## Index

### A
- acceptance/rejection sampling 345
- accuracy-weighted ensemble 203
- actionability 1, 930
- active
  - disks 6
  - learning 12, 1115
- AdaBoost 450
- adaptive
  - metric techniques 685
  - numerical value discretization method 816
  - sampling 346
  - Web sites 1223
- adaptivity 23
- adjacency lattice 61
- administration and management 17
- advising system 435
- aggregate conjunctive query 422
- aggregates 319
- aggregation 33, 469
- Akaike Information Criterion 465
- algebraic 197
- algorithm(s) 263, 504
- alternative document models 996
- analytical
  - information technologies 233
  - query 1055
- analytics 698
- annotation-free WI systems 679
- anomaly 78
  - detection 251, 384
- ANOVA F value 80
- antecedent 1098
- approximate 1196
  - query answering 990
- approximation space 659
- Apriori 59, 509, 957
  - based approach 1010
- AQ-PM Systems 203
- araneus data model (adm) 715

### archiving 321
ARIMA 1125
ARIS 118
artificial
  - intelligence 893, 1083
  - neural networks (ANN) 191, 273, 433
  - neuron 54
  - assertion between knowledge patterns 648
association
  - informatics 706
  - mining 560
  - rule 272, 773, 1217
    - discovery 795, 1245
    - extraction 752
    - mining 59, 65, 70, 403, 1098
    - rules 45, 74, 150, 288, 735, 746, 763, 859, 923, 925, 930, 941, 1222
    - utilization 705
attribute-oriented induction 74
audio retrieval by content 854
audit fees 1175
auditing 217
auditory scene analysis 854
authority 444, 1207
automated classification 1113
automatic
  - document classification 383
  - indexing and searching of audio files 86
  - music transcription 855
autonomous 1
averaging 355

### B
- background knowledge 494
- backpropagation learning 865
- bagging 355, 449
- base data 1054
- Bayesian 90
  - belief networks 146
  - criteria 465
  - inference 92
C

cache management 1221, 1223
CAD/CAM 324
cancer-testis (CT) antigen 551
candidate sequence 1029
canonical model 962
capital efficient marketing 912
car pooling 408
case studies 257
CASPian 1163
categorical attributes 34
categorization 129
causal relation 834
causality 263
CD3 203
cell 196
censoring 1077
census 94
center-based clustering 134
centering function 519
chat discussions 758
chemical, biological, radiological, nuclear 427
chemoinformatics 664
chi-squared distance 685
chunks 198
circulation 705
class-conditional distributions 34
classical probability 90
classification 141, 144, 272, 448, 622, 865, 922, 930, 1064, 1071, 1113, 1171, 1222
   ambiguity 1176
   and regression trees 353
   based on associations 774
   tree 141
classifier system 487
classifying 1155
clean data 1043
cleansing 454
clickstream 1216
data 1207, 1227
client log 438
closed
   feature set 514
   itemset(s) 150, 942
   sequential pattern 1029
cluster
   analysis 154, 159, 729, 1217
   centre 582
   sampling 345
   -support 516
clustering 165, 172, 176, 180, 272, 309, 477, 622, 659, 735, 820, 995, 1087, 1114, 1122, 1155, 1222
   algorithms 159
Index

and classification 263
methods 582
techniques 1201
co-training 1022
codes of ethics 832
cooperation 758
filtering (CF) 44, 124, 931, 1237
and personalized search 438
virtual environments (CVEs) 1190, 1193

collective dependence 938
collinearity 504
collocation 383
combinatorial optimization 482
committee methods 448
communities, informal 769
community mining 444
compact and generalized abstraction 902
compensation planning 262
competitive advantage 17, 382
complex data types 178
composite symbolic object 1089
compositional algorithms 191
compression 321
computational
biology 482
complexity 695
crimes

ethics 832
vision 965
-teaching 270
computer programming, the art of 59
computing 634
with words 562
concept drift 202
concepts 129, 150
conceptual
clustering for video data 1187
construction 293
maps 130
condensed representations 207
condition number 504
conditional
pattern bases 60
probabilities 89
probability 89
probability table (CPT) 90, 92
confidence 1098
confidentiality 832, 923, 1005
confusion matrix 466
congressional sampling 992
consent 832
consequent 1098
constraint-based mining 207
content

analysis 373
-image retrieval 837
context 514
ultra-sensitive approach 1238
continuous
association rule mining algorithm (CARMA) 61
auditing (CA) 217
contradictions 76
contrast sets 796
control of inductive bias 530
cooperation 487
cooperative 639
information system 297
core competence 387
corporate
financial distress 503
libraries 100
correlation 78, 1122
graph 957
cost
of turnover 262
projections 265
-sensitive 930
crisp 510
criteria based on score functions 465
critical support 510
CRM 1093
cross
-correlation 1127
-industry standard process (CRISP) 122
for data mining 894
cross-validation 466, 503, 1072
crossover 505
cryptographic techniques 926
Cube
lattice 197
operator processing 423
-by operator 469, 877
cubegrades 288
cuboid 196
curse 178, 684, 902, 1192
curve overlap 154
customer
classification 339
relationship management 59, 891, 950
customized prescriptions 239
cut point 392
CVFDT 203
cyber
attacks 251
security 567
D

