# CONTENTS

**NOTES ON CONTRIBUTORS**

**INTRODUCTION: EDUCATION AT COMPUTATIONAL CROSSROADS**  
Samira ElAtia, Donald Ipperciel, and Osmar R. Zaïane

## PART I

**AT THE INTERSECTION OF TWO FIELDS: EDM**

### CHAPTER 1  EDUCATIONAL PROCESS MINING: A TUTORIAL AND CASE STUDY USING MOODLE DATA SETS  
Cristóbal Romero, Rebeca Cerezo, Alejandro Bogarín, and Miguel Sánchez-Santillán

1.1 Background 5  
1.2 Data Description and Preparation 7  
1.2.1 Preprocessing Log Data 7  
1.2.2 Clustering Approach for Grouping Log Data 11  
1.3 Working with ProM 16  
1.3.1 Discovered Models 19  
1.3.2 Analysis of the Models’ Performance 23  
1.4 Conclusion 26  
Acknowledgments 27  
References 27

### CHAPTER 2  ON BIG DATA AND TEXT MINING IN THE HUMANITIES  
Geoffrey Rockwell and Bettina Berendt

2.1 Busa and the Digital Text 30  
2.2 Thesaurus Linguae Graecae and the Ibycus Computer as Infrastructure 32  
2.2.1 Complete Data Sets 33  
2.3 Cooking with Statistics 35  
2.4 Conclusions 37  
References 38

### CHAPTER 3  FINDING PREDICTORS IN HIGHER EDUCATION  
David Eubanks, William Evers Jr., and Nancy Smith

3.1 Contrasting Traditional and Computational Methods 42  
3.2 Predictors and Data Exploration 45  
3.3 Data Mining Application: An Example 50
3.4 Conclusions 52
References 53

CHAPTER 4  
EDUCATIONAL DATA MINING: A MOOC EXPERIENCE 55
Ryan S. Baker, Yuan Wang, Luc Paquette, Vincent Aleven, Octav Popescu, Jonathan Sewall, Carolyn Rosé, Gaurav Singh Tomar, Oliver Ferschke, Jing Zhang, Michael J. Cennamo, Stephanie Ogden, Therese Condit, José Diaz, Scott Crossley, Danielle S. McNamara, Denise K. Comer, Collin F. Lynch, Rebecca Brown, Tiffany Barnes, and Yoav Bergner

4.1 Big Data in Education: The Course 55
4.1.1 Iteration 1: Coursera 55
4.1.2 Iteration 2: edX 56
4.2 Cognitive Tutor Authoring Tools 57
4.3 Bazaar 58
4.4 Walkthrough 58
4.4.1 Course Content 58
4.4.2 Research on BDEMOOC 61
4.5 Conclusion 65
Acknowledgments 65
References 65

CHAPTER 5  
DATA MINING AND ACTION RESEARCH 67
Ellina Chernobilsky, Edith Ries, and Joanne Jasmine

5.1 Process 69
5.2 Design Methodology 71
5.3 Analysis and Interpretation of Data 72
5.3.1 Quantitative Data Analysis and Interpretation 73
5.3.2 Qualitative Data Analysis and Interpretation 74
5.4 Challenges 75
5.5 Ethics 76
5.6 Role of Administration in the Data Collection Process 76
5.7 Conclusion 77
References 77

PART II

PEDAGOGICAL APPLICATIONS OF EDM 79

CHAPTER 6  
DESIGN OF AN ADAPTIVE LEARNING SYSTEM AND EDUCATIONAL DATA MINING 81
Zhiyong Liu and Nick Cercone

6.1 Dimensionalities of the User Model in ALS 83
6.2 Collecting Data for ALS 85
6.3 Data Mining in ALS 86
6.3.1 Data Mining for User Modeling 87
6.3.2 Data Mining for Knowledge Discovery 88
6.4 ALS Model and Function Analyzing 90
## CONTENTS

