CRITICAL SUCCESS FACTORS FOR MANAGING DATA SCIENCE PROJECTS WITHIN AGILE METHODOLOGY

Garusin Arachchige Imali Limesha

199116P

Master of Business Administration in Information Technology Specialized in Business Analytics

Department of Computer Science Engineering

University of Moratuwa
Sri Lanka

July 2021

CRITICAL SUCCESS FACTORS FOR MANAGING DATA SCIENCE PROJECTS WITHIN AGILE METHODOLOGY

Garusin Arachchige Imali Limesha

199116P

Thesis Submitted in Partial Fulfilment of the Requirements for the Master of Business Administration in Information Technology specialized in Business Analytics.

Department of Computer Science Engineering

University of Moratuwa
Sri Lanka

July 2021

DECLARATION

This thesis contains, as far as I am aware, no material previously submitted for a Master of Business Administration in any other university or institute of higher education, nor does it include, without recognition, material previously published or written by someone other than when the recognition is provided by another individual.

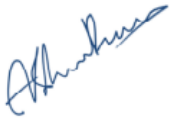
I also confer a non-exclusive right on University Moratuwa to reproduce and distribute my thesis in print, electronic or other media, wholesale or partial. I remain entitled to make full or part use of this material in future work (such as articles or books).

Signature: Date: 03rd July, 2021

The supervisor/s should certify the thesis/dissertation with the following Declaration.

The above candidate has carried out research for the MBA in Information Technology Dissertation under my supervision.

Name of the supervisor: Dr. Amal Shehan Perera

Signature of the supervisor:  Date: 03- 07- 2021

Abstract

Data science is an evolving discipline with a major emphasis on developing accessible data analysis techniques. But far less emphasis has been given on the factors which affect the success of data science projects under the Agile umbrella. In the field of software engineering agile approaches were initially developed and are characterized by their iterative software development approach. It is recommended to use process models or methodologies in literature to increase the success rate of data science projects however, organizations, which are perceived to be too restrictive and do not accept the traditional iteratives and transparent nature of data science projects, reluctant to use them. And there are some potential challenges which have been identified in the literature for using Agile methodologies in data science projects. The characteristics of possible critical success factor (CSF)s for Data Science projects have been established from the literature by updating Chow and Cao's list of success factors for agile software development projects in this research. The factors have been identified under five dimensions of organizational, people, process, technical and project. The findings of this study indicate team environment, team capacity, client engagement, project definition processes, agile software engineering techniques and project schedule as the factors that impact the success of data science projects within Agile methodology. Even though these factors were listed as important for managing data science projects within Agile methodology, the significance of these factors may vary according to the nature of the project that the team is involved in. Therefore, the team should always focus on these factors relative to the nature of the project.

Key words: Data Science, Project Management, Agile Methodology

Acknowledgement

I would like to thank all of those who helped me to carry out my research study on "Critical Success Factors for Managing Data Science Projects within Agile Methodology" successfully. First and foremost, I wish to thank my research supervisor Dr. Amal Shehan Perera Senior Lecturer of the Department of Computer Science and Engineering, University of Moratuwa, for the continuous support, encouragement and attention extended to me throughout my research.

And I would specially thank all the respondents who participated and spent their precious time giving their valuable responses to my online survey. Moreover, thanks to MBA batch mates I extend my commitment to the guidance and support for getting this study to a successful conclusion.

My dear husband, Shanaka Chathuranga, is most appreciative of all the motivation, care and unconditional help that I receive in good and bad times during this research study.

I would like to thank all those who helped me to carry this analysis to a successful conclusion.

