, A market insights report by ite build & interiors.

[3] Ministry of Works and Urban Development (MOWUD 2006).Plan for Accelerated and Sustained Development to end Poverity (PASDEP). Addis Ababa.

[4] Assefa, S. A., and Al.-Hejji., Sadiq (2006).Causes of delay in large construction projects. *International Journal of Project Management, vol. 24, no. 4, pp. 349–357.*

[5] Sriana,T., and Hayati, K.(2015).Time-Cost Relationship Model on the Construction of Education Building in Aceh Province. *Journal of Asian Scientific Research*. vol. 5, no. 7, pp. 328–339.

[6] Barrie, Donald S., and Paulson,Boyd C.(1992).Professional construction management : including C.M., design-construct, and general contracting /. McGraw-Hill Science/Engineering/Math.

[7] Bayram, S.(2016).Duration prediction models for construction projects: In terms of cost or physical characteristics?, *KSCE Journal of Civil Engineering*. vol. 21, no. 6, pp. 2049–2060.

[8] Sonmez,R.(2004).Conceptual cost estimation of building projects with regression analysis and neural networks, *Canadian Journal of Civil Engineering*. vol. 31, no. 4, pp. 677–683.

[9] Kim, Y.-J., Yeom, D.J., & Kim,Y. S. (2019). Development of construction duration prediction model for project planning phase of mixed-use buildings. *Journal of Asian Architecture and Building Engineering. Vol 18, no.6, pp. 586-598.*

[10] Bromillow. F J.(1969). Contract time performance expectations and the reality. *Building Forum. vol. 1, no. 3, pp. 70–80.*

[11] Love, Peter E.D., Tse,Raymond Y.C., and Edwards. David J.(2005).Time-cost relationships in Australian building construction projects. *Journal of Construction Engineering and Management*. vol. 131, no. 2, pp. 187–194.

[12] Choudhury, I. and Rajan, S. Sh.(2003).Time-cost relationship for residential construction in Texas. *20th International Conference on Construction IT,Construction IT Bridging the Distance, Waiheke Island, New Zealand*.

[13] Endut, I. R., Akintoye, A., and Kelly, J.(2006).Relationship between duration and cost

of Malaysian construction projects. *International Conference in the Built Environment in the 21st Century,Kuala Lumpur, Malaysia. vol. 1, pp. 299–309.*

[14] Rasdorf, W. J.,and Abudayyeh, O. Y.(1991).cost- and schedule-control integration: issues and n e e d s.*vol. 117, no. 3, pp. 486-502, 1991*.

[15] Laptali, E., Bouchlaghem, N., and Wild. S.(1996).An integrated computer model of time and cost optimisation, 12th Annual ARCOM Conference, Sheffield Hallam University, UK. pp. 133–139.

[16] Jarkas. Abdulaziz M.(2015).Predicting Contract Duration for Building Construction : Is Bromilow's Time-Cost Model a Panacea ?.

[17] Skitmore, R.M. and Ng, S.T. (2003). Forecast Models for Actual Construction Time and Cost. *Building and Environment. vol. 14, no. 4, pp. 1075-1083*.

[18] Owolabi, J. D.(2014).Causes and effect of delay on project construction delivery time. *International Journal of Education and Research*. vol. Vol. 2, no. No. 4, pp. 197–208.

[19] Mačková, D., and Bašková ,R.(2014).Applicability of Bromilow´s time-cost model for residential projects in Slovakia.*Selected Scientific Papers - Journal of Civil Engineering*.vol. Vol. 9, no. No. 2, pp. 05-12.

[20] Werku Koshe& Jha, K. N. (2016). Investigating Causes of Construction Delay in Ethiopian Construction Industries, *Journal of Civil, Construction and Environmental Engineering, 1(1), 18-29.*

[21] Skitmore, R.M. and Ng, S.T. (2001). Australian project time-cost analysis: Statistical analysis of intertemporal trends. *Construction Management and Economics.* vol. 19, no. No. 5, pp. 455–458.

[22] Dawit, B.(2020). Model for Estimating Construction Duration of Public Building Projects in Addis Ababa, Addis Ababa University School of Graduate Studies.

[23] Project Management Institute(2017). *A guide to the project management body of knowledge (PMBOK guide) / Project Management Institute.* Project Management Institute, publisher(Sixth Edition).

[24] Kam Shadan, P.E. Gannett Fleming I. "Construction Project Management Handbook." 2012.

