

Περιεχόμενα

Πρόλογος	i
Περιεχόμενα	v
1 ΕΙΣΑΓΩΓΗ	1
1.1 ΤΙ ΕΙΝΑΙ Η ΕΞΟΥΥΞΗ ΔΕΔΟΜΕΝΩΝ ;	3
1.2 ΕΝΔΙΑΦΕΡΟΥΣΕΣ ΠΡΟΚΛΗΣΕΙΣ	4
1.3 ΟΙ ΡΙΖΕΣ ΤΗΣ ΕΞΟΥΥΞΗΣ ΔΕΔΟΜΕΝΩΝ	5
1.4 ΕΡΓΑΣΙΕΣ ΤΗΣ ΕΞΟΥΥΞΗΣ ΔΕΔΟΜΕΝΩΝ	6
1.5 ΣΚΟΠΟΣ ΚΑΙ ΟΡΓΑΝΩΣΗ ΤΟΥ ΒΙΒΛΙΟΥ	10
1.6 ΒΙΒΛΙΟΓΡΑΦΙΚΕΣ ΑΝΑΦΟΡΕΣ	11
ΒΙΒΛΙΟΓΡΑΦΙΑ	13
1.7 ΑΣΚΗΣΕΙΣ	15
2 ΔΕΔΟΜΕΝΑ	17
2.1 ΤΥΠΟΙ ΔΕΔΟΜΕΝΩΝ	19
2.1.1 Χαρακτηριστικά και Μέτρηση	19
2.1.2 Τύποι συνόλων δεδομένων	25
2.2 ΠΟΙΟΤΗΤΑ ΤΩΝ ΔΕΔΟΜΕΝΩΝ	30
2.2.1 Θέματα μετρήσεων και συλλογής δεδομένων	30
2.2.2 Θέματα Σχετικά με Εφαρμογές	35
2.3 ΠΡΟΕΠΕΞΕΡΓΑΣΙΑ ΤΩΝ ΔΕΔΟΜΕΝΩΝ	36
2.3.1 Συνάθροιση	36
2.3.2 Δειγματοληψία	38
2.3.3 Μείωση των πολλών διαστάσεων	40
2.3.4 Επιλογή Υποσυνόλου Γνωρισμάτων	41
2.3.5 Δημιουργία Γνωρισμάτων	43
2.3.6 Διακριτοποίηση και Δυαδικοποίηση Δεδομένων	44
2.3.7 Μετασχηματισμοί Μεταβλητών	49
2.4 ΜΕΤΡΑ ΟΜΟΙΟΤΗΤΑΣ ΚΑΙ ΑΝΟΜΟΙΟΤΗΤΑΣ	50
2.4.1 Βασικές Έννοιες	50
2.4.2 Ομοιότητα και Ανομοιότητα μεταξύ Απλών Χαρακτηριστικών	52
2.4.3 Ανομοιότητες Μεταξύ Αντικειμένων Δεδομένων	53
2.4.4 Ομοιότητες μεταξύ Αντικειμένων Δεδομένων	55
2.4.5 Παραδείγματα Μέτρων Εγγύτητας	55
2.4.6 Αμοιβαία Πληροφορία	61
2.4.7 Συναρτήσεις Πυρήνα *	62
2.4.8 Απόκλιση Bregman*	65
2.4.9 Θέματα Υπολογισμών Εγγύτητας	66
2.4.10 Επιλέγοντας το Κατάλληλο Μέτρο Εγγύτητας	68
2.5 ΒΙΒΛΙΟΓΡΑΦΙΚΕΣ ΑΝΑΦΟΡΕΣ	69
2.6 ΑΣΚΗΣΕΙΣ	73
3 ΚΑΤΗΓΟΡΙΟΠΟΙΗΣΗ: ΒΑΣΙΚΕΣ ΕΝΝΟΙΕΣ ΚΑΙ ΤΕΧΝΙΚΕΣ	79
3.1 ΒΑΣΙΚΕΣ ΕΝΝΟΙΕΣ	80

3.2	ΓΕΝΙΚΟ ΠΛΑΙΣΙΟ ΚΑΤΗΓΟΡΙΟΠΟΙΗΣΗΣ	82
3.3	ΚΑΤΗΓΟΡΙΟΠΟΙΗΤΗΣ ΔΕΝΔΡΟΥ ΑΠΟΦΑΣΗΣ	84
3.3.1	Ένας Βασικός Αλγόριθμος Κατασκευής Ενός Δένδρου Απόφασης	85
3.3.2	Μέθοδοι Έκφρασης των Συνθηκών Ελέγχου Χαρακτηριστικού	87
3.3.3	Μέθοδοι Επιλογής μίας Συνθήκης Ελέγχου Χαρακτηριστικού	88
