



ΠΑΝΕΠΙΣΤΗΜΙΟ
ΠΑΤΡΩΝ
UNIVERSITY OF PATRAS

ΣΧΟΛΗ ΟΙΚΟΝΟΜΙΚΩΝ ΕΠΙΣΤΗΜΩΝ ΚΑΙ ΔΙΟΙΚΗΣΗΣ ΕΠΙΧΕΙΡΗΣΕΩΝ

ΤΜΗΜΑ ΔΙΟΙΚΗΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ
ΠΠΣ ΔΙΟΙΚΗΣΗΣ ΕΠΙΧΕΙΡΗΣΕΩΝ ΜΕΣΟΛΟΓΓΙ

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

ΕΦΑΡΜΟΓΗ ΤΕΧΝΙΚΩΝ ΕΞΟΡΥΞΗΣ ΓΝΩΣΗΣ
ΑΠΟ ΙΑΤΡΙΚΑ ΔΕΔΟΜΕΝΑ

Δημήτριος Καριωτάκης
Μαρία Βορύλλα

Μεσολόγγι 2022

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΑΤΡΩΝ

ΣΧΟΛΗ ΟΙΚΟΝΟΜΙΚΩΝ ΕΠΙΣΤΗΜΩΝ ΚΑΙ ΔΙΟΙΚΗΣΗΣ ΕΠΙΧΕΙΡΗΣΕΩΝ

ΤΜΗΜΑ ΔΙΟΙΚΗΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ
ΠΠΣ ΔΙΟΙΚΗΣΗΣ ΕΠΙΧΕΙΡΗΣΕΩΝ ΜΕΣΟΛΟΓΓΙ

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

ΕΦΑΡΜΟΓΗ ΤΕΧΝΙΚΩΝ ΕΞΟΡΥΞΗΣ ΓΝΩΣΗΣ
ΑΠΟ ΙΑΤΡΙΚΑ ΔΕΔΟΜΕΝΑ

Δημήτριος Καριωτάκης
Μαρία Βορούλλα

Επιβλέπων καθηγητής
Κωνσταντίνος Γιωτόπουλος
Επιτηρητές
Δημήτριος Μιχαλόπουλος
Νικόλαος Ζώτος

Μεσολόγγι 2022

UNIVERSITY OF PATRAS

SCHOOL OF ECONOMICS & BUSINESS

DEPARTMENT OF MANAGEMENT SCIENCE AND
TECHNOLOGY

**FORMER DEPARTMENT OF BUSINESS
ADMINISTRATION AT MESSOLONGHI**

THESIS

APPLICATION OF TECHNICAL KNOWLEDGE
DEVELOPMENT FROM MEDICAL DATA

Dimitrios Kariotakis

Maria Vorylla

Messolonghi 2022

Η έγκριση της πτυχιακής εργασίας από το Τμήμα Διοικητικής Επιστήμης και Τεχνολογίας του Πανεπιστημίου Πατρών, δεν υποδηλώνει απαραίτητως και αποδοχή των απόψεων του συγγραφέα εκ μέρους του Τμήματος.

Ευχαριστίες

Χωρίς την υποστήριξη και την παρουσία κάποιων ανθρώπων, δεν θα ήταν εφικτό να διεκπεραιωθεί η παρούσα πτυχιακή εργασία.

Πρώτα από όλους, θα θέλαμε να ευχαριστήσουμε τον επιβλέποντα καθηγητή μας Δρ. Κωνσταντίνο Γιωτόπουλο, του τμήματος Διοικητικής Επιστήμης και Τεχνολογίας, που ανέλαβε την πτυχιακή μας εργασία. Επίσης, όλους τους καθηγητές μας από το πρώην Τ.Ε.Ι. Δυτικής Ελλάδος στην Ιερά Πόλη Μεσολογγίου και ιδιαίτερα την Δρ. Αντωνία Στεφανή, τον κύριο Δημήτριο Ντόβα και τον Δρ. Ισίδωρο Περίκο. Επιπλέον, τον ενδοκρινολόγο κύριο Εμμανουήλ Βασιλείου, για την βοήθεια του όσον αφορά την ανάλυση μερικών μεταβλητών της νόσου του θειώδη.

Κλείνοντας, θέλουμε να ευχαριστήσουμε τους γονείς μας, όπως και την οικογένεια μας γενικότερα, καθώς και κάποιους ανθρώπους που δυστυχώς δεν είναι πια κοντά μας, αλλά παρόλ' αυτά η βοήθεια που μας προσέφεραν αυτά τα χρόνια ήταν πολύτιμη.

ΠΕΡΙΛΗΨΗ

Η εξόρυξη δεδομένων έχει πάρα πολλές πτυχές, οι οποίες δεν γίνεται να αποτυπωθούν σε μία και μόνο εργασία. Ο βασικός μας σκοπός σ' αυτή την εργασία είναι να μπορέσει ένας αρχάριος να ενημερωθεί, όσο το δυνατόν πιο εύκολα και κατανοητά, για βασικές λειτουργίες της εξόρυξης δεδομένων μέσα από παραδείγματα πάνω σε ένα συγκεκριμένο σύνολο δεδομένων.

Το σύνολο των δεδομένων που θα χρησιμοποιηθεί στην παρούσα εργασία είναι ιατρικά δεδομένα και σχετίζονται με την νόσο του θυρεοειδή. Στα επόμενα κεφάλαια θα ενημερωθείτε για πιο γενικούς όρους που σχετίζονται με την εξόρυξη, για προγράμματα εξόρυξης δεδομένων, για την εγκατάσταση αυτών, αναλόγως με το λογισμικό που έχει ο ηλεκτρονικός σας υπολογιστής, αλλά και για τις πληροφορίες που εξάγαμε, βήμα προς βήμα, για τη νόσο του θυρεοειδή, αναλύοντας 'τες μία προς μία.

ABSTRACT

There are many types of data mining facets, which can't be captured in a simple essay. Our main task in this essay is to enable a beginner to be informed, as easily and undestadably as possible, for basic functions of data mining following examples on a specific data set.

The information in this essay will have medical data, and are associated with the thyroid disease. In the following chapters you will learn about more general terms related to data mining, programs for data mining, and for the installation depending on the software your computer has, but also the information we extracted, step by step and one by one analyzing the thyroid disease.

ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ

ΠΕΡΙΛΗΨΗ.....	v
ABSTRACT	vi
ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ	viii
ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ	xi
ΚΑΤΑΛΟΓΟΣ ΕΙΚΟΝΩΝ	xii
ΚΑΤΑΛΟΓΟΣ ΔΙΑΓΡΑΜΜΑΤΩΝ	xviii
ΣΥΝΤΟΜΟΓΡΑΦΙΕΣ.....	xix
ΑΠΟΔΟΣΗ ΟΡΩΝ.....	xx
ΕΙΣΑΓΩΓΗ.....	xxii
1 ΕΞΟΥΥΞΗ ΔΕΔΟΜΕΝΩΝ	1
1.1 Τι είναι η εξόρυξη δεδομένων;.....	1
1.1.1 Ανάλυση των όρων του ορισμού.....	1
1.2 Πραγματικός στόχος εξόρυξης δεδομένων.	3
1.3 Ιστορική αναδρομή.....	4
1.4 Εφαρμογή της εξόρυξης δεδομένων στη σύγχρονη εποχή.....	4
1.4.1 Παραδείγματα εφαρμογής της εξόρυξης δεδομένων στο σήμερα.	5
1.5 Τεχνικές εξόρυξης γνώσης από βάσεις δεδομένων.....	5
1.6 Επικύρωση του αποτελέσματος της εξόρυξης δεδομένων.....	6
1.7 Μηχανική μάθηση.	7
1.8 Εξόρυξη δεδομένων και ανακάλυψη γνώσης.....	8
1.9 Συμπεράσματα.....	9
2 Αλγόριθμοι - Δέντρα Αποφάσεων	10
2.1 Αλγόριθμος J48	10
2.2 Αλγόριθμος Random Tree	12

2.3	Αλγόριθμος Random Forest	13
2.4	Αλγόριθμος REP Tree	14
3	Εργαλεία Εξόρυξης Δεδομένων	16
3.1	Machine Learning with Python	16
3.2	Scikit-learn	17
3.3	XGBoost	18
3.4	Python-recsys	19
3.5	MLlib	20
3.6	Eggplant.....	20
3.7	Mahout.....	21
3.8	Google Cloud TPU	22
3.9	WEKA	23
3.10	Σύγκριση Εργαλείων	23
4	Εργαλείο Εξόρυξης Δεδομένων Waikato Environment for Knowledge Analysis (WEKA). 25	
4.1	Περιγραφή WEKA.	25
4.2	Εγκατάσταση WEKA	25
4.2.1	Εγκατάσταση WEKA σε λειτουργικό σύστημα Windows	26
4.2.2	Εγκατάσταση WEKA σε λειτουργικό σύστημα Linux	30
4.3	Λειτουργίες WEKA.....	35
4.3.1	Weka Explorer.....	35
5	ΕΞΟΡΥΞΗ ΓΝΩΣΗΣ ΜΕ ΤΗΝ ΕΦΑΡΜΟΓΗ WEKA ΑΠΟ ΙΑΤΡΙΚΑ ΔΕΔΟΜΕΝΑ ΤΗΣ ΝΟΣΟΥ “ΘΥΡΕΟΕΙΔΗ”.....	40
5.1	Ανάλυση των ιατρικών δεδομένων	40
5.2	Ανάλυση αλγορίθμων που χρησιμοποιήθηκαν	71
5.2.1	Δέντρο απόφασης αλγόριθμου J48 με διακριτοποίηση τιμών	71

5.2.2	Δέντρο απόφασης αλγορίθμου J48 με διακριτοποίηση τιμών και επικύρωση..	77
5.2.3	Δέντρο απόφασης αλγορίθμου Random Tree με διακριτοποίηση τιμών	87
5.2.4	Δέντρο απόφασης αλγορίθμου Random Tree με διακριτοποίηση τιμών και επικύρωση	94
5.2.5	Δέντρο απόφασης αλγορίθμου Random Forest με διακριτοποίηση τιμών	104
5.2.6	Δέντρο απόφασης αλγορίθμου RandomForest με διακριτοποίηση τιμών και επικύρωση	111
5.2.7	Δέντρο απόφασης αλγορίθμου REP Tree με διακριτοποίηση τιμών	120
5.2.8	Δέντρο απόφασης αλγορίθμου REP Tree με διακριτοποίηση τιμών και επικύρωση	127
5.3	Συσχετίσεις Αλγορίθμων	136
ΣΥΜΠΕΡΑΣΜΑΤΑ		138
ΔΙΕΥΚΡΙΝΙΣΕΙΣ		139
Αναλυτική ανάλυση του αλγορίθμου J48		139
Αναλυτική ανάλυση του αλγορίθμου Random Tree		142
Αναλυτική ανάλυση του αλγορίθμου REP Tree		164
Βιβλιογραφία		167

ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ

Πίνακας 1: Παράδειγμα από ηλεκτρονική βιβλιοθήκη του weka.....	11
--	----

ΚΑΤΑΛΟΓΟΣ ΕΙΚΟΝΩΝ

Εικόνα 1: Λογότυπο WEKA	25
Εικόνα 2: Επίσημη ιστοσελίδα του weka.....	26
Εικόνα 3: Επιλογή λογισμικού windows.....	26
Εικόνα 4: Εγκατάσταση του Weka σε λειτουργικό σύστημα windows (1)	27
Εικόνα 5: Εγκατάσταση του Weka σε λειτουργικό σύστημα windows (2)	27
Εικόνα 6: Εγκατάσταση του Weka σε λειτουργικό σύστημα windows (3)	27
Εικόνα 7: Εγκατάσταση του Weka σε λειτουργικό σύστημα windows (4)	28
Εικόνα 8: Εγκατάσταση του Weka σε λειτουργικό σύστημα windows (5)	28
Εικόνα 9: Εγκατάσταση του Weka σε λειτουργικό σύστημα windows (6)	28
Εικόνα 10: Εγκατάσταση του Weka σε λειτουργικό σύστημα windows (7)	29
Εικόνα 11: Εγκατάσταση του Weka σε λειτουργικό σύστημα windows (8)	29
Εικόνα 12: Συντόμευση του weka στην επιφάνια εργασίας σας.....	30
Εικόνα 13: Επιλογή λογισμικού Linux.....	31
Εικόνα 14: Φάκελος weka-X-X-X	31
Εικόνα 15: Άνοιγμα στο τερματικό.....	32
Εικόνα 16: WEKA μέσω της ιστοσελίδας	32
Εικόνα 17: Εικονίδιο του weka	33
Εικόνα 18: WEKA μέσω του τερματικού	33
Εικόνα 19: Ubuntu Software	34
Εικόνα 20: Ubuntu Software weka.....	34
Εικόνα 21: Εγκατάσταση μέσω Ubuntu Software	35
Εικόνα 22: Weka Explorer	36
Εικόνα 23: Πίνακας προεπεξεργασίας	36
Εικόνα 24: Πίνακας Ταξινόμησης.....	37

Εικόνα 25: Visualize tree	39
Εικόνα 26: Στατιστική ανάλυση του χαρακτηριστικού «age»	41
Εικόνα 27: Στατιστική ανάλυση του χαρακτηριστικού «sex»	42
Εικόνα 28: Στατιστική ανάλυση του χαρακτηριστικού «on thyroxine»	43
Εικόνα 29: Στατιστική ανάλυση του χαρακτηριστικού «query on thyroxine»	44
Εικόνα 30: Στατιστική ανάλυση του χαρακτηριστικού «on antithyroid medication»	45
Εικόνα 31: Στατιστική ανάλυση του χαρακτηριστικού «sick»	46
Εικόνα 32: Στατιστική ανάλυση του χαρακτηριστικού «pregnant».....	47
Εικόνα 33: Στατιστική ανάλυση του χαρακτηριστικού «thyroid surgery»	48
Εικόνα 34: Στατιστική ανάλυση του χαρακτηριστικού « I131 treatment »	49
Εικόνα 35: Στατιστική ανάλυση του χαρακτηριστικού « query hypothyroid »	50
Εικόνα 36: Στατιστική ανάλυση του χαρακτηριστικού « query hyperthyroid ».....	51
Εικόνα 37: Στατιστική ανάλυση του χαρακτηριστικού « lithium »	52
Εικόνα 38: Στατιστική ανάλυση του χαρακτηριστικού « goitre »	53
Εικόνα 39: Στατιστική ανάλυση του χαρακτηριστικού « tumor »	54
Εικόνα 40: Στατιστική ανάλυση του χαρακτηριστικού « hypopituitary »	55
Εικόνα 41: Στατιστική ανάλυση του χαρακτηριστικού « psych »	56
Εικόνα 42: Στατιστική ανάλυση του χαρακτηριστικού « TSH measured ».....	57
Εικόνα 43: Στατιστική ανάλυση του χαρακτηριστικού « TSH »	58
Εικόνα 44: Στατιστική ανάλυση του χαρακτηριστικού « T3 measured »	59
Εικόνα 45: Στατιστική ανάλυση του χαρακτηριστικού « T3 »	60
Εικόνα 46: Στατιστική ανάλυση του χαρακτηριστικού « TT4 measured »	61
Εικόνα 47: Στατιστική ανάλυση του χαρακτηριστικού « TT4 ».....	62
Εικόνα 48: Στατιστική ανάλυση του χαρακτηριστικού « T4U measured »	63
Εικόνα 49: Στατιστική ανάλυση του χαρακτηριστικού « T4U »	64
Εικόνα 50: Στατιστική ανάλυση του χαρακτηριστικού « FTI measured »	65

