

MODELING PRIMARY SCHOOL
STUDENT ACADEMIC
PERFORMANCE USING DATA
MINING TECHNIQUE

MUHAMAD BIN MAT YAACUB

UNIVERSITI UTARA MALAYSIA 2011

MODELING PRIMARY SCHOOL
STUDENT ACADEMIC PERFORMANCE
USING DATA MINING TECHNIQUE

A project submitted to Dean of Research and Postgraduate Studies Office in partial
Fulfillment of the requirement for the degree
Master of Science Information Technology
Universiti Utara Malaysia

By
Muhamad Bin Mat Yaacob



**KOLEJ SASTERA DAN SAINS
(College of Arts and Sciences)
Universiti Utara Malaysia**

**PERAKUAN KERJA KERTAS PROJEK
(Certificate of Project Paper)**

Saya, yang bertandatangan, memperakukan bahawa
(I, the undersigned, certifies that)

MUHAMAD MAT YAACUB
(89435)

calon untuk Ijazah
(candidate for the degree of) **MSc. (Information Technology)**


telah mengemukakan kertas projek yang bertajuk
(has presented his/her project of the following title)

**MODELING PRIMARY SCHOOL STUDENT ACADEMIC
PERFORMANCE USING DATA MINING TECHNIQUE**


seperti yang tercatat di muka surat tajuk dan kulit kertas projek
(as it appears on the title page and front cover of project)

bahawa kertas projek tersebut boleh diterima dari segi bentuk serta kandungan
dan meliputi bidang ilmu dengan memuaskan.
(that this project is in acceptable form and content, and that a satisfactory
knowledge of the field is covered by the project).

Nama Penyelia
(Name of Supervisor) : **MR. MOHAMAD FARHAN MOHAMAD MOHSEN**

Tandatangan
(Signature) :  Tarikh (Date) : 3/3/2011

Nama Penilai
(Name of Evaluator) : **DR. SHAHRUL AZMI MOHD. YUSOF**

Tandatangan
(Signature) :  Tarikh (Date) : 3/3/11

PERMISSION TO USE

In presenting this project in partial fulfillment of the requirements for a postgraduate degree from Universiti Utara Malaysia, I agree that the University Library may make it freely available for inspection. I further agree that permission for copying of this project in any manner, in whole or in part, for scholarly purpose may be granted by my supervisor(s) or, in their absence by the Dean of Postgraduate and Research. It is understood that any copying or publication or use of this project or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to Universiti Utara Malaysia for any scholarly use which may be made of any material from my project.

Requests for permission to copy or to make other use of materials in this project, in whole or in part, should be addressed to

Dean of Research and Postgraduate Studies
College of Arts and Sciences
Universiti Utara Malaysia
06010 UUM Sintok
Kedah Darul Aman
Malaysia

ABSTRAK

Setiap kali peperiksaan dijalankan, perkara pertama yang dibincangkan oleh kebanyakan pihak adalah prestasi pencapaian pelajar. Begitulah pentingnya prestasi pencapaian pelajar untuk setiap peperiksaan bagi seseorang pelajar. Prestasi pencapaian sekolah rendah telah lama dititikberatkan oleh Kementerian Pendidikan dan juga Pejabat Pelajaran Daerah. Oleh yang demikian, prestasi pencapaian sekolah rendah telah menjadi satu tanggungjawab yang perlu dilaksanakan dengan penuh tanggungjawab oleh guru dan pihak pentadbiran sekolah rendah. Mereka perlu memastikan setiap kali peperiksaan diadakan, prestasi pelajar perlu dipertingkatkan dari semasa ke semasa. Oleh yang demikian, tujuan kajian ini adalah untuk membangunkan satu model prestasi pencapaian akademik pelajar sekolah rendah dengan menggunakan kaedah Data Mining. Untuk menjayakan kajian ini, data peperiksaan Sekolah Kebangsaan Kuala Teriang, Langkawi telah dijadikan data contoh. Data contoh yang mengandungi 393 rekod mewakili prestasi akademik pelajar tahun 4 hingga 5 untuk tahun 2010. Data contoh ini mengandungi 6 kriteria yang terdiri daripada jantina and 5 subjek utama peperiksaan yang dinamakan Pemahaman, Penulisan, Sains, Bahasa Inggeris dan Matematik telah dikategorikan kepada cemerlang, sederhana dan lemah dan telah diujikan dengan kaedah Association Rules berdasarkan algoritma Apriori untuk mengenalpasti gabungan kriteria yang dapat mempengaruhi prestasi pencapaian pelajar. Dapatan kajian menunjukkan gabungan kriteria yang terdiri daripada Pemahaman = A, Penulisan = A, Sains = A, Bahasa Inggeris = A dan Matematik = A atau B adalah untuk kumpulan pelajar yang cemerlang. Gabungan kriteria yang terdiri daripada Pemahaman = A atau B, Penulisan = A atau B, Sains = B atau C, Bahasa Inggeris = B dan Matematik = C atau D adalah untuk kumpulan pelajar yang sederhana dan gabungan kriteria yang terdiri daripada Pemahaman = B atau C atau D, Penulisan = C atau D, Sains = C atau D, Bahasa Inggeris = C atau D dan Matematik = D atau E adalah untuk kumpulan pelajar yang lemah. Berdasarkan dapatan tersebut, satu prototaip sistem prestasi pelajar sekolah rendah dibangunkan untuk membantu guru dan pentadbiran sekolah mengenal pasti status prestasi pelajar yang melibatkan kumpulan cemerlang, sederhana dan lemah.

