:B/DON/92/2019
IT 01/221

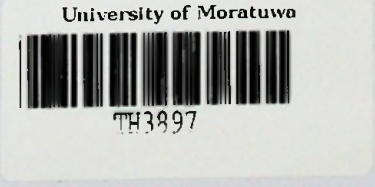
An Application of Data Mining for a Library Management System

LIBRARY
UNIVERSITY OF MORATUWA, SRI LANKA
MORATUWA

T.A.U.I. Ranaweera

169327K

004 "19"
004.(043)



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

TH3897
+
CD-ROM

February 2019

TH3897

Declaration

We declare that this thesis is our 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 Information derived from the published or unpublished work of others has been acknowledged in the text and a list of references is given.

T A U I Ranaweera

.....

Signature

Date: 04/04/2019

Supervised by

Mr. Saminda Premarathne

.....

Signature

Date: 04/04/2019

Department of Information Technology
04 APR 2019
University of Moratuwa

UNIVERSITY OF MORATUWA
LIBRARY
SRI LANKA

Dedication

To my family

Acknowledgement

Completing this research would be impossible without the support and guidance of whole bunch of people who were around me during this entire period. First of all I would like to express my deepest gratitude to Mr. Saminda Premarathne for being my supervisor and providing excellent guidance throughout the project. The knowledge I gained from you made me strong enough to complete this challenge. Furthermore, my gratitude goes to the lecturer panel of M.Sc. degree programme and I appreciate the effort they took to improve our subject knowledge throughout these two years.

Obtaining relevant data for the analysis was vital and I would like to thank Mr. Ananda Karunarathne, Librarian of University of Ruhuna for granting approval to access data from the library database. It's with great honor, I acknowledge the cooperation of Mr. Nimal Hettiarachi who provided necessary information on the library system of University of Ruhuna. Without your support, my efforts could have been in vain.

At last, but not least, I would like to thank my family members for being supportive as always. My family is the greatest strength of my life and I have no words to express my gratitude towards them.

List of Abbreviations

DDC – Dewy Decimal Classification

EG - Faculty of Engineering

HS- Faculty of Humanities and Social Science

MD - Faculty of Medicine

MF- Faculty of Management

PCA –Principle Component Analysis

SC – Faculty of Science

Abstract

Data mining has become an emerging concept in today's world. With the development of the technology in every field, a large amount of data can be collected and stored very easily. The challenge is to analyse these data from traditional analysisng techniques. The concept of data mining; which enables users to analyze data and draw conclusions through clearly defined procedures comes into play as a solution for this matter.

The purpose of this research is to test the adequacy of data mining techniques to improve the library usage in Sri Lankan State Universities. A data warehouse was designed and implemented using the most important variables in the raw dataset. After cleaning and preprocessing of data, association rule mining and clustering were basically used in the data analysis phase. Interesting rules were identified and the said results were used in the next stage of the study. A Book Recommendation System was implemented in Java based on the results obtained in the previous stage. The system enables users to select library materials according to their prior borrowing patterns. My SQL, R and Java were basically used in the analysis.

This research will be beneficial to enrich the library usage in state universities in Sri Lanka as well as the researchers those who are interested in Data Mining.

