INDEX

Page numbers in *italics* refer to figures, those in **bold** refer to tables.

**action research**
- challenges, 75–6
- curriculum leaders, 77
- data collection process, administration role, 76–7
- design methodology, 71–2
- and EDM, 68–9, 72
- goal, 67
- internal problem analysis, 69
- qualitative and quantitative approach, 71
- qualitative data analysis and interpretation, 74–5
- quantitative data analysis and interpretation, 73–4
- summative data analysis, 73
- teacher–researcher
  - committed teachers, 69
  - data interpretation, 73
  - focused topic, 70
  - literature review, 70
  - mixed methodology approach, 70
  - narrowing process, 70
- vs. traditional research, 68
- triangulation, 72

**adaptive coach for exploration (ACE), 102**

**adaptive eLearning laboratory environments**

**pathology, 139**

**potential limitations, 152–3**

**software, learning and teaching**
- Gene Suite, 147–52
- online practical lessons, 144–5
- online quizzes, 143–4
- reflective practice, ePortfolio, 141–2
- virtual laboratories, 145–6

**Adaptive eLearning Platform (AeLP), 140**

**adaptive hypermedia (AH) system, 81**

**adaptive learning**
- advantages of, 81
- environment, 82

**adaptive learning systems (ALS)**
- data collection for, 85–6
- data mining for
  - knowledge discovery, 88–90
  - multiple regression analyses, 88
  - predicting student performance, 87
- user modeling, 87–8
- future works, 94
- knowledge base quality, 86
- management module, 92
- nodes in working process, 93–4
- reasoning module, 92
- schematic illustration, 90, 91
- storage module, 91–2
- user model, dimensionalities of
  - classification, 84
  - cognitive ability, 85
  - differential model, 84
  - for disabled learners, 84
  - domain-specific and domain-independent information, 83–4
- ELM-ART, 83
- genetic model, 83–4
- individual model, 83
- learner’s performance, 85
- learner’s preferences, 84–5
- learner’s prior knowledge, 84
- overlay model, 83–4
adaptive learning systems (ALS) (cont’d)
  perturbation model, 84
  plan model, 84
  scalar model, 83
  stereotype model, 84
  student models, for intelligent tutoring systems, 83
  User Synopsis, 83
  working memory capacity, 84
ALS see adaptive learning systems (ALS)
Aquinas, Thomas, 30
association rule mining, 60
attrition predictors, 49, 49
automatic clustering method, 12
automatic speech recognition (ASR) programming, 270
Bayesian analysis, 104
Bayesian decision theory, 103–5
Bayesian Knowledge Tracing (BKT), 60, 87
Bayesian nonnegative matrix factorization (BNMF), 215
Bazaar tool, 58
BDEMOOC see Big Data in Education behaviors, natural clustering, 194–8
classification workflow, 195
cluster centroids distribution, 196
descriptive and performance values, 197
external validity of, 194
performance, 194
Bernoulli Naïve Bayes, 110–111
Big Data in Education
  Bazaar tool, 58
cognitive tutor authoring tools, 57–8
course content
  advanced topics, 61
  behavior detection, 59
diagnostic metrics, 59
  feature engineering, 59
  knowledge inference, 60
  prediction modeling, 58–9
  relationship mining, 60
  structure discovery, 61
  visualization, 60–61
Coursera platform, 55–6
edX platform, 56–7
RapidMiner, 58
research
  costs of MOOCs, 64
  MOOC learners’ motivations, 64–5
natural student communities, 63
negativity toward MOOC instructors, 63–4
post-MOOC participation, 61–2
predicting early stop-out, 62
social network analysis, 63
student language on discussion forum, 62–3
binary relevance classifier, 88
buggy messages, 57
Busa, Roberto, 29–31
causal data mining, 60
CCBC see Cooperative Children’s Book Center (CCBC)
“chance-corrected” measures, 43
close reading, 36
clustering techniques, 4
cognitive tutor authoring tools (CTAT), 57–8
comma-separated value (CSV) format, 9, 10
comments, online communities
  analyzing phrases, 240
  common phrases, 242
distribution, 241
  with like vote, 244, 245
  multidimensional scaling map, 242
Communications Security Establishment of Canada (CSEC) PowerPoint slide deck, 38
computational methods data mining, 42–4
computer-science data mining, 35
confusion matrix, for FUTACT12, 48, 48
co-occurrence patterns, grammatical errors, 163
Cooperative Children’s Book Center (CCBC), 248
Cooperative Institutional Research Program (CIRP), 50
corpus linguistics, 256
  big data purpose, 264–6
  ESL and EFL teaching and learning, 261–2
  language assessment, 263–4
  school-age language, 262–3
correlation mining, 60
Coursera commitment
  Honor code, 192
  selection of, 193
Coursera peer assessment tool, 181
INDEX 279