3.3.4	Αλγόριθμος Επαγωγής Δένδρου Απόφασης	95
3.3.5	Ένα Παράδειγμα: Ανίχνευση Προγραμμάτων Web Robot	96
3.3.6	Χαρακτηριστικά των Κατηγοριοποιητών Δένδρων Απόφασης	98
3.4	ΥΠΕΡΠΡΟΣΑΡΜΟΓΗ ΜΟΝΤΕΛΟΥ	102
3.4.1	Τα Αίτια της Υπερπροσαρμογής Μοντέλου	104
3.5	ΕΠΙΛΟΓΗ ΜΟΝΤΕΛΟΥ	108
3.5.1	Χρησιμοποίηση Συνόλου Επαλήθευσης	109
3.5.2	Συγχώνευση της Πολυπλοκότητας του Μοντέλου	110
3.5.3	Εκτίμηση Στατιστικών Ορίων	112
3.5.4	Επιλογή Μοντέλων για Δένδρα Απόφασης	113
3.6	ΕΚΤΙΜΗΣΗ ΜΟΝΤΕΛΟΥ	114
3.6.1	Εκτίμηση Μέσω Συνόλου Ελέγχου	115
3.6.2	Επαλήθευση με Διασταύρωση	115
3.7	Η ΠΑΡΟΥΣΙΑ ΓΕΝΙΚΩΝ ΠΑΡΑΜΕΤΡΩΝ	117
3.7.1	Επιλογή Γενικών Παραμέτρων	117
3.7.2	Εμφωλευμένη Επαλήθευση με Διασταύρωση	118
3.8	ΔΥΣΚΟΛΙΕΣ ΚΑΤΑ ΤΗΝ ΕΠΙΛΟΓΗ ΚΑΙ ΕΚΤΙΜΗΣΗ ΜΟΝΤΕΛΩΝ	119
3.8.1	Επικάλυψη Ανάμεσα στα Σύνολα Εκπαίδευσης και Ελέγχου	120
3.8.2	Χρήση του Σφάλματος Επαλήθευσης ως Σφάλμα Γενίκευσης	120
3.9	ΣΥΓΚΡΙΣΗ ΜΟΝΤΕΛΩΝ	120
3.9.1	Εκτίμηση ενός Διαστήματος Εμπιστοσύνης για την Ακρίβεια	121
3.9.2	Σύγκριση της Απόδοσης δύο Μοντέλων	122
3.10	ΒΙΒΛΙΟΓΡΑΦΙΚΕΣ ΑΝΑΦΟΡΕΣ	123
3.11	ΑΣΚΗΣΕΙΣ	129
4	ΚΑΤΗΓΟΡΙΟΠΟΙΗΣΗ: ΕΝΑΛΛΑΚΤΙΚΕΣ ΤΕΧΝΙΚΕΣ	137
4.1	ΤΥΠΟΙ ΚΑΤΗΓΟΡΙΟΠΟΙΗΤΩΝ	137
4.2	ΚΑΤΗΓΟΡΙΟΠΟΙΗΤΕΣ ΚΑΝΟΝΩΝ	138
4.2.1	Πως Λειτουργεί Ένας Κατηγοριοποιητής Κανόνων	139
4.2.2	Ιδιότητες Ενός Συνόλου Κανόνων	140
4.2.3	Άμεσες Μέθοδοι Εξαγωγής Κανόνων	141
4.2.4	Έμμεσες Μέθοδοι Εξαγωγής Κανόνων	145
4.2.5	Χαρακτηριστικά των Κατηγοριοποιητών Κανόνων	146
4.3	ΚΑΤΗΓΟΡΙΟΠΟΙΗΤΕΣ ΠΛΗΣΙΕΣΤΕΡΟΥ ΓΕΙΤΟΝΑ	147
4.3.1	Αλγόριθμος	148
4.3.2	Χαρακτηριστικά των Κατηγοριοποιητών Πλησιέστερου Γείτονα	149
4.4	Ο ΑΠΛΟΪΚΟΣ ΚΑΤΑ ΒΑΥΕΣ ΚΑΤΗΓΟΡΙΟΠΟΙΗΤΗΣ	150
4.4.1	Βασικά Στοιχεία Θεωρίας Πιθανοτήτων	151
4.4.2	Απλοϊκή Υπόθεση Bayes	154
