

Εισαγωγή στην Python®

για τις Επιστήμες Υπολογιστών και Δεδομένων

Εκμάθηση Προγραμματισμού με AI, Μεγάλα Δεδομένα και το Νέφος

Paul Deitel

Deitel & Associates, Inc.

Harvey Deitel

Deitel & Associates, Inc.

Απόδοση: **Αριστόνους Μιχ. Τροχάνης**

Πολιτικός Μηχανικός ΕΜΠ

PhD Carnegie-Mellon U., Pgh, USA

 **Εκδόσεις: Μ. Γκιούρδας**

Ζωοδόχου Πηγής 70-74 - Τηλ.: 210 3630219

106 81 Αθήνα, 2021

www.mgiurdas.gr

Τίτλος Πρωτοτύπου:

Intro to Python® for Computer Science and Data Science

ISBN-13: 978-0-13-540467-6

ISBN-10: 0-13-540467-3

Copyright © 2019 Pearson Education, Inc.

221 River Street, Hoboken, NJ 07030

Αποκλειστικότητα για την Ελληνική Γλώσσα

Εκδόσεις: **Μόσχος Γκιούρδας**



Ζωοδόχου Πηγής 70-74 - Τηλ.: 210 3630219

106 81 Αθήνα, 2021

www.mgiurdas.gr

ISBN: 978-960-512-744-2

Desktop Publishing: *male.gr* τηλ.: 210 7481081

Εκτύπωση & Βιβλιοδεσία: Ves Company, τηλ.: 210 5711700

Αναδημοσίευση του βιβλίου σε οποιαδήποτε μορφή, ολόκληρου ή μέρους, καθώς και των περιεχομένων προγραμμάτων, δεν επιτρέπεται χωρίς την έγγραφη εξουσιοδότηση του εκδότη.

*Στη μνήμη του Marvin Minsky,
ενός εκ των ιδρυτών της
τεχνητής νοημοσύνης*

*Ήταν τιμή μου να είμαι φοιτητής σας σε δύο μεταπτυχιακά
μαθήματα τεχνητής νοημοσύνης στο Μ.Ι.Τ. Εμπνεύσατε
τους φοιτητές σας να σκέπτονται πέρα από τα όρια.*

Harvey Deitel



Περιεχόμενα

Πρόλογος	xix
Πριν ξεκινήσετε	xlv
1 Εισαγωγή σε Υπολογιστές και Python	1
1.1 Εισαγωγή	2
1.2 Υλισμικό και Λογισμικό	3
1.2.1 Νόμος του Moore	4
1.2.2 Οργάνωση Υπολογιστών	4
1.3 Ιεραρχία Δεδομένων	6
1.4 Γλώσσες Μηχανής, Συμβολικές Γλώσσες και Γλώσσες Υψηλού Επιπέδου	9
1.5 Εισαγωγή στην Αντικειμενοστραφή Τεχνολογία	10
1.6 Λειτουργικά Συστήματα	13
1.7 Python	16
1.8 Είναι οι Βιβλιοθήκες!	18
1.8.1 Πρότυπη Βιβλιοθήκη Python	18
1.8.2 Βιβλιοθήκες Επιστήμης Δεδομένων	18
1.9 Άλλες Δημοφιλείς Γλώσσες Προγραμματισμού	20
1.10 Δοκιμές: Χρήση IPython και Τετραδίων Jupyter	21
1.10.1 Χρήση της Διαδραστικής Λειτουργίας του IPython ως Αριθμομηχανής	21
1.10.2 Εκτέλεση ενός Προγράμματος Python με τον Διερμηνέα IPython	23
1.10.3 Συγγραφή και Εκτέλεση Κώδικα Code σε ένα Τετράδιο Jupyter	24
1.11 Ίντερνερ και Παγκόσμιος Ιστός	29
1.11.1 Ίντερνερ: Ένα Δίκτυο Δικτύων	29
1.11.2 Παγκόσμιος Ιστός: Κάνοντας το Ίντερνερ Εύχρηστο	30
1.11.3 Το Νέφος	30
1.11.4 Διαδίκτυο των Πραγμάτων	31
1.12 Τεχνολογίες Λογισμικού	32
1.13 Πόσο Μεγάλα Είναι τα Μεγάλα Δεδομένα;	33
1.13.1 Ανάλυση Μαζικών Δεδομένων	38
1.13.2 Επιστήμη Δεδομένων και Μαζικά Δεδομένα Κάνουν τη Διαφορά: Περιπτώσεις Χρήσης	39
1.14 Μελέτη Περίπτωσης — Μία Εφαρμογή Μαζικών Δεδομένων για Κινητά	40
1.15 Εισαγωγή στην Επιστήμη Δεδομένων: Τεχνητή Νοημοσύνη — Στην Τομή Πληροφορικής και Επιστήμης Δεδομένων	42
2 Εισαγωγή στον Προγραμματισμό με Python	49
2.1 Εισαγωγή	50
2.2 Μεταβλητές και Εντολές Εκχώρησης	50