[25] What is Project Management? | Project Management Stages | AXELOS. *https://www.axelos.com/best-practice-solutions/prince2/what-is-project-management*.

[26] What is Project Management: Definition and Terms | Project Management Guide. *https://www.teamgantt.com/guide-to-project-management/what-is-project-*

management.

[27] Assudani, R.(2008) .Using Stakeholder and Social Network Theories to Improve Project Success.

[28] Baldwin, A., and Bordoli. D.(2014).A Handbook for Construction Planning and Scheduling.© 2014 by John Wiley & Sons, Ltd Registered.

[29] Mubarak, S.(2015).Construction Project Scheduling and Control. John Wiley & Sons, Inc., Hoboken, New Jersey.

[30] Othman.M.K.(2016).Cost Estimation.vol. 1, no, p. 45. https://www.researchgate.net/publication/29030608.

[31] Cost Estimate (Estimation) - The Project Definition. *https://www.theprojectdefinition.com/p-cost-estimate-estimation/*.

[32] Chapter 20 Project Development Cost Estimates| Course Hero. *https://www.coursehero.com/file/9734552/chapt20/*.

[33] The Ultimate Guide to Project Management - ProjectManager.com. *https://www.projectmanager.com/project-management.*

[34] Project Management for Construction: Cost Estimation. *https://www.cmu.edu/cee/projects/PMbook/05_Cost_Estimation.html*.

[35] Cost Estimation Techniques in Construction Projects - eSUB Construction Software. *https://esub.com/blog/cost-estimation-techniques-construction-projects/.*

[36] Hegazy, T.(1999).Optimization of construction time – cost trade-off.

[37] Liao, T. W. , Egbelu, P. J. , Sarker, B. R. and Leu, S. S.(2011). Metaheuristics for project and construction management – A state-of-the-art review. *Automation in Construction.* vol. 20, no. 5, pp. 491–505.

[38] Siemens, N.(1971). A Simple CPM Time- Cost Tradeoff Algorithm. *Management Science*.vol. 17, no. 6, 1971.

[39] *http://osp.mans.edu.eg/elbeltagi/CM%20CH8%20Time-Cost.pdf.*

[40] Project Management - ID:5c11f28fe37f6 https://docu.tips/documents/project-management-5c11f28fe37f6.

[41] Larson, E. W., and Gray, C. F.(2018).Project management: the managerial process (2018). Seventh Ed. McGraw-Hill Education.

[42] Bromilow, N. F., Hinds, F. J., and Moody, M. F.,(1980). Australian institute of quantity surveyors: survey of building contract time performance. *The Building*

Economist. Vol.19, no. 2, pp. 79–82.

[43] Ireland, V. (1985). The role of managerial actions in the cost, time and quality performance of high-rise commercial building projects. *Construction Management and Economics, Vol. 3, No. 4, pp. 59-87.*

[44] Kaka, A. P. and Price, A. D. F. (1991).Relationship between value and duration of construction projects.*Construction Management and Economics, Vol. 9, No. 4, pp. 383-400*.

[45] Kumaraswamy, M.M. and Chan, D.W.M. (1995), *Determinants of construction duration*. *Construction Management and Economics*, vol.13,no.3,pp.209-217.

[46] Love, P. E. D.(2002).Influence of project type and procurement method on rework costs in building construction projects. *Journal of Construction Engineering and Management*, vol. 128, no. 1, pp. 18–29, 2002.

[47] Hoffman, G. J., Thal, A. E., Webb, T. S., and Weir, J. D. (2007).Estimating performance time for construction projects.Journal of Management in Engineering (ASCE), Vol. 23, No. 4, pp. 193-199.

[48] Ogunsemi, D. R. and Jagboro, G. O. (2006).Time-cost model for building projects in Nigeria.*Construction Management and Economics*,Vol. 24, No. 3, pp. 253 258.

[49] Ameyaw, C; Mensah, S and Arthur, Y D (2012) Applicability of Bromilow's Time – Cost Model on Building Projects in Ghana In: *Laryea, S., Agyepong, S.A., Leiringer, R. and Hughes, W. (Eds) Procs 4th West Africa Built Environment Research (WABER) Conference, 24-26 July 2012, Abuja, Nigeria, 881-888.*

[50] Otali, M., and Adewuyi, T. O.(2018). Project duration prediction models using time-cost relationship for public building projects in north central. *Nigerian Journal of Agriculture, Food and Environment*,vol.11, no.1, pp:137-144.

[51] Abraham, A.(2008).Time – Cost Relationships for Public Road Construction Projects in Ethiopia. MSc Thesis,AAU, Ethiopia.

