FACULTY OF COMPUTING STAFF PUBLICATION DASHBOARD

ADIB BIN MORSHED

UNIVERSITI TEKNOLOGI MALAYSIA

PSZ 19:16 (Pind. 1/13)

**UNIVERSITI TEKNOLOGI MALAYSIA**

**DECLARATION OF THESIS / UNDERGRADUATE PROJECT REPORT AND COPYRIGHT**

Author’s full name : ADIB BIN MORSHED

Date of Birth : 19/09/2001

Title : FACULTY OF COMPUTING STAFF PUBLICATION DASHBOARD

Academic Session : 2022/23

I declare that this thesis is classified as:

|  |  |  |
| --- | --- | --- |
|  | CONFIDENTIAL | (Contains confidential information under the Official Secret Act 1972)\* |
|  |  |  |
|  | RESTRICTED | (Contains restricted information as specified by the organization where research was done)\* |
|  |  |  |
| ✓ | OPEN ACCESS | I agree that my thesis to be published as online open access (full text) |

I acknowledged that Universiti Teknologi Malaysia reserves the right as follows:

The thesis is the property of Universiti Teknologi Malaysia

The Library of Universiti Teknologi Malaysia has the right to make copies for the purpose of research only.

The Library has the right to make copies of the thesis for academic exchange.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  | Certified by: |  |

NOTES : If the thesis is CONFIDENTIAL or RESTRICTED, please attach with the letter from the organization with period and reasons for confidentiality or restriction

“I hereby declare that we have read this thesis and in my

opinion this thesis is suffcient in term of scope and quality for the

award of the degree of Bachelor of Computer Science (Computer Networks & Security)”

|  |  |  |
| --- | --- | --- |
| Signature | : | \_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_ |
| Name of Supervisor | : | **Prof. Madya. Ts. Dr. Mohd Shahizan bin Othman** |
| Date | : | 13th may,2023 |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

FACULTY OF COMPUTING STAFF PUBLICATION DASHBOARD

ADIB BIN MORSHED

A thesis submitted in fulfilment of the

requirements for the award of the degree of

Bachelor of Computer Science (Computer Networks & Security)

School of Computing

Faculty of Engineering

Universiti Teknologi Malaysia

JULY 2023

**DECLARATION**

I declare that this thesis entitled *“On-Line Recognition of Developing Control Chart Patterns”* is the result of my own research except as cited in the references. The thesis has not been accepted for any degree and is not concurrently submitted in candidature of any other degree.

|  |  |  |
| --- | --- | --- |
| Signature | : | .................. .................................. |
| Name | : | ADIB BIN MORSHED |
| Date | : | 13th May,2023 |
|  |  |  |
|  |  |  |

**DEDICATION**

This thesis is dedicated to my father, who taught me that the best kind of knowledge to have is that which is learned for its own sake. It is also dedicated to my mother, who taught me that even the largest task can be accomplished if it is done one step at a time.

**ACKNOWLEDGEMENT**

In preparing this thesis, I was in contact with many people, researchers, academicians, and practitioners. They have contributed towards my understanding and thoughts. In particular, I wish to express my sincere appreciation to my main thesis supervisor, Professor Dr. XX, for encouragement, guidance, critics and friendship. I am also very thankful to my co-supervisor Professor Dr YY and Associate Professor Dr. ZZZ for their guidance, advices and motivation. Without their continued support and interest, this thesis would not have been the same as presented here.

My fellow student should also be recognised for their support. My sincere appreciation also extends to all my colleagues and others who have provided assistance at various occasions. Their views and tips are useful indeed. Unfortunately, it is not possible to list all of them in this limited space. I am grateful to all my family member.

**ABSTRACT**

The purpose of this study is to investigate the application of genetic algorithm (GA) in modelling linear and non-linear dynamic systems and develop an alternative model structure selection algorithm based on GA. Orthogonal least square (OLS), a gradient descent method was used as the benchmark for the proposed algorithm. A model structure selection based on modified genetic algorithm (MGA) has been proposed in this study to reduce problems of premature convergence in simple GA (SGA). The effect of different combinations of MGA operators on the performance of the developed model was studied and the effectiveness and shortcomings of MGA were highlighted. Results were compared between SGA, MGA and benchmark OLS method. It was discovered that with similar number of dynamic terms, in most cases, MGA performs better than SGA in terms of exploring potential solution and outperformed the OLS algorithm in terms of selected number of terms and predictive accuracy. In addition, the use of local search with MGA for fine-tuning the algorithm was also proposed and investigated, named as memetic algorithm (MA). Simulation results demonstrated that in most cases, MA is able to produce an adequate and parsimonious model that can satisfy the model validation tests with significant advantages over OLS, SGA and MGA methods. Furthermore, the case studies on identification of multivariable systems based on real experiment t al data from two systems namely a turbo alternator and a continuous stirred tank reactor showed that the proposed algorithm could be used as an alternative to adequately identify adequate and parsimonious models for those systems. Abstract must be bilingual. For a thesis written in Bahasa Melayu, the abstract must first be written in Bahasa Melayu and followed by the English translation. If the thesis is written in English, the abstract must be written in English and followed by the translation in Bahasa Melayu. The abstract should be brief, written in one paragraph and not exceed one (1) page. An abstract is different from synopsis or summary of a thesis. It should states the field of study, problem definition, methodology adopted, research process, results obtained and conclusion of the research. The abstract can be written using single or one and a half spacing. Example can be seen in Appendix 1 (Bahasa Melayu) and Appendix J (English).

