

# **Under-graduate study recommendations using association rule mining**

*Thesis Submitted in Partial Fulfillment of the Degree Program*

**MS / M/Phil**

in

**Software Engineering**

*Under the supervision of:*

**Dr. Malik Tahir Hassan**

*Submitted By:*

**Hira Asim S2016114002**



*School Of Systems and Technology (SST)*

**University of Management and Technology**

Pakistan, Lahore

**Session: 2016 - 2018**

# Final Approval

## Panel of Examiners:

- **Thesis** **Advisor**  

---

Dr. Malik Tahir Hassan **Date:**  
HEC Approved PhD Supervisor  
Assistant Professor  
Department of Computer Science  
University of Management and Technology, Lahore
- **Director of Graduate Studies** 

---

**Date:**  

Dr. Muhammad Shoaib Farooq  
HEC Approved PhD Supervisor  
Associate Professor  
Department of Computer Science  
University of Management and Technology, Lahore
- **Dean SST** 

---

**Date:**  

Dr. Shaukat Iqbal  
Professor  
Department of Computer Science  
University of Management and Technology, Lahore

# Acknowledgement

---

The fruitful finish of this undertaking would be inadequate without recognizing the individuals who made it conceivable to design and develop and whose legitimate direction and consolation secured the achievement.

As a matter of first importance I wish to recognize the altruism of supreme Almighty ALLAH who gave me quality and valor to beat all snags.

With the significant feeling of appreciation and heartiest respect, I express my true sentiments of obligation to my advisor **Dr. Malik Tahir Hassan**, (Assistant Professor), Computer Science Department, University of Management and Technology, Lahore for his uplifting disposition, fabulous direction, consistent support, unmistakable fascination, precious co-operation, liberal state of mind or more every one of his gifts. He have been a wellspring of motivation for me.

Last yet not the slightest I might want to express my ardent on account of my guardians and companions who with their intriguing perspectives, veracity and entire hearted co-operation helped me in doing this research work.

# Dedication

---

I dedicate my work to the youth of Pakistan so that they may be able to compete with the external world in a better way than before. I am hopeful that this research initiative will help education seekers of our country to select suitable degree program as well as institutes of their own interest at under-graduate level.

*May this effort bring a lot of rise in literacy rate of Pakistan (AMEEN)*

# Abstract

---

In this research work, we are aimed to provide recommendations regarding degree programs and institutes at under-graduate level to students who have successfully passed their intermediate. In order to recommend we collected data via Google survey form regarding background study area, future goals and interests of individuals who are successfully done with their under-graduate studies for least. After the data has been collected it was being pre-proceed, visualized and analyzed for future use. Data mining technique named *Association Rule Mining* is applied on the data to find hidden patterns and trends in it. The results of the Association Rule Mining technique are the recommendations to the students regarding suitable under-graduate degree programs and educational institutes as per their background study stream, area of interest and future goals. Recommendations are based on the information collected where students and graduates have alike choices regarding study streams, area of interest and future goals.

We are hopeful that this contribution will be a huge step forward towards the betterment of educational environment as well as a source of proper guidance for the students who are done with their intermediate and now aimed to take admission in an under-graduate program in a well reputed university in Pakistan.

# Acronyms

---

<b>DM</b>	Data Mining	<b>BBA</b>	Bachelor of Business Administration
<b>EDM</b>	Educational Data Mining	<b>BCOM</b>	Bachelor of Commerce
<b>CF</b>	Collaborative Filtering	<b>BDS</b>	Bachelor of Dental Surgery
<b>PU</b>	Punjab University	<b>D Pharm</b>	Doctor of Pharmacy
<b>IT</b>	Information Technology	<b>DPT</b>	Doctor of Physical Therapy
<b>GCU</b>	Government College University	<b>Engr</b>	Engineering
<b>UOE</b>	University of Education	<b>LLB</b>	Bachelor of Laws
<b>NCA</b>	National College of Arts	<b>Com</b>	Commerce
<b>UCP</b>	University of Central Punjab	<b>MBBS</b>	Bachelor of Medicine/Surgery
<b>ICS</b>	Intermediate in Computer Science		
<b>IBA</b>	Institute of Business Administration		
<b>NUST</b>	National University of Sciences and Technology		
<b>UET</b>	University of Engineering and Technology		
<b>KEMU</b>	King Edward Medical University		
<b>UHS</b>	University of Health Sciences		
<b>UMT</b>	University of Management Technology		
<b>LUMS</b>	Lahore University of Management Sciences		
<b>FAST-NU</b>	National University of Computer and Emerging Sciences		

# Table of Contents

<b>Chapter 1: Introduction.....</b>	<b>12</b>
1.1 Motivation and Objectives.....	12
1.2 Background.....	13
1.3 Data Mining.....	13
1.4 Educational Data Mining.....	15
1.5 Recommender Systems.....	17
1.6 Association Rule Mining.....	18
1.7 Educational Environment.....	20
1.8 Student Performance Analysis.....	21
1.9 Research Question.....	21
1.10 Intended Audience.....	22
1.11 Thesis Outline.....	22
<b>Chapter 2: Literature Review.....</b>	<b>23</b>
2.1 Background of Recommender Systems.....	23
2.2 Recommender Systems in the field of Educational Data Mining.....	23
2.3 Background of Association Rule Mining.....	27
2.4 Literature Study of Association Rule Mining.....	28
2.5 Conclusion.....	29
<b>Chapter 3: Dataset Collection and Pre-Processing.....</b>	<b>30</b>
3.1 Introduction.....	30
3.2 Importance of Dataset.....	30
3.3 Data Gathering Phase.....	31
3.4 Data Pre-Processing Step.....	37
3.5 Conclusion.....	38