<table>
<thead>
<tr>
<th>Chapter 6</th>
<th>The “Geometry” of Naïve Bayes: Teaching Probabilities by “Drawing” Them</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.4.1</td>
<td>Introduction of Module Functions</td>
</tr>
<tr>
<td>6.4.2</td>
<td>Analyzing the Workflow</td>
</tr>
<tr>
<td>6.5</td>
<td>Future Works</td>
</tr>
<tr>
<td>6.6</td>
<td>Conclusions</td>
</tr>
<tr>
<td>Acknowledgment</td>
<td></td>
</tr>
<tr>
<td>References</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chapter 7</th>
<th>Examining the Learning Networks of a MOOC</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.1</td>
<td>Introduction</td>
</tr>
<tr>
<td>7.1.1</td>
<td>Main Contribution</td>
</tr>
<tr>
<td>7.1.2</td>
<td>Related Works</td>
</tr>
<tr>
<td>7.2</td>
<td>The Geometry of NB Classification</td>
</tr>
<tr>
<td>7.2.1</td>
<td>Mathematical Notation</td>
</tr>
<tr>
<td>7.2.2</td>
<td>Bayesian Decision Theory</td>
</tr>
<tr>
<td>7.3</td>
<td>Two-Dimensional Probabilities</td>
</tr>
<tr>
<td>7.3.1</td>
<td>Working with Likelihoods and Priors Only</td>
</tr>
<tr>
<td>7.3.2</td>
<td>De-normalizing Probabilities</td>
</tr>
<tr>
<td>7.3.3</td>
<td>NB Approach</td>
</tr>
<tr>
<td>7.3.4</td>
<td>Bernoulli Naïve Bayes</td>
</tr>
<tr>
<td>7.4</td>
<td>A New Decision Line: Far from the Origin</td>
</tr>
<tr>
<td>7.4.1</td>
<td>De-normalization Makes (Some) Problems Linearly Separable</td>
</tr>
<tr>
<td>7.5</td>
<td>Likelihood Spaces, When Logarithms make a Difference (or a SUM)</td>
</tr>
<tr>
<td>7.5.1</td>
<td>De-normalization Makes (Some) Problems Linearly Separable</td>
</tr>
<tr>
<td>7.5.2</td>
<td>A New Decision in Likelihood Spaces</td>
</tr>
<tr>
<td>7.5.3</td>
<td>A Real Case Scenario: Text Categorization</td>
</tr>
<tr>
<td>7.6</td>
<td>Final Remarks</td>
</tr>
<tr>
<td>References</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chapter 8</th>
<th>Exploring the Usefulness of Adaptive Elearning Laboratory Environments in Teaching Medical Science</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.1</td>
<td>Review of Literature</td>
</tr>
<tr>
<td>8.2</td>
<td>Course Context</td>
</tr>
<tr>
<td>8.3</td>
<td>Results and Discussion</td>
</tr>
<tr>
<td>8.4</td>
<td>Recommendations for Future Research</td>
</tr>
<tr>
<td>8.5</td>
<td>Conclusions</td>
</tr>
<tr>
<td>References</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chapter 9</th>
<th>Thuan Thai and Patsie Polly</th>
</tr>
</thead>
<tbody>
<tr>
<td>9.1</td>
<td>Introduction</td>
</tr>
<tr>
<td>9.2</td>
<td>Software for Learning and Teaching</td>
</tr>
<tr>
<td>9.2.1</td>
<td>Reflective Practice: ePortfolio</td>
</tr>
<tr>
<td>9.2.2</td>
<td>Online Quizzes</td>
</tr>
<tr>
<td>9.2.3</td>
<td>Online Practical Lessons</td>
</tr>
<tr>
<td>References</td>
<td></td>
</tr>
</tbody>
</table>
CHAPTER 10  INVESTIGATING CO-OCCURRENCE PATTERNS OF LEARNERS’ GRAMMATICAL ERRORS ACROSS PROFICIENCY LEVELS AND ESSAY TOPICS BASED ON ASSOCIATION ANALYSIS  

Yutaka Ishii

10.1 Introduction 157
10.1.1 The Relationship between Data Mining and Educational Research 157
10.1.2 English Writing Instruction in the Japanese Context 158
10.2 Literature Review 159
10.3 Method 160
10.3.1 Konan-JIEM Learner Corpus 160
10.3.2 Association Analysis 162
10.4 Experiment 1 162
10.5 Experiment 2 163
10.6 Discussion and Conclusion 164
Appendix A: Example of Learner’s Essay (University Life) 164
Appendix B: Support Values of all Topics 165
Appendix C: Support Values of Advanced, Intermediate, and Beginner Levels of Learners 168
References 169