2.3	Αριθμητική.....	52
2.4	Συνάρτηση <code>print</code> και Εισαγωγή σε Συμβολοσειρές σε Μονές και Διπλές Αποστρόφους.....	56
2.5	Συμβολοσειρές σε Τριπλές Αποστρόφους	58
2.6	Λήψη Δεδομένων από τον Χρήστη	59
2.7	Λήψη Αποφάσεων: Εντολή <code>if</code> και Τελεστές Σύγκρισης	61
2.8	Αντικείμενα και Δυναμική Πληκτρολόγηση.....	66
2.9	Εισαγωγή στην Επιστήμη Δεδομένων: Βασικά Περιγραφικά Στατιστικά Στοιχεία.....	68
2.10	Σύνοψη.....	70
3	Εντολές Ελέγχου και Ανάπτυξη Προγράμματος	73
3.1	Εισαγωγή.....	74
3.2	Αλγόριθμοι.....	74
3.3	Ψευδοκώδικας	75
3.4	Εντολές Ελέγχου	75
3.5	Εντολές <code>if</code>	78
3.6	Εντολές <code>if...else</code> και <code>if...elif...else</code>	80
3.7	Εντολή <code>while</code>	85
3.8	Εντολή <code>for</code>	86
3.8.1	Επαναλήψιμα, Λίστες και Επαναλήπτες.....	88
3.8.2	Ενσωματωμένη Συνάρτηση <code>range</code>	88
3.9	Επαυξημένες Εκχωρήσεις.....	89
3.10	Ανάπτυξη Προγράμματος: Επανάληψη Ελεγχόμενη Μέσω Ακολουθίας	90
3.10.1	Δήλωση Απαιτήσεων.....	90
3.10.2	Ψευδοκώδικας για τον Αλγόριθμο	90
3.10.3	Κωδικοποίηση του Αλγόριθμου σε Python.....	91
3.10.4	Εισαγωγή στις Μορφοποιημένες Συμβολοσειρές	92
3.11	Ανάπτυξη Προγράμματος: Επανάληψη Ελεγχόμενη Μέσω Τιμής Τερματισμού	93
3.12	Ανάπτυξη Προγράμματος: Εμφωλευμένες Εντολές Ελέγχου	97
3.13	Ενσωματωμένη Συνάρτηση <code>range</code> : Μία Βαθύτερη Ματιά.....	101
3.14	Χρήση του Τύπου <code>Decimal</code> για Νομισματικά Ποσά	102
3.15	Εντολές <code>break</code> και <code>continue</code>	105
3.16	Λογικοί Τελεστές <code>and</code> , <code>or</code> και <code>not</code>	106
3.17	Εισαγωγή στην Επιστήμη Δεδομένων: Μέτρα Κεντρικής Τάσης — Μέσος Όρος, Διάμεσος και Επικρατούσα Τιμή.....	109
3.18	Σύνοψη.....	111
4	Συναρτήσεις	119
4.1	Εισαγωγή.....	120
4.2	Ορισμός Συναρτήσεων.....	120
4.3	Συναρτήσεις με Πολλαπλές Παραμέτρους.....	123
4.4	Γέννηση Τυχαίων Αριθμών	125
4.5	Μελέτη Περίπτωσης: Ένα Παιχνίδι Τύχης.....	128
4.6	Πρότυπη Βιβλιοθήκη Python.....	131
4.7	Συναρτήσεις του Αρθρώματος <code>math</code>	132
4.8	Χρήση της Αυτόματης Συμπλήρωσης Tab του IPython για Ανακάλυψη.....	133