4.5	ΔΙΚΤΥΑ ΤΟΥ ΒΑΥΕΣ	160
4.5.1	Γραφική Αναπαράσταση	160
4.5.2	Εξαγωγή Συμπερασμάτων και Εκπαίδευση	164
4.5.3	Χαρακτηριστικά Δικτύων του Bayes	169
4.6	ΛΟΓΙΣΤΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ	170
4.6.1	Η Λογιστική Παλινδρόμηση ως Γενικευμένο Γραμμικό Μοντέλο	171
4.6.2	Μάθηση Παραμέτρων Μοντέλου	172
4.6.3	Χαρακτηριστικά της Λογιστικής Παλινδρόμησης	174
4.7	ΤΕΧΝΗΤΑ ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ	174
4.7.1	Ο Νευρώνας Perceptron	175
4.7.2	Τεχνητό Νευρωνικό Δίκτυο Πολλών Επιπέδων	177

4.7.3	Χαρακτηριστικά των Τεχνητών Νευρωνικών Δικτύων	182
4.8	ΒΑΘΙΑ ΜΑΘΗΣΗ	183
4.8.1	Χρησιμοποίηση Συνεργατικών Συναρτήσεων Απώλειας	184
4.8.2	Χρησιμοποίηση Ευαίσθητων Συναρτήσεων Ενεργοποίησης	185
4.8.3	Προσαρμογή	187
4.8.4	Αρχικοποίησης των Παραμέτρων του Μοντέλου	189
4.8.5	Χαρακτηριστικά της Βαθιάς Μάθησης	191
4.9	ΜΗΧΑΝΕΣ ΔΙΑΝΥΣΜΑΤΩΝ ΥΠΟΣΤΗΡΙΞΗΣ	192
4.9.1	Περιθώριο Διαχωριστικού Υπερεπιπέδου	192
4.9.2	Γραμμική Μηχανή Υποστήριξης Διανυσμάτων	194
4.9.3	SVM Απαλού Περιθωρίου.....	197
4.9.4	Μη Γραμμική SVM	202
4.9.5	Χαρακτηριστικά της SVM.....	205
4.10	ΜΕΘΟΔΟΙ ΟΜΑΔΩΝ	206
4.10.1	Αιτιολογία της Μεθόδου Ομάδων	207
4.10.2	Μέθοδοι Κατασκευής Ομάδας Κατηγοριοποιητών.....	208
4.10.3	Διαχωρισμός Μεροληψίας-Διακύμανσης	209
4.10.4	Εμφωλίαση.....	211
4.10.5	Ενίσχυση.....	212
4.10.6	Τυχαία Δάση	216
4.10.7	Εμπειρική Σύγκριση Ανάμεσα στις Μεθόδους Ομάδων	218
4.11	ΤΟ ΠΡΟΒΛΗΜΑ ΤΗΣ ΑΝΙΣΟΡΡΟΠΙΑΣ ΤΩΝ ΚΑΤΗΓΟΡΙΩΝ	218
4.11.1	Κατασκευή Κατηγοριοποιητών με Ανισορροπία Κατηγοριών	219
4.11.2	Εκτίμηση της Απόδοσης με Ανισορροπία Κατηγοριών	222
4.11.3	Εύρεση Ενός Βέλτιστου Κατωφλίου Βαθμού Κατηγοριοποίησης.....	224
