

Sadržaj

Predgovor	xi
1. Pregled sistema mašinskog učenja.....	1
Kada koristiti mašinsko učenje	3
Primeri upotrebe mašinskog učenja	9
Razumevanje sistema mašinskog učenja	12
Mašinsko učenje u istraživanju u odnosu na primenu	12
Sistemi mašinskog učenja nasuprot tradicionalnog softvera	22
Rezime	24
2. Uvod u projektovanje sistema mašinskog učenja	25
Poslovni i ML ciljevi	26
Zahtevi za ML sisteme	28
Pouzdanost	29
Skalabilnost	29
Održivost	31
Prilagodljivost	31
Iterativan proces	32
Formulisanje problema mašinskog učenja	34
Vrste ML zadataka	35
Ciljne funkcije	39
Um protiv podataka	42
Rezime	45
3. Osnove inženjeringa podataka.....	47
Izvori podataka	48
Formati podataka	51
JSON	52
Format po redovima nasuprot formata po kolonama	52
Tekstualni nasuprot binarnom formatu	55
Modeli podataka	56
Relacioni model	57
NoSQL	61
Strukturisani nasuprot nestrukturisanih podataka	64

Mašine za skladištenje podataka i obradu	65
Transakciona i analitička obrada	66
ETL: ekstrakcija, transformacija, učitavanje	68
Režimi protoka podataka	70
Prenos podataka kroz baze podataka	70
Prenos podataka kroz servise	71
Prenos podataka transportom u realnom vremenu	72
Paketna obrada nasuprot obrade u toku	76
Rezime	78
4. Podaci za obuku	81
Uzorkovanje	82
Uzorkovanje bez verovatnoće	83
Jednostavno nasumično uzorkovanje (Simple Random Sampling)	84
Slojevito uzorkovanje (Stratified Sampling)	84
Težinsko uzorkovanje (Weighted Sampling)	85
Uzorkovanje rezervoara (Reservoir Sampling)	86
Uzorkovanje po važnosti (Importance Sampling)	87
Označavanje (Labeling)	88
Ručne oznake	88
Prirodne oznake	91
Rešavanje nedostatka oznaka	94
Neravnoteža klasa	103
Izazovi neravnoteže klasa	103
Rešavanje problema neravnoteže klasa	105
Povećanje skupa podataka	114
Jednostavne transformacije koje čuvaju oznake	115
Perturbacija	115
Sinteza podataka	117
Rezime	119
5. Inženjering karakteristika	121
Naučene karakteristike nasuprot projektovanih karakteristika	121
Uobičajene operacije inženjeringa karakteristika	125
Rukovanje sa nedostajućim vrednostima	125
Normalizacija	128
Diskretizacija	130
Kodiranje kategorijskih karakteristika	131
Ukrštanje karakteristika	134
Diskretno i kontinuirano poziciono umetanje	135
Curenje podataka	137
Uobičajeni uzroci curenja podataka	139
Detektovanje curenja podataka	142

Inženjering dobrih karakteristika	142
Važnost karakteristika	143
Generalizacija karakteristika	146
Rezime	147
6. Razvoj modela i oflajn evaluacija	149
Razvoj i obučavanje modela	150
Evaluacija ML modela	150
Ansambli	156
Praćenje eksperimenata i verzionisanje	162
Distribuirana obuka	168
AutoML	172
Oflajn evaluacija modela	178
Osnove	179
Metode evaluacije	181
Rezime	189
7. Implementacija modela i servis za predikciju	191
Mitovi o implementaciji mašinskog učenja	193
Mit 1: Implementirate samo jedan ili dva ML modela odjednom	194
Mit 2: Ako ništa ne radimo, performanse modela ostaju iste	195
Mit 3: Nećete morati često ažurirati vaše modele	196
Mit 4: Većina ML inženjera ne mora brinuti o skaliranju	196
Paketno predviđanje nasuprot onlajn predviđanja	197
Od predviđanja u grupama ka onlajn predviđanju	201
Ujedinjavanje paketnog i strimovanog cevovoda	203
Kompresija modela	205
Faktorizacija niske rangiranosti	206
Destilacija znanja	207
Smanjenje dimenzija – orezivanje	207
Kvantizacija	208
ML na oblaku i na ograničenim resursima	211
Kompajliranje i optimizacija modela za uređaje kod korisnika	213
ML u veb čitačima	220
Rezime	222
8. Promena raspodele podataka i praćenje	225
Uzroci neuspeha ML sistema	226
Neuspesi softverskog sistema	227
Neuspesi specifični za ML	228
Promene u raspodeli podataka	236
Vrste promena u raspodeli podataka	237
Opšte promene raspodele podataka	241

Otkrivanje promena u raspodeli podataka	242
Adresiranje promene raspodele podataka	247
Praćenje i posmatranje	249
Metrike specifične za ML	251
Skup alata za praćenje	256
Posmatranje	259
Rezime	261
9. Neprekidno učenje i testiranje u proizvodnji	263
Neprekidno učenje	264
Obuka bez stanja nasuprot obuke sa stanjem	265
Zašto neprekidno učenje?	268
Izazovi neprekidnog učenja	270
Četiri faze neprekidnog učenja	275
Koliko često ažurirati modele	280
Testiranje u proizvodnji	282
Skrivena primena	283
A/B testiranje	283
Puštanje kanarinca	285
Prepletani eksperimenti	286
Banditi	288
Rezime	292
10. Infrastruktura i alati za MLOps	293
Skladištenje i računanje	297
Javni oblak nasuprot privatnih centara podataka	299
Okruženje za razvoj	302
Postavljanje okruženja za razvoj	303
Standardizacija razvojnih okruženja	306
Od razvoja do proizvodnje: kontejneri	308
Upravljanje resursima	311
Cron, planeri i orkestratori	311
Upravljanje radnim tokovima u nauci o podacima	314
ML Platforma	320
Primena modela	321
Skladište modela	321
Skladište karakteristika	326
Graditi ili kupiti	328
Rezime	330

11. Ljudski faktor u mašinskom učenju	333
Korisničko iskustvo	333
Obezbeđivanje doslednosti korisničkog iskustva	334
Borba protiv „uglavnom tačnih“ predviđanja	334
Neprimetno otkazivanje	336
Struktura tima	336
Saradnja funkcionalnih timova	337
Naučnik podataka vodi proces od početka do kraja	337
Odgovorna AI	341
Neodgovorna AI: studije slučajeva	342
Okvir za odgovornu AI	348
Rezime	355
Epilog	357
Indeks	359