

Data Mining for Students' Employability Prediction

S.M.Madhavie Malika

189467G

Master of Science in Information Technology

Faculty of Information Technology

University of Moratuwa

Sri Lanka

July 2021

Data Mining for Students' Employability Prediction

S.M.Madhavie Malika

189467G

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

July 2021

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 Information derived from the published or unpublished work of others has been acknowledged in the text and a list of references is provided.

Student Details:

Name of Student

Signature of Student

S.M.M.Malika

.....

Date:

Supervisor Details:

Name of Supervisor

Signature of Supervisor

Mr.S.C.Premarathne

.....

Date:

Acknowledgements

I would like to express my sincere gratitude towards my supervisor, Mr. S. C. Premaratne, Senior Lecturer at the Faculty of Information Technology, University of Moratuwa for his patient guidance, encouragement, supervision and advice as well as his precious time spent towards the success of this research project.

Further I'm grateful to my husband and my parents, for their enormous support and encouragement given to me during this time and who experienced all the ups and downs of my research.

Finally, I thank all the staff members of Faculty of Information Technology, University of Moratuwa.

Abstract

Assessing student employability enables a method of integrating student abilities and organizations requirements, which is an important aspect for educational institutions. Improving student-evaluation techniques for employability will assist students to have a better knowledge of business organizations and find the right career for them. As a result, improved student employability prediction can assist students in matching their desirability to company requirements and fitting the employment profile of the firm for which they are searching.

The data is gathered through a survey in which students are asked to fill out a questionnaire in which they may indicate their abilities and academic achievement. This information may be used to determine their competency in a variety of skill categories, including soft skills, problem-solving skills and technical abilities and so on.

Data mining has been used in a variety of fields to efficiently assess large volumes of data. The aim of this study to predict student employability by considering different factors such as skills that the students have gained during their diploma level and time duration with respect to the knowledge they have captured when they expect the placement at the end of graduation by using the data mining techniques. Further during this research most specific skills with relevant to each job category also was identified.

In this research for the prediction of the student employability Rapid Miner software has used and different data mining models such as such as KNN, Naive Bayer's, and Decision Tree were evaluated based on classification techniques. The best model was identified among these models for this institute's student's employability prediction. Further associated technique has been used to identify the most associated skills with respect to each job category. So in this research classification and association techniques were used and evaluated. This study will be expanded to get more data by using a qualitative research, and further the employer's aspects of employability will also consider.

Table of Contents

	Page
Acknowledgements	ii
Abstract	iii
Table of Contents	iv
List of Figures	vii
List of Tables	x
Chapter 1	1
Introduction	1
1.1 Prolegomena	1
1.2 Background and Motivation	2
1.3 Problem Statement	3
1.4 Aims and Objectives	3
1.4.1 Aim	3
1.4.2 Objectives	3
1.5 Proposed Solution	4
1.6 Structure of the Dissertation	4
1.7 Summary	4
Chapter 2	5
Review of Literature	5
2.1 Introduction	5
2.2 Data-Mining Techniques	5
2.3 Data-Mining Process	6
2.4 Review of Previous Work	8
2.4.1 Literature Related to Basics of Employability	8
2.4.2 Literature Related to Factors Affecting for Employability Prediction	9

2.4.3 Literature Related to Data Mining Techniques for Employability	
Prediction	12
2.5 Summary	14
Chapter 3	15
Technologies and Tools Adopted	15
3.1 Introduction	15
3.2 Data Mining Techniques Used	15
3.3 Rapid Miner Studio	16
3.4. Decision Tree	16
3.5 K-Nearest Neighbors	17
3.6 Naïve Bayes	17
3.7 Association techniques	18
3.8 Summary	18
Chapter 4	19
A Novel Approach to Predict Employability	19
4.1 Introduction	19
4.2 Hypothesis	19
4.3 Input	19
4.4 Output	24
4.5 Process and Features	27
4.6 Summary	27
Chapter 5	28
Analysis and Design	28
5.1 Introduction	28
5.2 Data Collection	28
5.3 Design the classification and Association Tools	28
5.4 Summary	29

Chapter 6	30
Implementation	30
6.1 Introduction	30
6.2 Collection of Data (and Basic Functionalities)	30
6.3 Data Pre-Processing	31
6.4 Classification Models	33
6.5 Association Analysis	35
6.6 Summary	36
Chapter 7	37
Evaluation	37
7.1 Introduction	37
7.2 Evaluation of Classification Techniques	37
7.3 Evaluation of Association Techniques	42
7.4 Summary	60
Chapter 8	61
Conclusion and Further Work	61
8.1 Introduction	61
8.2 Limitations	61
8.3 Future Developments	62
8.4 Summary	62
References	63