4.11.4	Αθροιστική Εκτίμηση Απόδοσης.....	225
4.12	ΠΡΟΒΛΗΜΑ ΠΟΛΛΩΝ ΚΑΤΗΓΟΡΙΩΝ	230
4.13	ΒΙΒΛΙΟΓΡΑΦΙΚΕΣ ΑΝΑΦΟΡΕΣ	232
4.14	ΑΣΚΗΣΕΙΣ.....	241
5	ΑΝΑΛΥΣΗ ΣΥΣΧΕΤΙΣΗΣ : ΒΑΣΙΚΕΣ ΕΝΝΟΙΕΣ ΚΑΙ ΑΛΓΟΡΙΘΜΟΙ	251
5.1	ΒΑΣΙΚΕΣ ΕΝΝΟΙΕΣ	252
5.2	ΠΑΡΑΓΩΓΗ ΣΥΧΝΩΝ ΣΤΟΙΧΕΙΟΣΥΝΟΛΩΝ	254
5.2.1	Η Εκ των Προτέρων Αρχή.....	256
5.2.2	Παραγωγή Συχνών Στοιχειοσυνόλων με τον Αλγόριθμο <i>Apriori</i>	257
5.2.3	Παραγωγή και Κλάδεμα Υποψηφίων	259
5.2.4	Μέτρηση Υποστήριξης	263
5.2.5	Πολυπλοκότητα Υπολογισμών	265
5.3	ΠΑΡΑΓΩΓΗ ΚΑΝΟΝΩΝ	268
5.3.1	Κλάδεμα Βάσει Εμπιστοσύνης	268
5.3.2	Παραγωγή κανόνων του Αλγορίθμου <i>Apriori</i>	268
5.3.3	Ένα Παράδειγμα: Εγγραφές Ψηφοφορίας του Κογκρέσο	269
5.4	ΣΥΝΤΟΜΗ ΑΝΑΠΑΡΑΣΤΑΣΗ ΤΩΝ ΣΥΧΝΩΝ ΣΤΟΙΧΕΙΟΣΥΝΟΛΩΝ	271
5.4.1	Μέγιστα Συχνά Στοιχειοσύνολα	271
5.4.2	Κλειστά Συχνά Στοιχειοσύνολα	271
5.5	ΕΝΑΛΛΑΚΤΙΚΕΣ ΜΕΘΟΔΟΙ ΠΑΡΑΓΩΓΗΣ ΣΥΧΝΩΝ ΣΤΟΙΧΕΙΟΣΥΝΟΛΩΝ	275
5.6	ΑΛΓΟΡΙΘΜΟΣ FP- ΑΝΑΠΤΥΞΗΣ	278
5.6.1	Αναπαράσταση FP-Δένδρου.....	278
5.6.2	Παραγωγή Συχνών Στοιχειοσυνόλων με τον Αλγόριθμο FP-Ανάπτυξης	280
5.7	ΕΚΤΙΜΗΣΗ ΤΩΝ ΥΠΟΔΕΙΓΜΑΤΩΝ ΣΥΣΧΕΤΙΣΗΣ	283
5.7.1	Αντικειμενικά Μέτρα Ενδιαφέροντος	283
5.7.2	Μέτρα Πέραν των Ζευγών Δυαδικών Μεταβλητών	292
5.7.3	Το παράδοξο του Simpson	293
5.8	ΕΠΙΔΡΑΣΗ ΤΗΣ ΜΗ ΟΜΟΙΟΜΟΡΦΗΣ ΚΑΤΑΝΟΜΗΣ ΥΠΟΣΤΗΡΙΞΗΣ	294
5.9	ΒΙΒΛΙΟΓΡΑΦΙΚΕΣ ΑΝΑΦΟΡΕΣ	298

5.10	ΑΣΚΗΣΕΙΣ.....	310
6	ΑΝΑΛΥΣΗ ΣΥΣΧΕΤΙΣΗΣ : ΠΡΟΧΩΡΗΜΕΝΕΣ ΕΝΝΟΙΕΣ	319
6.1	ΔΙΑΧΕΙΡΙΣΗ ΚΑΤΗΓΟΡΙΚΩΝ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ	319
6.2	ΔΙΑΧΕΙΡΙΣΗ ΣΥΝΕΧΩΝ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ	321
6.2.1	Μέθοδοι που Βασίζονται στη Διακριτοποίηση των Δεδομένων	322
6.2.2	Μέθοδοι που Βασίζονται στη Στατιστική.....	324
