

# Obsah

ABSTRACT .....	7
PŘEDMLUVA.....	9
<b>1 DATA MINING A STATISTIKA .....</b>	<b>11</b>
1.1 TECHNOLOGICKÝ ROZVOJ A DATA .....	11
1.2 DATA MINING A JEHO VÝVOJ .....	12
1.3 DATA MINING A TYPY ÚLOH .....	16
1.4 VÝVOJ PŘÍSTUPŮ KE STATISTICKÉ ANALÝZE DAT .....	17
1.5 ANALÝZA DAT A VLIV TECHNOLOGIÍ NA JEJÍ VÝVOJ .....	20
1.6 USPOŘÁDÁNÍ DAT A DATA MINING .....	23
1.7 VÝVOJ TYPŮ STATISTICKÉHO SOFTWARE .....	25
1.8 ANALYTICKÉ METODY DATA MINING .....	27
<b>2 VYHLEDÁVÁNÍ ZNALOSTÍ Z DAT A DATA MINING .....</b>	<b>31</b>
2.1 PROCES VYHLEDÁVÁNÍ ZNALOSTÍ Z DAT .....	32
2.2 PROBLÉMY S AUTOMATIZOVÁNÍM PROCESU KDD .....	34
2.3 MOŽNÉ ZDROJE CHYB V DATOVÝCH SOUBORECH .....	37
2.4 NÁROKY NA UŽIVATELE .....	39
2.5 MODELOVÁNÍ Z DAT .....	39
<b>3 DISKRIMINACE A KLASIFIKACE .....</b>	<b>43</b>
3.1 VYSVĚTLENÍ ÚLOHY KLASIFIKACE .....	43
3.2 DISKRIMINACE A PŘÍKLADY POUŽITÍ .....	45
3.3 POPIS ÚLOHY DISKRIMINACE .....	46
3.4 VYUŽITELNOST DAT PRO DISKRIMINAČNÍ ÚLOHY .....	48
3.5 BINÁRNÍ DISKRIMINACE A KLASIFIKACE .....	50
3.5.1 Fisherova lineární diskriminační funkce (FLDF) .....	51
3.5.2 Klasifikace při neshodné ceně chyb .....	57
3.5.3 Klasifikace podle Bayesova vzorce .....	58
3.5.4 Gaussova LDF – shodné kovarianční matice .....	59
3.5.5 Lineární diskriminační analýza a regresní model .....	63
3.5.6 Gaussova LDF – neshodné kovarianční matice .....	64
3.6 DISKRIMINACE MEZI DVĚMA NEBO VÍCE SKUPINAMI .....	65
3.6.1 Fisherova kanonická diskriminační analýza .....	66
3.6.2 Fisherova výběrová diskriminační funkce .....	69
3.7 KLASIFIKACE PŘI MINIMALIZACI CENY CHYB (ECM) PRO $G > 2$ .....	72
3.7.1 Klasifikace ECM pro vícerozměrné normální rozdělení .....	73

<b>4</b>	<b>LOGISTICKÁ REGRESE BINÁRNÍ.....</b>	<b>76</b>
4.1	REGRESNÍ MODEL A BINÁRNÍ VYSVĚTLOVANÁ PROMĚNNÁ .....	76
4.2	BINÁRNÍ LOGISTICKÁ REGRESE S JEDNOU VYSVĚTLUJÍCÍ PROMĚNNOU .....	77
4.3	LOGISTICKÁ REGRESE S VEKTOREM NEZÁVISLE PROMĚNNÝCH .....	80
4.3.1	<i>Logistická regrese s binomickou vysvětlující proměnnou</i> .....	81
4.4	INFERENCE V LOGISTICKÉM MODELU A REGRESNÍ DIAGNOSTIKA .....	82
4.4.1	<i>Doporučená kritéria</i> .....	86
<b>5</b>	<b>KLASIFIKAČNÍ METODY POČÍTAČOVÉHO UČENÍ .....</b>	<b>87</b>
5.1	NAIVNÍ BAYES.....	87
5.2	ROZHODOVACÍ STROMY .....	89
5.2.1	<i>Odhady chyb a rozsah stromu</i> .....	97
5.2.2	<i>Odhad přesnosti klasifikačního pravidla</i> .....	97
5.2.3	<i>Speciální problém klasifikace dvou tříd</i> .....	102
5.2.4	<i>Vytvoření exploračního stromu</i> .....	106
5.2.5	<i>Interpretace klasifikačního stromu</i> .....	106
5.2.6	<i>Modifikace algoritmů založených na rozhodovacích stromech</i> .....	108
5.3	MOŽNOSTI POROVNÁNÍ CHYB KLASIFIKACE .....	110
<b>6</b>	<b>PREDIKTIVNÍ KVALITA BINÁRNÍHO KLASIFIKAČNÍHO MODELU .....</b>	<b>111</b>
6.1	CHYBY KLASIFIKAČNÍHO MODELU .....	111
6.2	PREDIKTIVNÍ VLASTNOSTI MODELU A KLASIFIKAČNÍ MATICE .....	113
6.3	ODHAD PREDIKTIVNÍ CHYBY MODELU .....	116
6.4	ROC KŘIVKA .....	117
6.4.1	<i>Empirická ROC křivka</i> .....	119
6.4.2	<i>Vlastnosti ROC</i> .....	122
6.4.3	<i>Vyhrazení empirické ROC křivky</i> .....	122
6.4.4	<i>Odhad <math>b_1</math> – normální ROC</i> .....	123
6.4.5	<i>Jádrový odhad ROC křivky</i> .....	124
6.5	ODHADY AUC .....	126
6.6	DUÁLNÍ REPREZENTACE ROC KŘIVKY .....	130
6.7	KŘIVKY PR A DET .....	132
6.8	KONVEXNÍ OBAL ROCCH NAD PROSTOREM ROC .....	133
6.9	PŘEHLED CHARAKTERISTIK KVALITY KLASIFIKAČNÍHO MODELU .....	135
<b>7</b>	<b>ZÁVĚR .....</b>	<b>137</b>
<b>8</b>	<b>GLOSÁŘ ANGLICKÝCH POJMŮ A ZKRATEK .....</b>	<b>139</b>
	<b>LITERATURA .....</b>	<b>147</b>