AN APPROACH TO IMPROVE DATA QUALITY FROM VERACITY OF DATA ACCURACY FOR SENSITIVE COST AND TIME INDICATORS

Submitted by

Banan Aref Mohammad

Supervisor

Prof. Dr. Mohammed Alfayomi

Co-supervisor

Dr.Wael Alzyadat

This thesis was submitted in partial fulfillment of the requirements for the Master's Degree of Science in Software Engineering

Faculty of Information Technology

ISRA University

December 2018

The undersigned have examined the thesis entitled, An Approach To Improve Data Quality From Veracity Of Data Accuracy For Sensitive Cost And Time Indicators, presented by Banan Aref Mohammad, a candidate for the degree of Master of Science in Software Engineering and hereby certify that it is worthy of acceptance.

5/3/2019

Date

Prof.Dr. Mohammed Alfaiomi

5/3/2019

Date

Dr. Wael Zyadaat

5/3/2019

Date

Dr. Thamer Alrousaan

5/3/2019

Date

Dr.Bilal Ibrahim Sowan

DEDICATION

This thesis is dedicated

To my homeland Palestine and its capital city Jerusalem where I will go back one day.

To the kings of my life; my best parents who are the reason of my success in everything with all love and support to achieve my goals.

To my supervisor and co-supervisor for their patience, support and valuable advice and guidance.

To my sisters who shared with me all the steps and details of this work.

To my doctors in master degree study who granted me all their efforts during my study.

To my colleagues during my study for their support and encouragement.

To everyone who helped me even only by some word to convert my dream into reality.

I dedicate this achievement with sincere thanks from my heart

Banan Aref Mohammed

January.2019 .

ACKNOWLEDGMENT

All praise and thanks to Allah the almighty for His guidance and blessings and for granting me knowledge, patience and strength to accomplish this work successfully,

I would also like to kindly present my sincere thanks to my supervisor **Dr. Mohammad Al-Alfayomi,** who has been so helpful and cooperative in giving me support and guidance.

Massive Appreciation I'd like to express with very profound gratitude and personal thanks to my elegant **co-supervisor Dr.Wael Alzyadat** who encouraged and support me and took time to suggest changes, improvements, or clarifications. Although his suggestions caused more work for me, I believe that the end result is the best works I have ever done.

I owe everything to **my family** for being patient and understand while I set the normal flow of my life aside in order to focus on research. This research would have never been possible without the support of a number of people at ISRA University who helped me and provided me information, in addition to answering my questions in ordering to achieve my goal.

They all deserve to be acknowledged as well.

Table of Contents

I

LIST OF Tables

LIST OF FIGURES

Abstract

Big data is a term which describe the characteristics of a dataset, such as volume, value and veracity. There are many challenges which prevent proceeding and working with big data by using traditional techniques to extract value. Project management is a dynamic process that utilizes the appropriate resources of an organization in many phases by measuring in four factor: scope, time, cost and quality.

Improving data quality depends on relations between data and value, which are associated with veracity and accuracy of data and how we can get quality from value. In this research, we attempt to improve data quality from big data characteristics depending on trust of data by working with it in general and especially by using value, volume and veracity of data by finding out correlation statistical analysis results and distance equation. This approach was implemented by using IBM human resource scope with R framework through selecting "Deducer" package from R library.

Implementation and conducting experiment have been carried out by using three main factors: time, cost and scope in two types: product and project, where the strongest relation linking them starts with project scope as the strongest factor followed by cost, product and finally time which is the weakest factor among them. In the final form, we select the best quality using two sides generally: quality degree and middle-quality interval. Especially, relative distance is the strongest factor in the experiment between time and cost, where both sides lead to more trust of data and high accuracy in the chosen process.

Keywords: Big Data Characteristics, Project Management Perspective, Project Management Triangle, Sensitive Rule.