[52] Andinet, Behailu, and Derbachew (2006). Time Cost Relationship: Educational Building Projects in Ethiopia. Final year project, AAU, Ethiopia.

[53] Muluken,T.(2010).Time, Cost and floor area relationship for Educational buildings in Ethiopia. MSc Thesis ,Adama University,Ethiopia.

[54] Walker, D. H. T. (1995). An investigation into construction time performance *Construction Management and Economics*.vol. 13, no. 3, pp. 263–274.

[55] Mensah, I., Nani, G., Adjei-kumi, T., and Adinyira, E.(2017).Modelling Construction

duration : A comprehensive review of literature. https://www.researchgate.net/publication/322156390

[56] A. S. H. Samprit Chatterjee, - *Regression Analysis by Example*, Fifth Edit. John Wiley & Sons,Inc.,Hoboken,New Jersey, 2012.

[57] Yeom. D. J., Seo. H.M., Kim. Y.-J., Cho. C.S., and Kim. Y.development of an approximate construction duration prediction model during the project planning phase for general office buildings. *Journal of Civil Engineering and Management,* vol. 24, no. 3, pp. 238–253.

[58] Pushkar, A., Senthilvel, M., & Varghese, K. (2018). Automated progress monitoring of masonry activity using photogrammetric point cloud. 35th ISARC.

[59] Poh, C.Q.X., Ubeynarayana, C.U., & Goh, Y.M. (2018). Safety Leading Indicators for Construction Sites: A Machine Learning Approach, Automation in Construction, Vol. 93, pp. 375-386.

[60] Kononenko, I, & Kukar, M. (2007). Machine Learning and Data Mining: Introduction to Principles and Algorithms. Horwood publishing limited, Chapter 1, pp. 1–36.

[61] Siqueira, I. (1999). Neural Network Based Cost Estimating. Thesis in Department of Building and Civil Engineering, Concordia University Montreal, Quebec, Canada.

[62] Hammad ,A., Mohamed, Y., & AbouRizk, S. (2014). Application of KDD techniques to extract useful knowledge from labor resources data in industrial construction projects. Journal of Management Engineering. DOI: 10.1061/(ASCE)ME.19435479.0000280.

[63] Petruseva, S., Pancovska,V.Z., and Zujo, V.(2013).Neural Network Prediction Model for Construction Project Duration, vol. 2, no. 11, pp. 1646–1654.

[64] Wijaya, S.H.(2003).On representing factors influencing time performance of shop-house constructions in surabaya. *Civil Engineering Dimension*. vol. 5, no. 1, pp. 1–6, 2003.

[65] Ephrem Girma Sinesilassie, Syed Zafar Shahid Tabish, Kumar Neeraj Jha, "Critical factors affecting schedule performance: a case of Ethiopian public construction projects –engineers' perspective. Engineering, Construction and Architectural Management .

[66] Abdurezak, M.K. and Neway, S.(2019).Causes of Delay in Public Building Construction Projects : A Case of Addis Abeba Administration, Ethiopia. *Asian Journal of Managerial Science.*vol. Vol.8, no. No.2, pp. 4–9.

[67] Z. E, "Inflation in Construction 2019. What Should You Carry? « Construction

Analytics. https://edzarenski.com/2018/02/15/inflation-in-construction-2019-what-should-you-carry-1/

[68] Reshma Mary Johnson & Robin Itty Ipe Babu (2018): Time and cost overruns in the UAE construction industry: a critical analysis, International Journal of Construction Management, DOI: 10.1080/15623599.2018.1484864

[69] Windapo, A., and Cattell, K.(2012).Examining The Trends In Building Material Prices: Built Environment Stakeholders' Perspectives .

[70] Ethiopian Economic Association (2006). Report on the Ethiopian Economy; the Current State of the Construction Industry.

[71] National Planning Commission (2016).Growth and Transformation Plan II (GTP II).

[72] Mengistu, D.G., & Mahesh,G. (2019): Challenges in developing the Ethiopian construction industry, African Journal of Science, Technology, Innovation and Development, DOI: 10.1080/20421338.2019.1654252

[73] Aiyetan. A. O., Smallwood. J. J., and Shakantu. W.(2012).A linear regression modelling of the relationship between initial estimated and final achieved construction time in South Africa.*Acta Structilia*, vol. 19, no. 1, pp. 39–56, 2012.