6.2.3	Μέθοδοι μη Διακριτοποίησης.....	325
6.3	ΔΙΑΧΕΙΡΙΣΗ ΜΙΑΣ ΙΕΡΑΡΧΙΑΣ ΕΝΝΟΙΩΝ	326
6.4	ΑΚΟΛΟΥΘΙΑΚΑ ΥΠΟΔΕΙΓΜΑΤΑ	328
6.4.1	Βασικές Έννοιες	328
6.4.2	Ανακάλυψη Ακολουθιακών Υποδειγμάτων	330
6.4.3	Χρονικοί Περιορισμοί.....	334
6.4.4	Εναλλακτικές Μέθοδοι Μέτρησης	337
6.5	ΥΠΟΔΕΙΓΜΑΤΑ ΓΡΑΦΩΝ	339
6.5.1	Βασικές Έννοιες	340
6.5.2	Εξόρυξη Συχνών Υπογράφων	342
6.5.3	Παραγωγή Υποψήφιων Υπογράφων	344
6.5.4	Κλάδεμα Υποψηφίων	349
6.5.5	Μέτρηση Υποστήριξης	349
6.6	ΣΠΑΝΙΑ ΥΠΟΔΕΙΓΜΑΤΑ	349
6.6.1	Αρνητικά Υποδείγματα.....	350
6.6.2	Αρνητικά Συσχετιζόμενα Υποδείγματα	350
6.6.3	Συγκρίσεις Ανάμεσα σε Σπάνια Υποδείγματα, Αρνητικά Υποδείγματα και Αρνητικά Συσχετιζόμενα Υποδείγματα	351
6.6.4	Τεχνικές Εξόρυξης Ενδιαφέροντων Σπάνιων Υποδειγμάτων	352
6.6.5	Τεχνικές που Βασίζονται στην Εξόρυξη Αρνητικών Υποδειγμάτων	353
6.6.6	Τεχνικές που Βασίζονται στην Αναμενόμενη Υποστήριξη	355
6.7	ΒΙΒΛΙΟΓΡΑΦΙΚΕΣ ΑΝΑΦΟΡΕΣ	357
6.8	ΑΣΚΗΣΕΙΣ.....	361
7	ΑΝΑΛΥΣΗ ΣΥΣΤΑΔΩΝ : ΒΑΣΙΚΕΣ ΕΝΝΟΙΕΣ ΚΑΙ ΑΛΓΟΡΙΘΜΟΙ	371
7.1	ΠΕΡΙΛΗΨΗ.....	373
7.1.1	Τι Είναι η Ανάλυση Συστάδων;	373
7.1.2	Διαφορετικοί Τύποι Συσταδοποίησης	374
7.1.3	Διαφορετικοί Τύποι Συστάδων	375
7.2	ΣΥΣΤΑΔΟΠΟΙΗΣΗ K – ΜΕΣΩΝ	377
7.2.1	Ο Βασικός Αλγόριθμος K – Μέσων	377
7.2.2	Αλγόριθμος K – Μέσων: Επιπλέον Ζητήματα	384
7.2.3	Διχοτομικός Αλγόριθμος των K – Μέσων	386
7.2.4	Ο Αλγόριθμος K – Μέσων και Διαφορετικοί Τύποι Συστάδων	387
7.2.5	Δυνατά και Αδύνατα Σημεία	387
7.2.6	Ο Αλγόριθμος K – Μέσων ως Πρόβλημα Βελτιστοποίησης	388
7.3	ΣΥΣΣΩΡΕΥΤΙΚΗ ΙΕΡΑΡΧΙΚΗ ΣΥΣΤΑΔΟΠΟΙΗΣΗ	391
7.3.1	Βασικός Αλγόριθμος Συσσωρευτικής Ιεραρχικής Συσταδοποίησης	391
7.3.2	Εξειδικευμένες Τεχνικές	393
7.3.3	Ο Τύπος των Lance-Williams για την Εγγύτητα των Συστάδων	397