Table of Contents

DECLARATION iii

Abstract iv

Acknowledgement v

1.1 Introduction 1

1.2 Background and Motivation 2

1.3 Problem Statement 3

1.4 Objectives of Research 4

1.5 Outline 5

CHAPTER 2 - LITERATURE REVIEW 6

2.1 Introduction 6

2.2 Agile Software Development 6

2.2.1 Emergence of Agile 6

2.2.2 Drivers of Agile 7

2.3 Agile Frameworks used in Data Science Projects 9

2.3.1 SCRUM 9

2.3.2 CRISP-DM 9

2.3.3 Kanban 10

2.3.4 Team Data Science Process (TDSP) 11

2.3.5 SCRUM- DS 11

2.4 Challenges of Agile Frameworks for Data Science Projects 12

2.4.1 Challenges of SCRUM in Data Science projects 12

2.4.2 Challenges of Kanban in Data Science projects 12

2.4.3 Challenges of CRISP-DM in Data Science projects 13

2.4.4 Challenges of TDSP in Data Science projects 13

2.5 Critical Success Factors for Data Science Projects 13

2.6 Summary 16

CHAPTER 3 – Methodology 17

3.1 Introduction 17

3.2 Conceptual Framework 17

3.3 Variables on Relationships 18

3.3 Hypothesis Development 18
3.3.1 Organizational Factors 18
3.3.2 People Factors 20
3.3.3 Process Factors 20
3.3.4 Technical Factors 22
3.3.5 Project Factors 23
3.4 Questionnaire Instrument Development 24
3.5 Sample Selection 25
3.6 Method of Data Collection 26
3.7 Summary 26
CHAPTER 4 – Analysis and Interpretation 27
4.1 Introduction 27
4.2 Reliability Test 27
4.2.1 Preliminary Survey 27
4.2.1 Research Survey 29
4.3 Descriptive Statistics Analysis 30
4.3.1 Composition of the sample according to Age Group 30
4.3.2 Composition of the sample according to Gender 30
4.3.3 Composition of the sample according to Designation 30
4.3.4 Composition of the sample according to Working Experience. 33
4.3.5 Composition of the sample according to the used Agile approach. 34
4.4 Inferential Statistics – Inter Item Correlation Analysis 35
4.5 Hypothesis Testing – Pearson's Correlation Analysis 39
4.6 ANOVA Testing 47
4.8 Discussion on Results 47
4.9 Summary 49
CHAPTER 5 – Conclusion and Recommendation 50
5.1 Introduction 50
5.2 Research Implications 50
5.3 Research Limitation and Future Works 52
References 53

APPENDIX A 56
Questionnaire Instrument 56
APPENDIX B 61
Descriptive Statistics 61

Table of Figures

Figure 2. 1: Agile Manifesto 7
Figure 2. 2: Agile learning Curve 8
Figure 3. 1: Conceptual Framework 17
Figure 4.1: Frequency Distribution of Age Group 28
Figure 4.2: Frequency Distribution of Gender 29
Figure 4.3: Distribution of Job Designation 30
Figure 4.4: Distribution of Working Experience 31
Figure 4.5: Distribution of Agile methodologies used 32

Table of Tables

Table 3. 1: Summary of measurements in independent variables 24
Table 3. 2: Summary of measurements in dependent variable 25
Table 3. 3: Summary of measurements in Demographic factors 25

Table 4. 1: Cronbach's Alpha Coefficient for Preliminary Survey 30
Table 4. 2: Item-wise Cronbach's Alpha Coefficients 31
Table 4. 3: Cronbach's Alpha Coefficient for Research Survey 32
Table 4. 4: Cronbach's Alpha Coefficient for Dependent Variable 32
Table 4. 5: Single Item Variables 32
Table 4. 6: Age Distribution 33
Table 4. 7:Gender Distribution 34
Table 4. 8:Distribution of Job Designation 34