4.9	Προεπιλεγμένες Τιμές Παραμέτρων.....	135
4.10	Ορίσματα-Δεσμευμένες Λέξεις.....	136
4.11	Λίστες Αυθαίρετου Αριθμού Ορισμάτων.....	136
4.12	Μέθοδοι: Συναρτήσεις που Ανήκουν σε Αντικείμενα.....	138
4.13	Κανόνες Εμβέλειας.....	138
4.14	<code>import</code> : Μία Βαθύτερη Ματιά.....	140
4.15	Μεταβίβαση Ορισμάτων σε Συναρτήσεις: Μία Βαθύτερη Ματιά.....	142
4.16	Στοιβα Κλήσεων Συναρτήσεων.....	145
4.17	Προγραμματισμός Συναρτησιακού Ύφους.....	146
4.18	Εισαγωγή στην Επιστήμη Δεδομένων: Μέτρα Διασποράς.....	148
4.19	Σύνοψη.....	150
5	Ακολουθίες: Λίστες και Πλειάδες.....	155
5.1	Εισαγωγή.....	156
5.2	Λίστες.....	156
5.3	Πλειάδες.....	161
5.4	Αποσυσκευασία Ακολουθιών.....	163
5.5	Τεμαχισμός Ακολουθιών.....	166
5.6	Εντολή <code>del</code>	169
5.7	Μεταβίβαση Λιστών σε Συναρτήσεις.....	171
5.8	Ταξινόμηση Λιστών.....	172
5.9	Αναζήτηση σε Ακολουθίες.....	174
5.10	Άλλες Μέθοδοι Λιστών.....	176
5.11	Προσομοίωση Στοιβών με Λίστες.....	178
5.12	Κατανοήσεις Λίστας.....	179
5.13	Γεννήτριες Εκφράσεις.....	181
5.14	Διήθηση, Απεικόνιση και Αναγωγή.....	182
5.15	Άλλες Συναρτήσεις που Επεξεργάζονται Ακολουθίες.....	185
5.16	Δισδιάστατες Λίστες.....	187
5.17	Εισαγωγή στην Επιστήμη Δεδομένων: Προσομοίωση και Στατικές Απεικονίσεις.....	191
5.17.1	Δείγματα Γραφημάτων για 600, 60,000 και 6,000,000 Ρίψεις Ζαριού.....	191
5.17.2	Απεικόνιση Συχνοτήτων και Ποσοστών Εμφάνισης Πλευρών Ζαριού.....	193
5.18	Σύνοψη.....	199
6	Λεξικά και Σύνολα.....	209
6.1	Εισαγωγή.....	210
6.2	Λεξικά.....	210
6.2.1	Δημιουργία ενός λεξικού.....	210
6.2.2	Επαναλήψεις Διαμέσου Λεξικού.....	212
6.2.3	Βασικές Λειτουργίες Λεξικών.....	212
6.2.4	Μέθοδοι Λεξικών <code>keys</code> και <code>values</code>	214
6.2.5	Συγκρίσεις Λεξικών.....	216
6.2.6	Παράδειγμα: Λεξικό με Βαθμούς Φοιτητών.....	217
6.2.7	Παράδειγμα: Καταμέτρηση Λέξεων.....	218

6.2.8	Μέθοδος Λεξικών <code>update</code>	220
6.2.9	Κατανοήσεις Λεξικού.....	220
6.3	Σύνολα.....	221
6.3.1	Σύγκριση Συνόλων.....	223
6.3.2	Μαθηματικές Πράξεις Συνόλων.....	225
6.3.3	Τελεστές και Μέθοδοι Μεταβολής Συνόλων.....	226
6.3.4	Κατανοήσεις Συνόλου.....	228
6.4	Εισαγωγή στην Επιστήμη Δεδομένων: Δυναμικές Απεικονίσεις.....	228
6.4.1	Πώς Λειτουργεί η Δυναμική Απεικόνιση.....	228
6.4.2	Υλοποίηση μίας Δυναμικής Απεικόνισης.....	231
6.5	Σύνοψη.....	234
7	Προσανατολισμένος σε Ανάλυση Πινάκων Προγραμματισμός με τη NumPy	239
7.1	Εισαγωγή.....	240
7.2	Δημιουργία Πινάκων από Υφιστάμενα Δεδομένα.....	241
7.3	Χαρακτηριστικά Πινάκων.....	242
7.4	Συμπλήρωση Πινάκων με Συγκεκριμένες Τιμές.....	244
7.5	Δημιουργία Πινάκων από Διαμερίσεις.....	244
7.6	Απόδοση Λίστων έναντι Πινάκων: Εισαγωγή της <code>%timeit</code>	246
7.7	Τελεστές Πινάκων.....	248
7.8	Μέθοδοι Υπολογισμών της NumPy.....	250
7.9	Καθολικές Συναρτήσεις.....	252
7.10	Δεικτοδότηση και Τεμαχισμός.....	254
7.11	Όψεις: Ρηχά Αντίγραφα.....	256
7.12	Βαθιά Αντίγραφα.....	258
7.13	Ανασχηματισμός και Αναστροφή.....	259
7.14	Εισαγωγή στην Επιστήμη Δεδομένων: <code>pandas Series</code> και <code>DataFrame</code>	262
7.14.1	<code>pandas Series</code>	262
7.14.2	<code>DataFrames</code>	267
7.15	Σύνοψη.....	275
8	Συμβολοσειρές: Μία Βαθύτερη Ματιά	283
8.1	Εισαγωγή.....	284
8.2	Μορφοποίηση Συμβολοσειρών.....	285
8.2.1	Τύποι Παρουσίασης.....	285
8.2.2	Πλάτη Πεδίων και Στοίχιση.....	286
8.2.3	Αριθμητική Μορφοποίηση.....	287
8.2.4	Μέθοδος Συμβολοσειρών <code>format</code>	288
8.3	Συνένωση και Επανάληψη Συμβολοσειρών.....	289
8.4	Περικοπή Κενού Χώρου από Συμβολοσειρές.....	290
8.5	Εναλλαγή Πεζών – Κεφαλαίων Γραμμάτων.....	291
8.6	Τελεστές Σύγκρισης για Συμβολοσειρές.....	292
8.7	Αναζήτηση για Υποσυμβολοσειρές.....	292
8.8	Αντικατάσταση Υποσυμβολοσειρών.....	294