Coursera platform, 55–6
crosstabs, 45
data mining (DM), 100, 256
ALS
  knowledge discovery, 88–90
  multiple regression analyses, 88
  predicting student performance, 87
  user modeling, 87–8
  application of, 50–52
  classification problems, 45
  contrasting traditional and computational methods, 42–4
vs. educational research, 157–8
in educational settings (see educational data mining (EDM))
in higher education, 41–2
online communities, availability and popularity, 238
online news organizations, comments role, 238–9
predictors and data exploration, 45–50
research challenges, 3
data wrangling, 41
de-normalization, NB classifier
decision line, 112–14, 113–14
likelihood spaces, 115–16, 116
distant reading, 36
do-it-yourself (DIY) corpus-building, 261–2
DRGON see Dynamic Relational Graphical Ontological Network (DRGON)
Dynamic Language Learning Progressions (DLLPs), 255, 256, 268–9
Dynamic Relational Graphical Ontological Network (DRGON), 256, 258, 259, 267–8, 270–271
EDM see educational data mining (EDM)
educational data mining (EDM), 41, 99–100 see also massive open online courses (MOOCs) and action research, 68–9, 72
challenges, 75–6
data preprocessing, 70
design methodology, 71–2
e-learning, 72
ethics, 76
goals, 3
grouping students, process of, 6
IEDMS definition, 68
interactive cycle, 71
process mining, 4
qualitative and quantitative approach, 71–2
qualitative data analysis and interpretation, 74–5
quantitative data analysis and interpretation, 73–4
technological tools, 68
educational process mining (EPM) clustering techniques, 4
goal, 4
Moodle 2.0 course
  grouping log data, clustering approach for, 11–16
  preprocessing log data, 7–11
  with ProM, 16–25
for SRL with SCLEs, 4
edX platform, 56–7
English as a foreign language (EFL), 260
language corpora, 261–2
English as a second language (ESL), 260
language corpora, 261–2
English learner (EL) students, 256 see also corpus linguistics
  error, 266
  language abilities, progress in, 257
English writing
  association analysis, 162
  beginner-level learners, 164
  error analysis, 159–60
  error tags, 161
  grammatical errors, co-occurrence patterns, 163
  in Japanese context, instructions of, 158–9
  Konan-JIEM learner corpus, 160–162
  reliability coefficients, essay topics, 163
EPM see educational process mining (EPM)
epPortfolio, 141–2
error analysis (EA), 159–60
Error Detection and Correction Workshop (EDCW) 2012, 157
expectation–maximization (EM) algorithm, 15–16
Freshman Research Program (FRP), 50–52, 51
Frischer, Bernie, 32
FUTACT12
area under the curve, 48
confidence intervals for, 45–7, 46
collapse matrix for, 48, 48
distribution table for, 45, 46
ROC graph for, 47, 47

Gene Suite
electrophoretic mobility shift assay, 147
gel preparation, 149
polymerase chain reaction, 147
purpose of, 152
quantitative PCR, 147
sample loading, 150
Smart Sparrow AeLP authoring tool, 148, 151
vLabs, 147
Western blotting, 147, 149
Google Books, 29, 37
grammatical errors, co-occurrence patterns, 163

Helping Our Own (HOO) 2011 and 2012, 157
Heuristics Miner, 6
higher education
data mining in, 41–2
institutional effectiveness, 42
hint messages, 57