damage detection 245
techniques 246
data 1190
acquisition 454, 886
aggregation 1054
analysis 454
and model assembly 645
cleaning 1227
cleansing 996
collection methods 1000
compression 307
cube 28, 421, 470, 877
comparative computational approaches 469
dependency 947
envelopment analysis (DEA) 350
fusion 1227
imputation 595, 1044
integration 298
intensive Web sites 714
interoperability 94
management 566
marts 335
mining 112, 129, 159, 165, 190, 224, 239, 262, 272, 372, 388, 392, 403, 438, 443, 448, 454, 560, 621, 625, 630, 689, 728, 734, 746, 763, 785, 859, 921, 925, 947, 973, 990, 1048, 1206
algorithms 190
benchmarking association 265
by using database statistics 626
components 339
models 39, 339
technologies 233
missing
at random 594
completely at random 594
modeling 239
normalization 811
perturbation 926
preprocessing 1216
quality 297, 636, 1043, 1049
reduction 52, 307, 621
repository 647
security 921
sequence 1028
staging area (DSA) 312
stream(s) 59, 172
validation 1044
visualization 644, 1190
warehouse 259, 318, 312, 523, 826, 876, 906, 1054
development 18
team 17

warehouses 17, 101
warehousing 14, 94, 223, 263, 317, 630, 1048, 1092
-driven 223
-mining 141, 483
applications 378
data table 359
technology 433
database 443
machines 7
marketing 698
partitioning 319
systems 746
datamart 263
datasets of multidimensional 1196
decision
making 1135
rule 353
learning 145
rules from data 975
support 459, 990
system (DSS) 17, 233, 326, 469, 630
tree 394
algorithms 240
analysis 264
construction algorithm 141
induction 353
learning 144
trees 141, 184, 693, 735, 1175
-making 1211
decompositional algorithms 191
deductive data mining 562
deep Web mining 443
defense 94
software collaborators (DACS) 265
defining user requirements 459
degree of belief 90
dehooking 887
demand forecasting 1129
Dempster-Shafer Theory 1135, 1166
density
biased sampling 345
-based clustering 160
Department of
Agriculture 94
Defense 268
Transportation 94, 268
dependent variable 698
depth first 151
descriptive function 433
design 313
recovery 112
device profiles 25
diabetes 259, 359
Index

diagnosis 259
differential evolution 1168
digital
   forensics 568
      library 100, 758
      reference 102
dilemma 454
dimensional
   algebra 226
   attributes 654
   data model 223
   model 319
dimensionality reduction 843
dimensions 523
direct marketing 931
directed
   acyclic graph (DAG) 92, 674
   edges 92
disclosure 834
discovering knowledge 689
discovery 190
   informatics 387, 705
discrete
   attributes 397
      wavelet transform –(DWT) 307, 308
discretization 392
discriminant analysis 503
disease 259
disjoint clustering 515
dissimilarity measures 1088
distributed 2
   autonomous information systems 640
distributive 197
divide-and-conquer approach 1089
DNA (Deoxyribonucleic Acid) 735
   chip 106
document
   classification 1015
      cluster map (DCM) 1131
      clustering 555, 607, 1147
domain organization 1212
dot reduction 887
drug
   design 664
      discovery 239
dual problem 1074
dynamic
   bandwidth allocation 956
   clustering algorithm 1088
      graphs 543
   health strategies (DHS) 265
   sampling 346
   time warping 172
topic mining 608

