Ministry of Higher Education
and Scientific Research

University of Diyala

College of Science

Investment Projects Optimization using Data Mining Technique

A Thesis Submitted to Council of College of Science, University of Diyala in Partial Fulfillment of the Requirements for the Degree of Master of Computer Science

by

Atezaz Ahmed Abduljaleel

Supervised by

Prof.Dr Dhahir Abdulhade Abdulah

2020A.D IRAQ 1442 A.

بسم الله الرحمن الرحيم

وَقُلِ اعْمَلُوا فَسَيَرَى اللَّهُ عَمَلَكُمْ وَرَسُولُهُ وَالْمُؤْمِنُونَ ۖ

صدق الله العلي العظيم

[سورة التوبة105]

Dedication

I dedicate this research

To the Prophet And the savior of the nation Muhammad Abdullah Abdul-Muttalib,To the Commander of the Believers Ali bin Abi Talib, To our Lord Imam Al – Mahdi,To the mother of believers Fatima Zahra peace on them.

To my parents whose their prayers and words always inspired and encourage me to give more and pursuit of excellence.

To my beloved husband Ahmed for his support, encouragement to complete this thesis with his concern.

My Sister and Brothers , who always encourage me to give the best and supported me.

Acknowledgments

I would like to express my deepest thanks and gratitude to my supervisor Prof. Dr. Dhahir Abdulhade Abdulah. Through his vast experiences in this field of study, he has offered a valuable time, expertise, guidance, and encouragement.

I also would like to acknowledge the Department of computer science – University of Diyala for all of the support that they have offered.

Linguistic Certification

This is to certify that this thesis entitled

" Investment Projects Optimization using Data Mining Technique" was prepared under my linguistic supervision. It was amended to meet the style of English language**.**

Signature :
Name :
Date: / / 2020

Supervisor's Certification

I certify that this thesis entitled "Investment Projects Optimization using Data Mining Technique", was prepared under my supervision at Department of Computer Science\ College of Sciences\ University of Diyala by Atezaz Ahmed Abduljaleel, as a partial fulfillment of the requirements for the degree of Master of Science in Computer Science

(Supervisor)
Signature:
Name: Prof. Dr. Dhahir Abdulhade Abdulah
Date: / / 2020

Approved by University of a Diyala Faculty of Science Department of Computer Science.

Signature:

Name : Assist. Prof. Dr. Taha M. Hassan

Date : / / 2020

(Head of Computer Science Department)

ABSTRACT

Data mining processes are used to build machine learning models that power applications including search engine technology and website recommendation programs. The aim of this thesis is to form a system and regard an operative to accomplish the investment projects objective in term of both duration and charge by using the techniques: genetic algorithm (GA) and gravitational search algorithm (GSA).

To achieve the research aim of the research a system was built in order to optimize the solution for the investment projects problems and increase the number of projects to be implemented.

The results show that the there is confusion in the work of the Iraqi governments and the national Investment Commission in creating a favorable investment environment that holds the elements of a successful investment environment. GSA find the objective function according to its round from 4 to 5 iteration while GA more than 8 iteration.

GSA algorithm operative method in determining the greatest solution and it shorter time. GA technique consume a lot of time in finding the solution and managing the investment projects and sometime yield to non-optimal solution.

As comparing to GSA, the value of the objective function for the profit that is chosen is very low and leads to reduce the number of the project implemented , as GA is fall in under fitting and the steps as selection, cross over and mutation lead to this value. The difference between GA and GSA is also very

clear in the effectiveness of the solution, as xij also has lower value than GSA which lead to reduce the effectiveness. In most of the governorate the number of the projects that can be maximize increase more than its originals as they reach 91 projects as in Baghdad with profit more than 4 billion dollar. The number of project to be implemented by using GSA in National governorate about 101 projects comparing with GA which only give 51 projects.