Ibycus computer
complete data sets, 33–5
Thesaurus Linguae Graecae and, 32–5
Ibycus Scholarly Personal Computer (PSC), 32–3
Index Thomisticus project (Busa), 29–32
institutional data, mining of, 45
institutional effectiveness, higher education, 42
intelligent tutoring system, 140
interactive classification and extraction (ICE) approach, 118
interactive ML (IML), 101
International Educational Data Mining Society (IEDMS), 68
inter-rater agreement, 43–4, 44
Inter-Rater Facets, 43

Japanese context, English writing instructions, 158–9
Jockers, Matt, 36
Johnson, William, 32
Journal of Educational Data Mining, 41
knowledge points (KPs), 89
Konan-JIEM learner corpus, 160–6
labeling, 102
language abilities, monitoring progress in, 257
language assessment, 263–4
learning analytics (LA), 99–100
learning management systems (LMSs), 3
learning resources organization mode, 89
learning tool integration (LTI) activities, 185
learning to teach online (LTTO), 180
design and intent components, 181
course delivering, 183
course design, 180
course grading and certification, 183
course success, 184
curriculum design, 180
data sources, 185–6
engagement/participation, 184
evolution, 180
grading design, 183
hypothesis, 183
operationalize engagement, 184
personal success, 184
structure, 182
teaching, 180
in numbers, 186
Lempke, S.D., 249
Literary Data Processing (Tasman), 31
LTTO see learning to teach online (LTTO)
machine-learned latent response model, 88
machine learning (ML), 42, 100
management module, of ALS, 92
manual clustering method, 11–12
massive open online courses (MOOCs), 100, 121
“Aboriginal Worldviews and Education”
age of participants, 126, 127
course content, 124–5
education levels of participants, 126, 127
forum contributions vs. achievement, 130
forum posts and comments, 129
geographical location of participants, 126, 127
identity of participants, 126
participant gender, 126, 128, 128, 129
postcourse surveys, 131–2
structure and content, 132–3
challenges, 211
data analysis, sources, 185–6
data mining in
hypothosis, 179–80
issues, 179
references, 178, 179
demographics and engagement
patterns, 123
digital technology, 123
digital traces, 208
equity concerns, 124
examining dialogue, 213–14
experimentation, 216–17
future research recommendations, 133–4
globalization and technology, 123
impact, 176
interpretative models, 214–15
learning communities, 122
LTTO
characterizing patterns, 186–9
design and intent, 180–184
in numbers, 186
marginalization, 122
methodological and analytical
approaches, 210
mining behaviors and intents
natural clustering, behaviors, 194–8
participants’ intent and behaviors, 191–4
stated type, 198
openness and inclusivity, 125
participants
across educational sectors, 187
distribution, 187–8
funnel of, 187
navigation patterns, 177–8
pedagogy and course enhancement, 198–200
perceptions and challenges of, 176–7
pragmatic paradigm in, 209
redefining participation and engagement, 189–91
research methods, 210
structural connections
communication networks, 212
complementary approach, 213
social network analysis, 211
thread network visualizations, 212
time, 212
vulnerability of networks, 212
types of, 177
understanding experience, 215–16
mathematical notation, NB classifier, 102–3
MOOCs see massive open online courses (MOOCs)
Moodle logs, EPM
grouping log data, clustering
approach for
automatic clustering, 12
expectation–maximization algorithm, 15–16
manual clustering, 11–12
selected variables, 12, 13–14
Weka system, 15, 15
preprocessing log data, 7–11
comma-separated value format, 9, 10
event logs, 7
events in quiz view, 9
MXML file, 9, 11
ProM import tool, 9, 10
students’ activities, relevant actions
to, 8
variables of log file, 8
Moodle Quiz, 143
Moretti, Franco, 36
multidimensional scaling (MDS), 241
naïve Bayes (NB) approach, 109–10
naïve Bayes (NB) classifier, 101
decision line, 111–12
de-normalization, 112–14, 113–14
geometry of
Bayesian decision theory, 103–5
mathematical notation, 102–3
likelihood spaces, 114–15, 115
de-normalization in, 115–16, 116
new decision in, 116–17, 117
text categorization, 117, 118
two-dimensional probabilities, 105, 105–6, 106
Bernoulli Naïve Bayes, 110–111
de-normalizing probabilities, 108–9, 109
NB approach, 109–10
working with likelihoods and priors
only, 107, 107–8
National Center for Research on Evaluation, Standards, and Student Testing (CRESST), 267
natural language processing (NLP)
corpus linguistics, 255
problem identification, 256–61
NBTree classification algorithm, 88
network analysis, 60
nonsignature track (NST)
participants, 186
objective interestingness measures, 35
online communities
availability and popularity, 238
and comments
analyzing phrases, 240
common phrases, 242
distribution, 241
with like vote, 244, 245
multidimensional scaling map, 242
and common phrases, 242
from negative comments, 246, 247
from positive comments, 245, 246
positivity/negativity evaluation, 240
research questions, 239–40
WordStat, 241
online learning systems, 99
online practical lessons, 144–5
online quizzes, 143–4
Ontological Relations Builder (ORB)
system, 267
ontology, construction of, 42
oral and written explanation, 272