7.3.4	Βασικά Θέματα στην Ιεραρχική Συσταδοποίηση	397
7.3.5	Ακραίες Τιμές	398
7.3.6	Δυνατά Σημεία και Αδυναμίες	399
7.4	Ο ΑΛΓΟΡΙΘΜΟΣ DBSCAN.....	399
7.4.1	Πυκνότητα : Προσέγγιση Βάσει Κέντρου	399
7.4.2	Ο Αλγόριθμος DBSCAN	400
7.4.3	Δυνατά Σημεία και Αδυναμίες	402
7.5	ΕΚΤΙΜΗΣΗ ΣΥΣΤΑΔΩΝ	403
7.5.1	Επισκόπηση	404

7.5.2	Εκτίμηση Συστάδων Χωρίς Επίβλεψη Χρησιμοποιώντας τη Συνοχή και το Διαχωρισμό	406
7.5.3	Εκτίμηση Συστάδων Χωρίς Επίβλεψη Χρησιμοποιώντας τη Μήτρα Εγγύτητας	411
7.5.4	Εκτίμηση της Ιεραρχικής Συσταδοποίησης Χωρίς Επίβλεψη	414
7.5.5	Καθορισμός του Σωστού Πλήθους Συστάδων	414
7.5.6	Τάση Συσταδοποίησης	416
7.5.7	Μέθοδοι Εγκυρότητας Συστάδων Με Επίβλεψη	416
7.5.8	Αξιολόγηση της Σημασίας των Μέτρων Επαλήθευσης των Συστάδων	420
7.5.9	Επιλογή Μέτρου Εγκυρότητας Συστάδων	421
7.6	ΒΙΒΛΙΟΓΡΑΦΙΚΕΣ ΑΝΑΦΟΡΕΣ	422
7.7	ΑΣΚΗΣΕΙΣ	426
8	ΑΝΑΛΥΣΗ ΣΥΣΤΑΔΩΝ : ΕΠΙΠΡΟΣΘΕΤΑ ΘΕΜΑΤΑ ΚΑΙ ΑΛΓΟΡΙΘΜΟΙ	435
8.1	ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΤΩΝ ΔΕΔΟΜΕΝΩΝ, ΣΥΣΤΑΔΩΝ, ΚΑΙ ΑΛΓΟΡΙΘΜΩΝ ΣΥΣΤΑΔΟΠΟΙΗΣΗΣ	435
8.1.1	Παράδειγμα: Σύγκριση DBSCAN και K-μέσων	436
8.1.2	Χαρακτηριστικά των Δεδομένων	437
8.1.3	Χαρακτηριστικά Συστάδων	438
8.1.4	Γενικά Χαρακτηριστικά των Αλγόριθμων Συσταδοποίησης	439
8.2	ΣΥΣΤΑΔΟΠΟΙΗΣΗ ΒΑΣΕΙ ΠΡΟΤΥΠΟΥ	440
8.2.1	Ασαφής Συσταδοποίηση	440
8.2.2	Συσταδοποίηση με Συνδυαστικά Μοντέλα	443
8.2.3	Αυτοοργανωνόμενοι Χάρτες SOM	450
8.3	ΣΥΣΤΑΔΟΠΟΙΗΣΗ ΒΑΣΕΙ ΠΥΚΝΟΤΗΤΑΣ	455
8.3.1	Συσταδοποίηση Βάσει Πλέγματος	456
8.3.2	Συσταδοποίηση Υποχώρου	458
8.3.3	DENCLUE: Μια Στρατηγική Βάσει Πυρήνα για Συσταδοποίηση Βάσει Πυκνότητας	461
8.4	ΣΥΣΤΑΔΟΠΟΙΗΣΗ ΒΑΣΕΙ ΓΡΑΦΩΝ	464
8.4.1	Αραιώση	464
