

# Obsah

<b>Předmluva</b>	<b>13</b>
<b>O autorovi</b>	<b>15</b>
<b>Poděkování</b>	<b>16</b>
<b>O odborných korektorech</b>	<b>17</b>
<b>Úvod</b>	<b>19</b>
<b>Co kniha popisuje</b>	<b>19</b>
<b>Co budete potřebovat</b>	<b>20</b>
<b>Komu je kniha určena</b>	<b>20</b>
<b>Styly</b>	<b>21</b>
<b>Zpětná vazba od čtenářů</b>	<b>22</b>
<b>Errata</b>	<b>22</b>
KAPITOLA 1	
<b>Začínáme</b>	
<b>Informatika</b>	<b>23</b>
<b>Umělá inteligence</b>	<b>24</b>
<b>Strojové učení</b>	<b>24</b>
<b>Statistika</b>	<b>24</b>
<b>Matematika</b>	<b>25</b>
<b>Znalostní domény</b>	<b>25</b>
<b>Data, informace a znalosti</b>	<b>25</b>
<b>Povaha dat</b>	<b>26</b>

<b>Proces datové analýzy</b>	<b>27</b>
Problém	27
Příprava dat	28
Průzkum dat	28
Prediktivní modelování	28
Vizualizace výsledků	29
<b>Kvantitativní a kvalitativní datová analýza</b>	<b>29</b>
<b>Význam vizualizace dat</b>	<b>30</b>
<b>A co big data?</b>	<b>31</b>
Senzory a fotoaparáty	32
Analýza sociálních sítí	33
Nástroje a hračky v této knize	34
Proč Python?	34
Proč mlpy?	34
Proč D3.js?	35
Proč MongoDB?	35
<b>Shrnutí</b>	<b>36</b>

## KAPITOLA 2

<b>Práce s daty</b>	<b>37</b>
<b>Datové zdroje</b>	<b>38</b>
Otevřená data	39
Textové soubory	39
Soubory aplikace Excel	39
SQL databáze	40
NoSQL databáze	41
Multimédia	42
Data z webu	42
<b>Čištění dat</b>	<b>45</b>
Statistické metody	45
Rozložení textu	46
Převádění dat	47
<b>Datové formáty</b>	<b>47</b>
CSV	48
JSON	49
XML	50
YAML	51

<b>Začínáme s OpenRefine</b>	<b>52</b>
Textová fazeta	52
Clustering	53
Textové filtry	54
Číselné fazety	54
Převod dat	55
Export dat	56
Historie akcí	57
<b>Shrnutí</b>	<b>58</b>

## KAPITOLA 3

<b>Vizualizace dat</b>	<b>59</b>
<b>Dokumenty založené na datech (D3)</b>	<b>60</b>
HTML	60
DOM	61
CSS	61
JavaScript	61
SVG	61
<b>Začínáme s D3.js</b>	<b>62</b>
Sloupcový graf	62
Koláčový graf	67
Bodové vykreslení	70
Spojnicový graf	72
Spojnicový graf s více křivkami	76
<b>Interakce a animace</b>	<b>79</b>
<b>Shrnutí</b>	<b>81</b>

## KAPITOLA 4

<b>Třídění textu</b>	<b>83</b>
<b>Učení a třídění</b>	<b>83</b>
<b>Bayesiánské třídění</b>	<b>84</b>
Naivní Bayesův algoritmus	85
<b>Ověřování předmětu e-mailu</b>	<b>85</b>
Data	86