Packard, David, 32
paradoxes, 43
pathology, 139
Performance Factors Analysis (PFA), 60
pharmacy program
in Canada, 224, 224
data mining, admission record, 225
demographics
and average GPA, 227
predictors, 232–4
education of, 223
multiple linear regression, 226
prerequisite predictors, 230–232
raw data, 225–6
results, predictors, 228–9
in United States, 224
in University of Alberta, 225
unstandardized regression coefficients, 226, 229–30, 234–5
picture book concepts
CCBC, 248
definition, 249
differences in, 249
findings, 250–251
implications, 252
methodology, 250
research questions, 249
thematic differences, 251
predictor approach, 50
process mining (PM)
conformance checking, 5
description, 5
enhancement, 5
implementation, 4
process discovery, 5
ProM framework, 6
self-regulated learning models, 5–6
types of, 5
Project MORE, 263
ProM
description, 17
EPM with, 16
Heuristic Miner algorithm, 17–18, 18
interface for importing log file, 17
model analysis, 19–25
performance/fitness of models, 23–5
plug-ins, 18, 18
framework in process mining, 6
import tool, 9, 10

RapidMiner, 58
rater agreement, 43–5
reasoning module, of ALS, 92
receiver operator characteristic (ROC) graph, for FUTACT12, 47, 47
Renaissance humanism, 29
Ruhleder, Karen, 33

school-age language, 262–3
school-age learner corpus and digital data analytics system
DLLP project, 266, 268–9
DRGON, language measure, 267–8
modest data, 269–71
second language acquisition (SLA) teaching, 259
self-regulated learning (SRL) with student-centered learning environments (SCLEs), 4
Semantic analysis methods, 90
Semantic Web, 90
sequential pattern mining, 60
Snowden, Edward, 38
social network analysis (SNA), 211
storage module, of ALS, 91–2
subjective interestingness measures, 35
summative data analysis, 73
Survey of Text Mining: Clustering, Classification, and Retrieval, 90
susceptible-infected (SI) model, 213

Tasman, Paul, 31
test of English as a foreign language (TOEFL), 263
text mining, 29, 89–90
The Freshman Survey (TFS), 50–52, 51
Thesaurus Linguae Graecae/Thesaurus of the Greek Language and Ibycus computer, 32–5
complete data sets, 33–5
Treasury of the French Language (TLF), 35
two-dimensional probabilities
NB classifier, 105, 105–6, 106
Bernoulli Naïve Bayes, 110–111
de-normalizing probabilities, 108–9, 109
NB approach, 109–10
working with likelihoods and priors, 107, 107–8
UCLA Classicist’s Workbench, goal of, 32
user model, ALS
data mining for, 87–8
dimensionalities of classification, 84
cognitive ability, 85
differential model, 84
for disabled learners, 84
domain-specific and domain-independent information, 83–4
ELM-ART, 83
genetic model, 83–4
individual model, 83
learner’s performance, 85
learner’s preferences, 84–5
learner’s prior knowledge, 84
overlay model, 83–4
perturbation model, 84
plan model, 84
scalar model, 83
stereotype model, 84
student models, for intelligent tutoring systems, 83
User Synopsis, 83
working memory capacity, 84
module, 92
User Synopsis, 83

Valla, Lorenzo, 29
virtual laboratories (vLabs), 145–6, 152–3
Virtual Ryan, 58

Weka system, 15, 15
WordStat, 241
working memory capacity (WMC), 84