Web pages 714
   weighted majority 203
   weighting 134

E

e-commerce 1235
e-business (EB) 349
e-government 300
e-mail
   filing 768
   mining 768
e-metrics 1229
e-service 23
eature space 1074
ectastic algorithms 191
   “factless” fact 572
   efficiency 349
eigenfaces 967
electronic
   business (EB) 349
      customer relationship management
      (e-CRM) 892
      product code 1160
elementary granules 561
   elitism 505
   EM (expectation-maximization) algorithm 1115
   emantic associations 1011
   emerging patterns 796
   empirical risk 1064, 1065
   employee retention 262
   encoding schemes 680
   enhanced data mining 294
   ensemble 448
      methods 202, 355
      rule based classification methods 487
   enterprise resource planning (ERP) 118, 630
   Environmental Protection Agency 94
   episodal association rule r 1100
   episode 1099
   equal
      employment opportunity (EEO) 263
      frequency discretization method 398
      width interval binning 398
   ERR 888
   error 834
   estimated error 932
   ethics 454, 832
   ethnographic 459
   Euclidean distance 582
   event 1099
      type 1099
   evidence-based librarianship 103
   evolutionary algorithms 477, 483, 487
      approaches 184
computation 482
executive information system 269
expectation maximization 134, 1023
experimental stress tolerances 736
expertise 435
explanation 190
construction 494
evaluation 495
-oriented association mining 493
explicit 599
exponential weighting scheme 686
eXtensible Business Reporting Language (XBRL) 217
external
data 328, 826
segmentation 886
extraction
rules 13
extraction, transformation and loading (ETL) 312
extraction-transformation-loading 95
extractors 678
F
F value 80
face recognition 965
factor
analysis (FA) 498
loadings 499
facts 523
failure analysis (FA) 1077
false
acceptance rate (FAR or type II error) 874
discovery rate 850
negatives 848
positives 848
FAQ detecting 439
FAR 888
feature 283
construction 33, 530
element 773
-based image classification (FEBIC) 775
elements 774
extraction 530, 902
matrix 872
relevance 684
selection 285, 400, 902
and extraction 477
vectors 773
Federal 94
Aviation Administration 268
Credit Union 94
Government 268
Fellegi-Holt method 1044
field learning 611
filtering 887
financial ratios 503
finite closed interval 519
fitness landscapes 487
FLORA 203
forgeries 870
formal concept analysis (FCA) 150, 514
fragments 1059
framework 493
fraud 269
detection 269
free text 382
freeblock scheduling 8
frequency 1099
frequent
closed
episodes 1099
feature set 515
itemsets 754
itemsets 404, 555
mining 207
itemsets 150, 790, 859
pattern 941
-growth 60
-tree 60
structure mining 1140
subgraph mining 540
FRR 888
functional genomics 728
fuzzy
and interval set clustering 659
c-means 661
clustering 162, 180
histogram 520
logic 272
number 519
observations 294
patterns 294
rules 272
set theory 1175
transformation 295
G
Gaussian kernel 1074
gene
chip 106
clustering 283
expression microarray 283
ontology 812
selection 283
methods 107
general accounting office 268
generalized
additive multi-mixture models 355
Index

disjunction-free literal set model (GD 942)
projection (GP) 30, 423
generating-pruning 1029
generative model 1016
generators 942
  of frequent closed itemsets 754
genes 734
genetic 504
  algorithm 273, 378, 477, 779, 1201
  and evolutionary computation (GEC) 477
  programming (GP) 377, 529
genetically modified (GM) foods 736
genomics 482, 835
geographic information systems (GIS) 97, 409
geographical information systems 233
global
  features 872
  search 479
  -support 516
glycemic control 360
Gram-Schmidt Kernels 667
grammar-based 529
granular computing 560
graph 540, 1010
databases 1059
  grammar 542
  kernel 666
  mining 1140
  representation 526
  transaction 1141
  -based ARM 71
  data mining 1010
graphical
  models 935, 1232
  user interface (GUI) tools 326
greedy strategy 685
grid-based clustering 161
group
  differences 795
  health
    analysis 265
    management 265
group pattern discovery systems 546
Grubbs’ test 180
GT 324

H
hamming clustering 978, 981
health
  and urban development 94
  and wellness initiatives 265
  plan
    benefits 265

employer data and information set (HEDIS)
  265
Health Insurance Portability and Accountability Act
  265
heterogeneous gene expression data sets 550
heuristic rating algorithm 779
heuristics 1083
hidden
  Markov models 616
  Web 443
hierarchical
  clustering 159
  conceptual clustering 541
  methods 166
  partitioning 264
high
  frequency patterns 560
  pressure region image 872
histogram(s) 50, 309, 991
historical research 570
HITS 444
HMD 577
holistic 197
homeland security 566
homonymy 648
Hopfield neural network 295
horizontally partitioned database 927
HPR 872
HSOM 577
HTL 577
hubs 444, 1207
human
  resource information systems (HRIS) 262
  -computer interaction 644
  -computer interface 645
Human Resources Benchmarking Association™ 265
humanities
  data warehousing 570
  research 570
hybrid approaches 185
hyperbolic
  multidimensional scaling 577
  self-organizing map 577
  space 575
  tree layout 577
HyperCluster Finder 399
hyperlink analysis 1207
hyperlinked communities identification 1147
hyponym 648
hyperplane 1064
  classifier 1066
Hypersoap 966
hypertext
  analysis 384
model 714
view 710
hyponym 648
hypothesis 1064
testing 848

iceberg
cube 198
query 198
identifier attributes 34
identity theft 264
ideological 130
if-then rules 1175
IFL method 611
ignorance 1166, 1167
illumination 966
image
features and their relations 805
search engines 773
structure 805
impact factors 552
impoverish epistemology 224
imprecision 273
improving performance 190
inapplicable responses 593
inclusion 648
dependency mining 626
incomplete data 293
incremental
approaches 202
hypertext view maintenance 710
learning 529
mining 1030
text mining 607
independence principle 1104
index
trees 309
-term selection 439
indexing techniques 906
individualized protocols 243
inductive
database 207, 1010
logic
processing (ILP) 1010
programming 33, 70
reasoning 766, 978, 979
inexact
fielding learning (IFL) 611
rules 611
inference 92, 962
network model 961
information 382