Εικόνα 51: Στατιστική ανάλυση του χαρακτηριστικού « FTI ».....	66
Εικόνα 52: Στατιστική ανάλυση του χαρακτηριστικού « TBG measured »	67
Εικόνα 53: Στατιστική ανάλυση του χαρακτηριστικού « TBG ».....	68
Εικόνα 54: Στατιστική ανάλυση του χαρακτηριστικού « referral source »	69
Εικόνα 55: Στατιστική ανάλυση του χαρακτηριστικού « Class ».....	70
Εικόνα 56: Αλγόριθμος J48 -C 0,25 -M 2 με διακριτοποίηση 55%.....	71
Εικόνα 57: Αλγόριθμος J48 -C 0,25 -M 2 με διακριτοποίηση 60%.....	72
Εικόνα 58: Αλγόριθμος J48 -C 0,25 -M 2 με διακριτοποίηση 65%.....	73
Εικόνα 59: Αλγόριθμος J48 -C 0,25 -M 2 με διακριτοποίηση 70%.....	74
Εικόνα 60: Αλγόριθμος J48 -C 0,25 -M 2 με διακριτοποίηση 75%.....	75
Εικόνα 61: Αλγόριθμος J48 -C 0,25 -M 2 με διακριτοποίηση 80%.....	76
Εικόνα 62: Αλγόριθμος J48 -C 0,25 -M 2 με διακριτοποίηση 85%.....	77
Εικόνα 63: Αλγόριθμος J48 -C 0,25 -M 2 με επικύρωση 6.....	78
Εικόνα 64: Αλγόριθμος J48 -C 0,25 -M 2 με επικύρωση 6.....	79
Εικόνα 65: Αλγόριθμος J48 -C 0,25 -M 2 με επικύρωση 7.....	80
Εικόνα 66: Αλγόριθμος J48 -C 0,25 -M 2 με επικύρωση 8.....	81
Εικόνα 67: Αλγόριθμος J48 -C 0,25 -M 2 με επικύρωση 9.....	82
Εικόνα 68: Αλγόριθμος J48 -C 0,25 -M 2 με επικύρωση 10.....	83
Εικόνα 69: Αλγόριθμος J48 -C 0,25 -M 2 με επικύρωση 11.....	84
Εικόνα 70: Αλγόριθμος J48 -C 0,25 -M 2 με επικύρωση 12.....	85
Εικόνα 71: Αλγόριθμος Random Tree -K 0 -M 1,0 -V 0,001 -S 1 με διακριτοποίηση 55% .	88
Εικόνα 72: Αλγόριθμος Random Tree -K 0 -M 1,0 -V 0,001 -S 1 με διακριτοποίηση 60% .	89
Εικόνα 73: Αλγόριθμος Random Tree -K 0 -M 1,0 -V 0,001 -S 1 με διακριτοποίηση 65% .	90
Εικόνα 74: Αλγόριθμος Random Tree -K 0 -M 1,0 -V 0,001 -S 1 με διακριτοποίηση 70% .	91
Εικόνα 75: Αλγόριθμος Random Tree -K 0 -M 1,0 -V 0,001 -S 1 με διακριτοποίηση 75% .	92
Εικόνα 76: Αλγόριθμος Random Tree -K 0 -M 1,0 -V 0,001 -S 1 με διακριτοποίηση 80% .	93

Εικόνα 77: Αλγόριθμος Random Tree -K 0 -M 1,0 -V 0,001 -S 1 με διακριτοποίηση 85% .	94
Εικόνα 78: Αλγόριθμος Random Tree -K 0 -M 1,0 -V 0,001 -S 1 με επικύρωση 5	95
Εικόνα 79: Αλγόριθμος Random Tree -K 0 -M 1,0 -V 0,001 -S 1 με επικύρωση 6	96
Εικόνα 80: Αλγόριθμος Random Tree -K 0 -M 1,0 -V 0,001 -S 1 με επικύρωση 7	97
Εικόνα 81: Αλγόριθμος Random Tree -K 0 -M 1,0 -V 0,001 -S 1 με επικύρωση 8	98
Εικόνα 82: Αλγόριθμος Random Tree -K 0 -M 1,0 -V 0,001 -S 1 με επικύρωση 9	99
Εικόνα 83: Αλγόριθμος Random Tree -K 0 -M 1,0 -V 0,001 -S 1 με επικύρωση 10	100
Εικόνα 84: Αλγόριθμος Random Tree -K 0 -M 1,0 -V 0,001 -S 1 με επικύρωση 11	101
Εικόνα 85: Αλγόριθμος Random Tree -K 0 -M 1,0 -V 0,001 -S 1 με επικύρωση 12	102
Εικόνα 86: Αλγόριθμος RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1 με διακριτοποίηση 55%	105
Εικόνα 87: Αλγόριθμος RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1 με διακριτοποίηση 60%	106
Εικόνα 88: Αλγόριθμος RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1 με διακριτοποίηση 65%	107
Εικόνα 89: Αλγόριθμος RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1 με διακριτοποίηση 70%	108
Εικόνα 90: Αλγόριθμος RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1 με διακριτοποίηση 75%	109
Εικόνα 91: Αλγόριθμος RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1 με διακριτοποίηση 80%	110
Εικόνα 92: Αλγόριθμος RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1 με διακριτοποίηση 85%	111
Εικόνα 93: Αλγόριθμος RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1 με επικύρωση 5	112
Εικόνα 94: Αλγόριθμος RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1 με επικύρωση 6	113

Εικόνα 95: Αλγόριθμος RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0-V 0.001-S 1 με επικύρωση 7	114
Εικόνα 96: Αλγόριθμος RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0-V 0.001-S 1 με επικύρωση 8	115
Εικόνα 97: Αλγόριθμος RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0-V 0.001-S 1 με επικύρωση 9	116
Εικόνα 98: Αλγόριθμος RandomForest -P100 -I 100 -num-slots 1 -K 0-M 1.0-V 0.001-S 1 με επικύρωση 10	117
Εικόνα 99: Αλγόριθμος RandomForest -P100 -I 100 -num-slots 1 -K 0-M 1.0-V 0.001-S 1 με επικύρωση 11	118
Εικόνα 100: Αλγόριθμος RandomForest -P100 -I 100 -num-slots 1 -K 0-M 1.0-V 0.001-S 1 με επικύρωση 12	119
Εικόνα 101: Αλγόριθμος REPTree -M 2 -V 0,001 -N 3 -S 1 -L -1 -I 0.0 με διακριτοποίηση 55%.....	121
Εικόνα 102: Αλγόριθμος REPTree -M 2 -V 0,001 -N 3 -S 1 -L -1 -I 0.0 με διακριτοποίηση 60%.....	122
Εικόνα 103: Αλγόριθμος REPTree -M 2 -V 0,001 -N 3 -S 1 -L -1 -I 0.0 με διακριτοποίηση 65%.....	123
Εικόνα 104: Αλγόριθμος REPTree -M 2 -V 0,001 -N 3 -S 1 -L -1 -I 0.0 με διακριτοποίηση 70%.....	124
Εικόνα 105: Αλγόριθμος REPTree -M 2 -V 0,001 -N 3 -S 1 -L -1 -I 0.0 με διακριτοποίηση 75%.....	125
Εικόνα 106: Αλγόριθμος REPTree -M 2 -V 0,001 -N 3 -S 1 -L -1 -I 0.0 με διακριτοποίηση 80%.....	126
Εικόνα 107: Αλγόριθμος REPTree -M 2 -V 0,001 -N 3 -S 1 -L -1 -I 0.0 με διακριτοποίηση 85%.....	127
Εικόνα 108: Αλγόριθμος REPTree -M 2 -V 0,001 -N 3 -S 1 -L -1 -I 0.0 με επικύρωση 5 ..	128
Εικόνα 109: Αλγόριθμος REPTree -M 2 -V 0,001 -N 3 -S 1 -L -1 -I 0.0 με επικύρωση 6 ..	129

Εικόνα 110: Αλγόριθμος REPTree -M 2 -V 0,001 -N 3 -S 1 -L -1 -I 0.0 με επικύρωση 7 ..	130
Εικόνα 111: Αλγόριθμος REPTree -M 2 -V 0,001 -N 3 -S 1 -L -1 -I 0.0 με επικύρωση 8 ..	131
Εικόνα 112: Αλγόριθμος REPTree -M 2 -V 0,001 -N 3 -S 1 -L -1 -I 0.0 με επικύρωση 9 ..	132
Εικόνα 113: Αλγόριθμος REPTree -M 2 -V 0,001 -N 3 -S 1 -L -1 -I 0.0 με επικύρωση 11	133
Εικόνα 114: Αλγόριθμος REPTree -M 2 -V 0,001 -N 3 -S 1 -L -1 -I 0.0 με επικύρωση 12	134
Εικόνα 115: Αναλυτική ανάλυση του αλγοριθμου J48.....	142
Εικόνα 116: Αναλυτική ανάλυση του αλγοριθμου Random Tree.....	164
Εικόνα 117: Αναλυτική ανάλυση του αλγοριθμου REP Tree.....	166

ΚΑΤΑΛΟΓΟΣ ΔΙΑΓΡΑΜΜΑΤΩΝ

Διάγραμμα 1: Διάγραμμα αποτελεσμάτων αλγορίθμου J48	86
Διάγραμμα 2: Διάγραμμα αποτελεσμάτων αλγορίθμου J48 ταξινομημένες.....	87
Διάγραμμα 3: Διάγραμμα αποτελεσμάτων αλγορίθμου Random Tree.....	103
Διάγραμμα 4: Διάγραμμα αποτελεσμάτων αλγορίθμου Random Tree ταξινομημένες	104
Διάγραμμα 5: Διάγραμμα αποτελεσμάτων αλγορίθμου RandomForest	119
Διάγραμμα 6: Διάγραμμα αποτελεσμάτων αλγορίθμου RandomForest ταξινομημένες.....	120
Διάγραμμα 7: Διάγραμμα αποτελεσμάτων αλγορίθμου REPTree	135
Διάγραμμα 8: Διάγραμμα αποτελεσμάτων αλγορίθμου REPTree ταξινομημένες	136
Διάγραμμα 9: Διάγραμμα αποτελεσμάτων αλγορίθμων με τα μεγαλύτερα ποσοστά «Σωστά σύμφωνα με το πρότυπο»	137
Διάγραμμα 10: Διάγραμμα αποτελεσμάτων αλγορίθμων ταξινομημένες.....	137

ΣΥΝΤΟΜΟΓΡΑΦΙΕΣ

WEKA: Waikato Environment for Knowledge Analysis

ARFF: Attribute-Relation File Format

ASCII: American Standard Code for Information Interchange

ID3: Iterative Dichotomiser 3

RIPPER: Repeated incremental pruning to produce error reduction

ΑΠΟΔΟΣΗ ΟΡΩΝ

WEKA	Waikato(πόλη) Περιβάλλον για Ανάλυση Γνώσης
ARFF	Μορφή αρχείου συσχετισμού χαρακτηριστικών
ASCII	Αμερικανικός Πρότυπος Κώδικας για Ανταλλαγή Πληροφοριών
Missing	Αγνοούμενα
Unique	Μοναδικά
Decision Tree	Δέντρο αποφάσεων
ΕΦΓ	Επεξεργασία φυσικής γλώσσας
Filters	Φίλτρα
Classifiers	Ταξινομητές
RIPPER	Επαναλαμβανόμενο κλιμακωτό κλάδεμα για παραγωγή μείωσης σφάλματος
Clusterers	Συμπλέγματα
Associations	Συσχετίσεις
Attribute Selection	Επιλογή χαρακτηριστικών
Preprocess Panel	Πίνακας προεπεξεργασίας
Classifier Panel	Πίνακας ταξινόμησης
Cluster Panel	Πίνακας Ομαδοποίησης
Associate Panel	Πίνακας Συνεργάτης
Select Attributes Panel	Πίνακας Επιλογής χαρακτηριστικών
Visualize Panel	Πίνακας Οπτικοποίησης
Training set	Σετ προπόνησης
Supplied test set	Παρέχεται πακέτα δοκιμών
Cross-validation	Διασταυρωμένη επικύρωση

Percentage split

Ποσοστό διαχωρισμού

ΕΙΣΑΓΩΓΗ

Η εξόρυξη πληροφορίας (data mining) από ένα σύνολο δεδομένων, είναι η διαδικασία εύρεσης χρήσιμων μοτίβων στα δεδομένα. Το αντικείμενο της εξόρυξης πληροφορίας είναι να χρησιμοποιήσει τα μοτίβα που ανακαλύπτει, για να βοηθήσει στην εξήγηση της τρέχουσας συμπεριφοράς, ή να προβλέψει μελλοντικά ενδεχόμενα. Μπορούμε να μελετήσουμε πολλές πλευρές της διαδικασίας αυτής και κάποιες από αυτές είναι οι κάτωθι:

- Συλλογή και αποθήκευση δεδομένων.
- Επιλογή και προετοιμασία δεδομένων.
- Δημιουργία και έλεγχος μοντέλων.
- Ερμηνεία και έλεγχος εγκυρότητας αποτελεσμάτων.
- Εφαρμογή μοντέλων.

Μία μόνο εργασία δεν μπορεί να αναφερθεί αναλυτικά σε όλους τους τομείς της διαδικασίας εξόρυξης πληροφορίας από ένα σύνολο δεδομένων. Αν και παραθέτουμε λεπτομερίες για κάποιες από τις πλευρές της εξόρυξης και της ανακάλυψης γνώσης, ο πρωταρχικός μας σκοπός είναι να σας βοηθήσουμε να καταλάβετε καλύτερα τη διαδικασία της εξόρυξης πληροφορίας από ένα σύνολο δεδομένων.

1 ΕΞΟΡΥΞΗ ΔΕΔΟΜΕΝΩΝ

1.1 Τι είναι η εξόρυξη δεδομένων;

Ο πολύ μεγάλος όγκος διαθέσιμων δεδομένων, έχει ανάγκη μία μέθοδο, ώστε να μπορέσει να αναλυθεί. Η μέθοδος που θα μελετήσουμε είναι η «Εξόρυξη Δεδομένων». Κάνοντας χρήση πολλών αποδοτικών τεχνικών, μπορούμε να αναλύουμε τεράστιες συλλογές από δεδομένα και να εξάγουμε από αυτά χρήσιμες πληροφορίες.

Κάθε είδος φόρμας που εμπεριέχει μια τεράστια ποσότητα δεδομένων ή κάνει επεξεργασία μεγάλου όγκου δεδομένων, ονομάζεται εξόρυξη δεδομένων, δηλαδή κάνει συλλογή και εξαγωγή δεδομένων, warehouse, καθώς και ανάλυση δεδομένων και ανάλυση στατιστικής. Συμπεριλαμβάνει κάθε είδος συστήματος τεχνητής νοημοσύνης, μηχανικής εκμάθησης και επιχειρηματικής ευφυΐας. Η λέξη κλειδί στον όρο είναι η ανακάλυψη, δηλαδή η ανίχνευση κάτι καινούριου.

Ως «Εξόρυξη Δεδομένων» ορίζεται η εξεύρεση μιας πληροφορίας ή ενός προτύπου, μη προφανή και πιθανόν χρήσιμο, από μεγάλες βάσεις δεδομένων.

1.1.1 Ανάλυση των όρων του ορισμού.

- I. Η **πληροφορία** είναι μία διάταξη συνδυασμού διαφοροποιήσεων μέσα από εντροπικό πληροφοριακό εύρος. Στην κοινή γλώσσα, η πληροφορία αποτελείται από γεγονότα και απόψεις που ή προσφέρονται ή λαμβάνονται, είτε από ζωντανούς οργανισμούς, είτε από διάφορα μέσα μαζικής επικοινωνίας, είτε από ηλεκτρονικούς υπολογιστές μέσω της χρήσης του διαδικτύου, είτε από κάθε πάσης φύσεως φαινόμενα στο περιβάλλον. Τέλος, μία δεν αποτελεί απαραίτητα ένα δεδομένο και ούτε απαιτεί κάποια συγκεκριμένη γνώση.
- II. Ως **Βάση δεδομένων** ορίζεται μία συλλογή συστηματικά μορφοποιημένων δεδομένων που σχετίζονται μεταξύ τους, για τα οποία η ανάκτηση δεδομένων είναι εφικτή μέσω της αναζήτησης. Ο όρος βάσεις δεδομένων στην καθημερινή χρήση, αναφέρεται σε διακριτές και οργανωμένες συλλογές σχετιζόμενων μεταξύ τους δεδομένων, τόσο ηλεκτρονικά όσο και ψηφιακά αποθηκευμένων σε ένα λογισμικό.

Επίσης, μία βάση δεδομένων παρέχει τη δυνατότητα γρήγορης άντλησης και ανανέωσης των δεδομένων της.

III. Ως **αλγόριθμο** ονομάζουμε μία σειρά από εντολές που είναι σαφείς, έχουν αρχή και τέλος, και σκοπός τους είναι η επίλυση ενός προβλήματος. Η σημερινή της σημασία οφείλεται στη γρήγορη ανάπτυξη των ηλεκτρονικών υπολογιστών στα μέσα του 20ου αιώνα.

IV. Η **κατηγοριοποίηση** είναι μία τεχνική της εξόρυξης δεδομένων. Ο στόχος της είναι η ανάπτυξη ενός μοντέλου, το οποίο στη πορεία θα μπορεί να χρησιμοποιηθεί για την κατηγοριοποίηση δεδομένων. Στην βιβλιογραφία ο όρος κατηγοριοποίηση συναντάται και ως ταξινόμηση.