List of Figures

	Page
Figure 2 1. Data-mining process	6
Figure 3.1 Data-mining techniques.....	16
Figure 4.1 Questionnaire created on Google Forms	20
Figure 4.2 Professional courses followed by the students	21
Figure 4.3 Time duration to find a job after graduation	22
Figure 4.4 Job categories of business finance students.....	22
Figure 4.5 Types of skills and their level of importance	23
Figure 4.6 Types of skills and their development during within the ATI (institute)	24
Figure 4.7 Feedbacks for the survey questionnaire	25
Figure 4.8 Sample classification output.....	26
Figure 4.9 Sample classification test set output.....	26
Figure 5.1 Designing of pattern recognition using classification methods.....	29
Figure 6.1 Replace missing values.....	31
Figure 6 2 Discretization process.....	32
Figure 6.3 End of data preprocessing final processes.....	32
Figure 6.4 Decision tree model using the Rapid Miner	33
Figure 6.5 Decision tree model with split validation using the Rapid Miner	34
Figure 6.6 Decision tree model with split validation process.....	34
Figure 6.7 Decision tree model with split validation- training the model	35
Figure 6.8 Identifying associated skills for each job category.....	36
Figure 7.1 Accuracy of predicting the job category- using KNN algorithm	38
Figure 7.2 Accuracy of predicting the job category- using Naïve Bayes algorithm.....	39
Figure 7.3 Accuracy of predicting the job category- using Decision Tree algorithm	39
Figure 7.4 Accuracy of determining the time period to get a job – using KNN algorithm	40
Figure 7.5 Accuracy of determining the time period to get a job – using Naïve Bayes algorithm.....	40

Figure 7.6 Accuracy of determining the time period to get a job – using Decision Tree algorithm.....	41
Figure 7.7 Job categories mentioned in the survey questionnaire	42
Figure 7.8 Identify the five most important skills.....	43
Figure 7.9 Process diagram to identify most associated-skills	44
Figure 7.10 Process diagram to identify associated-skills cont..	44
Figure 7.11 Sample frequency set for accountancy job category in the Rapid Miner	45
Figure 7.12 Accountancy job category with one associated skill type	45
Figure 7.13 Accountancy job category with two associated skill types	46
Figure 7.14 Accountancy job category with three associated skill types	46
Figure 7.15 Accountancy job category with four associated skill types.....	47
Figure 7.16 Auditing job category with one associated skill type	47
Figure 7.17 Auditing job category with two associated skill types	47
Figure 7.18 Auditing job category with three associated skill types	48
Figure 7.19 The most associated four skill types for auditing job category	48
Figure 7.20 Banking and finance job category with one associated skill type	48
Figure 7.21 Banking and finance job category with two associated skill types	49
Figure 7.22 Banking and finance job category with three associated skill types	49
Figure 7.23 Banking and finance job category with four associated skill types.....	50
Figure 7.24 Customer service job category with one associated skill type	50
Figure 7.25 Customer service job category with two associated skill types	51
Figure 7.26 Customer service job category with three associated skill types	51
Figure 7.27 Customer service job category with four associated skill types	52
Figure 7.28 Management and supervision job category with one associated skill type ...	52
Figure 7.29 Management and supervision job category with two associated skill types .	52
Figure 7.30 Management and supervision job category with three associated skill types	53
Figure 7.31 Management and supervision job category with four associated skill types.	53
Figure 7.32 Marketing business services and public relations job category with one associated skill type	54
Figure 7.33 Marketing business services and public relations job category with two associated skill types.....	54

Figure 7.34 Marketing business services and public relations job category with three associated skill types.....	55
Figure 7.35 Marketing business services and public relations job category with four associated skill types.....	55
Figure 7.36 Marketing business services and public relations job category with five associated skill types.....	56
Figure 7.37 Administration job category with one associated skill type	56
Figure 7.38 Administration job category with two associated skill types	57
Figure 7.39 Administration job category with three associated skill types	57
Figure 7.40 Administration job category with four associated skill types	58
Figure 7.41 Administration job category with five associated skill types.....	58
Figure 7.42 Whole sale and retail trade job category with one associated skill type	58
Figure 7.43 Whole sale and retail trade job category with two associated skill types	59
Figure 7.44 Whole sale and retail trade category with three associated skill types.....	59
Figure 7.45 Whole sale and retail trade job category with four associated skill types	60

List of Tables

	Page
Table 7.1 Accuracy of prediction the job category according to the skills.....	39
Table 7.2 Accuracy of determining the time period to get a job	41