**ABSTRAK**

Kajian ini dilakukan bertujuan mengkaji penggunaan algoritma genetik (GA)

dalam pemodelan sistem dinamik linear dan tak linear dan membangunkan kaedah alternatif bagi pcmilihan struktur model menggunakan GA. Algorithma kuasa dua terkecil ortogon (OLS), satu kaedah penurunan kecerunan digunakan sebagai bandingan bagi kaedah yang dicadangkan. Pcmilihan struktur model mengunakan kaedah algoritma genetik yang diubahsuai (MGA) dicadangkan dalam kajian ini bagi

mengurangkan masalah konvergens pramatang dalam algoritma genetik mudah (SGA). Kesan penggunaan gabungan operator MGA yang berbeza ke atas prestasi model yang terbentuk dikaji dan keberkesanan serta kekurangan MGA diu t arakan. Kajian simulasi dilakukan untuk membanding SGA, MGA dan OLS. Dengan meggunakan bilangan parametcr dinamik yang setara kajian ini mendapati, dalam kebanyakan kes, prestasi MGA adalah lebih baik daripada SGA dalam mencari penyelesaian yang berpotensi dan lebih berkebolehan daripada OLS dalam menentukan bilangan sebutan yang dipilih dan ketcpatan ramalan. Di samping itu, penggunaan carian tcmpatan dalam MGA untuk menambah baik algorithma tersebut dicadang dan dikaji, dinamai sebagai algoritma mcmetic (MA). Hasil simulasi menunjukkan, dalam kebanyakan kes, MA berkeupayaan menghasilkan model yang bersesuaian dan parsimoni dan mcmenuhi ujian pengsahihan model di samping mcmperolehi beberapa kelebihan dibandingkan dengan kaedah OLS, SGA dan MGA. Tambahan pula, kajian kes untuk sistcm berbilang pcmbolehubah menggunakan data eksperimental sebenar daripada dua sistem iaitu sistem pengulang-alik turbo dan reaktor teraduk berterusan menunjukkan algoritma ini boleh digunakan sebagai alternatif untuk mcmperolehi model termudah yang memadai bagi sistcm tersebut.

**TABLE OF CONTENTS**

|  |  |  |
| --- | --- | --- |
|  | **TITLE** | **PAGE** |

Table of Contents

[CHAPTER 1 INTRODUCTION 1](#_Toc134837898)

[1.1 Introduction 1](#_Toc134837899)

[1.2 Problem Background 2](#_Toc134837900)

[1.3 Project Aim 3](#_Toc134837901)

[1.4 Project Objectives 3](#_Toc134837902)

[1.5 Project Scope 3](#_Toc134837903)

[1.6 Project Importance 4](#_Toc134837904)

[1.7 Report Organization 5](#_Toc134837905)

[CHAPTER 2 LITERATURE REVIEW 7](#_Toc134837906)

[2.1 Introduction 7](#_Toc134837907)

[2.2 Current System Analysis 7](#_Toc134837908)

[2.3 Comparison between existing systems 13](#_Toc134837909)

[2.4 Literature Review of Technology Used 14](#_Toc134837910)

[2.5 Chapter Summary 15](#_Toc134837911)

[CHAPTER 3 SYSTEM DEVELOPMENT METHODOLOGY 16](#_Toc134837912)

[3.1 Introduction 16](#_Toc134837913)

[3.2 Methodology Choice and Justification 16](#_Toc134837914)

[3.3 Phases of the Chosen Methodology 16](#_Toc134837915)

[3.4 Technology Used Description 17](#_Toc134837916)

[3.5 System Requirement Analysis 17](#_Toc134837917)

[3.6 Chapter Summary 18](#_Toc134837918)