Algoritmus	88
Přesnost třídění	91
<b>Shrnutí</b>	<b>94</b>
KAPITOLA 5	
<b>Rozpoznávání podobných obrázků</b>	<b>95</b>
<b>Vyhledávání obrázků na základě podobnosti</b>	<b>95</b>
<b>Dynamické borcení času (DTW)</b>	<b>96</b>
<b>Zpracování obrázků v datové sadě</b>	<b>98</b>
<b>Implementace dynamického borcení času</b>	<b>99</b>
<b>Analýza výsledků</b>	<b>101</b>
<b>Shrnutí</b>	<b>103</b>
KAPITOLA 6	
<b>Simulace cen akcií</b>	<b>105</b>
<b>Finanční časové řady</b>	<b>105</b>
<b>Simulace nahodilé chůze</b>	<b>106</b>
<b>Postupy Monte Carlo</b>	<b>107</b>
<b>Generování nahodilých čísel</b>	<b>108</b>
<b>Implementace v D3.js</b>	<b>109</b>
<b>Shrnutí</b>	<b>116</b>
KAPITOLA 7	
<b>Předvídání cen zlata</b>	<b>117</b>
<b>Práce s daty v časových řadách</b>	<b>117</b>
Komponenty časových řad	119
<b>Čištění časových řad</b>	<b>120</b>
<b>Data – historické ceny zlata</b>	<b>123</b>
<b>Nelineární regrese</b>	<b>123</b>
Hřebenová regrese jádra	123
Čištění časových řad cen zlata	125

Předvídání z vyčištěných časových řad	126
Srovnávání předpovězených hodnot	128
<b>Shrnutí</b>	<b>129</b>

## KAPITOLA 8

<b>Práce s podpůrnými vektorovými stroji</b>	<b>131</b>
<b>Úvod do datových sad s více proměnnými</b>	<b>132</b>
<b>Redukce dimenzionality</b>	<b>135</b>
Lineární diskriminanční analýza	136
Analýza hlavních komponent	136
<b>Začínáme pracovat s podpůrnými vektorovými stroji</b>	<b>139</b>
Funkce jádra	140
Problém dvojité spirály	140
Podpůrné vektorové stroje a knihovna mlpy	141
<b>Shrnutí</b>	<b>144</b>

## KAPITOLA 9

<b>Modelování infekčních nemocí prostřednictvím buněčných automatů</b>	<b>147</b>
<b>Úvod do epidemiologie</b>	<b>148</b>
Epidemiologický trojúhelník	149
<b>Epidemiologické modely</b>	<b>150</b>
Model SIR	150
Řešení běžných diferenciálních rovnic modlu SIR pomocí SciPy	151
Model SIRS	152
<b>Modelování pomocí buněčných automatů</b>	<b>153</b>
Buňka, stav, mřížka a sousedství	153
Globální stochastický kontaktní model	154
<b>Simulace modelu SIRS v CA prostřednictvím rozhraní D3.js</b>	<b>155</b>
<b>Shrnutí</b>	<b>163</b>

## KAPITOLA 10

<b>Práce se sociálními grafy</b>	<b>165</b>
<b>Struktura grafu</b>	<b>165</b>

Neorientovaný graf	166
Orientovaný graf	166
<b>Analýza sociálních sítí</b>	<b>167</b>
<b>Tvorba grafu z Facebooku</b>	<b>167</b>
Netvizz	167
<b>Reprezentace grafů v Gephi</b>	<b>170</b>
<b>Statistická analýza</b>	<b>172</b>
Poměr mužů a žen	172
<b>Rozložení směrů</b>	<b>174</b>
Histogram grafu	175
Středovost	176
<b>Převod GDF na JSON</b>	<b>177</b>
<b>Vizualizace grafu v D3.js</b>	<b>179</b>
<b>Shrnutí</b>	<b>183</b>

## KAPITOLA 11

<b>Analýza nálad na Twitteru</b>	<b>185</b>
<b>Anatomie dat z Twitteru</b>	<b>186</b>
Tweet	186
Sledující	186
Témata trendů	187
<b>Přístup k API Twitteru pomocí OAuth</b>	<b>187</b>
<b>Úvod do modulu Twython</b>	<b>189</b>
Jednoduché vyhledávání	190
Práce s časovými osami	193
Práce se sledujícími	195
Práce s místy a trendy	197
<b>Třídění nálad</b>	<b>198</b>
Pravidla pro anglická slova	199
Textový korpus	199
<b>Úvod do NLTK</b>	<b>200</b>
Dávka slov	201
Naivní Bayesův klasifikátor	201
Analýza nálad tweetů	203