Η κατηγοριοποίηση μπορεί να περιγραφεί ως μία διαδικασία δύο βημάτων:

1. Εκμάθηση (Learning): προσδιορίζεται το μοντέλο με βάση τα δεδομένα εκπαίδευσης. Αυτά για να σχηματιστεί το μοντέλο, αναλύονται από ένα αλγόριθμο κατηγοριοποίησης. Επειδή τα δεδομένα εκπαίδευσης ανήκουν σε μία προκαθορισμένη γνωστή κατηγορία, η κατηγοριοποίηση αποτελεί μέθοδο εποπτευομένης μάθησης. Τέλος, το μοντέλο, αναπαρίσταται με τη μορφή δέντρων απόφασης ή μαθηματικών τύπων ή κανόνων κατηγοριοποίησης.
2. Κατηγοριοποίηση (Classification): Εφόσον δημιουργηθεί το μοντέλο, στη συνέχεια έρχεται η αξιολόγηση του. Ο υπολογισμός της ακρίβειας του μοντέλου θα επιτευχθεί μόνο με τη χρήση δοκιμαστικών δεδομένων. Εφόσον πρώτα κατηγοριοποιήσει τα δοκιμαστικά δεδομένα, στη συνέχεια η κατηγορία που σχηματίστηκε θα συγκριθεί με την πρόβλεψη που έγινε με τα δεδομένα εκπαίδευσης. Τα δεδομένα εκπαίδευσης είναι εντελώς ανεξάρτητα από τα δεδομένα δοκιμής. Ο υπολογισμός της ακρίβειας του

μοντέλου γίνεται από το ποσοστό των δειγμάτων δοκιμής, τα οποία κατηγοριοποιήθηκαν σωστά.

Εάν το μοντέλο κριθεί αποδεκτό, τότε μπορεί να χρησιμοποιηθεί στην άγνωστη κατηγοριοποίηση μελλοντικών δειγμάτων δεδομένων.

V. Η **στατιστική** ήταν μία μεθοδική μαθηματική τεχνική ενώ σήμερα είναι μία μεθοδική μαθηματική επιστήμη η οποία προσπαθεί να εξαγάγει έγκυρη γνώση, κάνοντας χρήση διαφόρων εμπειρικών δεδομένων παρατήρησης ή πειράματος. Η συλλογή, η ταξινόμηση, η επεξεργασία, η παρουσίαση, η ανάλυση και η ερμηνεία διαφόρων δεδομένων είναι το κύριο αντικείμενο έρευνας και μελέτης της και απώτερος στόχος της είναι η εξαγωγή ασφαλών συμπερασμάτων με σκοπό τη λήψη ορθών αποφάσεων.

VI. Ο όρος **τεχνητή νοημοσύνη** διαιρείται στην συμβολική τεχνητή νοημοσύνη και στην υποσυμβολική τεχνητή νοημοσύνη. Η συμβολική είναι η εξομοίωση της ανθρώπινης νοημοσύνης αλγοριθμικά, χρησιμοποιώντας λογικούς κανόνες υψηλού επιπέδου και σύμβολα. Η υποσυμβολική τεχνητή νοημοσύνη, προσπαθεί να αναπαράγει την ανθρώπινη ευφυΐα κάνοντας χρήση στοιχειωδών αριθμητικών μοντέλων, τα οποία συνθέτουν νοήμονες συμπεριφορές με τη διαδοχική αυτό-οργάνωση απλούστερων δομικών συστατικών.

1.2 Πραγματικός στόχος εξόρυξης δεδομένων

Πραγματικός στόχος της εξόρυξης δεδομένων είναι η ημιαυτόματη ή αυτόματη ανάλυση τεράστιων ποσοτήτων δεδομένων, με σκοπό την εξαγωγή ενός ενδιαφέροντος προτύπου, το οποίο δεν ήταν γνωστό μέχρι εκείνη τη στιγμή.

Για παράδειγμα, πολλαπλά σύνολα δεδομένων, τα οποία μπορούν να χρησιμοποιηθούν ώστε να εξασφαλίσουν περισσότερο ακριβή αποτελέσματα από ένα σύστημα υποστήριξης αποφάσεων, μπορούν να προσδιοριστούν από την εξόρυξη δεδομένων.

1.3 Ιστορική αναδρομή

Εδώ και αιώνες γίνεται εξαγωγή προτύπων από δεδομένα. Οι παλινδρόμηση και η θεωρία “Bayes” ήταν οι πρώτες μέθοδοι για τον προσδιορισμό προτύπων. Η ευρεία διαθεσιμότητα και ο πολλαπλασιασμός των υπολογιστών, καθώς και η εξέλιξη της τεχνολογίας, έχουν αυξήσει την πολυπλοκότητα και τον όγκο συλλογής δεδομένων. Αυτό έχει ως αποτέλεσμα, την αύξηση στη ζήτηση πιο αποτελεσματικών και αποδοτικών χειρισμών επεξεργασίας δεδομένων, ώστε να γίνουν χρήσιμα και κατανοητά σε όλους τους χρήστες. Έτσι, γρήγορα η αυτόματη επεξεργασία δεδομένων αντικατέστησε την χειρωνακτική ανάλυση αυτών. Σε αυτό συνέβαλαν πολλές ανακαλύψεις πάνω στον τομέα των υπολογιστών. Κάποιες από αυτές είναι η συσταδοποίηση, τα δέντρα απόφασης, οι γενετικοί αλγόριθμοι και η μηχανή υποστήριξης διανυσμάτων. Πλέον, ως ρίζες της εξόρυξης δεδομένων θεωρούμε την μηχανική μάθηση και την στατιστική.

1.4 Εφαρμογή της εξόρυξης δεδομένων στη σύγχρονη εποχή

Η εξάπλωση του Internet έγινε λόγω της εξέλιξης της τεχνολογίας. Η πρόσβαση στο Internet, με το πέρασμα του χρόνου, έγινε προσιτή σε ολοένα και περισσότερους ανθρώπους. Αυτό, οδήγησε στην ανάπτυξη περισσότερων ιστότοπων καθώς και στη χρήση βάσεων δεδομένων, ώστε να γίνει η αποθήκευση αυτών. Τα πρώτα άλματα στις απαιτήσεις και τις ανάγκες, τόσο για διαχείριση μεγάλου όγκου δεδομένων όσο και για αποθήκευση αυτών, έγιναν υπαρκτά με τη δημιουργία εμπορικών και κοινωνικών ιστοσελίδων.

Πλέον, το πλήθος των διαθέσιμων δεδομένων αυξάνεται εκθετικά κάθε μέρα. Η μείωση στη δυσκολία και στο κόστος της συλλογής και της αποθήκευσης των δεδομένων, συνετέλεσε σημαντικά στην ανάπτυξη του πεδίου αυτού. Ο πολύ μεγάλος όγκος δεδομένων, ο οποίος συσσωρεύεται στις αποθήκες δεδομένων και στις βάσεις δεδομένων, δεν μπορεί να αξιοποιηθεί όπως είναι, πρέπει πρώτα να γίνουν κάποιες ενέργειες που θα δομήσουν κατάλληλα τα δεδομένα, ώστε στη συνέχεια να μπορούν να αξιοποιηθούν.

Καθημερινά, συναντάμε σχεδόν παντού την εξόρυξη δεδομένων. Μεγάλη χρησιμότητα έχει σε πολύ-καταστήματα ή αλυσίδες καταστημάτων, σε συναλλαγές μεταξύ τραπεζών και σε κοινωνικά δίκτυα. Επίσης, είναι πολύ χρήσιμη για τηλεσκοπία στον ουρανό, για

απομακρυσμένους αισθητήρες σε δορυφόρους, για επιστημονικές προσομοιώσεις, που παράγουν terabytes δεδομένων και για Microarrays τα οποία παράγουν γονιδιακά δεδομένα.

1.4.1 Παραδείγματα εφαρμογής της εξόρυξης δεδομένων στο σήμερα

- I. Ιατρική: η εξόρυξη δεδομένων χρησιμοποιείται ευρέως στους τομείς της ιατρικής, όπως στη βιοϊατρική, στο DNA, καθώς και στη γενετική και στη φαρμακευτική. Η εξόρυξη δεδομένων είναι σημαντικό εργαλείο το οποίο μπορεί να βοηθήσει στην βελτίωση της διάγνωσης, της πρόληψης και της θεραπείας των ασθενειών γενικότερα.

- II. Οικονομία: Τα οικονομικά δεδομένα συλλέγονται κυρίως διάφορους οικονομικούς οργανισμού, όπως για παράδειγμα από τις τράπεζες. Τα δεδομένα αυτά συνήθως έχουν υψηλή ποιότητα, είναι αξιόπιστα και ολοκληρωμένα και απαιτούν συστηματική μέθοδο για την ανάλυσή τους. Η εξόρυξη δεδομένων συνεισφέρει στην επιστήμη της οικονομίας όσον αφορά την συλλογή και την κατανόηση των δεδομένων, την βελτίωση αυτών, καθώς και την δημιουργία και την εκτίμηση ενός μοντέλου όπως και την ανάπτυξη αυτού.

- III. Τηλεπικοινωνία: Σήμερα, οι τηλεπικοινωνιακές υπηρεσίες έχουν επεκταθεί από τοπικές και μεγάλης απόστασης τηλεπικοινωνίες, στην χρήση φαξ, συσκευών τηλεειδοποίησης, κινητά τηλέφωνα και ηλεκτρονικά ταχυδρομεία. Εξαιτίας όλων αυτών των εξελίξεων, για να μπορέσουν να δουλέψουν αποτελεσματικά αυτές οι τεχνολογίες, οι τεχνικές της εξόρυξης δεδομένων ενσωματώνονται σε αυτές τις τεχνολογίες για να παράγουν αποδοτικά αποτελέσματα.

1.5 Τεχνικές εξόρυξης γνώσης από βάσεις δεδομένων

Η πολύ μεγάλη πρόοδος της πληροφορικής, αύξησε τη μαζική συλλογή και αποθήκευση δεδομένων σε όλους τους τομείς της ανθρώπινης δραστηριότητας. Η τεράστια πρόοδος όλων των μεγεθών και τύπων βάσεων δεδομένων, είναι ενδεικτική της ικανότητας συλλογής δεδομένων, αλλά ταυτόχρονα αυξάνει την ανάγκη για καλύτερες μεθόδους ανάλυσής τους. Η ανθρώπινη δυνατότητα, για τη διαχείριση των διαθέσιμων δεδομένων, δεν επαρκεί, όμως η τεχνολογία, δίνει πολλές δυνατότητες μεγιστοποίησης της ωφέλειας από τα δεδομένα, με οικονομικό και αρκετά γρήγορο τρόπο.

Η εξόρυξη δεδομένων περιλαμβάνει κάποιες από τις ακόλουθες τεχνικές διαδικασίες:

- Ανίχνευση ανωμαλιών: είναι ο προσδιορισμός ασυνήθιστων εγγραφών δεδομένων, τα οποία μπορεί να παρουσιάζουν λάθη στα δεδομένα ή κάποιο γενικό ενδιαφέρον, τα οποία απαιτούν περαιτέρω έρευνα.
- Κανόνες συσχέτισης: είναι οι αναζητήσεις για σχέσεις μεταξύ των μεταβλητών.
- Συσταδοποίηση: είναι η διαδικασία ανακάλυψης δομών και ομάδων ανάμεσα σε δεδομένα που είναι "παρόμοια", χωρίς να χρησιμοποιούνται γνωστές δομές στα δεδομένα.
- Κατηγοριοποίηση: είναι η διαδικασία γενίκευσης γνωστών δομών, ώστε να εφαρμοστούν πάνω σε νέα δεδομένα.
- Παλινδρόμηση: Προσπαθεί να βρει τη συνάρτηση με τα λιγότερα λάθη, η οποία μοντελοποιεί τα δεδομένα.

1.6 Επικύρωση του αποτελέσματος της εξόρυξης δεδομένων

Η επικύρωση των προτύπων που εξήχθησαν από τους αλγορίθμους της εξόρυξης δεδομένων που απευθύνονται σε ευρύτερο σύνολο δεδομένων, είναι το τελικό βήμα της ανακάλυψης γνώσης από δεδομένα. Δεν είναι απαραίτητα έγκυρα, όλα τα πρότυπα που βρέθηκαν. Είναι σύνηθες οι αλγόριθμοι της εξόρυξης δεδομένων να βρίσκουν διάφορα πρότυπα σε ένα σύνολο εκπαίδευσης, τα οποία όμως δεν υπάρχουν στο γενικό σύνολο δεδομένων. Αυτό ονομάζεται υπερφόρτωση και για να ξεπεραστεί αυτό, κατά τη διάρκεια της εκτίμησης χρησιμοποιείται ένα δοκιμαστικό σύνολο δεδομένων, στο οποίο δεν έχει εφαρμοστεί κανένας αλγόριθμος της εξόρυξης δεδομένων. Όλα τα πρότυπα που προέκυψαν, εφαρμόζονται σε αυτό το δοκιμαστικό σύνολο και το αποτέλεσμα που προκύπτει συγκρίνεται με το επιθυμητό αποτέλεσμα.

Στην περίπτωση που τα πρότυπα ανταποκρίνονται με τα επιθυμητά κριτήρια, το τελικό στάδιο είναι να ερμηνευτούν τα πρότυπα και να τα μετατρέψουμε σε γνώση. Στην αντίθετη περίπτωση που τα πρότυπα δεν ανταποκρίνονται με τα επιθυμητά κριτήρια, τότε είναι απαραίτητο να ξανά εκτιμηθεί και βέβαια να αλλαχθεί η προ-επεξεργασία και η εξόρυξη δεδομένων.

1.7 Μηχανική μάθηση

Στις αρχές της δεκαετίας του 1980 εμφανίστηκε η Μηχανική Μάθηση και αποτελεί έναν από τους βασικότερους τομείς της τεχνητής νοημοσύνης, καθώς και ένα επιστημονικό πεδίο, το οποίο μελετά τη σχεδίαση υπολογιστικών προγραμμάτων, τα οποία είναι ικανά να μαθαίνουν μέσω προηγούμενης γνώσης και εμπειρίας, ώστε να βελτιώνεται η απόδοσή τους.

Η έννοια της Μηχανικής Μάθησης ταυτίζεται με την απόκτηση γνώσης και εμπειρίας σε έναν τομέα. Κάθε νοήμων οντότητα, μπορεί να μαθαίνει, και αυτό έχει ως αποτέλεσμα, ή να αποκτά επιπλέον γνώση, ή να βελτιώνει προηγούμενη γνώση.

Το σύστημα για να εκτελέσει σωστά την επιθυμητή εργασία εφαρμόζεται στη γενική βάση δεδομένων και η διαδικασία αυτή διαιρείται σε δύο φάσεις:

1. Είναι η φάση της εκπαίδευσης, όπου χρησιμοποιούνται δεδομένα δειγματοληψίας, για τη δημιουργία αντιπροσωπευτικού μοντέλου.
2. Είναι η φάση του ελέγχου, όπου εφαρμόζεται το αντιπροσωπευτικό μοντέλο στα υπόλοιπα δεδομένα.

Έχουν αναπτυχθεί πολλές τεχνικές μηχανικής μάθησης, οι οποίες χρησιμοποιούνται ανάλογα με τη φύση του προβλήματος και κατατάσσονται σε τρία είδη:

- Μάθηση με επίβλεψη ή μάθηση με παραδείγματα, στην οποία το σύστημα τροφοδοτείται με διάφορα παραδείγματα αντικειμένων, που ανήκουν σε μία κατηγορία. Το ίδιο το σύστημα, καλείται να ανακαλύψει τις κοινές ιδιότητες των αντικειμένων αυτών.
- Μάθηση χωρίς επίβλεψη ή μάθηση από παρατήρηση, όπου μόνο του το σύστημα, βασισμένο στις δικές του ιδιότητες, καλείται να ανακαλύψει κατηγορίες αντικειμένων.
- Ενισχυτική μάθηση, όπου ένα πρόγραμμα υπολογιστή αλληλεπιδρά με ένα δυναμικό περιβάλλον, στο οποίο πρέπει να επιτευχθεί ένας συγκεκριμένος στόχος, χωρίς κάποιος να το ενημερώνει για το αν έχει φτάσει κοντά στο στόχο του.
- Ημί-επιτηρούμενη μάθηση, όπου ο χρήστης δίνει ένα σύνολο εκπαίδευσης με πολλά από τα αποτελέσματα να λείπουν και βρίσκεται Ανάμεσα στην επιτηρούμενη και την μη επιτηρούμενη μάθηση.