browsing 409
enhancing 546
extraction 278, 615, 678, 1146, 1206
integration system 678
management 1092
paralysis 891
retrieval 159, 278, 377, 382, 438, 443, 960, 1109, 1146
scientists 100
seeking 1000
system 2
technology 893
theory 763
visualization 1190
-retrieval 328
-theoretic 542
informational privacy 455
informed consent 833
instance
matching 625
selection 621
-based pattern synthesis 902
integrated library systems 101
integration of data sources 625
intelligence density (ID) 630
intelligent
data analysis 634
text mining 383
interaction 75
interactive visual data mining 644
interactivity 151
interdisciplinary study 634
interestingness 796
of rules 1
interface schema 445
interlibrary loan 102
Internal Revenue Service 94
Internet 328, 349
interpage structure 443
interpretability 1175
interschema properties 647
intersite schema matching 445
interstructure mining 372
intrapage structure 443
intrasite schema matching 445
infrastructure mining 372
intrusion detection 251, 568
system 251
isomorphic 561
isomorphism modulo a product 30
issues in mining digital libraries 279
item 1028
itemset 509, 795, 1028
**Index**

<table>
<thead>
<tr>
<th>J</th>
<th>K</th>
</tr>
</thead>
<tbody>
<tr>
<td>JDM 39, 1171</td>
<td>k-dimensional fuzzy vector 519</td>
</tr>
<tr>
<td>joint probability distribution (JPD) 429</td>
<td>k-free sets 942</td>
</tr>
<tr>
<td>k-harmonic means 134</td>
<td>k-lines 779</td>
</tr>
<tr>
<td>k-means 134, 166, 660, 1155</td>
<td>k-medoid 166</td>
</tr>
<tr>
<td>k-nearest neighbor queries 308 classifier 146</td>
<td>Kaplan-Meier method 1077</td>
</tr>
<tr>
<td>kernel 1068 function 1074 approach 1010</td>
<td>kernel 1068</td>
</tr>
<tr>
<td>functions 842, 1064 methods 664</td>
<td>partial least squares methods 687</td>
</tr>
<tr>
<td>partial least squares based methods 687</td>
<td>k-density-based clustering 173</td>
</tr>
<tr>
<td>keyword 1211</td>
<td>knowledge 382, 599</td>
</tr>
<tr>
<td>-based association analysis 383</td>
<td>creation 383, 600</td>
</tr>
<tr>
<td>knowledge discovery 372, 385, 388, 390, 482, 746, 1181</td>
<td>methods 639</td>
</tr>
<tr>
<td>process 593</td>
<td>management 94, 268, 387, 758</td>
</tr>
<tr>
<td>domain visualization 995</td>
<td>nuggets 76</td>
</tr>
<tr>
<td>management 94, 268, 387, 758</td>
<td>representations 184</td>
</tr>
<tr>
<td>nuggets 76</td>
<td>transformation 600</td>
</tr>
<tr>
<td>representations 184</td>
<td>-based recommenders 124</td>
</tr>
<tr>
<td>transformation 600</td>
<td>Kohonen self-organizing maps 661</td>
</tr>
</tbody>
</table>

| L | lifetime value 913 |
|---|lift chart 466|
| large datasets 1089 | linear discriminant analysis 685 |
| lasers 734 | models 503 |
| latent semantic indexing 807 | regression 309 |
| lattice | linguistics variables 1175 |
| of concepts 151 | linguistics computing 570 |
| theory 61 | link analysis 383, 566, 735 |
| layer-based strategy 151 | mining 1012 |
| learning algorithms 184 | local features 872 |
| theory 689, 1064, 1065 | pattern-analysis 546 |
| leave-one-out 503 | location services 965 |
| libraries 100 | log data 1216, 1226 |
| | files 1207 |
| | -based query expansion 438 |
| | clustering 438 |
| | logic data 693 |
| | synthesis 978, 981 |
| | logical analysis of data (LAD) 689 |
| | logistics support system 94, 268 |
| | loss function based criteria 466 |
| | low prediction accuracy (LPA) 611 |
| | lower bound 659 |

