Predicting absenteeism factors in the work place through data mining

S.P.Nishantha

169323U

Supervised By:

Dr. Chaman Wijesiriwardana
(Senior Lecturer)

Faculty of Information Technology

University of Moratuwa

Predicting absenteeism factors in the work place through data mining

S.P.Nishantha

169323U

Dissertation submitted to the Faculty of Information Technology, University of Moratuwa, Sri Lanka for the partial fulfillment of the requirements of Degree of Master of Science in Information Technology

August 2022

Declaration

I declare that this thesis is my own work and has not been submitted in any form for another degree or diploma at any university or other institution of tertiary education. Information derived from the published or unpublished work of others has been acknowledged in the text and a list of references is given.

Name of Student — Signature of Student

S.P.Nishantha ……..……

Date:

Supervised by

Name of Supervisor — Signature of Supervisor

Chaman Wijesiriwardana ……..……

Date:

Acknowledgement

I wish to express sincere appreciation to,

My research supervisor, Dr. Chaman Wijesiriwardana of the faculty of Information Technology of University of Moratuwa for dedicating his valuable time for guiding me throughout this.

My wife and parents for providing me with continuous encouragement throughout my years of study.

Finally, for my colleagues for their cooperation and support.

Thank you.

Abstract

Absenteeism is an employee's absence from work. Absences of employees can have a major effect on company strategies, finances, morale and other factors. Excessive absences may influence to decrease productivity of the company. Poorly performing employees cause significant losses to the organization, and absenteeism is considered one of the factors affecting performance. Therefore, understanding the causes of absenteeism can provide organizations with competitive advantage tools and open up research areas for computers and human resources fields. The purpose of this paper is to use computerized technology to discover the causes of employee absence. This study analyzes data from the absentee database and finds several factors that have a good correlation with absentees. In addition, two data mining techniques clustering and association rule mining are used to discover factors which cause in absenteeism with high accuracy. This research paper is to create association model to predict whether find the relationship of absenteeism of employee.

Table of Contents

Chapter 1……..…..……1

Introduction

1.1Introduction……..……..…….1

1.2Background and Motivation……...2

1.3 Problem Statement…….3

1.4Aim and Objectives…….3

1.4.1Aim……...……..…….……3

1.4.2Objectives……3

1.5 Proposed Solution……...…….…..4

1.6 Summary……...….5

Chapter2…….6

Review of the work……..6

2.1 Introduction……6

2.2 Absenteeism Prediction…….6

2.3 Classification comparison……..6

2.4 Using Association rule mining Algorithm……7

2.5 Using Clustering Algorithm…….…..8

2.6 Summery……..12

Chapter 3……..13

Technologies used for this work

3.1Introduction…….…….……13

3.2Rapid Miner…….…….13

3.3 Microsoft SQL Server…….13

3.4 Determining through Clustering……..……14

3.5 Determining through Association rule……14

3.6 Microsoft visual studio……15

3.7 C Sharp programming language……..……15

3.8Summary……...15

Chapter 4……..16

A novel approach for profile detection of absenteeism

4.1Introduction……...……...16

4.2Input……..…… 16

4.3Output……...……16

4.4Process……..……....17

4.5 Data Collection…….…..…….17

4.6 performing Clustering…….……...…..18

4.7 Performing Association Rules Mining……..…..….19

4.8 Summery……....…..19

Chapter 5……..…...……23

Absenteeism Profile detection-Analysis and Design

5.1. Introduction……..…….…….……20

5.2. Research Design……..…..…….…….20

5.3 Clustering analysis……..……....……23

5.4 Association rule mining……..…….…...25

5.4.1 Support and Confidence…….……..26

5.5 Summery……..……..…….27

Chapter 6...……...……28

Implementation

6.1 Introduction……..……...……28

6.2 Decision Support System……..……...28

6.3 Data Preprocessing……..…….29

6.3.1 Data Preprocessing for Clustering……,,……..…30

6.3.2 Data Preprocessing for Association Rules Mining……..…….30

6.4 Clustering Analysis……..……30

6.5 Associations Rules Mining…….…….32

6.6 Summery…….…….34

Chapter 7……..……35

Evaluation

7.1 Introduction……...….….….35

7.2 clustering……..….…..….35

7.3 Association Rule Mining……....…….36

7.3.1 Minimum support and Confidence…….……..37

7.4 Summary……..…...…….44

Chapter 8……...……...……44

Conclusion and Future Work

8.1 Introduction……..45

8.2 Conclusion…….……...……...……45

8.3 Future Work…….……...…….45

8.4Summery……...……46

References…….47

List of Tables

Table 5.1: Support Value Calculations……..26

Table 5.2: Confidence Value Calculations……27

Table 7.1: Confidence Value Calculations……40

Table 7.2: Confidence Value Calculations……42

Table 7.3: Confidence Value Calculations……43

List of Figures

Figure 1.1: Steps in Data mining process……4

Figure 5.1: Activity Diagram……...……..21

Figure 5.2: Research Design and Method……..22

Figure 6.1: Chart for Absenteeism hours Month of Absence……28

Figure 6.2: Pie Chart for Absenteeism hours for season……...……29

Figure 6.3: Clusterig Summery…….…….30

Figure 6.4: Clustering Diagram……...…….…….31

Figure 6.5: ClusterModel……...…….……...32

Figure 6.6: Sample Clustering Data…….……..32

Figure 6.7: Genarating Association Rule…….……..33

Figure 6.8: Sample data for Association……34

Figure 7.1: ClusterModel……...……35

Figure 7.2: Plot view of Cluster Analysis……..36

Figure 7.3: Association Rule……..36

List of Equations

Equation 5.1: Support……...…….26

Equation 5.2: Confidence……...……26