- Η διαδικασία εκμάθησης, η οποία μαθαίνει στην μηχανή να αναπτύσσει τις δικές της επαγωγικές μεθόδους, βασιζόμενο στην προηγούμενη εμπειρία.
- Η αναπτυξιακή μάθηση, η οποία έχει αναπτυχθεί για την εκμάθηση από ρομπότ. Δημιουργεί τη δική της ακολουθία μαθησιακών καταστάσεων, ώστε το ρομπότ να αποκτήσει ποικιλία δεξιοτήτων.

1.8 Εξόρυξη δεδομένων και ανακάλυψη γνώσης

Η εξόρυξη γνώσης από βάσεις δεδομένων, είναι μία νέα δυναμική τεχνολογία, η οποία βοηθάει τις επιχειρήσεις να εστιάσουν στη σημαντική πληροφορία, που βρίσκεται μέσα στις αποθήκες δεδομένων τους. Οι ικανότερες τεχνικές της, αναζητούν και βρίσκουν λεπτομερειακά και γρήγορα, βάσεις δεδομένων, για την αναζήτηση κρυμμένων προτύπων. Επομένως καταλήγουμε στο ότι η εξόρυξη γνώσης, είναι η διαδικασία εξαγωγής κρυμμένης πληροφορίας, από μεγάλες βάσεις δεδομένων. Εμπεριέχει πολλούς διαφορετικούς αλγόριθμους, ώστε να εκπληρωθούν διαφορετικές εργασίες, και οι οποίοι προσπαθούν να ταιριάξουν ένα μοντέλο στα δεδομένα. Επίσης, οι αλγόριθμοι καθορίζουν ένα μοντέλο, το οποίο είναι το πλησιέστερο στα χαρακτηριστικά των δεδομένων που εξετάζονται.

Οι αλγόριθμοι εξόρυξης γνώσης αποτελούνται από τρία μέρη:

1. Μοντέλο: Ο σκοπός του αλγορίθμου, είναι να ταιριάξει το μοντέλο στα δεδομένα.
2. Προτίμηση: χρησιμοποιούνται συγκεκριμένα κριτήρια, ώστε να ταιριάξει ένα μοντέλο έναντι ενός άλλου.
3. Αναζήτηση: Όλοι οι αλγόριθμοι, για να κάνουν αναζήτηση απαιτούν μια τεχνική.

Το μοντέλο που δημιουργείται είναι είτε προβλεπτικό, είτε περιγραφικό. Κάθε προβλεπτικό μοντέλο κάνει μία πρόβλεψη για τις τιμές των δεδομένων, κάνοντας χρήση γνωστών αποτελεσμάτων, τα οποία έχει βρει από άλλα δεδομένα. Η υλοποίηση μιας πρόβλεψης, γίνεται με βάση τη χρήση ιστορικών δεδομένων. Οι εργασίες εξόρυξης γνώσης από δεδομένα περιλαμβάνουν κατηγοριοποίηση, παλινδρόμηση, ανάλυση χρονολογικών σειρών και πρόβλεψη. Στόχος της εξόρυξης δεδομένων, είναι η ανακάλυψη νέας και χρήσιμης πληροφορίας σε βάσεις δεδομένων. Τα μέσα ή οι μέθοδοι για την επίτευξη του στόχου αυτού, μεταξύ άλλων, είναι η Στατιστική Ανάλυση, η Μηχανική Μάθηση, οι Βάσεις Δεδομένων, η

Αναγνώριση Προτύπων και η Γραφική Οπτικοποίηση. Όλες αυτές οι μέθοδοι χρησιμοποιούνται, για την εύρεση δομών, συσχετίσεων και προτύπων μέσα σε μεγάλες και ταυτόχρονα διαρκώς αυξανόμενες βάσεις δεδομένων.

1.9 Συμπεράσματα

Οι οικονομικές συναλλαγές σε συνδυασμό με τις μεγάλες αλλαγές τόσο στην υιοθέτηση όσο και στην χρήση των νέων τεχνολογιών, είτε σε μεγάλες είτε σε μικρές επιχειρήσεις, μας οδηγεί στη συγκέντρωση τεράστιου αριθμού δεδομένων. Γι' αυτό λοιπόν, ο αναλυτής έχει την ευθύνη να αναλύσει όλες αυτές τις συναλλαγές, καθώς και να εντοπίσει τόσο τις απάτες όσο και τα λάθη, μέσα σε αυτές. Οι τάσεις των επιχειρήσεων αλλάζουν συνεχώς, έτσι λοιπόν, είναι πολύ δύσκολο να επεξεργαστείς αλλά και να αναλύσεις τα δεδομένα με κάποια παλαιά μέθοδο και αυτό μας οδηγεί σε διάφορους περιορισμούς. Οι περιορισμοί αυτοί μας οδηγούν στην εκμετάλλευση των εργαλείων της εξόρυξης για περισσότερο αξιόπιστα και καλύτερα αποτελέσματα.

2 Αλγόριθμοι - Δέντρα Αποφάσεων

Σε αυτό το κεφάλαιο θα αναλύσουμε τους πιο σημαντικούς αλγορίθμους-δέντρα αποφάσεων (decision tree), βλέποντας μια πιο γενική-θεωρητική όψη τους. Γίνεται μεγαλύτερη ανάλυση αυτών, με παραδείγματα, σε επόμενο κεφάλαιο (5.2 Ανάλυση αλγορίθμων που χρησιμοποιήθηκαν).

Ένα δέντρο απόφασης χρησιμοποιείται ως προγνωστικό μοντέλο, το οποίο αντιστοιχίζει παρατηρήσεις, ως προς ένα στοιχείο, με συμπεράσματα, ως προς την τιμή-στόχο του αντικειμένου. Όλα τα Δέντρα Αποφάσεων παρακάτω χρησιμοποιούνται στην εφαρμογή Weka.

2.1 Αλγόριθμος J48

Το J48 είναι μια εφαρμογή Java, ανοιχτού κώδικα και χρησιμοποιείτε αποκλειστικά στο εργαλείο εξόρυξης δεδομένων Weka, το οποίο βασίζεται στον αλγόριθμο **C4.5** και μπορεί να δημιουργήσει δέντρα αποφάσεων με δύο τρόπους. Ένας τρόπος είναι σε κατάσταση δέσμης, όπου το πρόγραμμα δημιουργεί ένα μόνο δέντρο, χρησιμοποιώντας όλα τα διαθέσιμα δεδομένα, από ένα σύνολο δεδομένων εκπαίδευσης, κάνοντας χρήση της έννοιας της εντροπίας πληροφοριών¹. Τα δεδομένα εκπαίδευσης είναι ένα σύνολο $S = s_1, s_2, \dots, s_n$, ήδη ταξινομημένων δειγμάτων, όπου κάθε δείγμα s_i αποτελείται από ένα p -διαστατικό φορέα $(x_{1,i}, x_{2,i}, \dots, x_{p,i})$, όπου το x_i αντιπροσωπεύουν τιμές χαρακτηριστικών ή χαρακτηριστικά του δείγματος. Σε κάθε κόμβο του δέντρου, επιλέγει, εκείνο το χαρακτηριστικό των δεδομένων, το οποίο χωρίζει αποτελεσματικότερα το σύνολο δειγμάτων του σε υποσύνολα, εμπλουτισμένα σε μία κατηγορία.

Σε επαναληπτική λειτουργία, το πρόγραμμα ξεκινά με ένα τυχαία επιλεγμένο υποσύνολο των δεδομένων, δημιουργεί ένα δέντρο αποφάσεων δοκιμαστικής απόφασης, προσθέτει ορισμένα εσφαλμένα ταξινομημένα αντικείμενα και συνεχίζεται, έως ότου το δέντρο αποφάσεων δοκιμαστικής απόφασης να ταξινομήσει σωστά όλα τα αντικείμενα, που δεν βρίσκονται στο παράθυρο ή έως ότου φανεί ότι δεν έχει σημειωθεί κάποια πρόοδος. Δεδομένου ότι η

¹ Εντροπία πληροφοριών: είναι ένα μέτρο αβεβαιότητας, που διακατέχει ένα σύστημα.

επαναληπτική λειτουργία ξεκινά με ένα τυχαία επιλεγμένο υποσύνολο, πολλές δοκιμές με τα ίδια δεδομένα, μπορούν να χρησιμοποιηθούν για τη δημιουργία περισσότερων από ένα δένδρων. Ο αλγόριθμος C4.5 είναι δημιουργία του Ross Quinlan. Ο αλγόριθμος C4.5 βασίζεται στον αλγόριθμο **ID3** (Iterative Dichotomiser 3).

ΠΑΡΑΔΕΙΓΜΑ	ΣΥΜΒΟΛΙΣΜΟΙ	ΕΠΕΞΗΓΗΣΗ
@book{Quinlan1993,	S_i	Το book είναι ένα δείγμα
address = {San Mateo, CA},	X_1	Η τιμή του χαρακτηριστικού address είναι San Mateo, CA
author = {Ross Quinlan},	X_2	Η τιμή του χαρακτηριστικού author είναι Ross Quinlan
publisher = {Morgan Kaufmann Publishers},	X_3	Η τιμή του χαρακτηριστικού publisher είναι Morgan Kaufmann Publishers
title = {C4.5: Programs for Machine Learning},	X_4	Η τιμή του χαρακτηριστικού title είναι C4.5: Programs for Machine Learning
year = {1993}}	X_5	Η τιμή του χαρακτηριστικού year είναι 1993

Πίνακας 1: Παράδειγμα από ηλεκτρονική βιβλιοθήκη του weka

Ο αλγόριθμος **ID3** (Iterative Dichotomiser 3) χρησιμοποιείται συνήθως στην Μηχανική μάθηση και στην Επεξεργασία φυσικής γλώσσας (ΕΦΓ)².

$$I_E(i) = - \sum_{j=1}^m f(i,j) \log f(i,j)$$

Ο αλγόριθμος ID3 είναι βασισμένος στο Ξυράφι του Όκαμ (*Occam's Razor*).

² ΕΦΓ: είναι η επιστήμη της πληροφορικής, της τεχνητής νοημοσύνης και της υπολογιστικής γλωσσολογίας και ασχολείται με τις αλληλεπιδράσεις μεταξύ των υπολογιστών και των ανθρώπων (φυσικών) γλωσσών.

Πλεονεκτήματα:

- Εύκολη ερμηνεία.
- Έχει καλή απόδοση σε μεγάλα σύνολα δεδομένων.
- Απαιτεί ελάχιστο καθαρισμό δεδομένων.

Μειονεκτήματα:

- Επιρρεπής στην υπερβολική εφαρμογή.

2.2 Αλγόριθμος Random Tree

Το Random Tree είναι ένας αλγόριθμος μάθησης συνόλου, που δημιουργεί πολλά μεμονωμένα μοντέλα. Χρησιμοποιεί μια ιδέα συγκέντρωσης, ώστε να παραχθεί ένα τυχαίο σύνολο δεδομένων και να κατασκευαστεί ένα δέντρο αποφάσεων.

Στο τυπικό δέντρο, κάθε κόμβος χωρίζεται χρησιμοποιώντας την καλύτερη διαίρεση, μεταξύ όλων των μεταβλητών. Σε ένα τυχαίο δάσος, κάθε κόμβος χωρίζεται χρησιμοποιώντας το καλύτερο, μεταξύ του υποσυνόλου των προδιαθέτων, που επιλέχθηκαν τυχαία σε αυτόν τον κόμβο.

Τα τυχαία δέντρα έχουν εισαχθεί από τον Leo Breiman και Adele Cutler. Ο αλγόριθμος μπορεί να αντιμετωπίσει και τα δύο προβλήματα, ταξινόμησης και παλινδρόμησης. Τα τυχαία δέντρα είναι σύνολο προγνωστικών δέντρων, που ονομάζεται δάσος. Η ταξινόμηση λειτουργεί ως εξής: η τυχαία ταξινόμηση δέντρων παίρνει το διάνυσμα χαρακτηριστικών εισόδου, το ταξινομεί με κάθε δέντρο στο δάσος και βγάζει την ετικέτα της τάξης, που έλαβε την πλειοψηφία των «ψηφών». Σε περίπτωση παλινδρόμησης, η απόκριση του ταξινομητή είναι ο μέσος όρος των απαντήσεων, σε όλα τα δέντρα στο δάσος. Υπάρχουν δύο κατηγορίες δέντρων:

- Τα τυχαία δέντρα είναι ουσιαστικά ο συνδυασμός δύο υπαρχόντων αλγορίθμων στη μηχανική μάθηση, δηλαδή τα δέντρα με μονό μοντέλο συνδυάζονται με ιδέες τυχαίου δάσους.
- Τα πρότυπα δέντρα είναι δέντρα αποφάσεων, όπου κάθε φύλλο έχει ένα γραμμικό μοντέλο, το οποίο είναι βελτιστοποιημένο για τον τοπικό υποχώρο και περιγράφεται από αυτό το φύλλο.

Τα τυχαία δάση έχουν δείξει ότι βελτιώνουν σημαντικά την απόδοση των δέντρων μεμονωμένων αποφάσεων. Η ποικιλομορφία των δέντρων δημιουργείται από δύο τρόπους τυχαιοποίησης. Αρχικά, γίνεται δειγματοληψία των δεδομένων εκπαίδευσης, με αντικατάσταση, για κάθε ένα δέντρο, όπως στο Bagging. Δεύτερον, όταν μεγαλώνουμε ένα δέντρο, αντί να υπολογίζουμε πάντα το καλύτερο δυνατό διαχωρισμό για κάθε κόμβο, μόνο ένα τυχαίο υποσύνολο όλων των χαρακτηριστικών λαμβάνεται υπόψη και υπολογίζεται ο καλύτερος διαχωρισμός για αυτό το υποσύνολο. Τέτοια δέντρα προορίζονται για ταξινόμηση. Τα τυχαία δέντρα, για πρώτη φορά συνδυάζουν πρότυπα δέντρα και τυχαία δάση. Επίσης, χρησιμοποιούν αυτό το προϊόν για διαχωρισμένη επιλογή και έτσι προκαλούν λογικά ισορροπημένα δέντρα, όπου μια συνολική ρύθμιση για την τιμή της κορυφογραμμής, λειτουργεί σε όλα τα φύλλα, απλοποιώντας έτσι τη διαδικασία βελτιστοποίησης.

2.3 Αλγόριθμος Random Forest

Ο Random Forest είναι μια τεχνική που χρησιμοποιείται στη μοντελοποίηση προβλέψεων και ανάλυση συμπεριφοράς και βασίζεται σε δέντρα απόφασης. Περιέχει πολλά δέντρα απόφασης που αντιπροσωπεύουν μια ξεχωριστή περίπτωση της ταξινόμησης των δεδομένων που εισάγονται στο τυχαίο δάσος. Η τεχνική του τυχαίου δάσους εξετάζει τις περιπτώσεις μεμονωμένα, λαμβάνοντας ως επιλεγμένη πρόβλεψη εκείνη με την πλειοψηφία των ψήφων.

Κάθε δέντρο στις ταξινομήσεις λαμβάνει δεδομένα από δείγματα στο αρχικό σύνολο δεδομένων. Στη συνέχεια επιλέγονται τυχαία χαρακτηριστικά, τα οποία χρησιμοποιούνται για την ανάπτυξη του δέντρου σε κάθε κόμβο. Κάθε δέντρο στο δάσος δεν πρέπει να κόβετε μέχρι το τέλος όταν η πρόβλεψη έχει επιτευχθεί. Με αυτόν τον τρόπο, το τυχαίο δάσος επιτρέπει σε οποιονδήποτε ταξινομητή με συσχετισμό να δημιουργήσει έναν ισχυρό ταξινομητή.

Πλεονεκτήματα:

- Εκτιμήσεις για μεταβλητή σημασία.
- Προσφέρουν μια ανώτερη μέθοδο για εργασία με δεδομένα που λείπουν.
- Οι τιμές που λείπουν αντικαθίστανται από τη μεταβλητή που εμφανίζεται περισσότερο σε έναν συγκεκριμένο κόμβο.
- Μεταξύ όλων των διαθέσιμων μεθόδων ταξινόμησης, τα τυχαία δάση παρέχουν την υψηλότερη ακρίβεια.

