LB/DON/142/2016
IT 01/147

Mining Continuous Assessments Marks to Predict Final Results

Indika Peiris

129160

LIBRARY
UNIVERSITY OF MORATUWA, SRI LANKA
MORATUWA

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

University of Moratuwa
TH3220

March 2016

004 "16"
004 (043)
TH3220
+CD-ROM
TH 3220

DECLERATION

I hereby declare that the project work entitled "Mining Continuous Assessment Marks to Predict Final Results" submitted to the University of Moratuwa, Faculty of Information Technology, is a record of an original work done by me under the guidance of Mr. S. C. Premaratne, Senior Lecturer of Department of Information Technology, University of Moratuwa and this project work is submitted in the partial fulfillment of the requirements for the award of the degree of Master of Information Technology. The results embodied in this thesis have not been submitted to any other University or Institute for the award of any degree or Diploma.

Date: March 15, 2016

Indika Pe.

G.V.A.D.I.P. Peiris

MSCIT/11/062

Supervised by

Mr. Saminda Premaratne

UOM Verified Signature

Signature

...

Date

ACKNOWLEDGEMENT

I would like mostly thank to Mr. S. C. Premaratne for his boundless and instructive help for making this research project success. Without his supervision, guidance and expert knowledge in data mining, this research would have not been possible.

I would like to thank Mr. Fordhous for guiding me to decide the scope of the project and collect necessary data.

Further I would like to thank Ms. Thilini for assisting me by providing necessary students data which was very important for this research project.

Next I would like to thank the staff members of Faculty of Information Technology for motivating and assisting me to do this course of study in numerous ways.

I take this opportunity to extend my gratitude to thank many of the colleagues and friends who have provided help, advice and companionship during my M.Sc. research study.

Finally I thank my parents for their encouragement and patience during research period.

ABSTRACT

Educational Data Mining is used to study the data available in the Universities, Higher Educational Institutes and other educational fields and extract the knowledge from it. As a result of reducing the cost of processing data and storing data, data storage became more easy and cheaper. Education Institutions are facing important and fast growth of the volume of educational data.

Data mining also called as Knowledge Discovery in Database (KDD) and search for inter relationships and patterns that can find, but already hidden among the vast volume of educational data.

Classification methods like decision trees, rule mining, Bayesian network etc can be applied on the educational data for predicting the students performance in examinations. This prediction will help the lecturers, teachers, tutors and students themselves to identify students' performance in the end semester examination. It will help the intelligent students to motivate more to maintain higher standard of marks and motivate weak students score better marks.

The J48 decision tree algorithm is applied on students' internal assessment marks to predict the grade they would gain at the end semester examination. In order do more accurate prediction some personal attributes like gender, their academic district, Advanced level Stream had been considered. With this research, students' who are likely to get higher grade or lower grade will be predicted more accurately. Predicted results can be distributed among teachers and tutors and necessary steps can be taken to improve the performance of the students who will be predicted to get lower grade or fail.