Table 4. 9: Distribution of Working Experience 35
Table 4. 10: Distribution of Agile methodologies used. 36
Table 4. 11: Inter Item Correlation for Organizational Environment 37
Table 4. 12: Inter Item Correlation for Team Environment 37
Table 4. 13: Inter Item Correlation for Customer Involvement 37
Table 4. 14: Inter Item Correlation for Project Definition Process 38
Table 4. 15: Inter Item Correlation for Agile Software Engineering Techniques 38
Table 4. 16: Inter Item Correlation for Project Schedule 38
Table 4. 17: Inter Item Correlation for Success of the Data Science projects within the Agile Methodology. 39
Table 4. 18: Pearson's Correlation- Management Commitment and Success of the Data Science projects within the Agile Methodology. 40
Table 4. 19: Pearson's Correlation- Organizational Environment and Success of the Data Science projects within the Agile Methodology. 40
Table 4. 20: Pearson's Correlation- Team Environment and Success of the Data Science projects within the Agile Methodology. 41
Table 4. 21: Pearson's Correlation- Team Capability and Success of the Data Science projects within the Agile Methodology. 42
Table 4. 22: Pearson's Correlation- Customer Involvement and Success of the Data Science projects within the Agile Methodology. 42
Table 4. 23: Pearson's Correlation- Project Management Process and Success of the Data Science projects within the Agile Methodology. 43
Table 4. 24: Pearson's Correlation- Project Definition Process and Success of the Data Science projects within the Agile Methodology. 43
Table 4. 25: Pearson's Correlation- Improved Coordination and Success of the Data Science projects within the Agile Methodology. 44
Table 4. 26: Pearson's Correlation- Agile Software Engineering Techniques and Success of the Data Science projects within the Agile Methodology. 44
Table 4. 27: Pearson's Correlation- Delivery Strategy and Success of the Data Science projects within the Agile Methodology. 45
Table 4. 28: Pearson's Correlation- Project Type and Success of the Data Science projects within the Agile Methodology. 46

Table 4. 29: Pearson's Correlation- Project Schedule and Success of the Data Science projects within the Agile Methodology. 46
Table 4. 30: One-Way ANOVA- Agile methodologies used. 47

Table B. 1 Statistics: Management Commitment (MC) – Item No. 1 64
Table B. 2 Statistics: Management Commitment (MC) – Item No. 2 65
Table B. 3 Statistics: Organizational Environment (OE) – Item No. 3 66
Table B. 4 Statistics: Organizational Environment (OE) – Item No. 4 67
Table B. 5 Statistics: Team Environment (TE) – Item No. 5 67
Table B. 6 Statistics: Team Environment (TE) – Item No. 6 68
Table B. 7 Statistics: Team Capability (TC) – Item No. 7 69
Table B. 8 Statistics: Team Capability (TC) – Item No. 8 69
Table B. 9 Statistics: Customer Involvement (CI) – Item No. 9 70
Table B. 10 Statistics: Customer Involvement (CI) – Item No. 10 71
Table B. 11 Statistics: Project Management Process (PMP) – Item No. 11 71
Table B. 12 Statistics: Project Management Process (PMP) – Item No. 12 72
Table B. 13 Statistics: Project Definition Process (PDP) – Item No. 13 73
Table B. 14 Statistics: Project Definition Process (PDP) – Item No. 14 73
Table B. 15 Statistics: Improved Coordination (IC) – Item No. 15 73
Table B. 16 Statistics: Agile Software Engineering Techniques (ASE) – Item No. 16 74
Table B. 17 Statistics: Agile Software Engineering Techniques (ASE) – Item No. 17 74
Table B. 18 Statistics: Agile Software Engineering Techniques (ASE) – Item No. 18 75
Table B. 19 Statistics: Agile Software Engineering Techniques (ASE) – Item No. 19 75
Table B. 20 Statistics: Delivery Strategy (DS) – Item No. 20 75
Table B. 21 Statistics: Delivery Strategy (DS) – Item No. 21 76
Table B. 22 Statistics: Project Type (PT) – Item No. 22 77
Table B. 23 Statistics: Project Schedule (PS) – Item No. 23 77
Table B. 24 Statistics: Project Schedule (PS) – Item No. 24 78
Table B. 25 Statistics: Success of the Data Science projects within Agile methodology (SUCCESS) – Item No. 25 79

Table B. 26 Statistics: Success of the Data Science projects within Agile methodology (SUCCESS) – Item No. 26 79
Table B. 27 Statistics: Success of the Data Science projects within Agile methodology (SUCCESS) – Item No. 27 80