In Diyala Governorate, 5 projects were taken and each project has three significant problems with each one of them has three solutions, GSA found the best solution with effectiveness of 0.89 and the same for GA and its seem to be equal in number of problems which mean show an equal effect.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATION

Abbrev.	Total Name
AI	Artificial Intelligence
GSA	Gravitational Search Algorithm
GA	Genetic algorithm
PSO	Practical Swarm Optimization
ANN	Artificial Neural Network
PSPLIB	Project Scheduling Problem Library

CHAPTER ONE

INTRODUCTION

Chapter one

Introduction

1.1 An Overview

The process of mining the necessary information from data that are previously exist called data mining, in addition it was given different name as extraction of knowledge, detection of information, harvesting of information, data archeology, and processing of data pattern[1].

The main aim of data mining is to recognize a useable novel, likely consider valuable, and capable to be understood and there must be associations and existing data patterns [2]. The statisticians mostly used data mining term as well as researchers of database, and business communities as it aids them to get clear idea about hidden pattern and it is necessary for understanding.

There is a gap between the traditional construction industry and using data mining in construction as when it compares with other industries. This could be return to the reason that construction process is an impermanent and specific action that means the one project data can rarely be used for alternative project. But one can ask it may not be all the time true, as though products of the construction are exceptional but some resemblances can be exist among them and processes of construction and management skills are typically common to all projects[3].

In spite the fact that the construction industry did not applied the data mining application, researchers of construction management have been examining its ability to be implemented to different problem regions [4].The investment in the industry of construction face an issue in contemporary economic circumstances. Construction is a fragment of the

real sector economy, that point to the evolving capital construction position in the country. The fact that the economy of the country is going through of some difficulty there is an depletion of foreign investment[5].

Various project in the investment construction stages application require needs financial investments, that are resolute by the processes of cost management[6]. The preparation and conduction of business operations contain the following:

• Focus on approach of quantitative

• Focus on the process

• Focus on working groups [7] .

1.2 Literature Review

Below is a review of some researches which are related to our work:

Abraham Warszawski, F and Rafael Sacks [2014] [8] : Introduce a useful approach in the risk that originate in the investment construction project that can calculate with input information of different detail levels. The suggested multifactor approach comprises concern of project‘s risk factors interdependence.

Sofia Kaiafa and Athanasios P. Chassiakos [2015] [9] : They introduce a method for reducing the cost from over allocation of the resource project exceedance of the deadline, and fluctuations of resource. As there are large number of alternatives for activity implementation, a genetic algorithm was used for the optimization process. The algorithm

has been verified with number of test cases and there were comparison with those established by the Microsoft Project. The evaluation shows that the suggested algorithm can offer satisfactory and balanced solutions with respect to the three objectives and that these solutions are improved than those offered by commercial project scheduling software.

Jia Liu, et.al (2018) [10] : Introduce (GA) for the RCPSP. The suggested algorithm provide number of variations in the paradigm of the GA, as innovative selection operator to choose parents to rebound; an improved two-point operator of crossover with a precise crossover imperative. The suggested algorithm was verified by problems of standard level of size J30, J60, and J120 from "Project Scheduling Problem Library" (PSPLIB) and associated with 19 research in the met heuristics in the previous studies. The finding validate that the proposed algorithm is an inexpensive algorithm for RCPSP solving .

Panagiotis M. et.al [2018] [11]: developed optimization model of multi-objective dynamic for construction and moving facilities costs, resources transportation costs moving from different facilities or to workplaces, a genetic algorithm has been applied as it has the ability to strongly search within a great solution space. Numerous case studies and working scenarios have been applied through the software of Palisade‘s Evolver for analyzing the model and evaluation. The finding show acceptable model response to input of the time data with quality of solution and calculation time. The model can offer decision sustenance to managers of the site, allowing them to inspect substitute scenarios and best solutions conferring to their knowledge by producing necessary preferences.