8.9	Διαίρεση και Σύνδεση Συμβολοσειρών	294
8.10	Χαρακτήρες και Μέθοδοι Ελέγχου Χαρακτήρων	297
8.11	Ανεπεξέργαστες Συμβολοσειρές	298
8.12	Εισαγωγή στις Κανονικές Εκφράσεις	299
8.12.1	Άρθρωμα <code>re</code> και Συνάρτηση <code>fullmatch</code>	300
8.12.2	Αντικατάσταση Υποσυμβολοσειρών και Διαίρεση Συμβολοσειρών	303
8.12.3	Άλλες Συναρτήσεις Αναζήτησης· Προσπέλαση Ταυτίσεων	304
8.13	Εισαγωγή στην Επιστήμη Δεδομένων: Pandas, Κανονικές Εκφράσεις και Προεπεξεργασία Δεδομένων	307
8.14	Σύνοψη	312
9	Αρχεία και Εξαιρέσεις	319
9.1	Εισαγωγή	320
9.2	Αρχεία	321
9.3	Επεξεργασία Αρχείων Κειμένου	321
9.3.1	Εγγραφή σε Αρχείο Κειμένου: Εισαγωγή της Εντολής <code>with</code>	322
9.3.2	Ανάγνωση Δεδομένων από Αρχείο Κειμένου	323
9.4	Ενημέρωση Αρχείων Κειμένου	325
9.5	Σειριοποίηση με JSON	327
9.6	Εστίαση στην Ασφάλεια: Σειριοποίηση και Αποσειριοποίηση <code>pickle</code>	330
9.7	Πρόσθετες Παρατηρήσεις για Αρχεία	330
9.8	Χειρισμός Εξαιρέσεων	331
9.8.1	Διαίρεση με Μηδέν και Μη Έγκυρα Δεδομένα	332
9.8.2	Εντολές <code>try</code>	332
9.8.3	Σύλληψη Πολλαπλών Εξαιρέσεων σε μία Πρόταση <code>except</code>	335
9.8.4	Τι Εξαιρέσεις Εγείρουν μία Συνάρτηση ή μία Μέθοδος;	336
9.8.5	Τι Κώδικας Πρέπει να Περιλαμβάνεται σε μία Πλοκάδα <code>try</code> ;	336
9.9	Πρόταση <code>finally</code>	336
9.10	Ρητή Έγερση Εξάιρεσης	339
9.11	(Προαιρετική) Εκτύλιξη Στοιβάς και Ίχνη	339
9.12	Εισαγωγή στην Επιστήμη Δεδομένων: Εργασία με Αρχεία CSV	342
9.12.1	Άρθρωμα <code>csv</code> της Πρότυπης Βιβλιοθήκης Python	342
9.12.2	Ανάγνωση Αρχείων CSV σε Pandas <code>DataFrame</code>	344
9.12.3	Ανάγνωση του Συνόλου Δεδομένων Titanic Disaster	346
9.12.4	Απλή Ανάλυση Δεδομένων με το Σύνολο Δεδομένων Titanic Disaster	347
9.12.5	Ιστόγραμμα Ηλικιών Επιβατών	348
9.13	Σύνοψη	349
10	Αντικειμενοστραφής Προγραμματισμός	355
10.1	Εισαγωγή	356
10.2	Εξειδικευμένη Κλάση <code>Account</code>	358
10.2.1	Δοκιμές της Κλάσης <code>Account</code>	358
10.2.2	Ορισμός της Κλάσης <code>Account</code>	360
10.2.3	Σύνθεση: Αναφορές σε Αντικείμενα ως Μέλη Κλάσεων	361
10.3	Έλεγχος Πρόσβασης σε Χαρακτηριστικά	363

