



**UNIVERSITY OF SCIENCE AND TECHNOLOGY  
COLLEGE OF GRADUATE STUDIES AND ACADEMIC  
ADVANCEMENT**

Faculty of Computer Science and Information  
Technology

**Enhancement of Data to Support Decision Using Data  
Quality Technique**

by

Modwi Abdaletif Hassan Osman

A Thesis

Submitted to the College of Graduate Studies and Academic Advancements  
in Partial Fulfillment of the Requirement for the Degree of Master of Science  
in Information Technology

Supervisor

Dr: Atif Ail Mohammed

September 2015

## Abstract

The research problem is in need in institutions across the board minutes for the data to be utilized in the form required taking decisions Correct that the use of data quality techniques provides institutions across the board the ability to explore and focus on the most important information in databases , and data quality as well as focus on building future forecasts and explore the behavior and trends , which allows to take the right decisions and take timely that the use of data quality provides for institutions and companies in all areas follows. , ability to explore and focus on the most important information in databases, the ability to build future forecasts and explore the behavior and trends The ability taking and the right decisions at the right time, Revision of the data and prepare the data is incomplete for taking the right decision, get ready and incomplete information to process data analysis to take advantage of them. In this research were applied techniques data cleaning products and problem-free data can help in decision-making.

كلية علوم الحاسوب وتقانة المعلومات  
الدراسات العليا

## المستخلص

إن مشكلة البحث هي حوجة المؤسسات في جميع المجالات لبيانات دقيقة لكي تتم الاستفادة منها بالشكل المطلوب في اتخاذ القرارات الصحيحة ان استخدام تقنيات جودة البيانات يوفر للمؤسسات في جميع المجالات القدرة على استكشاف و التركيز على أهم المعلومات في قواعد البيانات، كما تركز جودة البيانات كذلك على بناء التنبؤات المستقبلية و استكشاف السلوك و الاتجاهات مما يسمح باتخاذ القرارات الصحيحة و اتخاذها في الوقت المناسب ان استخدام جودة البيانات يوفر للمؤسسات و الشركات في جميع المجالات الاتي: القدرة علي استكشاف و التركيز علي اهم المعلومات في قواعد البيانات القدرة علي بناء التنبؤات المستقبلية و استكشاف السلوك و الاتجاهات القدرة باتخاذ القرارات الصحيحة واتخاذها في الوقت المناسب تنقيح البيانات و تحضير البيانات بشكل مكتمل لاتخاذ القرار الصحيح الحصول علي معلومات مكتملة وجاهزة لعملية تحليل البيانات للاستفادة منها في هذا البحث تم تطبيق تقنيات تنظيف البيانات والحصول علي بيانات خالية من المشاكل يمكن ان تساعد في اتخاذ القرار.

كلية علوم الحاسوب وتقانة المعلومات  
الدراسات العليا

## Introduction

This chapter gives general review for the Research containing background about the Research field. Also it covers the problem statement and objectives research questions, and at the end mention from Research.

### 1.1 Background

We live in an era of unprecedented data abundance and aggregation. The sheer variety of new information available on the Internet, in databases, and from other sources has changed the way we conduct business, undertake research, and communicate. Most of the changes are positive. Yet, increased reliance upon networked data has also introduced new challenges. One serious problem we need to address is that of incorrect data missing or inaccurate information that resides in (and, indeed, frequently results from) the abundance and aggregation of data in our lives today Dirty data can have several pernicious effects. In particular, it:

1. Impacts the quality of care.
2. Introduces privacy and other civil liberty concerns.
3. Increases costs and inefficiencies.
4. Creates liability risks and Undermines the reliability and benefits of information technology (IT) investments, including the potential to streamline service delivery, accounting, and billing. Wrong data to system which will effect in business, prevent system itself from unauthorized access.

### 1.2 Problem Statement

The important of data quality has a high priority now a days so the problem of the issues is missing data, data uncompleted, data duplication, data heterogeneity. The effect of using data without cleaning.

### 1.3 Main Objectives

This main objective of the research is to applying data cleaning tools in raw data to enhance its quality for using data in supporting Decision.

## 1.4 Key Research Question

We attempt to enhance data quality on this the key research questions came as follow:

1. How we can improve the data to provide the proper decision making.
2. What is data quality?
3. Which tool are suitable to enhance data quality?
4. How implement data quality steps using weka Data mining tool?
5. How to enhance data quality and audit system?

## 1.5 Research Organization

Chapter two illustrates the pervious researches in the same filed; Chapter three states the Preprocessing work and method/methodology which used in the researches, Chapter Four Implementation and Results and discussion, Chapter Five conclusion and recommendations.

كلية علوم الحاسوب وتقانة المعلومات  
الدراسات العليا