<table>
<thead>
<tr>
<th>M</th>
<th>machine learning 124, 234, 272, 390, 443, 448, 674, 693, 1015, 1083, 1113</th>
</tr>
</thead>
<tbody>
<tr>
<td>magnum opus 796</td>
<td>maintenance algorithm 712</td>
</tr>
<tr>
<td></td>
<td>mal-processes 118</td>
</tr>
<tr>
<td></td>
<td>marked-up language 384</td>
</tr>
<tr>
<td></td>
<td>market basket analysis 59, 925</td>
</tr>
<tr>
<td></td>
<td>efficiency theory 779</td>
</tr>
<tr>
<td></td>
<td>research 698</td>
</tr>
<tr>
<td></td>
<td>segmentation 339</td>
</tr>
<tr>
<td></td>
<td>-basket analysis 653</td>
</tr>
<tr>
<td></td>
<td>Markov chain models 1231</td>
</tr>
<tr>
<td></td>
<td>matching parameters 303</td>
</tr>
<tr>
<td></td>
<td>material acquisition 705</td>
</tr>
<tr>
<td></td>
<td>categories 706</td>
</tr>
<tr>
<td></td>
<td>materialization 710</td>
</tr>
<tr>
<td></td>
<td>materialized hypertext view 714</td>
</tr>
<tr>
<td></td>
<td>view(s) 29, 717, 906, 1055</td>
</tr>
<tr>
<td></td>
<td>mathematical programming 1072</td>
</tr>
</tbody>
</table>
maximal-profit item selection (MPIS) 67
maximum
a posteriori or MAP decision rule 89
entropy 763
MDLPC 398
mean filter 871
median filter 871
medical
digital libraries 278
informatics 243
transactional database 359
MEDLINE 616
membership functions 1175
memorized 55
Mercer’s condition 1074
message understanding conferences 615
meta-
classification 552
data 329, 826
learner 15
learning 12
recommenders 46
metabolic pathways 736
metadata 298, 382, 1048, 1211
database 1050
model 1048
standards 1049
metainformation 1048
method schema 1000
methods for gene selection 107
metric-driven 223
micro-array data 687
microarray 106, 728, 810, 848
data sets 107
image analysis 735
microarrays 734
military services 268
min-max basis for
approximate association rules 755
exact association rules 755
minconf 509
mine rule 740
minimal
non-redundant association rule 755
occurrence 1100
minimum
cost satisfiability problem 694
description length 541
support 1028
mining 1120
biomedical literature 280
of association rules 740
question-answer pairs 769
sentences 770
MINSAT 695
minsup 509
MIS 1092
missing values 1044
misuse detection 251
mixture of curves 154
mobile user trajectories 956
modal data 247
model
combiners 448
identification 820
-based
clustering 161, 134
pattern synthesis 902
modeling 313, 434
models of the data 340
molecular biology 735
monotonicity of the query 198
Monte Carlo method 414
moral 454
morphology 384
mosaic-based retrieval 838
multi-
agent system 23
criterion optimization 478
dimensional 523
dimensional sequential pattern mining 1031
link lookahead search 938
method 185
relational 540
view learning 12
multidimensional
cubes 1054
database 17
datasets 1196
on-line analytical processing (MO 877
multifactorial diseases 483
multimedia 965
metadata 842
multimodal analysis 843
multiple
data source mining 546
hypothesis testing 848
relations 71
site link structure analysis 996
multivariate discretization 398
multiview learning 1022
music information retrieval 854
mutation 505
N
n-dimensional cube 196
Naive Bayes 1115
Naive
### Index

Bayesian classification 89  
Bayesian classifier 145  
-Bayes 394  
National Committee for Quality Assurance (NCQA) 265  
national security 832, 1012  
natural language  
   - document 1147  
   - processing 384, 443, 1109  
navigational patterns 1218  
NCR Teradata database system 746  
nearest neighbor  
   - classifier 904  
   - rule 684  
   - and correlation-based recommender 45  
negative  
   - association rules 860  
   - itemsets 860  
nested objects 715  
networked data 34  
near neural networks 146, 245, 272, 507, 735, 779, 865  
neuro-fuzzy computing 274  
noise  
   - clustering 180, 582  
   - magnitude 154  
   - reduction 887  
   - removal 871  
noisy link 444  
non-derivable itemsets 942  
non-Ignorable Missing Data 594  
nonlinear dynamics 234  
nonparametric classifier 904  
normalization 570, 871, 887  
null hypothesis 849  
number of intervals 397  
numerical dependencies 31  

### O

object  
   - classification 1166  
   - identification 302  
objective 492  
   - function-based clustering 582  
observability 1162  
ODM Implementation 892  
OLAP 17, 319, 413, 1196  
   - systems 422  
OLE DB for DM 39, 1171  
on-line analytical processing (OLAP) 876, 891  
online  
   - aggregation 991  
   - analytical processing (OLAP) 233  
   - 196, 335, 469, 876, 1083  
   - technologies 325  