ABSTRACT

Every time when an examination comes, the first issue discussed by all parties is student performance. That scenario describes the importance of the student performance during the examination for student. Primary school student performances have long time focused by the Ministry of Education. Thus, the primary school academic performance becomes a problem that must be solved by teachers and school administrations. They responsible to improve the students' performance from time to time for every examination held. Thus, the aim of this study is to develop a model for primary school student academic performance using data mining technique. To achieve this, a student academic profile from Sekolah Kebangsaan Kuala Teriang, Langkawi was chosen as a case study. The dataset that consist of 393 records represent the academic performance of student from standard 4 to 6 in the year 2010. The dataset consists of 6 attributes that are gender and 5 core subjects namely *Pemahaman*, *Penulisan*, Sciences, English and Mathematics which were then grouped into excellent, fair and weak group and was mined using association rules technique based on Apriori algorithm to find interesting rules which can influence student academic performance. The finding indicates that the combination attributes of *Pemahaman* = A, *Penulisan* = A, Sciences = A, English = A and Mathematics = A or B is for excellent group. The combination attributes of *Pemahaman* = A or B, *Penulisan* = A or B, Sciences = B or C, English = B and Mathematics = C or D is for fair group and the combination attributes of *Pemahaman* = B or C or D, *Penulisan* = C or D, Sciences = C or D, English = C or D or E and Mathematics = D or E is for weak group. From the finding, a prototype of the primary school academic performance was developed to help the teachers and school administration identify their student academic performance status which involves excellent, fair and weak groups.

ACKNOWLEDGEMENT

In the name of Allah that the most merciful and most of compassionate. Thanks to Allah for granted me strength, patience and motivation to complete this study.

Firstly, I would like to thank to my Supervisor for giving me his support to complete this study, Mr. Mohamad Farhan B. Mohamad Mohsin for providing brilliant ideas, guidelines, comment and moral support throughout this study. Besides that, special thanks to Head Master, teachers and students of Sekolah Kebangsaan Kuala Teriang for providing me with the student academic performance dataset which was used in this study.

Finally, to anyone who has helped me in completing this study, thank you very much and may Allah bless you.

TABLE OF CONTENTS

	Page
PERMISSION TO USE	I
ABSTRACT (BAHASA)	II
ABSTRACT	III
ACKNOWLEDGMENTS	IV
TABLE OF CONTENTS	V
LIST OF TABLES	XIII
LIST OF FIGURES	X
CHAPTER I : INTRODUCTION	1
1.1 Problem Statement	5
1.2 Objective	6
1.3 Scope	7
1.4 Research Outcome	8
1.5 Significant and Contribution	8
1.6 Organization of the Thesis	9
CHAPTER II: LITERATURE REVIEW	10
2.1 Student Performance	10
2.2 Student Academic Performance	11
2.3 Academic Achievement in School	12
2.4 Data Mining	13
2.5 Association Rules	14
2.6 Data Mining in Education	18
2.7 Chapter Summary	19

CHAPTER III : METHODOLOGY	20
3.1 Data Collection	21
3.1.1 Instrument and Respondent	23
3.1.2 Target	24
3.2 Data Preparation	24
3.3 Data Mining Approach	28
3.3.1 Parameter Setting for Mining	29
3.4 Model Development	32
3.5 Prototype Development	34
3.6 Chapter Summary	37
CHAPTER IV : RESULTS AND FINDINGS	38
4.1 Excellent Group Mining Process	39
4.2 Fair Group Mining Process	49
4.3 Weak Group Mining Process	61
4.4 Primary School Student Academic Performance Model	75
4.5 Prototype Development	78
4.6 Chapter Summary	82
CHAPTER V : CONCLUSION AND RECOMMENDATION	83
5.1 Conclusion of the Study	83
5.2 Limitation	85
5.3 Recommendation and Future Work	86
REFERENCES	87
APPENDICES	89
Appendix A Sample Raw Data	90