- Μπορεί να χειριστεί μεγάλα δεδομένα με πολλές μεταβλητές που ανέρχονται σε χιλιάδες.
- Μπορεί να εξισορροπήσει αυτόματα σύνολα δεδομένων όταν μια κλάση είναι πιο σπάνια από άλλες κλάσεις στα δεδομένα.
- Η μέθοδος χειρίζεται τις μεταβλητές γρήγορα, καθιστώντας την κατάλληλη για περίπλοκες εργασίες.

Μειονεκτήματα:

- Αυτή η τεχνική δεν είναι κατάλληλη για ταξινόμηση θορυβώδες σύνολο δεδομένων.
- Το τυχαίο δάσος είναι ευνοϊκό για χαρακτηριστικά με περισσότερα επίπεδα για κατηγορικές μεταβλητές δεδομένων.

2.4 Αλγόριθμος REP Tree

Το REP Tree χρησιμοποιεί τη λογική του δέντρου παλινδρόμησης και δημιουργεί πολλά δέντρα σε διαφορετικές επαναλήψεις. Μετά από αυτό, επιλέγει το καλύτερο από όλα τα δέντρα που δημιουργήθηκαν, το οποίο θα θεωρηθεί ως εκπρόσωπος. Κατά την επιλογή του δέντρου, το μέτρο που χρησιμοποιείται, είναι το μέσο τετραγωνικό σφάλμα στις προβλέψεις που γίνονται από το δέντρο.

Το Basically Reduced Error Pruning Tree ("REPT") είναι γρήγορη εκμάθηση δέντρων αποφάσεων και δημιουργεί ένα δέντρο αποφάσεων, με βάση την απόκτηση πληροφοριών ή τη μείωση της διακύμανσης.

Το REP Tree είναι ένα γρήγορο μοντέλο μάθησης δέντρων αποφάσεων, που δημιουργεί ένα δέντρο αποφάσεων, χρησιμοποιώντας το κέρδος πληροφοριών, ως κριτήριο διάσπασης και με την μείωση των 'κλαδιών' του δέντρου, μειώνονται και τα σφάλματα. Ταξινομεί μόνο τιμές για αριθμητικά χαρακτηριστικά και μόνο μία φορά. Οι τιμές που λείπουν αντιμετωπίζονται, χρησιμοποιώντας τη μέθοδο του C4.5, για τη χρήση κλασματικών εμφανίσεων.

Πλεονεκτήματα:

- Διαχειρίζεται κατηγορικές αξίες.
- Ευκολότερη κατανόηση της περίπλοκης σχέσης μεταξύ μεταβλητών.

- Ελαχιστοποιεί το αποτέλεσμα λανθασμένου ή ελλείποντος τιμές στην τελική αναπαράσταση του δέντρου.

Μειονεκτήματα:

- Το δέντρο είναι ασταθές ακόμα και για μικρές αλλαγές εισαγωγή δεδομένων.
- Τα μεγάλα μοντέλα δέντρων είναι δύσκολο να αναλυθούν.

3 Εργαλεία Εξόρυξης Δεδομένων

3.1 Machine Learning with Python

Η Python είναι μία γλώσσα προγραμματισμού που δημιουργήθηκε το 1989 από τον Guido van Rossum στο ερευνητικό κέντρο Centrum Wiskunde & Informatica (CWI). Η γλώσσα προγραμματισμού ABC φαίνεται να λειτούργησε ως πρότυπο για τη δημιουργία της Python.

Το μεγάλο πλεονέκτημα και ο βασικός λόγος για την επιτυχία της Python είναι το ότι μπορεί να επεκτείνεται με μεγάλη ευκολία, σύμφωνα με το δημιουργό της. Η Python έχει σχεδιαστεί με τέτοιο τρόπο ώστε να επεκτείνεται εύκολα, να δίνει τη δυνατότητα ενσωματωμένων στοιχείων όπως εντολές ή τύπους αντικειμένων, αλλά και να μπορούν οι προγραμματιστές να προσθέτουν τα δικά τους στοιχεία, αναλόγως τις ανάγκες τους. Το 1991 κυκλοφόρησε. Φαίνεται να πήρε το όνομά της από μια κωμική σειρά του '70, το «Monty Python's Flying Circus» του BBC της Μεγάλης Βρετανίας. Ο Guido van Rossum ήθελε δώσει ένα μικρό, μοναδικό και μυστήριο όνομα στη γλώσσα προγραμματισμού του. Επομένως, η ονομασία της δεν φαίνεται να έχει σχέση με τον πύθωνα, το γνωστό φίδι.

Η Python εξελίσσεται με γοργούς ρυθμούς, με καινούριες εκδόσεις και με βασικό εργαλείο τα Python Enhancement Proposals ("PEPs"), που είναι αναγκαία για να δίνουν γενικές πληροφορίες, οδηγίες, προτάσεις και περιγραφές καινούρια χαρακτηριστικά της γλώσσας που μπορεί να προκύψουν. Πολλές άλλες εφαρμογές έχουν δημιουργηθεί χρησιμοποιώντας ως πρότυπο τις εκδόσεις της. Πρόκειται για μία γλώσσα προγραμματισμού σχετικά εύκολη και η ποικιλία αποκλειστικών αναλυτικών βιβλιοθηκών που διαθέτει δίνει τη δυνατότητα στους επιστήμονες να μπορούν να βρίσκουν δεδομένα σχετικά με κάθε επιστημονικό τομέα, που είναι προσαρμοσμένα στις ανάγκες τους και έτοιμα για λήψη. Επιλέγεται κυρίως από προγραμματιστές που επιθυμούν να εφαρμόσουν στατιστικές τεχνικές ή ανάλυση δεδομένων στην εργασία τους, αλλά και από επιστήμονες που ασχολούνται με εφαρμογές του διαδικτύου.

Πρόκειται για μία γλώσσα προγραμματισμού γενικής χρήσης, απλή και μπορεί να την μάθει πολύ εύκολα κάποιος. Είναι δυνατόν να χρησιμοποιηθεί από αρχάριους, αλλά και από έμπειρους προγραμματιστές. Τέλος, μπορεί να χρησιμοποιηθεί και για εκπαιδευτικούς σκοπούς, αλλά και για την κατασκευή διάφορων εφαρμογών.

3.2 Scikit-learn

Η Scikit-learn είναι μια βιβλιοθήκη Python για μηχανική μάθηση βασίζεται στη NumPy και τη SciPy. Η Scikitlearn παρέχει τη δυνατότητα να εφαρμοστούν άμεσα γνωστοί αλγόριθμοι στο σύνολο δεδομένων. Ακόμη, έχει ποικίλα εργαλεία που χρησιμοποιούνται για πολλές εργασίες εκμάθησης μηχανών, αλλά και εξόρυξης δεδομένων. Αρχικά, δημιουργήθηκε από τον David Courvapeau ως έργο κώδικα καλοκαιριού της εταιρείας Google το 2007. Αργότερα, ο Matthieu Brucher άρχισε να το χρησιμοποιεί εκτός από τη διατριβή του. Το 2010 κυκλοφόρησε δημοσίως για πρώτη φορά (έκδοση v0.1 beta), στα τέλη Ιανουαρίου 2010. Το scikit - learn έχει τώρα περισσότερους από 30 ενεργούς συντελεστές και είχε χορηγίες από την INRIA, την Google, την Tinyclues και το ίδρυμα Λογισμικού Python. Το Scikit-learn παρέχει μια σειρά από εποπτευόμενους και χωρίς επίβλεψη αλγόριθμους μάθησης μέσω μιας σταθερής διεπαφής με την Python. Έχει αδειοδοτημένη απλοποιημένη άδεια BSD και διανέμεται με πολλές διανομές Linux, ενθαρρύνοντας την ακαδημαϊκή και εμπορική χρήση. Η βιβλιοθήκη είναι βασισμένη στο SciPy (Scientific Python) το οποίο είναι απαραίτητο να εγκατασταθεί για να μπορέσει κάποιος να χρησιμοποιήσει το scikit-learn. Αυτή η στοίβα που περιλαμβάνει:

- NumPy: Βάση n -διαστάσεων πακέτου πίνακα
- SciPy: Θεμελιώδης βιβλιοθήκη για επιστημονική υπολογιστική
- Matplotlib: Ολοκληρωμένη σχεδίαση 2D/3D
- IPython: Βελτιωμένη διαδραστική κονσόλα
- SymPy: Συμβολικά μαθηματικά
- Pandas: Δομές δεδομένων και ανάλυση
- Επεκτάσεις ή ενότητες για τη φροντίδα SciPy που ονομάζονται συμβατικά SciKits

Ως εκ τούτου, η ενότητα παρέχει αλγόριθμους μάθησης και ονομάζεται scikit-learn. Μερικές δημοφιλείς ομάδες μοντέλων που παρέχονται από το scikit-learn περιλαμβάνουν:

- Ομαδοποίηση: για ομαδοποίηση δεδομένων χωρίς ετικέτα, όπως το KMeans.
- Διασταυρούμενη επικύρωση: για την εκτίμηση της απόδοσης των εποπτευόμενων μοντέλων σε αόρατα δεδομένα.
- Σύνολα δεδομένων: για δοκιμαστικά σύνολα δεδομένων και για τη δημιουργία συνόλων δεδομένων με συγκεκριμένες ιδιότητες για τη διερεύνηση της συμπεριφοράς του μοντέλου.

- Μείωση διαστάσεων: για τη μείωση του αριθμού των χαρακτηριστικών στα δεδομένα για σύνοψη, οπτικοποίηση και επιλογή χαρακτηριστικών, όπως η ανάλυση κύριων στοιχείων.
- Μέθοδοι συνόλου: για συνδυασμό των προβλέψεων πολλαπλών εποπτευόμενων μοντέλων.
- Εξαγωγή χαρακτηριστικών: για τον καθορισμό χαρακτηριστικών στα δεδομένα εικόνας και κειμένου.
- Επιλογή χαρακτηριστικών: για τον προσδιορισμό σημαντικών χαρακτηριστικών από τα οποία δημιουργούνται εποπτευόμενα μοντέλα.
- Ρύθμιση παραμέτρων: για να αξιοποιήσετε στο έπακρο τα εποπτευόμενα μοντέλα.
- Manifold Learning: Για σύνοψη και απεικόνιση σύνθετων πολυδιάστατων δεδομένων.
- Εποπτευόμενα Μοντέλα: μια τεράστια συλλογή που δεν περιορίζεται σε γενικευμένα γραμμικά μοντέλα, ανάλυση διακρίσεων, αφελείς κόλποι, τεμπέλικες μέθοδοι, νευρωνικά δίκτυα, μηχανές διάγνωση υποστήριξης και δέντρα αποφάσεων.

3.3 XGBoost

Το XGBoost είναι ένας αλγόριθμος μηχανικής μάθησης που βασίζεται σε δέντρα και χρησιμοποιεί ένα πλαίσιο ενίσχυσης κλίσης. Σε προβλήματα πρόβλεψης που περιλαμβάνουν μη δομημένα δεδομένα (εικόνες, κείμενο, κ.λπ.) τα τεχνητά νευρωνικά δίκτυα τείνουν να υπερτερούν όλων των άλλων αλγορίθμων ή πλαισίων. Ωστόσο, όταν πρόκειται για μικρά έως μεσαία δομημένα/πίνακες δεδομένων, οι αλγόριθμοι που βασίζονται σε δέντρα αποφάσεων θεωρούνται οι καλύτεροι στην κατηγορία αυτή τη στιγμή. Ο αλγόριθμος XGBoost αναπτύχθηκε ως ερευνητικό έργο στο Πανεπιστήμιο της Ουάσινγκτον. Ο Tianqi Chen και ο Carlos Guestrin παρουσίασαν το έγγραφό τους στο συνέδριο SIGKDD το 2016 και έπιασαν τον κόσμο της μηχανικής μάθησης. Από την εισαγωγή του, αυτός ο αλγόριθμος δεν έχει πιστωθεί μόνο ότι κέρδισε πολυάριθμους διαγωνισμούς Kaggle, αλλά και ότι ήταν η κινητήριος δύναμη κάτω από το καπό για πολλές βιομηχανικές εφαρμογές αιχμής. Ως αποτέλεσμα, υπάρχει μια ισχυρή κοινότητα επιστημόνων δεδομένων που συμβάλλουν στα έργα ανοιχτού κώδικα XGBoost με 350 ~ συνεισφέροντες και 63.600 δεσμεύσεις στο GitHub. Ο αλγόριθμος διαφοροποιείται με τους εξής τρόπους:

- Ένα ευρύ φάσμα εφαρμογών: Μπορεί να χρησιμοποιηθεί για την επίλυση προβλημάτων παλινδρόμησης, ταξινόμησης, κατάταξης και καθορισμού από το χρήστη.
- Φορητότητα: Λειτουργεί ομαλά σε Windows, Linux και OS X.
- Γλώσσες: Υποστηρίζει όλες τις κύριες γλώσσες προγραμματισμού, συμπεριλαμβανομένων των C ++, Python, R, Java, Scala και Julia.
- Cloud Integration: Υποστηρίζει συμπλέγματα AWS, Azure και Yarn και λειτουργεί καλά με Flink, Spark και άλλα οικοσυστήματα

3.4 Python-recsys

Ο στόχος του RecSys είναι να προτείνει σχετικά στοιχεία στους χρήστες, με βάση τις προτιμήσεις τους. Η προτίμηση και η συνάφεια είναι υποκειμενικά και συνάγονται γενικά από στοιχεία που οι χρήστες είχαν καταναλώσει προηγουμένως.

Οι κύριες οικογένειες μεθόδων για το RecSys είναι:

- Συνεργατικό φιλτράρισμα: Αυτή η μέθοδος κάνει αυτόματες προβλέψεις (φιλτράρισμα) για τα ενδιαφέροντα ενός χρήστη συλλέγοντας προτιμήσεις ή γευστικές πληροφορίες από πολλούς χρήστες (συνεργαζόμενοι). Η βασική υπόθεση της συνεργατικής προσέγγισης φιλτραρίσματος είναι ότι εάν ένα άτομο A έχει την ίδια γνώμη με το άτομο B για ένα σύνολο στοιχείων, το A είναι πιο πιθανό να έχει τη γνώμη του B για ένα δεδομένο στοιχείο παρά αυτή ενός τυχαία επιλεγμένου ατόμου.
- Φιλτράρισμα βάσει περιεχομένου: Αυτή η μέθοδος χρησιμοποιεί μόνο πληροφορίες σχετικά με την περιγραφή και τα χαρακτηριστικά των στοιχείων που οι χρήστες είχαν καταναλώσει προηγουμένως για να προτιμήσουν τις προτιμήσεις του χρήστη. Με άλλα λόγια, αυτοί οι αλγόριθμοι προσπαθούν να προτείνουν στοιχεία που είναι παρόμοια με αυτά που άρεσε στο χρήστη στο παρελθόν (ή εξετάζει στο παρόν). Συγκεκριμένα, διάφορα υποψήφια στοιχεία συγκρίνονται με στοιχεία που είχαν βαθμολογηθεί προηγουμένως από τον χρήστη και συνιστώνται τα στοιχεία που ταιριάζουν καλύτερα.
- Υβριδικές μέθοδοι: Πρόσφατη έρευνα απέδειξε ότι μια υβριδική προσέγγιση, που συνδυάζει συνεργατικό φιλτράρισμα και φιλτράρισμα βάσει περιεχομένου θα μπορούσε να είναι πιο αποτελεσματική από τις καθαρές προσεγγίσεις σε ορισμένες περιπτώσεις. Αυτές οι μέθοδοι μπορούν επίσης να χρησιμοποιηθούν για να

ξεπεραστούν ορισμένα από τα κοινά προβλήματα στα συστήματα προτάσεων, όπως το κρύο ξεκίνημα και το πρόβλημα αραιότητας.