[74] Deb, D., Dey. R., and Balas. V. E.(2019).*Engineering Research Methodology. A Practical Insight for Researchers*. Springer Nature Singapore Pte Ltd,vol. 153.

[75] Kothari, C.R.(2004). *Research Methodology Methods and Techniques*. New Age International (P) Ltd., Publishers.

[76] Johnson, B., and Christensen. L.(2014).Research design: qualitative, quantitative, and mixed methods approaches.Sixth Edition.,SAGE Publications.

[77] Kumar. R.(2011).Research methodology a step-by-step guide for beginners. *SAGE Publications Ltd.*

[78] Han, J., Pei, J., & Kamber, M. (2006). Data Mining - Concepts and Techniques (2nd Ed). ISBN 13: 978-1-55860-901-3.

[79] Manikandan, S. (2011). Measures of Dispersion, Journal of Pharmacology &Pharmacotherapeutics, Vol. 2, no.4, pp. 315–316).

[80] Guyon, I., & Elisseeff, A. (2003). An Introduction to Variable and Feature Selection, Journal of Machine Learning Research. (Vol. 3, pp. 1157-118).

[81] Jovic, A., Brkic, K., & Bogunovic, N. (2015). A review of feature selection methods with applications. Opatija, Croatia. DOI: 10.1109/MIPRO.2015.7160458

[82] F. Richard and L. Anita, *Research Methods for Construction*, Fourth edi. John Wiley

& Sons, Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, United Kingdom., 2015.

[83] Williams, R. C.(2008).The Development of Mathematical Models for Preliminary Prediction of Highway Construction Duration. Ph.D. thesis, Virginia Polytechnic Institute and State University, Blacksburg, VA.

[84] Gurney, K. (1997). An Introduction to Neural Networks (1st Ed.). Master ISBN: 0-203-45151-1.

[85] Ciaburro, G., & Venkateswaran, B. (2017). Neural Networks with R: Smart models using CNN, RNN, deep learning and artificial Intelligence principles (1st Ed.). ISBN: 978-1-78839-787-2.

[86] Saleh, H. (2018). Machine Learning Fundamentals: Use Python and scikit- learn to get up and running with the hottest developments in machine learning. ISBN: 978-1-78980-355-6.

[87] Noriega, L. (2005).Multilayer Perceptron Tutorial. School of Computing, Staffordshire University, Beaconside Staffordshire ST18 0DG.

[88] Liou, J. D., Borcherding, F.S.(1986).Work sampling can predict unit rate productivity. *Journal of Construction Engineering and Management*, vol. 112, no. 1, pp. 90–103.

[89] http://www.linkedin.com/pulse.com/what-mape-mad-msd-time-series-allameh- statistics/.

[90] Rocha, A., Adeli, H., Reis, L. P., & Costanzo, S. (2018). Trends and Advances in Information Systems and Technologies. Advances in Intelligent Systems and Computing. (Vol. 3). DOI: https://doi.org/10.1007/978-3-319-77700-9.

APPENDICES

Project data collection format

An application form for the data required from the project

---

1. Stakeholders participated in construction

1.1. Client……………………………………………………………………………..

1.2. Contractor………………………………………………………………………..........

1.3. Consultant……………………………………………………….……………………

2. Project characteristics

2.1. Project location……………………………………………………………………...

2.2. Building Type ……………………………………………………………………........

2.3. Procurement method…………………………………………………………………..

3. Project scope

3.1.Gross floor area ……………………………………………………………………...

3.2.Total construction area…...

3.3.Building height……………………………………………………………………….

