

Career and skills recommendations using data mining technique:  
matching right people for right profession, in Pakistani context

*Thesis Submitted In Partial Fulfillment of the Degree Program*

**MS**

in

**Software Engineering**

*Under the supervision of:*

Dr. Malik Tahir Hassan

*Submitted By:*

Hafiza Maria Kiran 15007114008



*School of Systems and Technology (SST)*

**University of Management and Technology**

Pakistan, Lahore

**Session: 2016 – 2018**

# Final Approval

## Panel of Examiners:

1) **Thesis Advisor** \_\_\_\_\_

Dr. Malik Tahir Hassan  
Assistant Professor  
Department of Computer Science  
University of Management and Technology, Lahore

2) **Director of Graduate Studies** \_\_\_\_\_

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

3) **DeanSST, Professor** \_\_\_\_\_

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

# Acknowledgement

---

First of all I am grateful to Allah who gave me sound mind and sound health to accomplish my research.

I would like to thank all the wonderful people whose support and encouragement made this research a possibility.

Special thanks go to my affectionate supervisor, **Dr. Malik Tahir Hassan**, (Assistant Professor), Computer Science Department, University of Management and Technology, Lahore for his kind supervision, encouragement, continuous support, motivation and patience without which it would not be possible to complete it.

Last but not least, I humbly extend thanks to my family and friends, for their love, support and co-operation in this regard.

# Dedication

---

This dissertation is dedicated to the people of Pakistan who have completed their studies or currently studying and are trying to figure out the best profession for them so that they can apply for the right job which actually suit them. We hope the proposed model for profession recommendations will help jobseekers in Pakistan by recommending them multiple professions according to their studies.

# Abstract

---

Nowadays, recommendation systems are commonly used by the people for finding the products which best match with their individual preferences. In the context of profession and recommendations, a lot of recommendation systems are available on the internet for the help of jobseekers. The systems are performing well and generating job recommendations for people jobs but they have some serious problems that are faced by jobseekers in Pakistan. They are not much intelligent, require a lot of user's time in filling long forms for registration so that they can get recommendations. Moreover they do not give suggestions that which skills will be suitable for a specific profession. A user must have to spend too much time for applying a job and still he doesn't know the skill that is more valuable for him and is it the best job for his skills on which he is going to apply? The problem is that people are not clear in which field they should start or switch working. Actually there is a point that first of all one should be clear about his/her profession and important skills regarding selected profession in which he/she wants to start a career, then he/she should start finding job related to the selected profession. Based on above issues, there is a need to design such a system that can overcome the problem of profession selection and skills suggestions so that it can be easy for a jobseeker to apply for a specific job. In this research, the problem which is discussed above regarding profession and skills recommendations is resolved by proposing a model by using *Association Rules Mining*, a data mining technique. In this model, professions are recommended to job seekers by matching the profile of applicant or job seeker with those persons who have same profile like educational background, professional skills and the type of jobs which they are doing. The data collected for this research itself is a major contribution as we collected it from different sources. We will make this data publically available for others so that they can use for further research.

# Acronyms

---

**CV** Curriculum Vitae

**DM** Data Mining

**HRM** Human Resource Management

**CF** Collaborative Filtering

**TF-IDF** Term Frequency-Inverse Document Frequency

**OCS** Office of Career services

**ARM** Association Rule Mining

**AR** Association Rule

**IDE** Integrated Development Environment

**JRS** Job Recommender System

**UMT** University of Management and Technology

**UOG** University of Gujrat

**NUST** National University of Science and Technology

**PUP** Punjab University

**UOS** University of Sargodha

**BZU** Bahauddin Zakariya University

**UOL** University of Lahore

**UOE** University of Education

**LUMS** Lahore University of Management Sciences

**NA** Not Available

# Table of Contents

---

<b>Chapter 1: Introduction</b> .....	11
1.1 Background .....	12
1.2 E-Recruitment .....	13
1.3 Recommender System.....	14
1.4 Data Mining Techniques for Recommender System .....	15
1.5 Research Questions .....	16
1.6 Intended Audience: .....	17
1.7 Structure of thesis .....	17
<b>Chapter 2: Literature Review</b> .....	18
<b>Chapter 3: Data Collection Technique &amp; Approaches</b> .....	26
3.1 Introduction.....	26
3.2 Type of Data Required.....	26
3.3 Google Survey Form.....	27
3.4 Data Collection Sources.....	32
3.5 Data Collection Approaches .....	33
3.6 Data Pre-Processing .....	34
<b>Chapter 4: Data Analysis &amp; Visualization</b> .....	36
4.1 Introduction.....	36
4.2 Data Visualization.....	36
4.3 Graphical Representation of All Attributes with Response .....	37
4.4 Attribute’s Relationship .....	55
3.1 Scatter Plots for Relationship between Professions and Educational Backgrounds .....	55
<b>Chapter 5: Recommendations Technique &amp; Results</b> .....	66
5.1 Introduction.....	66
5.2 Recommendations Technique .....	66
5.2.1 Association Rule Mining (ARM).....	66
5.3 Recommendation Process Step by Step .....	67
5.3.1 Tool Selection for Recommendations.....	67
5.3.2 Association Rule Mining Using R Studio.....	67
5.4 Recommendations to Jobseekers .....	70