Appendix B	Dataset	91
Appendix C	Excellent group mining process output	94
Appendix D	Fair Group mining process output	99
Appendix E	Weak group mining process output	104
Appendix F	Source Code for Prototype System	109

LIST OF TABLES

No	Title	Page
1.1	Sekolah Kuala Teriang Student Academic Performance	4
3.1	Distribution of Student	23
3.2	Student Classification	24
3.3	Sample of Excellent, Fair and Weak Group Dataset	27
3.4	Apriori Algorithm Options	30
4.1	Result for 50 Best Rules for Excellent Group	39
4.2	Detail Description about the interesting rules of 50 best rules	40
4.3	Result of 80 best rules for Excellent Group	43
4.4	Detail Description about the interesting rules of 80 best rules	44
4.5	Result of 100 best rules for Excellent Group	45
4.6	Detail description about the interesting rules of 100 best rules	46
4.7	Differentiate between Folds for Excellent Group	48
4.8	Result of 50 best rules for Fair Group	50
4.9	The 50 best rules for fair group	50
4.10	Detail Description of 50 best rules for fair group	52
4.11	Result of 80 best rules for fair group	53
4.12	Detail description for 80 best rules for fair group	54
4.13	Result of 100 best rules for fair group	56
4.14	Detail description of 100 best rules for fair group	57
4.15	The Differentiate between fold for fair group	59
4.16	Result of 50 best rules for weak group	61
4.17	Result detail of 50 best rules for weak group	62

4.18	Detail description of 50 best rules for weak group	63
4.19	Result of 80 best rules for weak group	64
4.20	Detail description of 80 best rules for weak group	66
4.21	Result of 100 best rules for weak group	67
4.22	Detail description of 100 best rules for weak group	69
4.23	The differentiate between fold for weak group	70
4.24	Summarization of interesting best rules found	74
4.25	Student Academic performance best rules found	76

LIST OF FIGURES

No	Title	Page
1.1	Scope of the primary school student performance system	8
2.1	Example of Association rules	16
2.2	Simple patterns	17
3.1	Research Methodology adapted from Usama	21
3.2	Sample of raw data	22
3.3	Data preparation process	25
3.4	Sample dataset	26
3.5	Attributes of Apriori Algorithm in WEKA	29
3.6	Excellent group dataset for 50 best rules	32
3.7	Model Development	33
3.8	RAD methodology for prototype	35
3.9	Modules of the system	36
4.1	Result of 50 best rules for excellent group	40
4.2	Interesting rules appeared in 50 best rules	42
4.3	Interesting rules appeared in 80 best rules	44
4.4	Interesting rules appeared in 100 best rules	47
4.5	Interesting rules appeared in 50 best rules for fair group	52
4.6	Interesting rules appeared in 80 best rules	55
4.7	Interesting rules appeared in 100 best rules	58
4.8	Interesting rule appeared in 50 best rules for weak group	63
4.9	Interesting rules appeared in 80 best rules	66

4.10	Interesting rules appeared in 100 best rules	69
4.11	Model for Primary school student academic Performance	77
4.12	Login menu interface	78
4.13	Main Menu Interface	79
4.14	Student performance identification interface	80
4.15	Student Reporting interface	81

CHAPTER 1

INTRODUCTION

Primary school students' academic performance becomes a serious issue that frequently discussed among primary school teachers and school administrator. The discussion basically related to the topics that concentrate to improve the students' academic performance achievement. According to examination report, every year the primary school students' academic performance shown an inconsistent result which is differs from year to year. Those phenomenon was worried a lot of parties especially teachers and school administrator, parents itself and also Ministry of education even though it was discussed and study from time to time.

The primary school students' academic performance achievement becomes a big challenge for country in order to produce the knowledgeable generation. The knowledgeable generation that will replace the country leadership depends on this generation especially in developing an educated society that capable to achieve the future development. Thus, primary school student academic performance becomes a basic education factor that determines the students' performance for secondary and higher learning education.