3.5 MLlib

Η βιβλιοθήκη μηχανικής μάθησης Apache Spark (MLlib) έχει σχεδιαστεί για απλότητα, επεκτασιμότητα και εύκολη ενσωμάτωση με άλλα εργαλεία. Με την επεκτασιμότητα, τη συμβατότητα γλώσσας και την ταχύτητα του Spark, οι επιστήμονες δεδομένων μπορούν να επικεντρωθούν στα προβλήματα και τα μοντέλα δεδομένων τους αντί να λύσουν τις πολυπλοκότητες που περιβάλλουν τα καταμεμημένα δεδομένα (όπως υποδομή, διαμορφώσεις κ.ο.κ.). Χτισμένο πάνω από το Spark, το MLlib είναι μια κλιμακούμενη βιβλιοθήκη μηχανικής εκμάθησης που αποτελείται από κοινούς αλγόριθμους μάθησης και βοηθητικά προγράμματα, συμπεριλαμβανομένης της ταξινόμησης, της παλινδρόμησης, της ομαδοποίησης, του συνεργατικού φιλτραρίσματος, της μείωσης των διαστάσεων και των βασικών πρωτόγονων βελτιστοποίησης. Το Spark MLlib ενσωματώνεται απρόσκοπτα με άλλα εξαρτήματα Spark, όπως το Spark SQL, το Spark Streaming και το DataFrames και είναι εγκατεστημένο στο χρόνο εκτέλεσης των Databricks. Η βιβλιοθήκη μπορεί να χρησιμοποιηθεί σε Java, Scala και Python ως μέρος των εφαρμογών Spark, ώστε να μπορείτε να την συμπεριλάβετε σε πλήρεις ροές εργασιών. Το MLlib επιτρέπει την προεπεξεργασία, την επεξεργασία, την εκπαίδευση μοντέλων και την πραγματοποίηση προβλέψεων σε κλίμακα σε δεδομένα. Μπορείτε ακόμη να χρησιμοποιήσετε μοντέλα εκπαιδευμένα στο MLlib για να κάνετε προβλέψεις στη δομημένη ροή. Το Spark παρέχει ένα εξελιγμένο API μηχανικής μάθησης για την εκτέλεση ποικίλων εργασιών μηχανικής μάθησης, από την ταξινόμηση έως την παλινδρόμηση, την ομαδοποίηση έως τη βαθιά μάθηση.

3.6 Eggplant

Το Eggplant χρησιμοποιεί τεχνητή νοημοσύνη και μηχανική μάθηση για να αυτοματοποιήσει τις δοκιμές, βοηθώντας τις ομάδες σε ρυθμιζόμενα περιβάλλοντα να παραδώσουν το DevOps σε κλίμακα και να εκπληρώσουν τις νομοθετικές απαιτήσεις. Προτάθηκε πρόσφατα από το χαρτί του Defense Science Board στο σχεδιασμό και την απόκτηση λογισμικού για αμυντικά συστήματα.

Χρησιμοποιεί τη γλώσσα SenseTalk, η οποία μοιάζει με την αγγλική. Οι βασικές εντολές και λειτουργίες της είναι οι παρακάτω:

- Οι συναρτήσεις έχουν πάντα τιμές επιστροφής, ενώ οι εντολές δεν είναι απαραίτητο να έχουν.
- Οι συναρτήσεις χρησιμοποιούνται συνήθως ως παράμετροι σε εντολές.
- Οι εντολές πρέπει να γράφονται μόνες τους, όχι ως μέρος μιας δήλωσης.
- Οι συναρτήσεις πρέπει να αποτελούν μέρος μιας δήλωσης.
- Οι εντολές και οι συναρτήσεις στην Eggplant χρησιμοποιούν μια ποικιλία τύπων δεδομένων: αριθμούς, συντεταγμένες, συμβολοσειρές και λίστες ιδιοτήτων.

3.7 Mahout

“Mahout” είναι αυτός που οδηγεί έναν ελέφαντα ως κύριός του. Το όνομα προέρχεται από τη στενή σχέση που έχει με το Apache Hadoop, το οποίο έχει για λογότυπο έναν ελέφαντα. Το Mahout είναι ένα πλαίσιο ανοιχτού κώδικα από το Apache, το οποίο επιτρέπει την αποθήκευση και επεξεργασία μεγάλων δεδομένων σε κατανεμημένο περιβάλλον σε ομάδες υπολογιστών χρησιμοποιώντας απλά μοντέλα προγραμματισμού. Το Apache Mahout είναι ένα έργο ανοιχτού κώδικα που χρησιμοποιείται κυρίως για να δημιουργούνται κλιμακούμενοι αλγόριθμοι μηχανικής μάθησης. Εφαρμόζει δημοφιλείς τεχνικές μηχανικής μάθησης όπως:

- Σύσταση
- Ταξινόμηση
- Ομαδοποίηση

Το Apache Mahout ξεκίνησε ως υπο-έργο του Apache's Lucene το 2008. Το 2010, το Mahout έγινε έργο κορυφαίου επιπέδου του Apache.

Χαρακτηριστικά του Mahout είναι:

- Οι αλγόριθμοι του Mahout είναι γραμμένοι πάνω από το Hadoop, έτσι λειτουργεί καλά σε κατανεμημένο περιβάλλον.
- Το Mahout χρησιμοποιεί τη βιβλιοθήκη Apache Hadoop για αποτελεσματική κλιμάκωση στο cloud.
- Το Mahout προσφέρει στον κωδικοποιητή ένα έτοιμο για χρήση πλαίσιο για την εκτέλεση εργασιών εξόρυξης δεδομένων σε μεγάλο όγκο δεδομένων.
- Το Mahout επιτρέπει στις εφαρμογές να αναλύουν μεγάλα σύνολα δεδομένων αποτελεσματικά και γρήγορα.

- Περιλαμβάνει αρκετές εφαρμογές ομαδοποίησης με δυνατότητα MapReduce, όπως k-means, fuzzy k-means, Canopy, Dirichlet και Mean-Shift.
- Υποστηρίζει τις κατανεμημένες Naive Bayes και τις συμπληρωματικές εφαρμογές ταξινόμησης Naive Bayes.
- Έρχεται με κατανεμημένες δυνατότητες λειτουργίας φυσικής κατάστασης για εξελικτικό προγραμματισμό.
- Περιλαμβάνει βιβλιοθήκες μήτρας και διανυσμάτων.

Εταιρείες όπως η Adobe, το Facebook, το LinkedIn, το Foursquare, το Twitter και το Yahoo χρησιμοποιούν το Mahout εσωτερικά. Το Foursquare βοηθά το άτομο να βρει μέρη, τρόφιμα και ψυχαγωγία διαθέσιμα σε μια συγκεκριμένη περιοχή χρησιμοποιώντας το Mahout. Το Twitter χρησιμοποιεί το Mahout για μοντελοποίηση ενδιαφερόντων χρηστών. Το Yahoo! χρησιμοποιεί το Mahout για εξόρυξη προτύπων.

3.8 Google Cloud TPU

Μια προσαρμοσμένη μηχανική εκμάθηση ASIC ή ένα cloud TPU τροφοδοτεί τα προϊόντα της Google, όπως το Assistant, το Gmail, την Αναζήτηση και τη Μετάφραση. Τα δεύτερης γενιάς Cloud TPU της Google φέρνουν ταχύτερη υπολογιστική απόδοση στην κοινότητα TensorFlow. Το χαμηλό κόστος και η γρήγορη απόδοση του Google Cloud TPU το καθιστούν ιδανικό για τις ομάδες που μαθαίνουν τα προγράμματα μηχανών για καλύτερη απόδοση. Ένας χρήστης μπορεί επίσης να προσαρμόσει τις δικές του λύσεις μηχανικής μάθησης για παραδείγματα πραγματικής ζωής. Μπορεί να γίνει απλώς από τον χρήστη εάν φέρει τα δικά του δεδομένα, κατεβάσει ένα μοντέλο αναφοράς, το οποίο είναι βελτιστοποιημένο από την Google και ξεκινήσει την εκπαίδευση. Το Cloud TPU διαθέτει τις ακόλουθες τέσσερις προσφορές:

- Το Cloud TPU v2- Διαθέτει 180 teraflops με 64 GB μνήμης εύρους ζώνης υψηλής ταχύτητας (HBM).
- Το Cloud TPU v3- Διαθέτει 420 teraflops με 128GB HBM.
- Το Cloud TPU v2 Pod- Διαθέτει 11,5 petaflops με 4 TB HBM και 2-D toroidal mesh mesh.
- Το Cloud TPU v3 Pod- Διαθέτει 100 petaflops με 32 TB HBM και 2-D toroidal mesh mesh (Bisong, E. , 2019).

Το Cloud TPU έχει τις ακόλουθες δυνατότητες:

Μια πρότυπη βιβλιοθήκη η οποία αποτελείται από τη συλλογή βελτιωμένων μοντέλων του cloud-TPU που παρέχουν ακρίβεια, εξαιρετική απόδοση και ποιότητα στην ανίχνευση αντικειμένων, την ταξινόμηση εικόνας, την αναγνώριση ομιλίας, τη μοντελοποίηση γλώσσας και πολλά άλλα.

Η σύνδεση των Cloud-TPU με προσαρμοσμένους τύπους μηχανών είναι εκεί που ο χρήστης μπορεί να συνδεθεί σε cloud-TPU από προσαρμοσμένους τύπους εικόνας VM με πλατφόρμα AI Platform Deep Learning, η οποία εξισορροπεί τη μνήμη, τις ταχύτητες επεξεργαστή και τους πόρους αποθήκευσης υψηλής απόδοσης.

Τα δεδομένα και οι αναλυτικές υπηρεσίες των Google Cloud και Cloud TPU είναι πλήρως ενσωματωμένες με άλλες υπηρεσίες Google Cloud Platform. Έτσι, ο χρήστης μπορεί να λάβει το όφελος λόγω των τεχνολογιών δικτύωσης, αποθήκευσης και ανάλυσης δεδομένων.

Τα Preemptible clouds είναι συγκριτικά μια οικονομική επιλογή από τα παραδείγματα κατά παραγγελία. επομένως πολλά χρήματα μπορούν να εξοικονομηθούν.

3.9 WEKA

Το Waikato Environment for Knowledge Analysis (Weka) είναι μια συλλογή εργαλείων οπτικοποίησης αλγορίθμων, για ανάλυση δεδομένων και προγνωστική μοντελοποίησης.

Παρουσιάζεται αναλυτικότερα στο κεφάλαιο 4 (σελ. 25), διότι είναι το εργαλείο που θα χρησιμοποιήσουμε για την εξόρυξη δεδομένων της παρούσας εργασίας.

3.10 Σύγκριση Εργαλείων

ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ								
A/A	ΟΝΟΜΑ	Μέτριο ή Μεγάλο Κόστος	Εύχρηστο	Γρήγορη Απόδοση	Γλώσσα Λειτουργίας	Ποικιλία Εργαλείων (εντολές, κλπ.)	Εφαρμογή Γνωστών Αλγορίθμων	Δυνατότητα Επεκτασιμότητας
1	Machine Learning with Python		✓	✓	ΑΓΓΛΙΚΑ	✓		✓
2	Scikit-Learn			✓	ΑΓΓΛΙΚΑ	✓	✓	
3	XGBoost	✓	✓	✓	ΑΓΓΛΙΚΑ	✓		
4	Python reclsys	✓		✓	ΑΓΓΛΙΚΑ			✓
5	Mllib		✓		ΑΓΓΛΙΚΑ	✓		✓
6	Eggplant	✓		✓	ΑΓΓΛΙΚΑ	✓		
7	Mahout		✓	✓	ΑΓΓΛΙΚΑ	✓		
8	Google Cloud TPU	✓		✓	ΑΡΚΕΤΕΣ		✓	
9	WEKA		✓	✓	ΑΓΓΛΙΚΑ		✓	

A/A	ΟΝΟΜΑ	Αποθήκευση - Επεξεργασία μεγάλου όγκου δεδομένων	Δυνατότητα Ομαδοποίησης δεδομένων					
1	Machine Learning with Python							
2	Scikit-Learn							
3	XGBoost	✓						
4	Python reclsys		✓					
5	Mllib							
6	Eggplant							
7	Mahout	✓	✓					
8	Google Cloud TPU							
9	WEKA	✓	✓					

Με βάση τον ανωτέρω πίνακα, καταλήγουμε στο συμπέρασμα ότι τόσο το «Mahout», όσο και το «WEKA» είναι τα ιδανικότερα για έναν αρχάριο χρήστη των εργαλείων του κεφαλαίου, αλλά ταυτόχρονα έχει και πολλές δυνατότητες και με μικρό ή καθόλου κόστος.

4 Εργαλείο Εξόρυξης Δεδομένων Waikato Environment for

Knowledge Analysis (WEKA)

4.1 Περιγραφή WEKA

Το Waikato Environment for Knowledge Analysis (Weka) είναι μια συλλογή εργαλείων οπτικοποίησης και αλγορίθμων, για ανάλυση δεδομένων και προγνωστική μοντελοποίηση. Είναι γραμμένο με την γλώσσα προγραμματισμού Java και αναπτύχθηκε στο Πανεπιστήμιο του [Waikato](#) της Νέα Ζηλανδία.

Η πρώτη έκδοση ήταν το 1993 και συντάχθηκε με τις γλώσσες Tcl/Tk³, C και Makefiles. Το 1997, αποφασίστηκε να αναπτυχθεί ξανά το WEKA από την αρχή, όμως, αυτή την φορά, μόνο με Java, αλλά μαζί με τους αλγορίθμους μοντελοποίησης. Η σημερινή πλήρης έκδοση είναι η WEKA 3.8.4, όπου και θα την εγκαταστήσουμε στην επόμενη ενότητα (Εγκατάσταση WEKA).



Εικόνα 1: Λογότυπο WEKA

4.2 Εγκατάσταση WEKA

Για να εγκαταστήσουμε το weka στον υπολογιστή μας θα πρέπει να μεταβούμε στην ιστοσελίδα www.cs.waikato.ac.nz/ml/weka/.

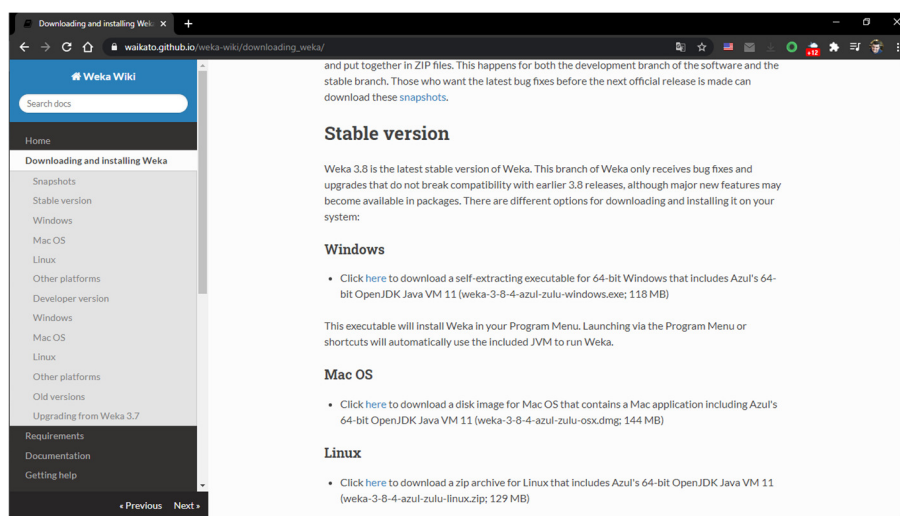
³ Tcl/Tk είναι μια γλώσσα προγραμματισμού σεναρίων



Εικόνα 2: Επίσημη ιστοσελίδα του weka

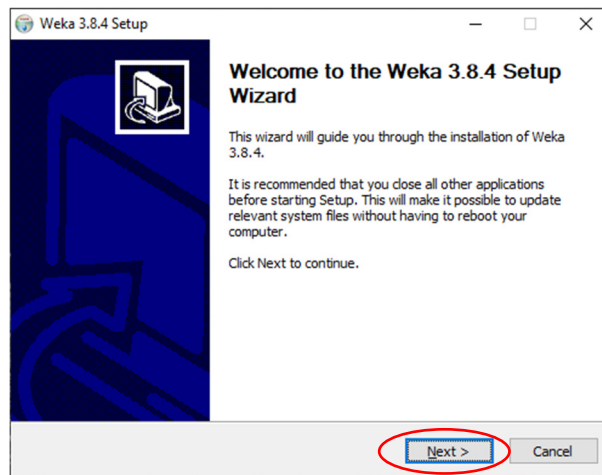
Έπειτα επιλέγετε ‘Download’, σας εμφανίζει την παρακάτω σελίδα και κατεβαίνοντας προς τα κάτω, θα βρείτε το Stable version, από εκεί επιλέγετε το λογισμικό που τρέχει η συσκευή σας (Windows, Mac OS, Linux) και ακολουθείτε τις οδηγίες.