### P

page  
   - class 714  
   - scheme 715  
   - vectorization 1202  
PageRank 444  
pageview 1227  
paradigm shift 239  
partial  
   - cube 199  
   - memory approaches 202  
partition  
   - search space 791  
   - based pattern synthesis 903  
partitioning 543  
   - algorithm 817  
   - and parallel processing 906  
   - clustering 160  
path mining 896  
pattern  
   - classification 684  
   - count tree 904  
   - decomposition 790, 791
<table>
<thead>
<tr>
<th>Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>discovery</td>
</tr>
<tr>
<td>finding</td>
</tr>
<tr>
<td>recognition</td>
</tr>
<tr>
<td>problem</td>
</tr>
<tr>
<td>synthesis</td>
</tr>
<tr>
<td>with negation</td>
</tr>
<tr>
<td>-projection</td>
</tr>
<tr>
<td>patterns</td>
</tr>
<tr>
<td>Pearsonian correlation coefficient</td>
</tr>
<tr>
<td>pedagogical algorithms</td>
</tr>
<tr>
<td>peer-to-peer</td>
</tr>
<tr>
<td>perceptron</td>
</tr>
<tr>
<td>performance</td>
</tr>
<tr>
<td>personal digital assistants (PDAs)</td>
</tr>
<tr>
<td>personalization</td>
</tr>
<tr>
<td>personnel hiring records</td>
</tr>
<tr>
<td>pharmaceutical industry</td>
</tr>
<tr>
<td>piece-wise affine (PWA) identification</td>
</tr>
<tr>
<td>pixel</td>
</tr>
<tr>
<td>plant</td>
</tr>
<tr>
<td>geonomics</td>
</tr>
<tr>
<td>microarray databases</td>
</tr>
<tr>
<td>-made-pharmaceuticals</td>
</tr>
<tr>
<td>plausibility</td>
</tr>
<tr>
<td>PMML</td>
</tr>
<tr>
<td>Poincaré disk</td>
</tr>
<tr>
<td>point density</td>
</tr>
<tr>
<td>policy learning</td>
</tr>
<tr>
<td>polynomial kernel</td>
</tr>
<tr>
<td>pooled information</td>
</tr>
<tr>
<td>post-pattern-analysis</td>
</tr>
<tr>
<td>pragmatics</td>
</tr>
<tr>
<td>pre-calculation</td>
</tr>
<tr>
<td>predicate tree (P-tree)</td>
</tr>
<tr>
<td>prediction</td>
</tr>
<tr>
<td>join</td>
</tr>
<tr>
<td>predictive</td>
</tr>
<tr>
<td>function</td>
</tr>
<tr>
<td>model markup language (PMML)</td>
</tr>
<tr>
<td>preprocessing</td>
</tr>
<tr>
<td>primary Web data</td>
</tr>
<tr>
<td>principal</td>
</tr>
<tr>
<td>component analysis</td>
</tr>
<tr>
<td>direction divisive partitioning</td>
</tr>
<tr>
<td>priori algorithm</td>
</tr>
<tr>
<td>privacy</td>
</tr>
<tr>
<td>issues</td>
</tr>
<tr>
<td>preserving data mining</td>
</tr>
<tr>
<td>probabilistic</td>
</tr>
<tr>
<td>decision table</td>
</tr>
<tr>
<td>graphical models</td>
</tr>
<tr>
<td>reasoning</td>
</tr>
<tr>
<td>probability distributions</td>
</tr>
<tr>
<td>probably approximately correct (PAC) learning</td>
</tr>
<tr>
<td>PROC OPTEX</td>
</tr>
<tr>
<td>procedural factors</td>
</tr>
<tr>
<td>process</td>
</tr>
<tr>
<td>management</td>
</tr>
<tr>
<td>modeling methodology</td>
</tr>
<tr>
<td>of discretization</td>
</tr>
<tr>
<td>product assortment decisions</td>
</tr>
<tr>
<td>professionalism</td>
</tr>
<tr>
<td>profit mining</td>
</tr>
<tr>
<td>program</td>
</tr>
<tr>
<td>analysis</td>
</tr>
<tr>
<td>mining</td>
</tr>
<tr>
<td>progressive</td>
</tr>
<tr>
<td>sampling</td>
</tr>
<tr>
<td>progressively</td>
</tr>
<tr>
<td>projected databases</td>
</tr>
<tr>
<td>projection indexes</td>
</tr>
<tr>
<td>proportional hazards model</td>
</tr>
<tr>
<td>propositional logic</td>
</tr>
<tr>
<td>propositionalization</td>
</tr>
<tr>
<td>protein-protein interaction graph</td>
</tr>
<tr>
<td>prototype(s)</td>
</tr>
<tr>
<td>proxy</td>
</tr>
<tr>
<td>log</td>
</tr>
<tr>
<td>server</td>
</tr>
<tr>
<td>pruning</td>
</tr>
<tr>
<td>algorithm</td>
</tr>
<tr>
<td>search space</td>
</tr>
<tr>
<td>pseudo-independent (Pl) models</td>
</tr>
<tr>
<td>public health</td>
</tr>
<tr>
<td>publish/subscribe</td>
</tr>
<tr>
<td>push-cache</td>
</tr>
<tr>
<td>Q</td>
</tr>
<tr>
<td>quadratic time</td>
</tr>
<tr>
<td>qualitative data</td>
</tr>
<tr>
<td>quality of care</td>
</tr>
<tr>
<td>quantified beliefs</td>
</tr>
<tr>
<td>quantitative</td>
</tr>
<tr>
<td>association rules</td>
</tr>
<tr>
<td>data</td>
</tr>
<tr>
<td>SAR (QSAR)</td>
</tr>
<tr>
<td>query</td>
</tr>
<tr>
<td>answering system</td>
</tr>
<tr>
<td>evaluation</td>
</tr>
<tr>
<td>language engines</td>
</tr>
<tr>
<td>log</td>
</tr>
<tr>
<td>mining</td>
</tr>
<tr>
<td>preprocessing</td>
</tr>
<tr>
<td>optimization</td>
</tr>
<tr>
<td>parallelism</td>
</tr>
<tr>
<td>reformulation</td>
</tr>
<tr>
<td>session clustering</td>
</tr>
<tr>
<td>sessions</td>
</tr>
</tbody>
</table>
Index