PART III

EDM AND EDUCATIONAL RESEARCH 173

CHAPTER 11  MINING LEARNING SEQUENCES IN MOOCs: DOES COURSE DESIGN CONSTRAIN STUDENTS’ BEHAVIORS OR DO STUDENTS SHAPE THEIR OWN LEARNING? 175

Lorenzo Vigentini, Simon McIntyre, Negin Mirriaahi, and Dennis Alonzo

11.1 Introduction 175
11.1.1 Perceptions and Challenges of MOOC Design 176
11.1.2 What Do We Know About Participants’ Navigation: Choice and Control 177
11.2 Data Mining in MOOCs: Related Work 178
11.2.1 Setting the Hypotheses 179
11.3 The Design and Intent of the LTTO MOOC 180
11.3.1 Course Grading and Certification 183
11.3.2 Delivering the Course 183
11.3.3 Operationalize Engagement, Personal Success, and Course Success in LTTO 184
11.4 Data Analysis 184
CHAPTER 12
UNDERSTANDING COMMUNICATION PATTERNS IN MOOCs: COMBINING DATA MINING AND QUALITATIVE METHODS

Rebecca Eynon, Isis Hjorth, Taha Yasseri, and Nabeel Gillani

12.1 Introduction 207
12.2 Methodological Approaches to Understanding Communication Patterns in MOOCs 209
12.3 Description 210
12.3.1 Structural Connections 211
12.4 Examining Dialogue 213
12.5 Interpretative Models 214
12.6 Understanding Experience 215
12.7 Experimentation 216
12.8 Future Research 217
References 218

CHAPTER 13
AN EXAMPLE OF DATA MINING: EXPLORING THE RELATIONSHIP BETWEEN APPLICANT ATTRIBUTES AND ACADEMIC MEASURES OF SUCCESS IN A PHARMACY PROGRAM

Dion Brocks and Ken Cor

13.1 Introduction 223
13.2 Methods 225
13.3 Results 228
13.4 Discussion 230
13.4.1 Prerequisite Predictors 230
13.4.2 Demographic Predictors 232
13.5 Conclusion 234
Appendix A 234
References 236

CHAPTER 14
A NEW WAY OF SEEING: USING A DATA MINING APPROACH TO UNDERSTAND CHILDREN’S VIEWS OF DIVERSITY AND “DIFFERENCE” IN PICTURE BOOKS

Robin A. Moeller and Hsin-liang Chen

14.1 Introduction 237
14.2 Study 1: Using Data Mining to Better Understand Perceptions of Race 238
14.2.1 Background 238
CHAPTER 15
DATA MINING WITH NATURAL LANGUAGE PROCESSING AND CORPUS LINGUISTICS: UNLOCKING ACCESS TO SCHOOL CHILDREN’S LANGUAGE IN DIVERSE CONTEXTS TO IMPROVE INSTRUCTIONAL AND ASSESSMENT PRACTICES

Alison L. Bailey, Anne Blackstock-Bernstein, Eve Ryan, and Despina Pitsoulakis

15.1 Introduction 255
15.2 Identifying the Problem 256
15.3 Use of Corpora and Technology in Language Instruction and Assessment 261
15.3.1 Language Corpora in ESL and EFL Teaching and Learning 261
15.3.2 Previous Extensions of Corpus Linguistics to School-Age Language 262
15.3.3 Corpus Linguistics in Language Assessment 263
15.3.4 Big Data Purposes, Techniques, and Technology 264
15.4 Creating a School-Age Learner Corpus and Digital Data Analytics System 266
15.4.1 Language Measures Included in DRGON 267
15.4.2 The DLLP as a Promising Practice 268
15.5 Next Steps, “Modest Data,” and Closing Remarks 269
Acknowledgments 271
Appendix A: Examples of Oral and Written Explanation Elicitation Prompts 272
References 272

INDEX 277