10.4	Ιδιότητες για Προσπέλαση Δεδομένων	364
10.4.1	Δοκιμές της Κλάσης Time	364
10.4.2	Ορισμός της Κλάσης Time	366
10.4.3	Σημειώσεις Σχεδιασμού του Ορισμού της Κλάσης Time	370
10.5	Προσομοίωση «Ιδιωτικών» Χαρακτηριστικών	371
10.6	Μελέτη Περίπτωσης: Προσομοίωση Ανακατέματος και Μοιράσματος Τράπουλας	373
10.6.1	Δοκιμές των Κλάσεων Card και DeckOfCards	373
10.6.2	Κλάση Card — Εισαγωγή Χαρακτηριστικών της Κλάσης	375
10.6.3	Κλάση DeckOfCards	377
10.6.4	Προβολή Εικόνων Χαρτιών με τη Matplotlib	378
10.7	Κληρονομικότητα: Βασικές Κλάσεις και Υποκλάσεις	382
10.8	Δημιουργία Ιεραρχίας Κληρονομικότητας: Εισαγωγή του Πολυμορφισμού	384
10.8.1	Βασική Κλάση CommissionEmployee	384
10.8.2	Υποκλάση SalariedCommissionEmployee	387
10.8.3	Πολυμορφική Επεξεργασία CommissionEmployee και SalariedCommissionEmployee	391
10.8.4	Σημείωση για τον Προγραμματισμό με Βάση τα Αντικείμενα και τον Αντικειμενοστραφή Προγραμματισμό	391
10.9	Duck Typing και Πολυμορφισμός	392
10.10	Υπερφόρτωση Τελεστών	393
10.10.1	Δοκιμές της Κλάσης Complex	394
10.10.2	Ορισμός της Κλάσης Complex	395
10.11	Ιεραρχία Κλάσεων Εξαιρέσεων και Προσαρμοσμένες Εξαιρέσεις	397
10.12	Ονοματισμένες Πλειάδες	399
10.13	Σύντομη Εισαγωγή στις Νέες Κλάσεις Δεδομένων της Python 3.7	400
10.13.1	Δημιουργία μίας Κλάσης Δεδομένων Card	401
10.13.2	Χρήση της Κλάσης Δεδομένων Card	403
10.13.3	Πλεονεκτήματα Κλάσεων Δεδομένων έναντι Ονοματισμένων Πλειάδων	405
10.13.4	Πλεονεκτήματα Κλάσεων Δεδομένων έναντι Παραδοσιακών Κλάσεων	406
10.14	Δοκιμές Μονάδας με Συμβολοσειρές Τεκμηρίωσης και doctest	406
10.15	Χώροι Ονομάτων και Εμβέλεις	411
10.16	Εισαγωγή στην Επιστήμη Δεδομένων: Χρονοσειρές και Απλή Γραμμική Παλινδρόμηση ..	414
10.17	Σύνοψη	423
11	Πληροφορική Σκέψη: Αναδρομή, Αναζήτηση, Ταξινόμηση και Big O	431
11.1	Εισαγωγή	432
11.2	Παραγοντικά	433
11.3	Παράδειγμα Αναδρομικού Παραγοντικού	433
11.4	Παράδειγμα Αναδρομικής Σειράς Fibonacci	436
11.5	Αναδρομή έναντι Επανάληψης	439
11.6	Αναζήτηση και Ταξινόμηση	440
11.7	Γραμμική Αναζήτηση	440
11.8	Αποδοτικότητα Αλγορίθμων: Big O	442
11.9	Διαδική Αναζήτηση	444
11.9.1	Υλοποίηση της Διαδικής Αναζήτησης	445

11.9.2	Big O της Δυναμικής Αναζήτησης	447
11.10	Αλγόριθμοι Ταξινόμησης	448
11.11	Ταξινόμηση με Επιλογή	448
11.11.1	Υλοποίηση της Ταξινόμησης με Επιλογή	449
11.11.2	Βοηθητική Συνάρτηση <code>print_pass</code>	450
11.11.3	Big O της Ταξινόμησης με Επιλογή	451
11.12	Ταξινόμηση με Εισαγωγή	451
11.12.1	Υλοποίηση της Ταξινόμησης με Εισαγωγή	452
11.12.2	Big O της Ταξινόμησης με Εισαγωγής	453
11.13	Ταξινόμηση με Συγχώνευση	454
11.13.1	Υλοποίηση της Ταξινόμησης με Συγχώνευση	454
11.13.2	Big O της Ταξινόμησης με Συγχώνευση	459
11.14	Σύνοψη Big O για τους Αλγόριθμους Αναζήτησης και Ταξινόμησης Αυτού του Κεφαλαίου	459
11.15	Οπτικοποίηση Αλγορίθμων	460
11.15.1	Γεννήτριες Συναρτήσεων	462
11.15.2	Υλοποίηση Κινούμενης Απεικόνισης της Ταξινόμησης με Επιλογή	463
11.16	Σύνοψη	468
12	Επεξεργασία Φυσικής Γλώσσας (NLP)	477
12.1	Εισαγωγή	478
12.2	TextBlob	479
12.2.1	Δημιουργία μίας TextBlob	481
12.2.2	Χωρισμός Κειμένου σε Προτάσεις και Λέξεις	482
12.2.3	Επισημείωση των Μερών του Λόγου	482
12.2.4	Εξαγωγή Ονοματικών Φράσεων	483
12.2.5	Ανάλυση Συναισθήματος με τον Προεπιλεγμένο Αναλυτή Συναισθήματος της TextBlob	484
12.2.6	Ανάλυση Συναισθήματος με τον Αναλυτή NaiveBayesAnalyzer	486
12.2.7	Αναγνώριση και Μετάφραση Γλώσσας	487
12.2.8	Κλίση: Μετατροπή σε Πληθυντικό και Ενικό	489
12.2.9	Έλεγχος και Διόρθωση Ορθογραφίας	489
12.2.10	Κανονικοποίηση: Αποκοπή Θέματος και Λημματοποίηση	490
12.2.11	Συχνότητες Λέξεων	491
12.2.12	Λήψη Ορισμών, Συνωνύμων και αντωνύμων από τη WordNet	492
12.2.13	Αφαίρεση Κοινών Λέξεων	494
12.2.14	ν-γράμματα	496
12.3	Οπτικοποίηση Συχνοτήτων Λέξεων με Ραβδογράμματα και Σύννεφα Λέξεων	497
12.3.1	Οπτικοποίηση Συχνοτήτων Λέξεων με την Pandas	497
12.3.2	Οπτικοποίηση Συχνοτήτων Λέξεων με Σύννεφα Λέξεων	500
12.4	Αξιολόγηση αναγνωσιμότητας με την Textatistic	503
12.5	Αναγνώριση Ονοματικών Οντοτήτων με τη spaCy	505
12.6	Εντοπισμός Ομοιότητας με τη spaCy	507
12.7	Άλλες Βιβλιοθήκες και Εργαλεία NLP	509
12.8	Εφαρμογές Μηχανικής Μάθησης και Βαθιάς Μάθησης Φυσικής Γλώσσας	509
12.9	Σύνολα Δεδομένων Φυσικής Γλώσσας	510
12.10	Σύνοψη	510