3.4. Floor numbers……………………………………………………………………….

3.4.1. Number of floors above ground floors……………………………………………….

3.4.2. Number of basement below ground floors……………………………………………

4. Contract and Actual Performance of the project

4.1. Date of signing the contract………………………………………………………

4.2. Contract Amount in ETB…………………………………………………………...

4.3. Date of provisional acceptance...…………………………………………………….

4.4. Actual cost project completed in ETB……….……………………………………………..

5. The amount of the date project is halted (stopped) without construction in days…..................

Private mixed use buildings data used for modelling

P/No.	Project Name	EPD (Days)	APD (Days)	EPC in M ETB	APC in M ETB	Adjusted Actual cost in M ETB	GFA	TCA	BH	F	NFAG	NFBG
1	Guest entertainment mixed use complex building	365	537	80	116.1	192.3	12856	725	53	14	12	2
2	Africa Insurance	545	1142	130	129.2	247.6	6900	507	54.87	16	13	3
3	Lion Insurance	415	1402	39.9	35.3	74.4	5628	585	29.4	9	8	1
4	Amhara National Democratic Movement(ANDM)	450	1257	107	113.5	169.2	7098	1318	22.4	7	5	2
5	Awash bank	450	1394	121.2	138.5	206.5	7600	1600	32	10	10	0
6	Dashan Bank SC, A.A	365	480	133.7	138.5	206.5	7500	1625	17.4	5	5	0
7	1B+G+13 Mixed use building	720	2200	342	379.4	565.6	20678	1920.1	53	14	13	1
8	B+G+6 Mixed use building (BRD construction plc)	800	760	149.8	119.9	208.9	10500	585	22.5	8	7	1
9	Ethiopian orthodox church B+G+5 Multi- purpose building	856	1394	151.4	103.5	171.4	9860	450	24.86	7	6	1
10	G+9 building construction	700	1401	119.9	145.6	282.2	12000	1600	32.5	10	10	0
11	G+4 Mixed- use building construction	365	485	53.3	58.3	96.6	3800	1235	13.5	5	5	0
12	Mekonnen Bitew Mixed use building(2B+G+14)	480	817	41.7	48.8	93.5	6239	525	40.3	17	15	2
13	2B+G+10 Mixed Used Building	730	1110	86.5	85.9	142.3	7600	1200	35	13	11	2
14	Tigat Mixed use building building complex	1496	2855	130	160	238.5	21500	1746	48.66	16	14	2
15	Oromia Development Association multi Purpose building	548	2170	70.5	99.6	250.9	11382	1850	29	8	7	1
16	Mixed use building for Military Tera Merchants	1305	1815	223.2	239.4	396.5	30866	4000	29	9	7	2
17	S.A Bagerisa Mixed use	620	1760	145	169	323.9	11270	1750	42	17	15	2
18	Ato Gashaw Mixed use building	620	2100	220	265.2	508.3	21934	2806	36	11	10	1
19	2B+G+9 Ethiopian Orthodox thewahdo church Mixed Used Building	730	1279	218.6	141.3	234.0	11484	1032	37	12	10	2

20	B+G+8 Kimem Mixed use Building	540	1730	75.9	79.7	152.8	7500	1071	31.4	10	9	1
21	Tigray Development Association	365	2098	130.13	121.3	255.7	14145	1600	37.5	12	10	2
22	B+G+11 Mixed use building	450	720	25	27	51.7	4450	380	39.6	12	11	1
23	Solomon Kitew B+G+9 Mixed-use Building	730	755	18.5	24	41.8	3467	450	34.78	11	10	1
24	ETCON B+G+7 Mixed use building	290	490	12.52	13.6	26.1	2100	274	28.6	9	8	1
25	AMT 3B+G+15 Mixed use Building	365	730	37.9	40.9	78.4	5965	410	57	19	16	3
26	Lentebur Multi market Center SH.Co	510	1309	73.3	80.6	140.4	11500	950	48.6	16	14	2
27	Fezak B+G+7 Mixed use Building	360	936	35.4	38.9	67.8	5690	850	24.3	8	7	1
28	Giant Eagle Trading PLC	438	942	36.9	39.9	76.5	5850	560	47.1	15	15	0
29	W/RO Huluye Enaeayehu 2B+G+12 Mixed use Building	548	1308	21.5	24	41.8	3998	305	45.9	15	13	2
30	3B+G+19 Mixed use building(gong construction)	913	1379	157.6	176.2	307.0	19450	1208	67.4	23	20	3
31	2B+G+19 Mixed use building	664	431.6	341.3	366.5	702.4	58500	3000	61.75	21	19	2
32	2B+G+14 Apartment and commercial	464	1090	56.8	57.6	110.4	9450	1320	56.4	17	15	2
33	2B+G+10 Awash Bank Balcha Aba Nefso Branch office	1095	1245	257.6	257.6	384.0	24107	1980	40.6	13	11	2
34	B+G+M+5 Mixed use Building	540	1009	109	109	172.7	10960	1507	25.7	8	7	1
35	2B+G+5 Shopping Mall building project	365	400	41	45.6	75.5	6800	1300	28.5	9	7	2
36	B+G+8 Mixed use Building	410	493	22	24.1	46.2	3280	412	29.7	10	9	1
37	Nib Bank	540	970	21.4	23.5	35.0	3011	510	25.2	8	8	0
38	Dashan Bank SC	365	910	38.7	41.6	79.7	3058	560	19.5	6	6	0
39	B+G+9 W/ro Zehara Yesuf Mixed use building	450	650	21.6	23.5	40.9	3600	540	31	11	10	1
40	B+G+9 Wendafrash Bekele Mixed use building	450	700	31.6	34.7	60.5	4500	650	32.5	11	10	1