Besides that, primary school students are main assets for the country. The primary school students' academic performance plays an important role in producing the best quality of society

The contents of
the thesis is for
internal user
only

REFERENCES

- Baker, R.S.J.d. (2008). Data Mining for Education. To appear in McGaw,B., Peterson, P., Baker, E. (Eds.) *International Encyclopedia of Education (3rd edition)*. Oxford, UK: Elsevier.
- Beal, C. R. & Cohen, P. R. (2006). Temporal Data Mining for Educational Applications. *Proceeding of International Workshop on Applying Data Mining in e-Learning (ADML '06)*. pp. 3-12.
- Chansarkar, B.A. and Michaeloudis, A (2001). Student profiles and factors affecting performance. *International Journal Mathematics Education Science & Technology*, 32 (1). 97-104.
- Delavari, N., Beikzadeh, M.R. (2004). A New Model for Using Data Mining in Higher Educational System, *5th International Conference on Information Technology based Higher Education and Training: ITEHT '04, Istanbul, Turkey*, 31st May-2nd Jun 2004.
- Emmanuel N.Ogor. (2007). Student Academic Performance Monitoring and Evaluation Using Data Mining Techniques. *Fourth Congress of Electronics, Robotics and Automotive Mechanics. Department of Natural Sciences, Turks & Caicos Island Community College, Turkey*.
- Fomby, T (2008). Market Basket Analysis. *Proceeding of OCLC Systems & Services*. University of Joensuu, Finland 15(2), pp. 81-90.
- Hamalainen, W., Laine, T. H. & Suitinen, E. (2004). Data Mining in Personalizing. *Proceeding of Educational Data Mining Workshop*. Marina Del Rey, CA. USA. pp. 40-49.
- Jeffery, L & Lonnie, D. (1998). *System Analysis and Design*, (4th edition). New York: Mc Graw-Hill.
- Luan, J. (2001). Data mining as driven by knowledge management in higher education: Oxford. Persistence clustering and prediction. *Proceeding of Keynote for SPSS Public Conference, UCSF*. pp. 1-16.

- Muslihah, W. Norshahriah, W. & Fatimah, A. (2009). *Predicting NDUM Student's Academic Performance Using Data Mining Techniques*. 2nd International Conference on Computer and Electronic Engineering. National Defence University of Malaysia.
- Mohsin, M.F., Abd Wahab, M.H., Zaiyadi, M.F., Md Norwwi, N., and Hibadullah, F. (2010). *An Investigation Into Influence Factor of Student Programming Grade Using Association Rule Mining*. *Advance in Information Sciences and Service Siences*, 2, Number 2, June 2010
- Ong, L. C, Chandran V, Lim Y. Y, Chen A. H and Poh B. K (2010). *Factor Associated With Poor Academic Achievement Among Urban Primary School Children In Malaysia*. *Singapore Med J 2010, 51(3), 247*.
- Patricia L. H., Ching-Huei, C. Su-Hua, H. Chen-Ting, C. Fen-Lan, J. and Leslie, W (2006). *Factors Affecting High School Students' Academic Motivation in Taiwan*. *Asia Pacific Journal of Education*. 26 (2), 189-207.
- Raghavan, V. & Hafez, A. (2000). *Dynamic Data Mining*. University of Louisiana, USA. pp. 1-10.
- Rasmaswami, M and Bhaskaran, R (2010). *A Chaid Based Performance Prediction Model in Educational Data Mining*. *IJCSI In ternational Journal of Computer Science Issues*. 7(1), 11 - 15..
- Seifert, J. W. (2004). *Data mining: An overview*. *Proceeding of CRS Report for Congress*. East Lansing, USA. pp. 1-16.
- Tanimoto, S. T. (2007). *Improving the prospects for educational data mining*. Pp. 1-6. *The KDD Process for Extracting Useful Knowledge from Volumes of Data*, COMMUNICATIONS OF THE ACM November 996/Vol. 39, No. 11.
- Usama, F, & Gregory, P. S & (1996). *The KDD Process for Extracting Useful Knowledge from Volumes of Data*, Communication of ACM November 1996/Vol. 39, No. 11
- Vranic. M, Pintar. D and Skocir. Z. (2007). *The Use of Data Mining in Education Environment*,. *9th International Conference on Telecommunications*. Faculty of Electrical Engineering and Computing, University of Zegreb, Croatia.

Yahya. A & Abdul. K. S (2003). *Hubungan Gaya Pembelajaran Dengan Pencapaian Akademik Pelajar Di Tingkatan Empat Sekolah Menengah Teknik N.Sembilan*. National Seminar Memperkasakan Sistem Pendidikan, Puteri Pan Pacific, Johor Bahru.

Yvonne. B. W & K. Soyibo (2001). An Analysis of High School Students' Performance on Five Integrated Science Process Skills. *Research in Science & Technical Education*, 19 (2), 133 – 145.