

Předmluva.....	11
I. Dobývání znalostí z databází	
1 Dobývání znalostí z databází.....	15
1.1 Úlohy.....	18
1.2 Metodiky.....	22
1.2.1 Metodika 5A.....	22
1.2.2 Metodika SEMMA.....	23
1.2.3 Metodika CRISP-DM.....	24
Literatura.....	28
II. Tři zdroje	
2 Databáze.....	33
2.1 Relační databáze.....	33
2.2 EIS.....	35
2.3 OLAP.....	35
2.4 Datové sklady a datová tržiště.....	41
2.5 Dotazovací jazyky pro dobývání znalostí z databází.....	43
Literatura.....	45
3 Statistika.....	46
3.1 Kontingenční tabulky.....	46
3.2 Regresní analýza.....	49
3.3 Diskriminační analýza.....	53
3.4 Shluková analýza.....	55
Literatura.....	59
4 Strojové učení.....	60
4.1 Základní pojmy.....	60
4.2 Učení jako prohledávání.....	69
4.3 Učení jako aproximace funkcí.....	78
Literatura.....	81
III. Proces dobývání znalostí	
5 Modelování.....	85
5.1 Rozhodovací stromy.....	86
5.1.1 Základní algoritmus.....	86
5.1.2 Převod stromu na pravidla.....	93
5.1.3 Prořezávání stromů.....	94
5.1.4 Numerické atributy.....	95
5.1.5 Chybějící hodnoty.....	98
5.1.6 Ceny atributů.....	98
5.1.7 Regresní stromy.....	99
5.1.8 Systémy.....	100
5.1.9 Použití rozhodovacích stromů.....	101
5.2 Asociační pravidla.....	102

5.2.1	Základní charakteristiky pravidel	103
5.2.2	Generování kombinací	106
5.2.3	Počet kombinací	107
5.2.4	Algoritmus apriori	109
5.2.5	Zobecněná asociační pravidla	111
5.2.6	Pravidla s výjimkami	113
5.2.7	Časové sekvence	114
5.2.8	Více tabulek	116
5.2.9	Implikace, dvojité implikace a ekvivalence	118
5.2.10	Metoda GUHA	121
5.2.11	Kombinační analýza dat	125
5.2.12	Chybějící hodnoty	128
5.3	Rozhodovací pravidla	130
5.3.1	Pokrývání množin	130
5.3.2	Rozhodovací seznam	134
5.3.3	Pravděpodobnostní pravidla	138
5.3.4	Algoritmus ESOD	139
5.3.5	Chybějící hodnoty	147
5.3.6	Numerické atributy	147
5.3.7	Numerické třídy	149
5.3.8	Koncepty proměnlivé v čase	150
5.3.9	Integrace znalostí	153
5.3.10	Hierarchie hodnot atributů	155
5.4	Neuronové sítě	157
5.4.1	Model jednoho neuronu	157
5.4.2	Perceptron	163
5.4.3	Topologie soudobých sítí	166
5.4.4	Metoda SVM	171
5.4.5	Neuronové sítě a dobývání znalostí z databází	173
5.5	Evoluční algoritmy	176
5.5.1	Základní podoba genetických algoritmů	177
5.5.2	Použití genetických algoritmů	180
5.5.3	Genetické programování	181
5.6	Bayesovská klasifikace	182
5.6.1	Základní pojmy	182
5.6.2	Naivní bayesovský klasifikátor	185
5.6.3	Bayesovské sítě	187
5.6.4	Systémy a aplikace	196
5.7	Metody založené na analogii	197
5.7.1	Podobnost mezi příklady	198
5.7.2	Podobnost mezi časovými řadami a sekvencemi	201
5.7.3	Učení založené na instancích	203
5.7.4	Nejbližší soused	205
5.7.5	Případové usuzování	209
5.7.6	Systémy IBL	210
5.8	Induktivní logické programování	211
5.8.1	Základní pojmy	212
5.8.2	Systémy ILP	214
	Literatura	217
6	Vyhodnocení výsledků	223
6.1	Testování modelů	224
6.1.1	Celková správnost	227
6.1.2	Správnost pro jednotlivé třídy	227