<b>Chapter 4: Getting Insight from Data.....</b>	<b>39</b>
4.1 Introduction.....	39
4.2 Data Visualization.....	39
4.3 Data Analysis.....	39
4.4 Data Visualization and Data Analysis on dataset .....	40
4.4.1 Step 1: Selection of an appropriate tool.....	40
4.4.2 Step 2: Data Visualization and Data Analysis on MATLAB Tool .....	41
4.5 Conclusion .....	65
<b>Chapter 5: Recommendation Methodology and Results.....</b>	<b>66</b>
5.1 Recommendation Process .....	66
5.2 Recommendations to Students.....	68
5.2.1 Under-Graduate Program Recommendations.....	70
5.2.2 Under-Graduate Institute Recommendations .....	84
5.3 Limitations of proposed recommendations.....	96
5.4 Conclusion .....	96
<b>Conclusion.....</b>	<b>97</b>
<b>Future Work.....</b>	<b>98</b>
<b>References.....</b>	<b>99</b>
<b>Appendix.....</b>	<b>105</b>

# List of Figures

---

Figure 1: The KDD Process: A step by step approach [34].....	14
Figure 2: Educational Data Mining (EDM) Components [37] .....	16
Figure 3: Responses collected after processing .....	38
Figure 4: MATLAB R2017a Interface .....	41
Figure 5: Histogram analysis of attribute Gender.....	42
Figure 6: Histogram analysis of attribute Matriculation Stream .....	42
Figure 7: Histogram analysis of attribute Intermediate Stream .....	43
Figure 8: Histogram analysis of attribute Under-Graduate Program.....	44
Figure 9: Histogram analysis of attribute Under-Graduate Institute .....	45
Figure 10: Histogram analysis of attribute Area of Interest .....	46
Figure 11: Histogram analysis of attribute Future Goals.....	47
Figure 12: Choice of Under Graduate Programs among Males and Females.....	52
Figure 13: Inclination of graduates towards different Under Graduate Programs on the basis of their Matriculation Stream .....	53
Figure 14: Choice of Under Graduate Programs on the basis of Intermediate Stream studied ....	53
Figure 15: Trend of selection of Under Graduate Institute on the basis of choice of Under Graduate Program .....	55
Figure 16: Bonding between Under Graduate Programs and Areas of Interest .....	60
Figure 17: Selection of different Future Goals by graduates on the basis of Under Graduate Program Selected .....	61
Figure 18: Trend of choice of Under Graduate Institutes by graduates by taking in to account their Matriculation Stream .....	62
Figure 19: Choices of Under Graduate Institutes made by graduates on the basis of their Intermediate Stream.....	63
Figure 20: Inclination of graduates towards different Areas of Interest on the basis of Under Graduate Institute selected by them.....	64
Figure 21: Recommendation Process: A step by step approach.....	67

# List of Tables

---

Table 1: List of values of attribute Gender .....	32
Table 2: List of values of attribute Matriculation Stream.....	32
Table 3: List of values of attribute Intermediate Stream .....	33
Table 4: List of values of attribute Under Graduate Program .....	34
Table 5: List of values of attribute Under Graduate Institute .....	35
Table 6: List of values of attribute Area of Interest.....	36
Table 7: List of values of attribute Future Goals .....	36
Table 8: Summary of attribute Gender .....	48
Table 9: Summary of attribute Matriculation Stream.....	48
Table 10: Summary of attribute Intermediate Stream.....	48
Table 11: Summary of attribute Under Graduate Program.....	49
Table 12: Summary of attribute Under Graduate Institute .....	50
Table 13: Summary of Attribute Area of Interest.....	51
Table 14: Summary of Attribute Future Goals .....	52
Table 15: Most frequent occurrence of value against each attribute .....	65
Table 16: Threshold value of support assigned to each rule of different length against Under Graduate Program.....	70
Table 17: Under Graduate Program recommendations made against different length rules .....	71
Table 18: Under Graduate Program Recommendations on selecting any attribute.....	71
Table 19: Under Graduate Program Recommendations on selecting any two attributes .....	72
Table 20: Under Graduate Program Recommendations on selecting any three attributes .....	74
Table 21: Under Graduate Program Recommendations on selecting any four attributes.....	75
Table 22: Under Graduate Program Recommendations on selecting any five attributes .....	77
Table 23: Under Graduate Program Recommendations on selecting all attributes .....	83
Table 24: Threshold Value of support assigned to each rule of different length against Under Graduate Institute.....	84
Table 25: Under Graduate Institute Recommendations made against different length rules .....	85
Table 26: Under Graduate Institute Recommendations on selecting any attribute .....	86
Table 27: Under Graduate Institute Recommendations on selecting any two attributes .....	87

Table 28: Under Graduate Institute Recommendations on selecting any three attributes ..... 89  
Table 29: Under Graduate Institute Recommendations on selecting any four attributes ..... 90  
Table 30: Under Graduate Institute Recommendations on selecting any five attributes..... 91  
Table 31: Under Graduate Institute Recommendations on selecting all attributes..... 95

# Chapter 1

## Introduction

---

*We are drowning in data but starving for knowledge. John Naisbitt [23]*

Selection of a right degree program from a well reputed institute at under-graduate level is a very important decision to be made by any student. Students after their intermediate are not mature enough to select a right pathway for their future which may cause a mismatch between their educational background and interest. As far as parents and teachers are concerned, they don't have any proper metrics on the basis of which appropriate recommendations to the students can be made. It can be concluded that there is an absence of proper guidance for learners and no trustworthy source is available for their help as well. Keeping in view the whole situation, recommender system find its way out in the field of Educational Data Mining (EDM).