13	Εξόρυξη Δεδομένων Twitter	515
13.1	Εισαγωγή.....	516
13.2	Επισκόπηση των Twitter API.....	518
13.3	Δημιουργία ενός Λογαριασμού Twitter	519
13.4	Λήψη Διαπιστευτηρίων Twitter — Δημιουργία ενός App	520
13.5	Τι Περιλαμβάνει ένα Tweet;	521
13.6	Tweepy.....	525
13.7	Πιστοποίηση στο Twitter Μέσω Tweepy.....	525
13.8	Λήψη Πληροφοριών για ένα Λογαριασμό Twitter.....	527
13.9	Εισαγωγή στα Tweepy Cursor: Λήψη των Ακόλουθων και των Φίλων ενός Λογαριασμού.....	529
13.9.1	Προσδιορισμός των Ακόλουθων ενός Λογαριασμού.....	529
13.9.2	Προσδιορισμός των Ακολουθούμενων από ένα Λογαριασμό.....	532
13.9.3	Λήψη των Πρόσφατων Tweets ενός Χρήστη.....	532
13.10	Αναζήτηση σε Πρόσφατα Tweets	534
13.11	Εντοπισμός Δημοφιλών Θεμάτων: Twitter Trends API.....	536
13.11.1	Μέρη με Δημοφιλή Θέματα	536
13.11.2	Λήψη Λίστας Δημοφιλών Θεμάτων.....	537
13.11.3	Δημιουργία Σύννεφου Λέξεων από Δημοφιλή Θέματα.....	539
13.12	Καθαρισμός/Προεπεξεργασία Tweets για Ανάλυση.....	541
13.13	Twitter Streaming API.....	542
13.13.1	Δημιουργία μίας Υποκλάσης της StreamListener.....	543
13.13.2	Έναρξη Επεξεργασίας της Ροής	545
13.14	Ανάλυση Συναισθήματος Tweets	547
13.15	Γεωκωδικοποίηση και Χαρτογράφηση.....	551
13.15.1	Λήψη και Χαρτογράφηση Tweets.....	552
13.15.2	Βοηθητικές Συναρτήσεις στο tweetutilities.py	556
13.15.3	Κλάση LocationListener	558
13.16	Τρόποι Αποθήκευσης Tweets	559
13.17	Twitter και Χρονοσειρές	560
13.18	Σύνοψη	560
14	IBM Watson και Γνωσιακή Υπολογιστική	565
14.1	Εισαγωγή: IBM Watson και Γνωσιακή Υπολογιστική.....	566
14.2	Λογαριασμός και Κονσόλα IBM Cloud.....	568
14.3	Υπηρεσίες Watson	568
14.4	Πρόσθετες Υπηρεσίες και Εργαλεία.....	572
14.5	Watson Developer Cloud Python SDK	573
14.6	Μελέτη Περίπτωσης: Μεταφραστική Εφαρμογή ως Ταξιδιωτικός Βοηθός.....	574
14.6.1	Πριν να Εκτελέσετε την Εφαρμογή	575
14.6.2	Δοκιμή της Εφαρμογής	576
14.6.3	Περιήγηση στο Σενάριο SimpleLanguageTranslator.py	577
14.7	Πόροι Watson.	587
14.8	Σύνοψη	589