subsumption 31
-by-humming 855
-by-sample-image 837
-by-sketch 837
questionnaires 1045

R
radio frequency identification (RFID) 326
random forest 355
range
  query 49
  -aggregate queries 1196
ranking 838
  function discovery 378
ratio 79
raw data 436
receiver operating characteristic 466
recommendation 930
recommender systems 44, 124, 758, 931, 1235, 1246
reconciliation 300
record linkage 302
recursive partitioning 354
reengineering 947
refinement 190
reflexive relationships 70
refusal of response 593
regression 141, 144, 264, 272, 735
  clustering 134
  statistics 1064
reinforcement learning 529
relational 540
  data 70
    mining 72
database systems 262
databases 33
join operation 71
learning 33
modeling 570
on-line analytical processing (ROLAP) 877
privacy 455
queries 808
relationship 1120
relative
  advantage 1162
  frequencies 90
relevance feedback 438, 837
reporting 334
representation
  of frequent patterns 942
  space 397
  episodal association rules 1100
representative samples 413
resampling 503, 886
reservoir sampling 345
result schema 445
reuse 529
risk
  exposure 265
  models 265
rough
classification 611
decision table 974
set and genetic algorithm 272
sets 659, 973
-set genome 660
rule
  extraction 191
generation 821
    method(s) 978
mining 509
    -based classification models 773
S
salient variables 820
sample 413
class prediction 283
clustering 283
sampling 310, 622, 991
satisfiability problem 694
scalability 636, 1181
  and efficiency 178
scalable data-mining tools 434
scanned images 734
schema 523
  abstraction 647
  integration 647
  matching 625
  theory 477
scheme 131
science 234
scientific
  applications 1196
  research 492
Web Intelligence 995
scientometrics 995
score function 516
search
  engine 328, 443, 1146, 1211
  situation 1001
    transition 1000
  space 790
    transition 1001
secondary Web data 786
secure multiparty computation 1005
security
  definitions 1005
  management 18
segmentation 353, 444, 1217
selectivity 413, 991
self
organizing map (SOM) 273
-insured 265
-organizing maps 294
semantic
context 1012
data mining 1010
dimensions 740
metadata 1011
Web 1012, 1214
-based retrieval 805
semantics 384
of attributes 1
semi-
structured
data 372
documents 678
supervised
clustering 1114
learning 12, 1022
semiotics 129
semistructured pages 445
sensor network 965
sentence completion 770
separating set 694
sequence
analysis 383
rules 1231
pattern 1028
discovery 1245
sequential random sampling 345
server log 438
session 1221
sessionization 1227
set theoretic interval 660
shared
dimensions 199
knowledge 130
signature 251
significant genes 551
similarity 1120
detection 383
simple random 345
sampling 344
simplex plot 1168
single-
graph 1141
-link lookahead search 938
singular value decomposition 307
site structure 714
skeletonization 872
SMC 405
SMO 1075
smoothening 872, 887
social
data mining system 46
network 1152
SODAS software 1089
soft computing 272
software
analysis 1033
classification/clustering 1034
engineering 112
filtering 1034
integration 1034
mining 1034
transformation 1034
warehouse 1033
SOK 334
sound 84
analysis 84
space
models 165
standardization 871
spam filtering 768
spanning tree 151
sparse data cubes 423
spherical clusters 582
split selection algorithm 142
splitting predicates 141
spurious rules 797
SQL/MM DM 39, 1171
SRM 1093, 1097
SSP 334
STAGGER 203
-concepts 204
standalone information system 640
standardization 1171
star
schema 877
-snowflake 526
state of the variable 674
static Features 887
statistical
analysis 737
data editing 1043
hypothesis testing 464
information systems (SIS) 1048
significance 848
statistics 74, 634, 1109
utilization 705
step-wise regression 80
STICH (Space Models Identification Through Hierarc
165
stock technical analysis 779
strategic HR decisions 262
Index