[CHAPTER 4 REQUIREMENT ANALYSIS AND DESIGN 19](#_Toc134837919)

[4.1 Introduction 19](#_Toc134837920)

[4.2 Requirement Analysis 19](#_Toc134837921)

[4.3 Project Design 19](#_Toc134837922)

[4.4 Database Design 19](#_Toc134837923)

[4.5 Interface Design 20](#_Toc134837924)

[4.6 Chapter Summary 20](#_Toc134837925)

[CHAPTER 5 IMPLEMENTATION AND TESTING 22](#_Toc134837926)

[5.1 Introduction 22](#_Toc134837927)

[5.2 Coding of System Main Functions 22](#_Toc134837928)

[5.3 Interfaces of System Main Functions 22](#_Toc134837929)

[5.4 Testing 23](#_Toc134837930)

[5.4.1 Black box Testing 23](#_Toc134837931)

[5.4.2 White box Testing 24](#_Toc134837932)

[1.2.2 User Testing 25](#_Toc134837933)

[5.5 Chapter Summary 25](#_Toc134837934)

[CHAPTER 2 CONCLUSION 26](#_Toc134837935)

[2.2 Introduction 26](#_Toc134837936)

[2.3 Achievement of Project Objectives 26](#_Toc134837937)

[2.4 Suggestions for Future Improvement 26](#_Toc134837938)

**LIST OF TABLES**

|  |  |  |
| --- | --- | --- |
| **TABLE NO.** | **TITLE** | **PAGE** |

[Table 2.1](#_1ci93xb) Example Repeated Header Table 6

**LIST OF FIGURES**

|  |  |  |
| --- | --- | --- |
| **FIGURE NO.** | **TITLE** | **PAGE** |

[Figure 2.1](#_2xcytpi) Continuous variability reduction using SPC chart (Revelle and Harrington, 1992) 5

[Figure 3.1](#_2grqrue) Example of Formatting Method 10

[Figure A.1](#_25b2l0r) xxxxxxxxxxxxxxxx 23

**LIST OF ABBREVIATIONS**

|  |  |  |
| --- | --- | --- |
| ANN | - | Artificial Neural Network |
| GA | - | Genetic Algorithm |
| PSO | - | Particle Swarm Optimization |
| MTS | - | Mahalanobis Taguchi System |
| MD | - | Mahalanobis Distance |
| TM | - | Taguchi Method |
| UTM | - | Universiti Teknologi Malaysia |
| XML | - | Extensible Markup Language |
| ANN | - | Artificial Neural Network |
| GA | - | Genetic Algorithm |
| PSO | - | Particle Swarm Optimization |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

**LIST OF SYMBOLS**

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

**LIST OF APPENDICES**

|  |  |  |
| --- | --- | --- |
| **APPENDIX** | **TITLE** | **PAGE** |

[Appendix A](#_3q5sasy) Mathematical Proofs 23

[Appendix B](#_34g0dwd) Psuedo Code 25

[Appendix C](#_1jlao46) Time-series Results Long 26

# INTRODUCTION

## Introduction

Over the years, technology has advanced to a stage where we’re able to make everything organized into a single platform with different components that have been able to solve many real-life problems. Data is one of the most important factors in the advancement of technology in the modern world. Modern technology is dependent on the data. It is without a doubt that technology has simplified our lives and has made changes in the way people used to work or think before. Without proper data collection, no system can be developed whether it’s related to technology, science, or any other relevant fields. Data is extracted and refined using certain techniques and methods. Web Scraping is one such technology. Web Scraping is the process involving the extraction of data from a source using automated software and tools. It has grown more popular over the years since its utilization boosts the businesses of different organizations or individuals. In recent years, it has helped different organizations and individuals gain access to useful information that can be used for a wide range of tasks like marketing, research, analysis, etc. For businesses looking to gain a competitive edge in the industry, web scraping has become an essential tool for them.

Web scraping-related technology has advanced significantly over time. Web scraping was a labour-intensive manual procedure in the early days of the internet that entailed manually copying and pasting data from webpages into spreadsheets. The technology evolved in tandem with the growth and complexity of the internet. Various organizations and companies use this technology to analyze their data over the years and work on improving their system in the future. For universities like UTM which is a research-intensive institute, it is very important to keep track of the publications made by the staff of different faculties. One of the main faculties of UTM is the faculty of computing. Web Scraping is one of the very good approaches to keeping track of the publications made by the faculty of computing staff. With this approach, we can analyze the research output and evaluate the current state. A dashboard will be more insightful for viewing the outputs of research and publications made by the staff from the faculty of computing,

In a nutshell, web scraping can assist faculty members with keeping track of faculty publications, identifying patterns in research output, and automating the process of updating publication data by giving them a comprehensive and up-to-date perspective of research output.