15	Μηχανική Μάθηση: Ταξινόμηση, Παλινδρόμηση και Ομαδοποίηση	593
15.1	Εισαγωγή στη Μηχανική Μάθηση	594
15.1.1	Scikit-Learn	595
15.1.2	Είδη Μηχανικής Μάθησης	596
15.1.3	Σύνολα Δεδομένων Συσχευμένα με τη Scikit-Learn	598
15.1.4	Βήματα μίας Τυπικής Μελέτης της Επιστήμης Δεδομένων	599
15.2	Μελέτη Περίπτωσης: Ταξινόμηση με Αλγόριθμο k-Πλησιέστερων Γειτόνων και το Σύνολο Δεδομένων Digits, Μέρος 1	599
15.2.1	Αλγόριθμος των k-Πλησιέστερων Γειτόνων	601
15.2.2	Φόρτωση του Συνόλου Δεδομένων	602
15.2.3	Απεικόνιση των Δεδομένων	606
15.2.4	Διαχωρισμός των Δεδομένων για Εκπαίδευση και Έλεγχο	608
15.2.5	Δημιουργία του Μοντέλου	609
15.2.6	Εκπαίδευση του Μοντέλου	610
15.2.7	Πρόβλεψη Κατηγοριών Ψηφίων	610
15.3	Μελέτη Περίπτωσης: Ταξινόμηση με Αλγόριθμο k-Πλησιέστερων Γειτόνων και το Σύνολο Δεδομένων Digits, Μέρος 2	612
15.3.1	Μετρικές Ακρίβειας του Μοντέλου	612
15.3.2	Διασταυρωμένη Επικύρωση k-Πτυχών	616
15.3.3	Δοκιμή Πολλαπλών Μοντέλων για την Εύρεση του Βέλτιστου	617
15.3.4	Ρύθμιση Υπερπαραμέτρων	619
15.4	Μελέτη Περίπτωσης: Χρονοσειρά και Απλή Γραμμική Παλινδρόμηση	620
15.5	Μελέτη Περίπτωσης: Πολλαπλή Γραμμική Παλινδρόμηση με το Σύνολο Δεδομένων California Housing	625
15.5.1	Φόρτωση του Συνόλου Δεδομένων	626
15.5.2	Εξερεύνηση των Δεδομένων με την Pandas	628
15.5.3	Απεικόνιση των Χαρακτηριστικών	630
15.5.4	Διαχωρισμός Δεδομένων για Εκπαίδευση και Έλεγχο	634
15.5.5	Εκπαίδευση του Μοντέλου	634
15.5.6	Έλεγχος του Μοντέλου	635
15.5.7	Απεικόνιση Αναμενόμενων έναντι Προβλεφθεισών Τιμών	636
15.5.8	Μετρικές Μοντέλων Παλινδρόμησης	637
15.5.9	Επιλογή του Βέλτιστου Μοντέλου	638
15.6	Μελέτη Περίπτωσης: Μη Επιτηρούμενη Μηχανική Μάθηση, Μέρος 1 – Μείωση της Διασπασιμότητας	639
15.7	Μελέτη Περίπτωσης: Μη Επιτηρούμενη Μηχανική Μάθηση, Μέρος 2 – Ομαδοποίηση k-Μέσων	642
15.7.1	Φόρτωση του Συνόλου Δεδομένων Iris	644
15.7.2	Εξερεύνηση του Συνόλου Δεδομένων Iris: Περιγραφικά Στατιστικά με την Pandas	646
15.7.3	Απεικόνιση του Συνόλου Δεδομένων μ' ένα Seaborn pairplot	647
15.7.4	Χρήση ενός Εκτιμητή kMeans	650
15.7.5	Μείωση της Διασπασιμότητας με Ανάλυση Κύριων Συνιστωσών	652
15.7.6	Επιλογή του Βέλτιστου Εκτιμητή Ομαδοποίησης	655
15.8	Σύνοψη	656