8.4.2	Συσταδοποίηση Ελάχιστου Δένδρου Κάλυψης- ΕΔΚ	465
8.4.3	Δίδελφος: Βέλτιστος Διαχωρισμός Αραιών Ομοιοτήτων με Χρήση της Τεχνικής METIS	467
8.4.4	Χαμαιλέον: Ιεραρχική Συσταδοποίηση με Δυναμική Μοντελοποίηση	467
8.4.5	Φασματική Συσταδοποίηση	471
8.4.6	Ομοιότητα Κοινού Πλησιέστερου Γείτονα	477
8.4.7	Ο Αλγόριθμος συσταδοποίησης Jarvis-Patrick	479
8.4.8	Πυκνότητα ΚΠΓ	480
8.4.9	Συσταδοποίηση Βάσει της Πυκνότητας ΚΠΓ	481
8.5	ΚΛΙΜΑΚΟΥΜΕΝΟΙ ΑΛΓΟΡΙΘΜΟΙ ΣΥΣΤΑΔΟΠΟΙΗΣΗΣ	483
8.5.1	Κλιμάκωση: Γενικά Θέματα και Προσεγγίσεις	483
8.5.2	Η Τεχνική BIRCH	484
8.5.3	Ο Αλγόριθμος CURE	486
8.6	ΠΟΙΟΣ ΑΛΓΟΡΙΘΜΟΣ ΣΥΣΤΑΔΟΠΟΙΗΣΗΣ ΕΠΙΛΕΓΕΤΑΙ;	488
8.7	ΒΙΒΛΙΟΓΡΑΦΙΚΕΣ ΑΝΑΦΟΡΕΣ	490
8.8	ΑΣΚΗΣΕΙΣ	495
9	ΑΝΙΧΝΕΥΣΗ ΑΝΩΜΑΛΙΩΝ	499
9.1	ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΤΩΝ ΠΡΟΒΛΗΜΑΤΩΝ ΑΝΙΧΝΕΥΣΗΣ ΑΝΩΜΑΛΙΩΝ	500
9.1.1	Ορισμός της Ανωμαλίας	500
9.1.2	Φύση των Δεδομένων	501
9.1.3	Πως Χρησιμοποιείται η Ανίχνευση Ανωμαλιών	502
9.2	ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΜΕΘΟΔΩΝ ΑΝΙΧΝΕΥΣΗΣ ΑΝΩΜΑΛΙΩΝ	502
9.3	ΣΤΑΤΙΣΤΙΚΕΣ ΠΡΟΣΕΓΓΙΣΕΙΣ	503
9.3.1	Χρήση Παραμετρικών Μοντέλων	503
9.3.2	Χρήση Μη Παραμετρικών Μοντέλων	505
9.3.3	Μοντελοποίηση Ομαλών και Ανώμαλων Κατηγοριών	507
9.3.4	Αξιολόγηση της Στατιστικής Σημαντικότητας	508
9.3.5	Δυνατά Σημεία και Αδυναμίες	509
9.4	ΠΡΟΣΕΓΓΙΣΕΙΣ ΒΑΣΕΙ ΕΓΓΥΤΗΤΑΣ	509

9.4.1	Βαθμός Ανωμαλίας Βάσει Απόστασης	510
9.4.2	Βαθμός Ανωμαλίας Βάσει Πυκνότητας	510
9.4.3	Σχετικός Βαθμός Ανωμαλίας Βάσει Πυκνότητας	512
9.4.4	Δυνατά Σημεία και Αδυναμίες	513
9.5	ΠΡΟΣΕΓΓΙΣΕΙΣ ΒΑΣΕΙ ΣΥΣΤΑΔΟΠΟΙΗΣΗΣ	514
9.5.1	Εύρεση Ανώμαλων Συστάδων	514
9.5.2	Εύρεση Ανώμαλων Συστάδων	514
9.5.3	Αξιολόγηση του Βαθμού στον Οποίο ένα Αντικείμενο Ανήκει σε μια Συστάδα	514