## Problem Background

One of the cornerstones of UTM is the Faculty of Computing. The staff of the faculty of computing not only excels in academic teaching but also in research. University Teknologi Malaysia (UTM) is a research-intensive university. The Faculty of Computing at the Universiti Teknologi Malaysia (UTM) has many faculty members who publish their research papers, articles, and other academic works. Keeping track of all the publications from each faculty member can be a challenging task, particularly when this information is dispersed over numerous websites and databases. It is important to keep track of the publications to do an analysis of the results from the staff of faculty of computing staff. Through this approach, the faculty members can keep track of their latest research updates and can analyze them accordingly to make better progress in the future. It can be challenging to obtain an accurate picture of the faculty's research output because of discrepancies in data and inaccuracies caused by manual reliance on individual faculty members. Through this, faculty members may risk missing out on funding or partnership opportunities as well as possible avenues for research advancement without a thorough and current perspective of research output.

## Project Aim

The purpose of this project is to develop a web-based dashboard system that’ll highlight publications, indexed publications, non-indexed publications, and citations to improve the decision-making process and can assist faculty members with keeping track of faculty publications, identifying patterns in research output, and automating the process of updating publication data by giving them a thorough and current perspective of research output.

## Project Objectives

The objectives of the project are given below:

1. To gather and analyze data on the publication output of faculty members in the Faculty of Computing at UTM.
2. To provide a centralized and easily accessible platform for faculty members to track their own publication output as well as the output of their colleagues.
3. To enable the faculty to identify research strengths and areas for improvement within the department.
4. To support accreditation and performance evaluations by providing reliable and up-to-date data on faculty publication output.
5. To provide a comprehensive and user-friendly platform for tracking and analyzing publication output within the Faculty of Computing at UTM, with the goal of improving research outcomes and informing decision-making.

## Project Scope

The scope of the web scraping project for the faculty of computing staff journal publication at UTM includes the collection and analysis of publication data from various sources and the creation of a user-friendly dashboard. The project will prioritize data security and privacy while also enabling the faculty to track their own publication output, identify research strengths, and inform decision-making.

1. Web scraping data from various sources including the UTM website, Google Scholar, and other academic databases.
2. Creating a dashboard that allows users to view and analyze publication data based on various criteria such as author, publication year, journal, and citation count.
3. The project will use appropriate web scraping and data visualization tools based on the requirements and constraints of the project.

## Project Importance

The main idea of this project is to provide real-time information on faculty activities using web scraping techniques. The main aim of this project is to give an idea to improve the decision-making process by providing accurate and timely information to the administration and faculty. In today’s world, many organizations use the data acquisition technique to make future decisions on the improvement of the system. The type of data that’ll be collected and included are publications, indexed publications, non-indexed publications, and citations. This sort of data is important for analyzing the effectiveness of the research published by the faculty of computing staff.

With web scraping, the analysis of the information gathered can be used to identify trends, patterns, and insights to make informed decisions. The Dashboard developed during the project can be used for a variety of purposes, including monitoring faculty performance, identifying areas for improvement, tracking research activity, and fostering collaboration among faculty. For example, a dashboard can provide information on the number of papers each faculty authored, the number of indexed and non-indexed publications, and the number of citations received. This data will help to identify highly productive faculty members and will guide the allocation of R&D resources.

Additionally, the Dashboard can also be used to monitor the status of grant applications. This information can facilitate effective communication and collaboration among faculty members, leading to successful research projects. The dashboard will also provide information about other resources such as income generation, training conducted, and conferences organized. This information will help the faculty members to work on their professional development and can increase the visibility of goals set by the faculty. The administration can also allocate resources more effectively. The dashboard will highlight the networking opportunities which will help the faculty members to recognize the collaboration standards and work on them.

In summary, the project is an important attempt to boost decision-making processes, foster collaboration among faculty members, and enable effective resource management. This project uses web scraping techniques to collect real-time information from various sources that can be used to identify trends, patterns, and insights.

## Report Organization

Chapter 1 provides a summary of the project's background, including the introduction, problem statement, project aim, objectives, scopes, project importance, and report structure. Chapter 2 elaborates on the literature review for this project. which will be used to develop a dashboard containing research publications using web scraping. Chapter 3 describes the methodology of the system development process as well as the hardware and software required to develop the project. All of the system's requirement analyses and designs are shown in Chapter 4. The system's implementation (code) and testing are discussed in Chapter 5 to ensure quality. Finally, chapter 6 is the conclusion, which includes a review of the conclusions as well as suggestions for further studies.