6.1.3 Přesnost a úplnost.....	228
6.1.4 Senzitivita a specifická.....	228
6.1.5 Spolehlivost klasifikace.....	229
6.1.6 Křivka učení.....	230
6.1.7 Křivka navýšení.....	232
6.1.8 Křivka ROC.....	233
6.1.9 Analýza DEA.....	235
6.1.10 Numerické predikce.....	235
6.2 Vizualizace.....	236
6.2.1 Vizualizace modelů.....	236
6.2.2 Vizualizace klasifikací.....	238
6.3 Porovnávání modelů.....	239
6.3.1 t-test.....	239
6.3.2 Použití křivek ROC.....	240
6.3.3 Occamova břitva.....	241
6.4 Volba nejvhodnějšího algoritmu.....	241
6.4.1 STATLOG.....	242
6.4.2 METAL.....	243
6.5 Kombinování modelů.....	243
Literatura.....	245
7 Příprava dat.....	247
7.1 Strukturovaná data.....	247
7.2 Více vzájemně propojených tabulek.....	250
7.3 Odvozené atributy.....	251
7.4 Data s příliš mnoha objekty.....	252
7.5 Data s příliš mnoha atributy.....	253
7.6 Numerické atributy.....	257
7.7 Kategoriální atributy.....	266
7.8 Chybějící hodnoty.....	267
7.9 Závěr.....	267
Literatura.....	267
IV. Systémy a úlohy	
8 Systémy pro dobývání znalostí z databází.....	271
8.1 Clementine.....	272
8.2 Enterprise Miner.....	275
8.3 Intelligent Miner.....	277
8.4 Systém Kepler.....	278
8.5 KnowledgeSTUDIO.....	280
8.6 LISp-Miner.....	281
8.7 MineSet.....	284
8.8 Statistica Data Miner.....	286
8.9 Weka.....	287
8.10 Který systém zvolit?.....	289
Literatura.....	290
9 Dobývání znalostí v praxi.....	291
9.1 Příklad úlohy.....	291
9.1.1 Porozumění problematice.....	291
9.1.2 Porozumění datům.....	291
9.1.3 Příprava dat.....	295
9.1.4 Modelování.....	296
9.1.5 Vyhodnocení výsledků.....	300

9.1.6 Využití výsledků	301
9.2 Obecné zkušenosti	302
Literatura	303
10 Nové směry	304
10.1 Dobývání znalostí z textů	304
10.1.1 Reprezentace dokumentu	304
10.1.2 Podobnost dokumentů	306
10.1.3 Typy úloh	307
10.1.4 Systémy	312
10.2 Dobývání znalostí z webu	312
10.2.1 Obsah webu	313
10.2.2 Struktura webu	319
10.2.3 Používání webu	320
10.3 Co bude dál ?	322
Literatura	322
Příloha	
A Stručný popis PMML	327
A.1 Pravidla pro zápis syntaxe	327
A.2 Struktura dokumentu PMML	329
A.2.1 Element Header	329
A.2.2 Element DataDictionary	329
A.2.3 Element TransformationDictionary	330
A.2.4 Element pro popis modelu	330
A.3 Příklady dokumentů PMML	331
A.3.1 Rozhodovací strom	331
A.3.2 Asociační pravidla	332
A.3.3 Neuronové sítě	334
A.3.4 Naivní bayesovský klasifikátor	337
A.3.5 Model k-NN	340
A.4 Úplný popis DTD	341
B Obsah CD	353
B.1 Systémy dobývání znalostí	353
B.1.1 Systémy na CD	353
B.1.2 Informace o dalších nekomerčních systémech	354
B.1.3 Informace o komerčních systémech	355
B.2 Data a úlohy	356
B.2.1 PKDD Discovery Challenge – finanční data	357
B.2.2 Soutěžní úlohy dobývání znalostí	357
B.2.3 Referenční data	358
B.3 Výzkumné projekty EU	358
B.3.1 Networks of Excellence	358
B.3.2 Konkrétní výzkumné projekty	359
B.4 Zdroje z Internetu	359
B.4.1 Dokumenty na CD	359
B.4.2 Informační portály	360
B.4.3 Katalogy ve vyhledávačích	360
B.4.4 Časopisy	361
B.4.5 Různé	361
Rejstřík	363