4.2.1 Εγκατάσταση WEKA σε λειτουργικό σύστημα Windows

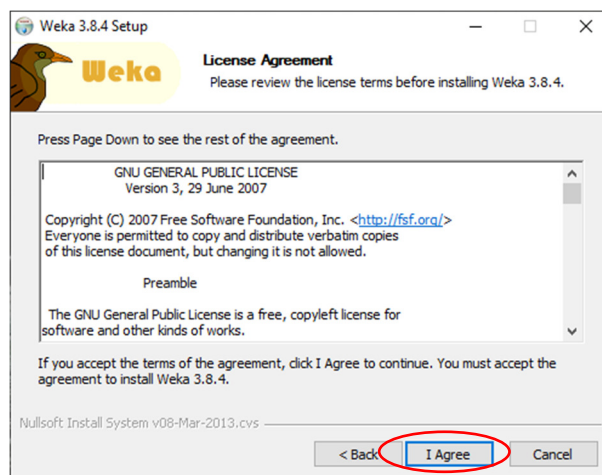


Εικόνα 3: Επιλογή λογισμικού windows

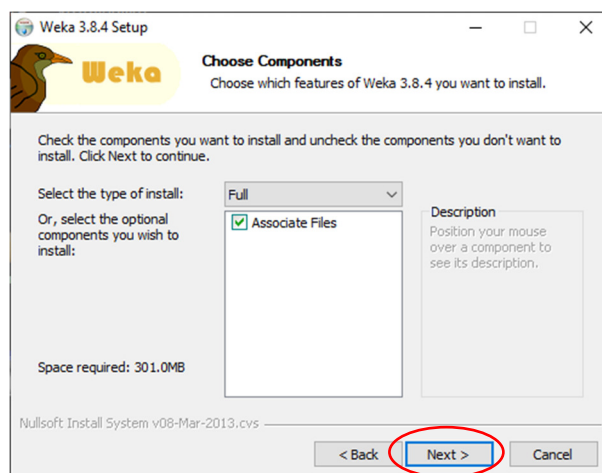
Παρακάτω βλέπουμε την εγκατάσταση σε λειτουργικό σύστημα Windows και με κόκκινο τις ενέργειες που ακολουθούμε για την εγκατάσταση.



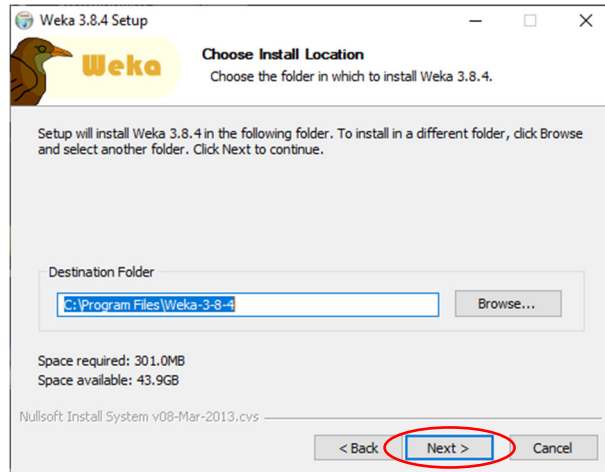
Εικόνα 4: Εγκατάσταση του Weka σε λειτουργικό σύστημα windows (1)



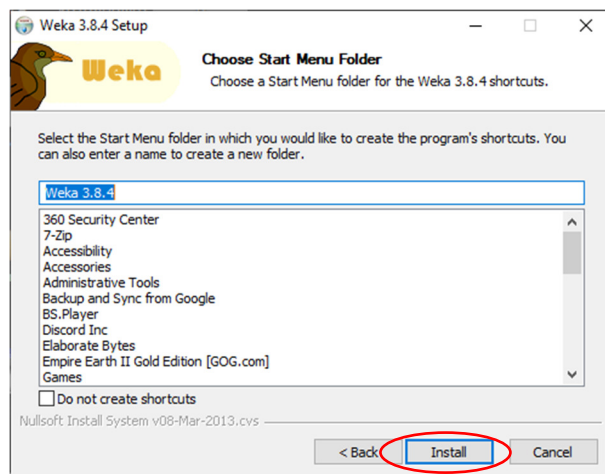
Εικόνα 5: Εγκατάσταση του Weka σε λειτουργικό σύστημα windows (2)



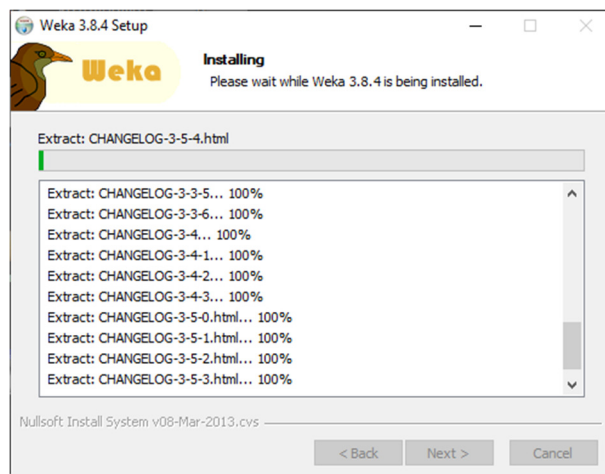
Εικόνα 6: Εγκατάσταση του Weka σε λειτουργικό σύστημα windows (3)



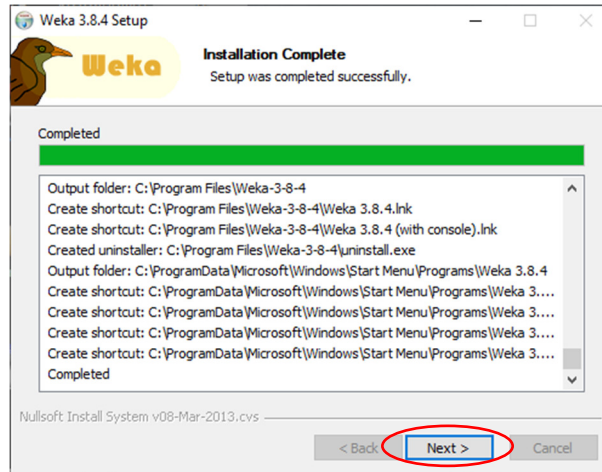
Εικόνα 7: Εγκατάσταση του Weka σε λειτουργικό σύστημα windows (4)



Εικόνα 8: Εγκατάσταση του Weka σε λειτουργικό σύστημα windows (5)



Εικόνα 9: Εγκατάσταση του Weka σε λειτουργικό σύστημα windows (6)



Εικόνα 10: Εγκατάσταση του Weka σε λειτουργικό σύστημα windows (7)



Εικόνα 11: Εγκατάσταση του Weka σε λειτουργικό σύστημα windows (8)

Πλέον έχει εγκατασταθεί με επιτυχία το weka στον υπολογιστή σας. Θα σας εμφανιστεί μια συντόμευση του προγράμματος στην επιφάνεια εργασίας σας, καθώς και στην έναρξη.



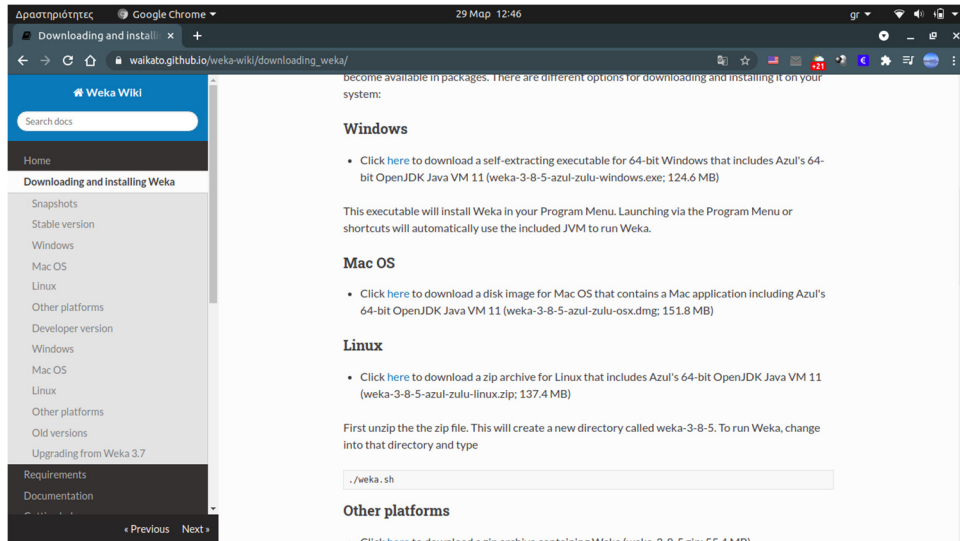
Εικόνα 12: Συντόμευση του weka στην επιφάνεια εργασίας σας

4.2.2 Εγκατάσταση WEKA σε λειτουργικό σύστημα Linux

Η συγκεκριμένη εγκατάσταση θα πραγματοποιηθεί στο ελεύθερο και δωρεάν λειτουργικό σύστημα Ubuntu στην έκδοση 20.04. Θα δούμε τρεις τρόπους με τους οποίους μπορούμε να εγκαταστήσουμε το weka.

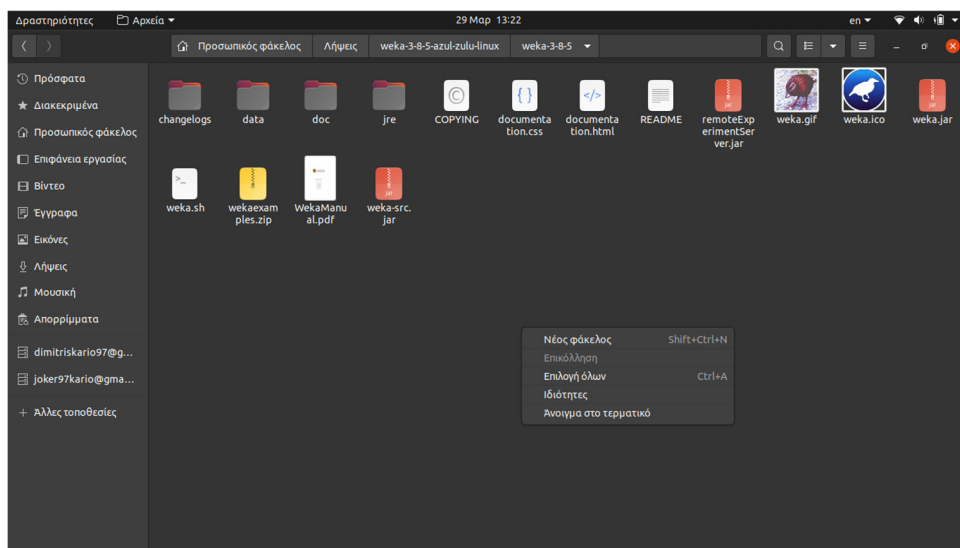
4.2.2.1 Εγκατάσταση WEKA μέσω της ιστοσελίδας

Μεταβαίνοντας στην ιστοσελίδα της εγκατάστασης του weka (βλέπω Εγκατάσταση WEKA σελ. 25) πηγαίνουμε στην παράγραφο για το Linux και πατάμε πάνω στο « Click [here](#) to download a zip archive for Linux ».



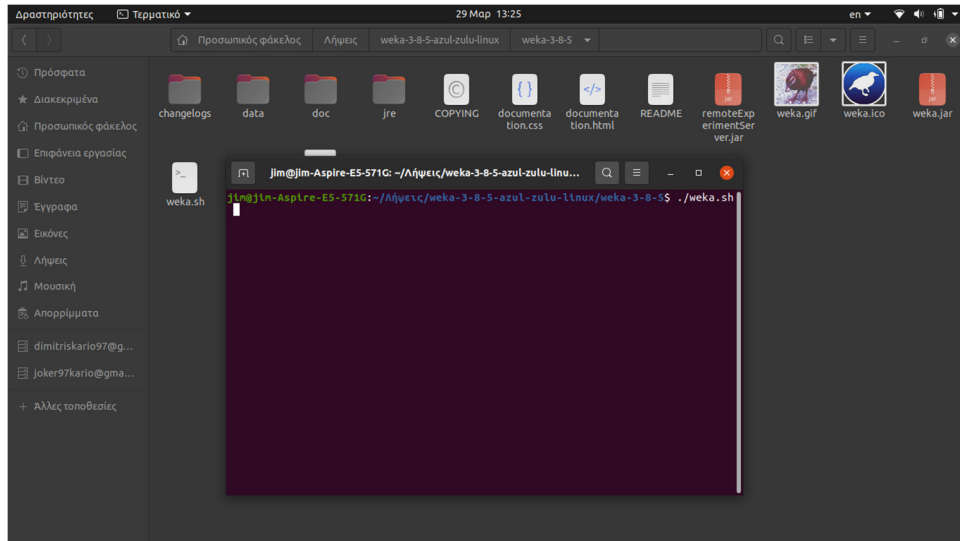
Εικόνα 13: Επιλογή λογισμικού Linux

Όταν κατέβει το αρχείο, πηγαίνουμε στις λήψεις και αποσυμπιέζουμε το αρχείο «weka-X-X-X-azul-zulu-linux.zip». Έπειτα το ανοίγουμε, πατάμε το weka-X-X-X, μετά πατάμε δεξί κλικ και επιλέγουμε «άνοιγμα στο τερματικό».



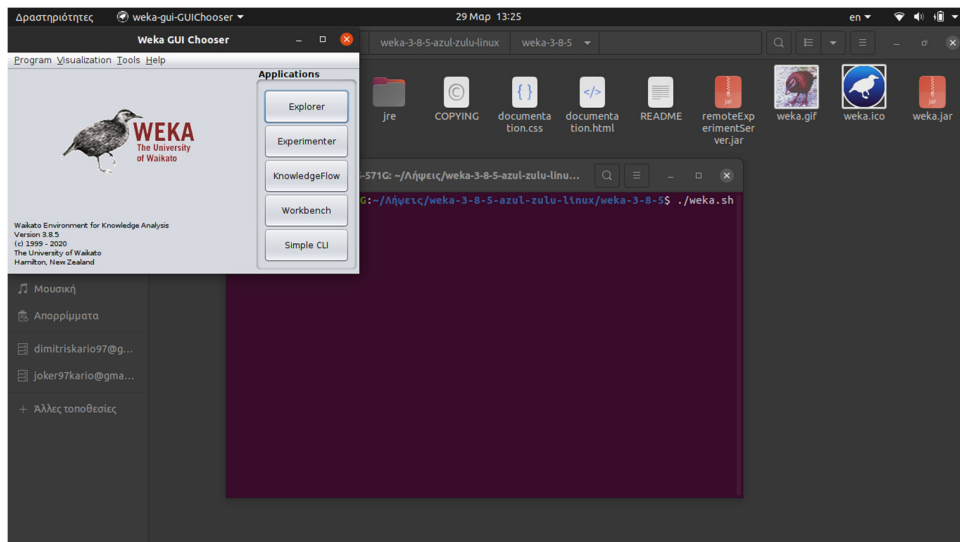
Εικόνα 14: Φάκελος weka-X-X-X

Όταν ανοίξει πληκτρολογούμε το ακόλουθο «./weka.sh».



Εικόνα 15: Ανοίγμα στο τερματικό

Τώρα μπορούμε να τρέξουμε το weka.



Εικόνα 16: WEKA μέσω της ιστοσελίδας

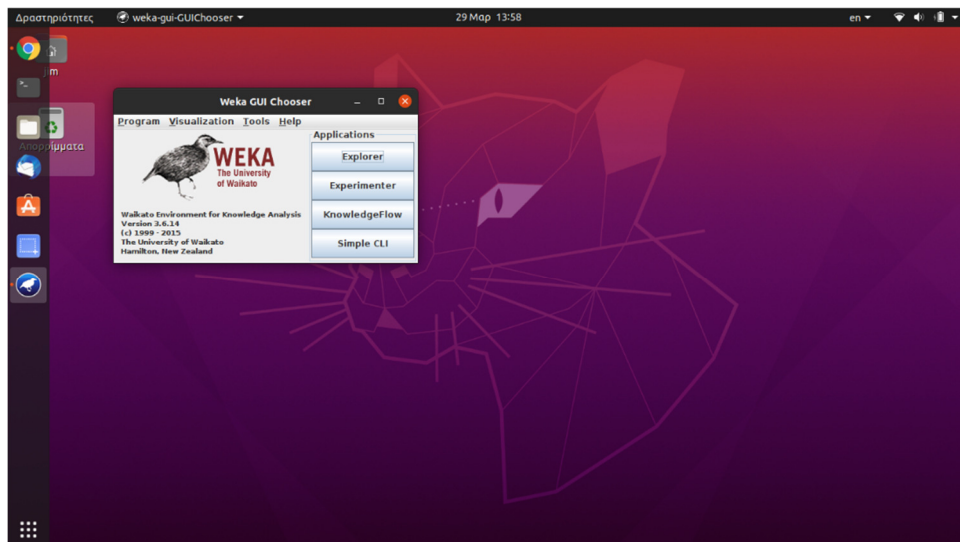
4.2.2.2 Εγκατάσταση WEKA μέσω του τερματικού

1. Πατάμε ταυτόχρονα τα πλήκτρα Ctrl + Alt + T, ώστε να ανοίξει το τερματικό.
2. Πληκτρολογούμε την ακόλουθη εντολή «sudo apt update».
3. Πληκτρολογούμε την ακόλουθη εντολή «sudo apt install weka -y».