Public office building data used for modelling

P/No.	Project Name	EPD (Days)	APD (Days)	EPC in M ETB	APC in M ETB	Adjusted Actual cost in M ETB	GFA	TCA	BH	F	NFAG	NFBG
1	Bole Sub-city	558	1760	227.13	234.77	338.1	20,658	1,589.10	45.7	13	12	1
2	Kirkos Sub-city	730	1317	297.22	280.25	403.6	20,658	1,589.10	45.7	13	12	1
3	Gulele Sub-city	558	1035	202.76	197.62	284.6	20,658	1,589.10	45.7	13	12	1
4	Addis Ketema Sub-	558	1245	228.16	237.36	341.8	20,658	1,589.10	45.7	13	12	1
5	Arada Sub-city	730	1192	290.08	220.51	317.5	20,658	1,589.10	45.7	13	12	1
6	Kolfe Sub-city	730	1223	250.48	231.92	334.0	20,658	1,589.10	45.7	13	12	1
7	Yeka Sub-city	558	1095	240.61	220.34	317.3	20,658	1,589.10	45.7	13	12	1
8	Akaki Kality Sub-city	558	1095	229.01	225.9	325.3	20,658	1,589.10	45.7	13	12	1
9	Yeka Sub-city Police	530	763	60.14	54.56	78.6	2,218	554.6	14.3	4	4	0
10	Addis Ketema Sub-city Police	530	560	63.97	53.16	76.6	2,218	554.6	14.3	4	4	0
11	Polytechnic (Tegbared) Building	365	1095	53.94	37.76	54.4	5,249	583.2	34.14	9	8	1
12	Akaki TVET Building	365	740	18.58	19.07	27.5	2,333	583.2	14.3	4	4	0
13	Entoto Technical	180	410	11.08	8.13	11.7	2,218	443.7	17.75	5	5	0
14	Kotebe UniversityCollege Building Project 1	180	977	8.99	9.4	13.5	2,500	500	16.5	5	5	0
15	Kotebe University College Building Project 2	180	702	8.2	8.27	11.9	2,500	500	16.5	5	5	0
16	Bole Sub-city Fire and Rescue Agency Building	365	365	28.46	23.87	34.4	3,166	791.5	14.3	4	4	0
17	Addis Ketema Sub- city Fire and RescueAgency Building	180	768	10.96	11.19	16.1	3,166	791.5	14.3	4	4	0
18	Gulele Sub-city Fire and Rescue Agency Building	365	949	36.51	25.56	36.8	3,166	791.5	14.3	4	4	0
19	Yekatit 12 Hospital Office Building	540	1580	178.48	197.61	284.6	10,188	1,132.00	33.41	9	8	1
20	Zewditu Hospital Office Building	365	480	8.19	11.56	16.6	2,162	432.5	17.75	5	5	0

21	Justice Bureau Building	540	803	66.64	61.42	88.4	8,870	633.6	48.54	14	13	1
22	Environment Protection Office Building	485	1760	43.51	46.75	67.3	3,857	964.2	15	4	4	0
23	Addis Ababa Construction Bureau Laboratory Building	180	730	13.48	14.12	20.3	2,728	454.6	20.7	6	5	1
24	Education Bureau Building	540	1085	60.34	56.47	81.3	8,870	633.6	48.54	14	13	1
25	Ethiopian Sport Commision Office Building	540	1420	45.16	51.63	74.3	8,870	633.6	48.54	14	13	1
26	Urban Integrated Land Information System Development Office Building	540	1,273.00	166.26	160.42	231.0	13,500	900	46.5	15	13	2
27	Ministry of Science and Technology Office Building	540	1,385.00	30.56	35.23	50.7	7,200	800	26.7	9	9	0
28	Industry Minister Office Building	540	1,547.00	30.48	33.76	48.6	8,800	800	34.5	11	11	0
29	Central Statistics Agency Office Building	555	1,273.00	23.65	27.35	39.4	5,853	731.6	28.38	8	7	1
30	Ethiopian Roads Authority Office Building	590	1,742.00	39.66	33.88	48.8	5,600	800	22.92	7	7	0
31	Ministry of Foreign Affairs Conference Hall Project	426	1,505.00	162.61	190.27	274.0	12,400	3,100.00	16.3	4	3	1