Table of Contents

Chapter 1 1

Introduction 1

1.1. Introduction 1

1.2. Background 1

1.3. Statement of Research Problem 2

1.4. Aim 3

1.5. Specific Objectives 3

1.6. Solution 3

Chapter 2 4

Literature Review 4

2.1. Knowledge Discovery Process 4

2.2 Data Mining 5

2.3 Applications of Data Mining 6

2.4 Data Mining in Library Management Systems 7

2.5 Association Rule Mining for Library Systems 9

2.6 Clustering for Library Systems 11

2.7 Book Recommendation Systems 12

Chapter 3 14

Technology Adapted 14

3.1 Theoretical Background 14

3.1.1. Association Rule Mining 14

3.1.1.1. Rule Assessment Measures 15

3.1.2. K-means Clustering 17

3.1.2.1. Cluster Validation 17

3.2. Technologies Used for the Application 18

3.2.1. Java 18

3.2.2. R Statistical Software 18

3.2.3. My SQL 18

Chapter 4 19

Methodology 19

4.1 Data Selection Procedure 19

4.2. Data Cleaning and Preprocessing 22
4.3. Building the Data Warehouse 23
4.4. Creation of the Connection between R and Other Sources 24
4.5. Establishment of the Connection between R and MySQL 25
4.6. Establishment of the Connection between R and Java 25
Chapter 5 26
Analysis and Design 26
5.1. Identifying Borrowing Patterns of Patrons 26
5.1.1 Identify Frequent Patterns of Borrowing Books by the Users 26
5.1.1.1 Data Preparation 26
5.1.1.2 Test for Statistical Significance of Rules 27
5.2.2 Identify User Behavior when Lending Books 28
5.2. Clustering Library Users with k-means Clustering 30
5.2.1 Clustering the Library Users without Dimension Reduction 30
5.2.2 Clustering the Library Users by Dimension Reduction Using Principle Components (Two Step Clustering) 31
Chapter 6 33
Implementation of the Book Recommendations System 33
6.1. Providing Recommendations for the Users without Logging-in 33
6.1.1. Recommendations Based on the Support Value of the Dewy Numbers of Borrowed Books 33
6.1.1.1. Recommendations Based on Association Rules between Faculty and Dewy Numbers of Borrowed Books. 34
6.1.1.2. Recommendations Based on Association Rules between Dewy Numbers of Borrowed Books 35
6.2 Providing Recommendations for the Users after Logging-in 37
6.2.1. Recommendations Based on the Most Popular Books in the Cluster 37
6.2.2. Recommendations Based on Association Rules between Dewy Numbers of Borrowed Books within the Cluster 38
Chapter 07 40
Results of the Study 40
7.1 Descriptive Analysis 40
7.1.1 Composition of Student Users in 2013 40
7.2 Association Rule Mining 42
7.2.1 Mining Associations between the Dewy Numbers of Borrowed Books 42

7.2.2 Association Rules between Faculty and Dewy Decimal Number 50
7.3 Clustering the Library Users (Students) 52
7.3.1 Clustering the Library Users in the Faculty of Science 52
7.4 Book Recommendation System 54
7.4.1 Main Interface 54
7.4.2 Interface after Logging-in 55
7.5 Evaluation 56
Chapter 7 57
Conclusion & Further Work 57
7.1 Conclusions 57
7.2 Limitations 59
7.3 Future Work 59
References 61

UNIVERSITY OF MORATUWA
LIBRARY
SRI LANKA

List of Tables

Table 4.1: Table of "dewycodes" 20
Table 4.2: Tables Selected from the Database for the Study 20
Table 4.3: Summary Table of Selected Variables 21
Table 4.4: Summary of New Variables 22
Table 5.1: Arrangement of Data in the Database 26
Table 5.2: Layout of the Rearranged Data 27
Table 7.1: Summary of Association Rules 45
Table 7.2: Summary of Association Rules 46
Table 7.3: Summary of Association Rules 47
Table 7.4: Summary of Association Rules 48
Table 7.5: Summary of Association Rules 50
Table 7.6: Summary of Association Rules 51

List of Figures

Figure 2.1: Knowledge Discovery Process 4
Figure 4.1: Implementation of "transactionall" Table 24
Figure 5.1: Flow Chart of Generating Association Rules 28
Figure 5.2: Flow Chart for Generating Associations Considering More Than One Variable . 29
Figure 5.3 : Flow Chart for the Process of Clustering 31
Figure 5.4: Flow Chart for the Process of Two Step Clustering 32
Figure 6.1 : Process for the Method 01 34
Figure 6.2 : Process of the Method 02 35
Figure 6.3: Process of the Method 03 36
Figure 6.4: Process of the Recommendations after Logging-in 37
Figure 6.5: Process of Recommendations when Searching for a Certain Book 38
Figure 7.1: Composition of Library Users (Students) 40
Figure 7.2: Faculty-wise Transactions according to the Month 41
Figure 7.3: Item Frequency Plot 43
Figure 7.4: Graph Based Visualization Rules 44
Figure 7.5: Plot of Number of Clusters vs. Within Groups Sum of Squares 52
Figure 7.6 : Cluster Validation 52
Figure 7.7: Main Interface of the Book Recommendation System 54
Figure 7.8: Interface after Logging-in to the User Account 55