Victoria Borkovskaya, et.al [2019] [12] : they presented an approach to improve the investment construction efficiency of given endless danger of new risks. examine the kinds and reasons of risks, make conclusions about the risks rank related with the investment attraction of projects. Development and vicissitudes in the surrounding area are the leading direction in the modern society development**.**

Jingkuang Liu et.al [2019][13]: GA algorithm was used for the choice of site of a recycling plant. The study reveal the best solution gained on enhancing the choice of the site of a construction and destruction plant of waste recycling by GA imitates the real examination. The best solution by GA was gained after 200 repetitions, at point the value of fitness converges at a steady value of 1.8×10^{-5}, which verifies the judiciousness and workability of the model of the site-selection optimization. Though, given the slow speed of the evolutionary GA, it is easy to drop into a local finest.

Seyed Kamal Chaharsooghi et.al [2019] [14] : the project cash flow problem was solved by Genetic algorithm optimization. The key hypothesis that project inputs and outputs can be assessed in terms of project cash flow. the model, the objective function is present value of project cash flow maximization . Then, after addition of the constraints. It is supposed that the suggested approach may also be valuable for both managing of project cash flow and for control of project.

Peng Rong and others [2020] [15] : provide model from project cost analysis and data mining depend on the estimate of massive feasibility study, initial enterprise estimate and final explanation data of power grid infrastructure projects, backup the amount estimate of the investment in

the stage of planning, the estimate review of the investment in the stage of the feasibility study, and the prediction of the investment control target in the stage of planning, so as to enhance the investment efficacy of electric grid project.

Naji Mutar et.al [2020] [16] : they study the issues through the pre-construction phase by using PSO and GSA. The finding show GSA and PSO are both used and show outstanding results in solving problems, the algorithm of GSA shows improved results in both the velocity and in the accuracy.

Mohammed K. Al Mhdawi [2020] [17] : Improve methodology for integrated decision support for risk factors managing in Iraq oil and gas construction projects.. The suggested methodology involves of the following phases: analysis of risk using a model of multi-criteria risk analysis based on fuzzy set theory, prediction of risk effect on project time and cost using artificial neural network (ANN), risk response selection using (GSA) optimization technique. The implemented methodology will enable decision makers to evaluate the risky events of oil and gas projects, provision their decision through planning and stages of work implementation.

The previous studies show that the investment projects are mostly is been managed without using data mining techniques or decision making techniques but algorithm as GA and GSA are rarely used. Therefore the authors used data mining techniques as tool for decision making in investment projects to increase the profitability of the projects.

1.3 Problem Statement

In Iraq, investment projects aspect several issues as the most projects are being built in an unbalanced situation in term of price and

period and thus most of the investment projects are either not be awarded or stooped in early stages which lead to rise of finical crises.

The thesis problem can be explained as follows:

1. Iraqi instruction and rules regard the investment cause undesirable environment to the investors in term of profit which require a system to manage the investment projects.
2. The size of investment projects and its difficulty need to the existence of a system in order to increase the awarded projects .
3. The unstable environment of the investment make the investor without any profit which requires a system to increase the profit for the investors.
4. The cost and time of the discovery of the optimal solution for the problems of investment projects is high and consequently it needs an integrated system.

1.4 Thesis Aim and Objectives

The aim of this thesis is to form a system and regard an operative to accomplish the investment projects objective in term of both time and cost. Attaining the aim, there are some purposes should be gained as follows:

1- Investigate the number of stooped investment projects.

2- Examination and documentation of the difficulties in construction projects.

3- Determining the consequence of these difficulties on cost and time of the investment projects.

4- Find the solutions for these problems with effect on both cost and time using GA and GSA.

5- Maximize the profit for the investors using GA and GSA.

6- Build system to accomplish these solutions by using the techniques: GSA and GA.

1.5 Thesis Structure:

The research includes several chapters, which are:

Chapter Two: Deals with the investment definition, investment projects, their type the problems associated with the investment along with the techniques that used.

Chapter Three: Presents the proposed system in the design of the system to manage investment project problems, the system consist of several parts and two type of algorithms which are GSA and these will be explained in this chapter.

Chapter Four: The results obtained from applying the proposed algorithms for the investment projects in terms of number of the project to be proceed ,profit gain and the solution for the problems.

Chapter Five: This chapter comprises conclusions that obtain by the researcher as well as commendations and proposals for subsequent research.