strategies 636
stratified
  random sampling 344
  sampling 992
stream data mining 1030
streaming 991
  ensemble algorithm 203
strong or unique identifiers 303
structural
  health monitoring system 245
  risk 1064
structure-activity relationship (SAR) 664
structured
  data 443
  documents 1015
  information retrieval 960
STUCCO 796
subdue system 540
subgraph mining 1059
subjective probability 90
subschema similarity 648
substructure 541
summarization 251, 272
supervised
  data mining 811
  discretization 398
  inductive learning 477
  learning 54, 144, 494, 542, 865, 1064, 1071, 1105
  method 303
supply chain 323
  management 325, 326
support 1231, 1098
  threshold 60
  vector machine 117, 664, 1064, 1071
  vectors 1067, 1074
suppress redundancies 76
surveillance 835
survival
  analysis (SA) 1077
  data mining (SDM) 1077
SVM 1075
SW-Cube 1034
SWAMS (software warehouse application management) 1034
symbiotic
  data mining 1083
  data analysis 1087
symbolic
  descriptions 1087
  objects 1087
synonymy 648
syntax 384
system LERS 974
systems analysis 234

T
‘t’ value 80
tacit 600
taxonomies 392
taxonomy 615, 1113
  tree 1212
TDT (topic detection and tracking) 608
technology trends 8
temporal
  characteristics 870
  data mining 662
  recommenders 47
tenfold cross-validation 1072
termination reports 262
terrorist information awareness 268
test set 1071
text
  classification 13, 768
  mining 278, 382, 606, 1130, 1206
    -machine learning 1109
    -mining 1113
textual messages 758
thematic search engine 1201
thermometer Code 980
thinning 872
threats 566
three-dimensional visualization 1193
time series 172
time series 991, 1125
top-down computation 197
topic
  drifting 444
  hierarchy 1113
  map authoring 1130
  maps 1130
  ontology 608
total information awareness 268
training set 1071
transaction
  record 653
    -based
      data 325
      sequences and event sequences 1098
transactional
  data 323
  database 325
  supply chain data 323
transactions 947
transferable belief model 1135
transformations 1051
tree
  algorithms 190
  isomorphism 1143
Index

mining 1140
pattern matching 1143
trend 1120
trialability 1162
Trie, Patricia 60
tumor classification 550
two-dimensional scatter plot 154
type
  a error 695
  b error 695
  conflict 648
  I error 888
  II error 888

U

U.S.
  Army 94
  Coast Guard 94, 268
  Navy 94
uncertainty 834, 960, 1135
ungeneralizable patterns 263
unified theory of data 226
uniform sampling 413
units 54
univariate sampling 346
universal itemset 59
unlabeled data 1022
unstructured data 384
unsupervised
  data mining 811
  discretization 398
  learning 555, 865, 1065
  mining 1155
upper bound 659
usage
  information 102
  pattern 1151
user
  information 101
  modeling 1218
  preferences 440
  profile(s) 23, 440, 1236
uses for mining medical digital libraries 280
utility(ies) 630, 930

V

validation 190
value
  density 816
  distance 816
variable
  precision rough set model 974
  selection 735
  VC-Dimension 1066
  vector
    space model 1104
      -characterizing function 519
  version
    space 208
  vertical data mining 1181
  vertically partitioned data 927
Veteran administrations 268
vibration data 247
video
  association mining 1187
  classification 1186
  clustering 1186
  data mining 1185
  databases 966
view
  dependency graph 711
  maintenance 718
  monotonicity of a query 198
  selection 717
  usability problem 30
views 12
virtual
  environment (VE) 1193
  private network 1161
  reality 1190, 1193
  reference 102
vision-based page 444
visual data mining 1190
visualization 995, 1190

W

wavelet transformation 1196
wavelets 991, 1196
weak identifiers 303
Web
  analytics 1208, 1217, 1229
  clickstream analysis 1231
  communities 1207
  content 1146
    mining 800, 950, 1103, 1206
  databases 443
  graph mining 443
  IE 443
  information
    extraction 443
    searching 1000
  intelligence 995
  linking 996
  log 1151, 1217
  file 1243
  mining 950, 1103, 1146, 1151, 1201, 1206
  techniques 438, 1211
Index

pages representation 1104
personalization 1218, 1223, 1248
processes 896
robot 1228
search 438
server logs 1221
structure 1146
    mining 800, 1103, 1206
usage mining 438, 800, 950, 1103, 1151, 1206, 1216, 1221, 1226, 1236, 1242
-enabled e-business 785
weighted-majority 203
well-formed structures 525
wild point correction 887
window 1099
    adjustment heuristic 203
wireless
    ad-hoc sensor grids 427
    mobile networks 955
    telephony 912
word cluster map (WCM) 1131
workflow management systems 896
workflows 896
workload 992
World Wide Web 59, 1211
data mining 1242
wrapper induction 13, 445, 678
wrappers 678

X

XML 372, 787, 800, 965, 1015
    documents 372
    mining 372
    sources 648
    tags 384
XyDiff 801

Y

y 440

Z

z-score models 503