16	Βαθιά Μάθηση	665
16.1	Εισαγωγή.....	666
16.1.1	Εφαρμογές Βαθιάς Μάθησης	668
16.1.2	Προγράμματα Επίδειξης Βαθιάς Μάθησης	669
16.1.3	Πόροι Keras	669
16.2	Σύνολα Δεδομένων Ενσωματωμένα στην Keras	669
16.3	Προσαρμοσμένα Περιβάλλοντα Anaconda	670
16.4	Νευρωνικά Δίκτυα	672
16.5	Τανυστές.....	674
16.6	Συνελκτικά Νευρωνικά Δίκτυα για Όραση· Πολυταξινόμηση με το Σύνολο Δεδομένων MNIST	676
16.6.1	Φόρτωση του Συνόλου Δεδομένων MNIST	677
16.6.2	Εξερεύνηση Δεδομένων	678
16.6.3	Προετοιμασία Δεδομένων	680
16.6.4	Δημιουργία του Νευρωνικού Δικτύου	682
16.6.5	Εκπαίδευση και Αξιολόγηση του Μοντέλου	691
16.6.6	Αποθήκευση και Φόρτωση ενός Μοντέλου	696
16.7	Οπτικοποίηση Εκπαίδευσης Νευρωνικού Δικτύου με το TensorBoard	697
16.8	ConvnetJS: Εκπαίδευση και Οπτικοποίηση Βαθιάς Μάθησης Μέσω Φυλλομετρητή	700
16.9	Αναδρομικά Νευρωνικά Δίκτυα για Ακολουθίες· Ανάλυση Συναισθήματος με το Σύνολο Δεδομένων IMDb	701
16.9.1	Φόρτωση του Συνόλου Δεδομένων Κριτικών Ταινιών IMDb	702
16.9.2	Εξερεύνηση Δεδομένων	703
16.9.3	Προετοιμασία Δεδομένων	705
16.9.4	Δημιουργία του Νευρωνικού Δικτύου	706
16.9.5	Εκπαίδευση και Αξιολόγηση του Μοντέλου	709
16.10	Ρύθμιση Μοντέλων Βαθιάς Μάθησης	710
16.11	Μοντέλα Convnet Προεκπαιδευμένα στο ImageNet	711
16.12	Ενισχυτική Μάθηση	712
16.12.1	Βαθιά Q-Μάθηση	713
16.12.2	OpenAI Gym	713
16.13	Σύνοψη.....	714
17	Μαζικά Δεδομένα: Hadoop, Spark, NoSQL και IoT	723
17.1	Εισαγωγή.....	724
17.2	Σχισιακές Βάσεις Δεδομένων και Δομημένη Γλώσσα Ερωτημάτων (SQL)	728
17.2.1	Βάση Δεδομένων books	730
17.2.2	Ερωτήματα SELECT	734
17.2.3	Πρόταση WHERE	734
17.2.4	Πρόταση ORDER BY	736
17.2.5	Συγχώνευση Δεδομένων από Πολλαπλούς Πίνακες: INNER JOIN	737
17.2.6	Εντολή INSERT INTO	738
17.2.7	Εντολή UPDATE	739
17.2.8	Εντολή DELETE FROM	739

17.3	Βάσεις Μαζικών Δεδομένων NoSQL και NewSQL: Σύντομη Περιήγηση.....	741
17.3.1	NoSQL Βάσεις Δεδομένων Κλειδιού-Τιμής	741
17.3.2	NoSQL Βάσεις Δεδομένων Εγγράφων.....	742
17.3.3	NoSQL Βάσεις Δεδομένων σε Στήλες.....	742
17.3.4	NoSQL Βάσεις Δεδομένων Γράφων	743
17.3.5	NewSQL Βάσεις Δεδομένων.....	743
17.4	Μελέτη Περίπτωσης: Μία MongoDB Βάση Δεδομένων Εγγράφων JSON.....	744
17.4.1	Δημιουργία του Συμπλέγματος MongoDB Atlas.....	745
17.4.2	Ροή από Tweets προς τη MongoDB	746
17.5	Hadoop.....	755
17.5.1	Επισκόπηση του Hadoop	755
17.5.2	Ανασκόπηση Μηκών Λέξεων στο <i>Romeo and Juliet</i> μέσω MapReduce.....	758
17.5.3	Δημιουργία ενός Συμπλέγματος Apache Hadoop στη Microsoft Azure HDInsight.....	758
17.5.4	Συνεχής Ροή Hadoop	760
17.5.5	Υλοποίηση Χαρτογράφου	760
17.5.6	Υλοποίηση Αναγωγέα	761
17.5.7	Προετοιμασία Εκτέλεσης του Παραδείγματος MapReduce.....	762
17.5.8	Εκτέλεση της Εργασίας MapReduce	763
17.6	Spark.....	766
17.6.1	Επισκόπηση του Spark	766
17.6.2	Docker και Στοιβές Docker της Jupyter	767
17.6.3	Μέτρηση Λέξεων με Spark.....	770
17.6.4	Μέτρηση Λέξεων με Spark στη Microsoft Azure.....	773
17.7	Συνεχής Ροή Spark: Μέτρηση Ετικετών Twitter με τη Στοιβα Docker pyspark-notebook.....	777
17.7.1	Αποστολή Συνεχούς Ροής Tweets σε μία Υποδοχή.....	777
17.7.2	Ανασκόπηση Ετικετών Tweets· Εισαγωγή της Spark SQL.....	780
17.8	Διαδίκτυο των Αντικειμένων και Πίνακες Ελέγχου.....	786
17.8.1	Δημοσίευση και Συνδρομή	788
17.8.2	Απεικόνιση Δείγματος Ζωντανής Ροής PubNub μ' έναν Πίνακα Ελέγχου Freeboard	788
17.8.3	Προσομοίωση ενός Συνδεδεμένου στο Διαδίκτυο Θερμοστάτη σε Python	790
17.8.4	Δημιουργία του Πίνακα Ελέγχου με το Freeboard.io	792
17.8.5	Δημιουργία ενός Συνδρομητή Python στην PubNub	794
17.9	Σύνοψη.....	798