---

<b>Shrnutí</b>	<b>204</b>
KAPITOLA 12	
<b>Zpracování a agregace dat v MongoDB</b>	<b>205</b>
<b>Úvod do databáze MongoDB</b>	<b>206</b>
Databáze	206
Kolekce	207
Dokument	208
Příkazový řádek Mongo	208
Vkládání, aktualizace a odstraňování	209
Dotazy	209
<b>Příprava dat</b>	<b>211</b>
Převod dat pomocí OpenRefine	211
Vkládání dokumentů pomocí PyMongo	214
<b>Skupina</b>	<b>217</b>
<b>Agregační rozhraní</b>	<b>219</b>
Kanály	220
Výrazy	221
<b>Shrnutí</b>	<b>222</b>
KAPITOLA 13	
<b>Práce s modelem MapReduce</b>	<b>223</b>
<b>Úvod do modelu MapReduce</b>	<b>224</b>
<b>Programovací model</b>	<b>225</b>
<b>Model MapReduce a databáze MongoDB</b>	<b>226</b>
Funkce map	226
Funkce reduce	227
Příkazový řádek Mongo	227
Práce s nástrojem UMongo	229
Práce s nástrojem PyMongo	232
<b>Filtrování vstupní kolekce</b>	<b>233</b>
<b>Seskupování a agregace</b>	<b>234</b>
<b>Vizualizace nejběžnějších slov ve tweetech v mraku slov</b>	<b>236</b>
<b>Shrnutí</b>	<b>241</b>

## KAPITOLA 14

<b>Online analýza dat s nástroji Ipython a Wakari</b>	<b>243</b>
<b>Úvod do nástroje Wakari</b>	<b>244</b>
Založení účtu ve Wakari	244
<b>Úvod do nástroje IPython Notebook</b>	<b>246</b>
Vizualizace dat	248
<b>Úvod do zpracování obrázků knihovnou PIL</b>	<b>250</b>
Otevření obrázku	250
Histogram obrázku	251
Filtrování	252
Operace	254
Převody	256
<b>Úvod do knihovny Pandas</b>	<b>257</b>
Práce s časovými řadami	257
Práce s datovými sadami s více proměnnými a s objektem DataFrame	261
Seskupování, agregace a korelace	264
<b>Multiprocesing s nástrojem IPython</b>	<b>266</b>
Pool	267
<b>Sdílení poznámkového bloku</b>	<b>267</b>
Data	267
<b>Shrnutí</b>	<b>270</b>

## PŘÍLOHA

<b>Zavádění infrastruktury</b>	<b>271</b>
<b>Instalace a spuštění prostředí Python 3</b>	<b>272</b>
Instalace a spuštění prostředí Python v Ubuntu	272
Instalace a spuštění prostředí IDLE v Ubuntu	272
Instalace a spuštění prostředí Python 3.2 ve Windows	273
Instalace a spuštění prostředí IDLE ve Windows	274
<b>Instalace a spuštění knihovny NumPy</b>	<b>275</b>
Instalace a spuštění knihovny NumPy v Ubuntu	275
Instalace a spuštění knihovny NumPy ve Windows	276
<b>Instalace a spuštění knihovny SciPy</b>	<b>277</b>
Instalace a spuštění knihovny SciPy v Ubuntu	277
Instalace a spuštění knihovny SciPy ve Windows	278



---

<b>Instalace a spuštění knihovny mlpy</b>	<b>279</b>
Instalace a spuštění knihovny mlpy v Ubuntu	279
Instalace a spuštění knihovny mlpy ve Windows	279
<b>Instalace a spuštění nástroje OpenRefine</b>	<b>280</b>
Instalace a spuštění nástroje OpenRefine v Linuxu	280
Instalace a spuštění nástroje OpenRefine ve Windows	281
<b>Instalace a spuštění databáze MongoDB</b>	<b>281</b>
Instalace a spuštění databáze MongoDB v Ubuntu	282
Instalace a spuštění databáze MongoDB ve Windows	283
Připojení Pythonu k databázi MongoDB	285
<b>Instalace a spuštění nástroje UMongo</b>	<b>286</b>
Instalace a spuštění nástroje Umongo v Ubuntu	287
Instalace a spuštění Umongo ve Windows	288
<b>Instalace a spuštění Gephi</b>	<b>289</b>
Instalace a spuštění Gephi v Linuxu	290
Instalace a spuštění Gephi ve Windows	290
<b>Rejstřík</b>	<b>291</b>