5.4.1	Profession Recommendations .....	71
5.4.2	Profession Recommendations Visualization.....	84
5.4.3	Skills Recommendations.....	86
5.4.4	Skills Recommendations Visualization.....	88
5.5	Limitations of Proposed Model.....	89
<b>Conclusion .....</b>		<b>90</b>
<b>Future Work.....</b>		<b>91</b>

# List of Figures

---

Figure 1: Major Categories in which recommender systems generate Recommendations .....	14
Figure 2: Sources of Data Collection.....	33
Figure 3: View of Data after Data Preprocessing.....	35
Figure 4: Response against Gender Attribute .....	38
Figure 5: Response against Under-Graduate Stream.....	39
Figure 6: Response against Under-Graduate Institute.....	41
Figure 7: Response against Master's Stream.....	43
Figure 8: Response against Master's Institute .....	44
Figure 9: Response against Graduate Stream .....	45
Figure 10: Response against Graduate Institute.....	46
Figure 11: Response against Doctorate Stream attribute.....	48
Figure 12: Response against Doctorate Institute .....	49
Figure 13: Response for Working Experience (In Years) .....	50
Figure 14: Response against Professions.....	51
Figure 15: Response against Skills .....	53
Figure 16: Relationship b/w Professions & Under Graduate Studies.....	56
Figure 17: Professions for the people having Master's degree .....	59
Figure 18: Relationship b/w Professions & Graduate Studies.....	62
Figure 19: A step by step Association Rules generation process .....	68
Figure 20: Dataset Loading in R Studio .....	68
Figure 21: Installation of arules package into R Studio.....	69
Figure 22: Implementation of Apriori algorithm on dataset.....	69
Figure 23: Sorting & Filtering of Association Rules .....	69
Figure 24: Save Result of Apriori Algorithm in txt file.....	70
Figure 25: Association Rules Visualization.....	85
Figure 26: Profession Recommendations Visualization.....	86
Figure 27: Skills Recommendations Visualization.....	88

# List of Tables

Table 1: Possible values for the attribute Gender in survey form.....	27
Table 2: Possible values for attribute Under-Graduate Stream .....	27
Table 3: Names of Popular Institutes for Under-Graduate Studies .....	28
Table 4: Options for Working Experience in Years .....	30
Table 5: Popular Professions in Market.....	31
Table 6: Skills Set .....	31
Table 7: Possible values of Gender attribute .....	37
Table 8: Possible options for Under-Graduate Stream attribute.....	38
Table 9: Response against Under-Graduate Stream .....	40
Table 10: Response against Under-Graduate Institute.....	42
Table 11: Masters Stream along with response percentage .....	43
Table 12:Response against Master’s Institute.....	44
Table 13: Response against Graduate Stream.....	46
Table 14: Graduation Institute attribute and response percentage.....	47
Table 15: Response Rate of Doctrate against All Options .....	48
Table 16: Responses against Institutes for Doctorate .....	49
Table 17: Employees’ Working Experience .....	50
Table 18: Percentage Response for all professions.....	51
Table 19: Skill set and Employees’ response .....	53
Table 20: Support Threshold Values Against Length of Rule.....	71
Table 21: Number of Recommendations Generated Against Each Length Rule .....	71
Table 22: Profession Recommendation by Selecting 2 Attributes .....	72
Table 23: Profession Recommendations on three attributes.....	73
Table 24: Profession Recommendations by Selecting Four Attributes .....	74
Table 25: Profession Recommendations by Selecting Five Attributes.....	75
Table 26: Profession Recommendations by Selecting Six Attributes .....	76
Table 27: Profession Recommendations by Selecting Seven Attributes .....	77
Table 28: Profession Recommendations by Selecting Eight Attributes .....	78
Table 29: Profession Recommendations by Selecting nine Attributes .....	79
Table 30: Profession Recommendations by Selecting Ten Attributes.....	81
Table 31: Profession Recommendations by Selecting Eleven Attributes.....	82
Table 32: Profession Recommendations by Selecting Twelve Attributes.....	83
Table 33: No. of rules generated for Skills Recommendations .....	87
Table 34: Skills Recommendations .....	87

# Chapter 1: Introduction

To get a good job is a dream of everyone after completion of studies. At the early stage when a person starts to find a job, he does not have much knowledge and information about the job searching sites and popular areas that actually match with educational background where he/she should apply. Most of the job seekers waste their important time in the searching and applying for jobs. They have lack of proper guidance of finding the dreamed job. It's the era of global world village where people seek to have everything automated. In Recruitment, there are many global as well as local specialized systems available like Rozee.pk, RightJobs.pk, jobee.pk, Indeed.com.pk, etc. All are doing well but they are lacking artificial intelligence in their systems.