9.5.4	Δυνατά Σημεία και Αδυναμίες	517
9.6	ΠΡΟΣΕΓΓΙΣΕΙΣ ΒΑΣΕΙ ΑΝΑΚΑΤΑΣΚΕΥΗΣ	517
9.6.1	Δυνατά Σημεία και Αδυναμίες	519
9.7	ΜΟΝΑΔΙΑΙΑ ΚΑΤΗΓΟΡΙΟΠΟΙΗΣΗ	519
9.7.1	Χρήση των Πυρήνων	520
9.7.2	Το Κόλπο της Αρχής των Αξόνων	521
9.7.3	Δυνατά Σημεία και Αδυναμίες	523
9.8	ΠΡΟΣΕΓΓΙΣΗ ΘΕΩΡΙΑΣ ΤΗΣ ΠΛΗΡΟΦΟΡΙΑΣ	524
9.8.1	Δυνατά Σημεία και Αδυναμίες	525
9.9	ΑΞΙΟΛΟΓΗΣΗ ΤΗΣ ΑΝΙΧΝΕΥΣΗΣ ΑΝΩΜΑΛΙΩΝ	525
9.10	ΒΙΒΛΙΟΓΡΑΦΙΚΕΣ ΑΝΑΦΟΡΕΣ	527
9.11	ΑΣΚΗΣΕΙΣ	533

10 ΑΠΟΦΥΓΗ ΕΣΦΑΛΜΕΝΩΝ ΑΝΑΚΑΛΥΨΕΩΝ 537

10.1	ΒΑΣΙΚΑ ΣΤΟΙΧΕΙΑ : ΣΤΑΤΙΣΤΙΚΟΙ ΕΛΕΓΧΟΙ	538
10.1.1	Έλεγχος Σημαντικότητας	538
10.1.2	Έλεγχοι Υποθέσεων	541
10.1.3	Έλεγχοι Πολλαπλών Υποθέσεων	545
10.1.4	Παγίδες του Στατιστικού Ελέγχου	551
10.2	ΜΟΝΤΕΛΟΠΟΙΗΣΗ ΜΗΔΕΝΙΚΩΝ ΚΑΙ ΕΝΑΛΛΑΚΤΙΚΩΝ ΚΑΤΑΝΟΜΩΝ	553
10.2.1	Παραγωγή Συνόλων Συνθετικών Δεδομένων	554
10.2.2	Τυχαιοποίηση Ετικετών Κατηγοριών	555
10.2.3	Επαναδειγματοληψία Στιγμιότυπων	555
10.2.4	Μοντελοποίηση της Κατανομής του Στατιστικού Ελέγχου	556
10.3	ΣΤΑΤΙΣΤΙΚΟΙ ΕΛΕΓΧΟΙ ΓΙΑ ΤΗΝ ΚΑΤΗΓΟΡΙΟΠΟΙΗΣΗ	556
10.3.1	Εκτίμηση της Απόδοσης της Κατηγοριοποίησης	556
10.3.2	Η Δυαδική Κατηγοριοποίηση ως Έλεγχος Πολλαπλών Υποθέσεων	557
10.3.3	Έλεγχος Πολλαπλών Υποθέσεων στην Επιλογή Μοντέλου	558
10.4	ΣΤΑΤΙΣΤΙΚΟΙ ΕΛΕΓΧΟΙ ΓΙΑ ΤΗΝ ΑΝΑΛΥΣΗ ΣΥΣΧΕΤΙΣΗΣ	559
10.4.1	Χρήση Στατιστικών Μοντέλων	559
10.4.2	Χρήση Μεθόδων Τυχαιοποίησης	563
10.5	ΣΤΑΤΙΣΤΙΚΟΙ ΕΛΕΓΧΟΙ ΓΙΑ ΤΗΝ ΑΝΑΛΥΣΗ ΣΥΣΤΑΔΩΝ	564
10.5.1	Παραγωγή Μίας Μηδενικής Κατανομής για Εσωτερικούς Δείκτες	564
10.5.2	Παραγωγή Μίας Μηδενικής Κατανομής για Εξωτερικούς Δείκτες	565
10.5.3	Εμπλουτισμός	566
10.6	ΣΤΑΤΙΣΤΙΚΟΙ ΕΛΕΓΧΟΙ ΓΙΑ ΤΗΝ ΑΝΙΧΝΕΥΣΗ ΑΝΩΜΑΛΙΩΝ	567
10.7	ΒΙΒΛΙΟΓΡΑΦΙΚΕΣ ΑΝΑΦΟΡΕΣ	569
10.8	ΑΣΚΗΣΕΙΣ	573

ΕΥΡΕΤΗΡΙΟ 579