Table of Contents

Chapter 1 1

Introduction 1

1.1 Introduction 1

1.2 Background and Motivation 2

1.3 Problem Definition 2

1.4 Objectives of the Project 2

1.5 Structure of the Thesis 3

1.6 Summary 3

Chapter 2 4

Literature Review 4

2.1 Introduction 4

2.2 Data Mining Concept 4

2.3 Educational Data Mining 4

2.4 Data Mining Techniques 4

2.5 Data Mining Tools 5

2.6 Related Work 6

2.7 Summary 6

Chapter 3 7

Analysis and Technology Adapted towards Prediction 7

3.1 Introduction 7

3.2 Selection of inputs for the system 7

3.3 Process 8

3.4 Data Mining Techniques used in this Project 8

3.5 Expected Output 9

3.6 Data Analysis Process 9

3.6.1 Data understanding and Pre-Process 9

3.6.2 Visualizing Data 10

3.6.3 Creating Model for Data Analysis 10

3.6.4 Apply the Model for Data Analysis 10

3.6.5 Validate the Model 11

3.7 Calculating the Accuracy 14

3.8 Summary 14

Chapter 4 15

Development Approach 15

4.1 Introduction 15
4.2 Application of Model and Validation 15
4.2.1 Prediction of Results Based on One Attribute – Subject3 15
4.2.2 Summary of Prediction using one Attribute –Subject 3 18
4.2.3 Prediction of Results Based on Two Attributes – Subject 3 18
4.2.4. Summary of Prediction Using Two Attributes-Subject3 21
4.2.5 Prediction with three attributes – Subject3 22
4.2.6 Application of Model – Subject3 22
4.2.7 Validation of Model – Subject3 23
4.2.8 Calculating the Accuracy – Subject3 24
4.2.9. Prediction of Results based on One Attribute – Subject 2 24
4.2.10 Prediction of Results based on Two Attributes – IN 1310 31
4.2.11 Summary of Correctly classified rate for single attribute (Subject 2 B12 & B13) 33
4.2.12 Summary of Correctly classified rate for two attributes (Subject 2 B12 & B13) 33
4.2.13 Summary of Correctly classified rate for two attributes (Subject 2 B12 + B13) 34
4.2.14 Application of Model Subject 2(B12 + B13) 34
4.2.15 Validation of Model Subject 2(B12 + B13) 35
4.2.16 Calculating Accuracy subject 2(B12 + B13) 35
4.2.17 Prediction of Results based on one attribute - subject 1 36
4.2.18 Prediction of Results based on two attributes - subject 1 37
4.2.19 Prediction of Results based on three attributes - subject 1 37
4.3 Gender → Results Analysis 38
4.4 Summary 39
Chapter 5 40
Design 40
5.1 Introduction 40
5.2 Introducing Changes to Rules 40
5.2 Summary 44
Chapter 6 45
Implementation 45
6.1 Introduction 45
6.2 Proposed System 45
6.2.1 Inputs for the proposed system 45
6.2.2 Process 45
6.2.3 Output of the proposed system 45

6.3 Prerequisites for the proposed system 45
6.4 Proposed System Implementation 46
6.4 Results Prediction Process 47
6.5 Testing 51
6.6 Summery 52
Chapter 7 53
Conclusion and Further Work 53
7.1 Introduction 53
7.2 Conclusion 53
7.3 Further work 54
References 55
Appendix A 57

Table of Figures

Figure3. 1 Data mining Techniques and Methods 9
Figure3. 2 Creation of Model for prediction 10
Figure3. 3 Application of Model 11
Figure3. 4 Validation of Model step1 12
Figure3. 5 Validation of Model Step 2 12
Figure3. 6 Validation of Model step 3 13
Figure3. 7 Validation of Model step 4 13

Figure 4. 1 Decision Tree-CA:Gender 18
Figure 4. 2 Decision Tree CA=A: District 19
Figure 4. 3 Decision Tree Z score: Results 20
Figure 4. 4 Decision Tree CA: AL Stream 21
Figure 4. 5 Z-score → results (B12) subject2 24
Figure 4. 6 Z-score → results (B13) Subject 2 25
Figure 4. 7 Z-score → results (B12+B13) Subject 2 26
Figure 4. 8 District → results (B12) Subject 2 27
Figure 4. 9 District → results (B13) - Subject 2 27
Figure 4. 10 District → results (B12+B13) - Subject 2 28
Figure 4. 11 Gender → results (B12) Subject 2 28
Figure 4. 12 Gender-> results B12 subject 2 29
Figure 4. 13 Gender-> Results (both Batches) subject 2 29
Figure 4. 14AL stream → results (B12+B13) Subject 2 30
Figure 4. 15 CA → results (B12+B13) Subject 2 30
Figure 4. 16 Decision Tree CA, AL → Results 31
Figure 4. 17 Decision Tree CA, District → Results 32
Figure 4. 18validation of model subject 2 35
Figure 4. 19 Confusion Matrix subject 2 35
Figure 4. 20 Subject, Gender, Grade Distribution 38

Figure 5. 1 result prediction when CA=B and District=Kandy 41
Figure 5. 2result prediction when CA=B and District=Kurunegala 43

Figure 6. 1 Predicator Interface 46
Figure 6. 2 Interface to Import data file 47
Figure 6. 3 After data importing to the predicator interface 48
Figure 6. 4 Predicted Results 49
Figure 6. 5 Save predicted results to the local machine 50
Figure 6. 6 open the excel file saved in the local machine 50
Figure 6. 7 Excel file with Predicted Results 51

Table of Tables

Table 3. 1 Confusion Matrix 14

Table 4. 1 Summary of Prediction- subject 3 18
Table 4. 2 summary of prediction with two attributes 21
Table 4. 3 Application of Model - subject 3 23
Table 4. 4 Validation of Predicted Results - subject 3 23
Table 4. 5 Accuracy Calculation with Confusion Matrix - subject 3 24
Table 4. 6 summary of correctly classified instance rate -one attribute 36
Table 4. 7 summary of correctly classified instance rate -two attribute 37
Table 4. 8summary of correctly classified instance rate -two attribute subject1 37
Table 4. 9 subject, gender, grade distribution 39

Table 5. 1 results prediction when CA=B and AL=Biology 40
Table 5. 2 result prediction when CA=B and AL=Biology and District=Kandy R=A 41
Table 5. 3 results prediction when CA=B and AL=Biology 42
Table 5. 4 result prediction when CA=B and AL=Biology and District=Kurunegala R=A 43