Πλέον είναι εγκατεστημένο στον υπολογιστή.



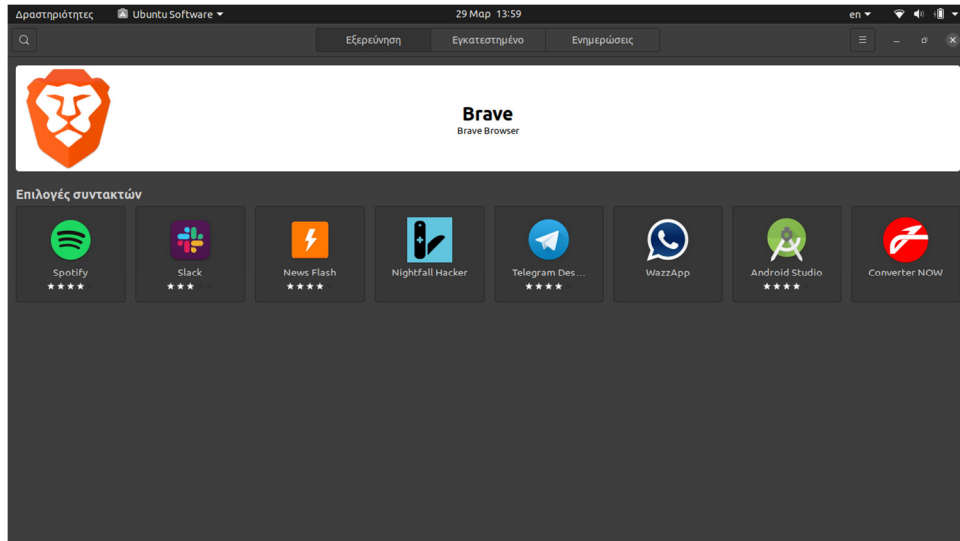
Εικόνα 17: Εικονίδιο του weka



Εικόνα 18: WEKA μέσω του τερματικού

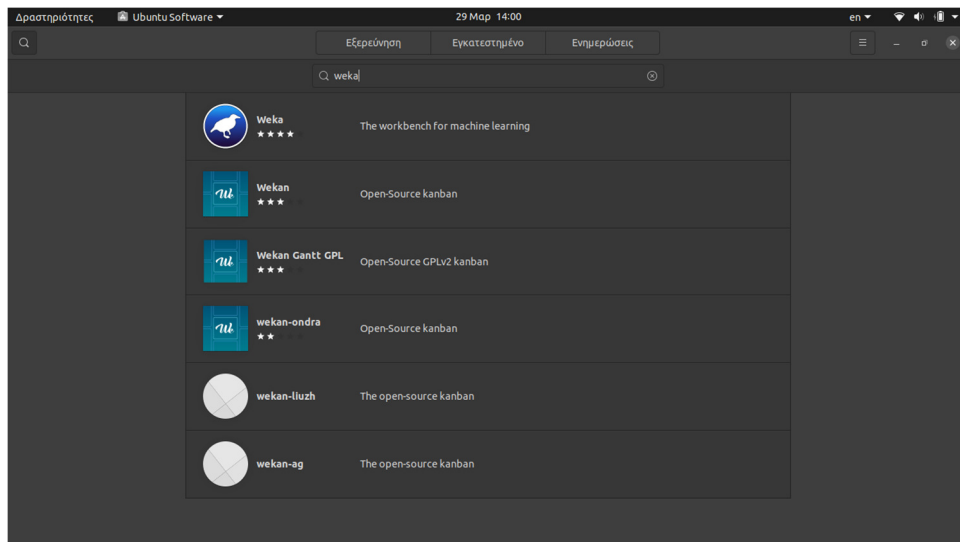
4.2.2.3 Εγκατάσταση WEKA μέσω του Ubuntu Software

Μεταβαίνουμε στην εφαρμογή Ubuntu Software



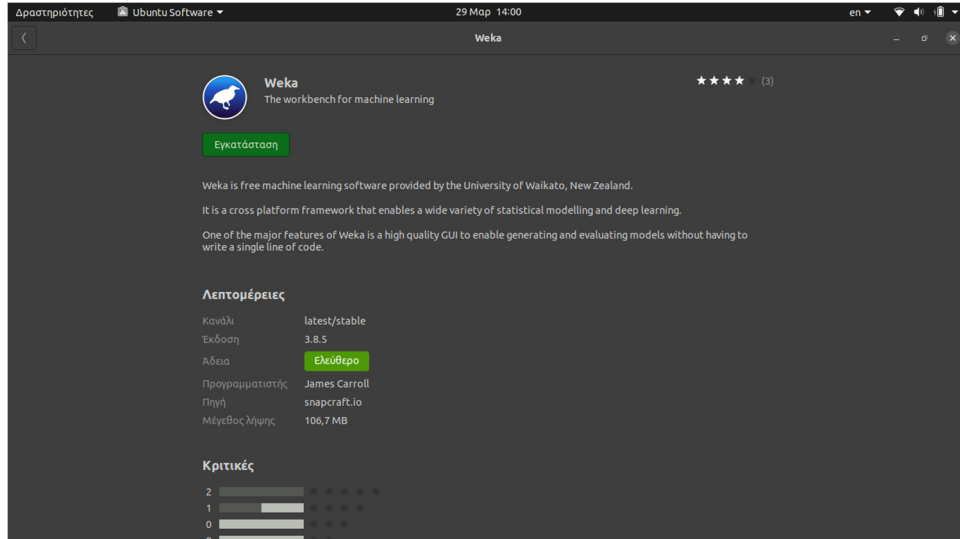
Εικόνα 19: Ubuntu Software

Μπορείτε στην αναζήτηση να πληκτρολογήσετε weka



Εικόνα 20: Ubuntu Software weka

Επιλέγετε το πρώτο αποτέλεσμα όπου είναι το weka και έπειτα το κάνετε εγκατάσταση.



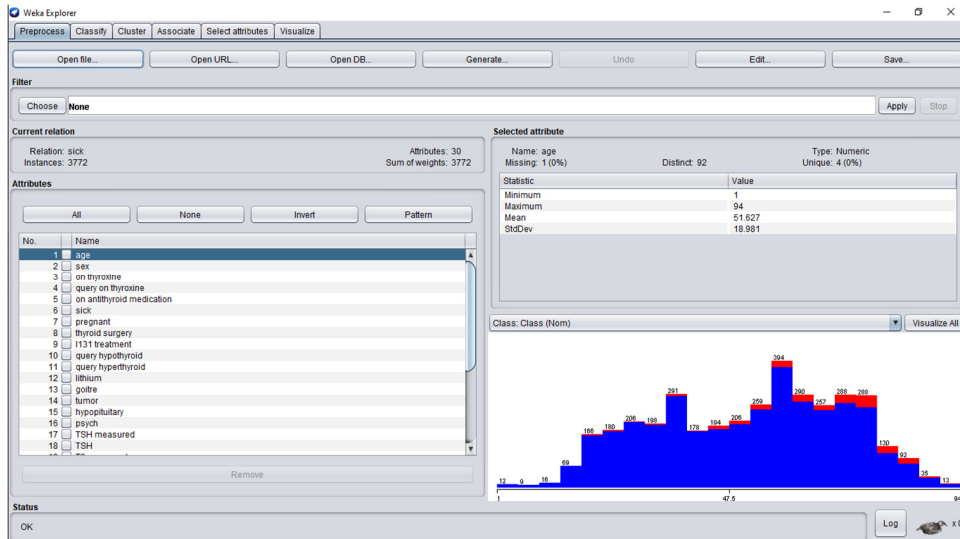
Εικόνα 21: Εγκατάσταση μέσω Ubuntu Software

4.3 Λειτουργίες WEKA

Παρακάτω αναφερόμαστε σε μερικές βασικές λειτουργίες, που θα χρησιμοποιηθούν και στο πέμπτο κεφάλαιο.

4.3.1 Weka Explorer

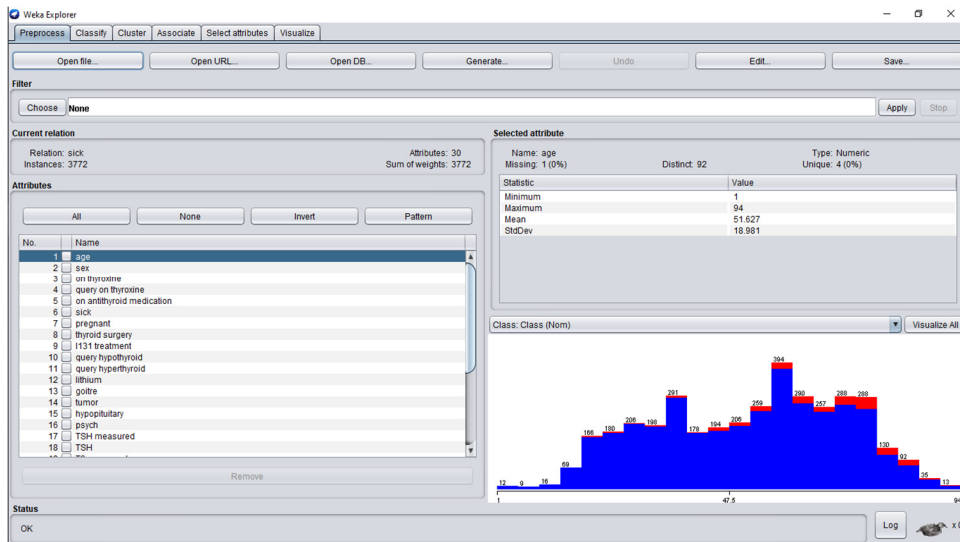
Το Weka Explorer είναι ένα εύκολο στη χρήση γραφικό περιβάλλον. Κάθε ένα από τα μεγάλα πακέτα weka (Filters, Classifiers, Clusterers, Associations, Attribute Selection), παρουσιάζεται στον χρήστη μαζί με ένα εργαλείο οπτικοποίησης, το οποίο επιτρέπει στις προβλέψεις των ταξινομητών και των συμπλεγμάτων, να απεικονίζονται σε δύο διαστάσεις.



Εικόνα 22: Weka Explorer

4.3.1.1 Πίνακας προεπεξεργασίας

Ο πίνακας προεπεξεργασίας είναι το σημείο εκκίνησης, για την εξερεύνηση γνώσεων. Από αυτόν τον πίνακα μπορείτε να φορτώσετε σύνολα δεδομένων, να περιηγηθείτε στα χαρακτηριστικά των χαρακτηριστικών και να εφαρμόσετε οποιονδήποτε συνδυασμό των μη επιτηρούμενων φίλτρων, στα δεδομένα της Weka.

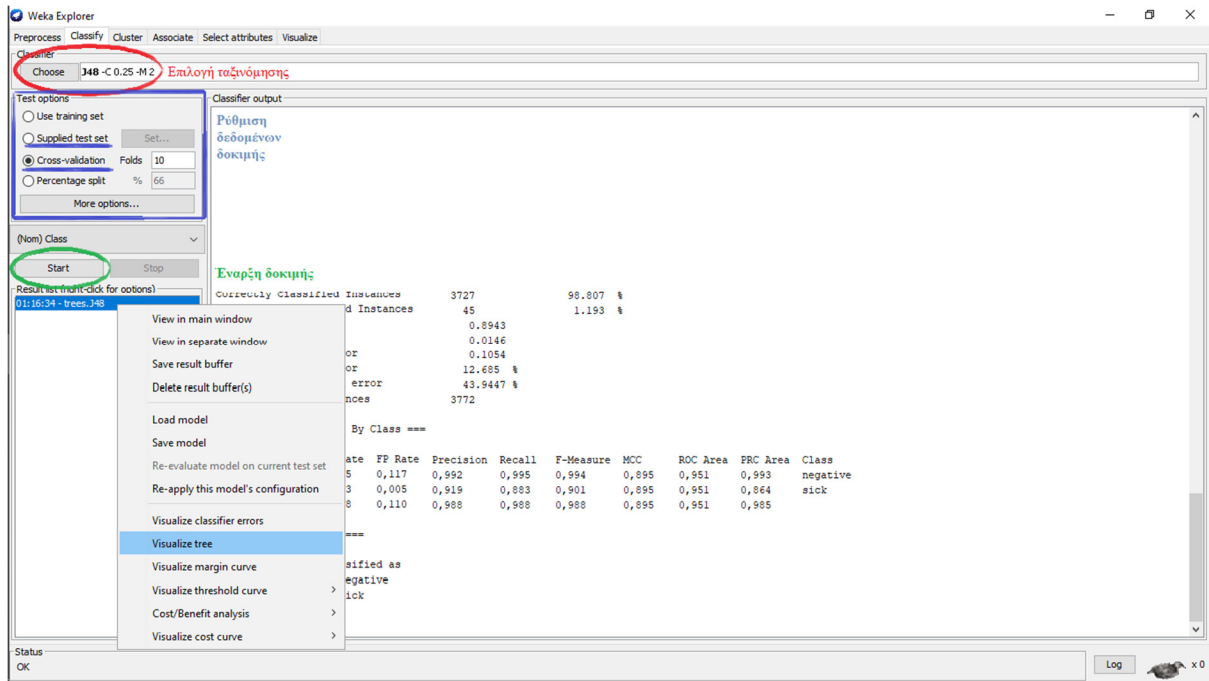


Εικόνα 23: Πίνακας προεπεξεργασίας

4.3.1.2 Πίνακας ταξινόμησης

Ο πίνακας ταξινόμησης, σας επιτρέπει να διαμορφώσετε και να εκτελέσετε οποιονδήποτε από τους ταξινομητές weka, στο τρέχον σύνολο δεδομένων. Μπορείτε να επιλέξετε, να

πραγματοποιήσετε δοκιμή, σε ξεχωριστό σύνολο δεδομένων. Τα σφάλματα ταξινόμησης μπορούν να απεικονιστούν σε ένα αναδυόμενο εργαλείο οπτικοποίησης δεδομένων. Εάν ο ταξινομητής παράγει ένα δέντρο αποφάσεων, μπορεί να εμφανίζεται γραφικά σε ένα αναδυόμενο δέντρο απεικόνισης.



Εικόνα 24: Πίνακας Ταξινόμησης

4.3.1.2.1 Ταξινομητές weka

Παρακάτω, θα αναφέρουμε μερικούς από τους βασικούς ταξινομητές του weka.

4.3.1.2.1.1 Κανόνες ταξινόμησης - Classifiers rules

- **ZeroR** είναι η απλούστερη μέθοδος ταξινόμησης, που βασίζεται στον στόχο και αγνοεί όλους τους προγνωστικούς παράγοντες. Ο ταξινομητής ZeroR, απλώς προβλέπει την κατηγορία πλειοψηφίας (class). Αν και δεν υπάρχει δυνατότητα προβλεψιμότητας στο ZeroR, είναι χρήσιμο για τον καθορισμό μιας απόδοσης βάσης, ως σημείο αναφοράς για άλλες μεθόδους ταξινόμησης.
- **OneR** (One Rule), είναι ένας απλός, αλλά ακριβής, αλγόριθμος ταξινόμησης που δημιουργεί έναν κανόνα, για κάθε προγνωστικό παράγοντα στα δεδομένα και στη συνέχεια επιλέγει τον κανόνα με το μικρότερο συνολικό σφάλμα, ως τον "έναν κανόνα". Για να δημιουργήσουμε έναν κανόνα για μια πρόβλεψη, κατασκευάζουμε έναν πίνακα συχνοτήτων, για κάθε προγνωστικό παράγοντα, έναντι του στόχου.

- **Decision table** δημιουργεί μοντέλα ταξινόμησης ή παλινδρόμησης με τη μορφή μιας δομής δέντρου. Διασπά ένα σύνολο δεδομένων, σε μικρότερα και μικρότερα υποσύνολα, ενώ ταυτόχρονα αναπτύσσεται σταδιακά ένα σχετικό δέντρο αποφάσεων. Το τελικό αποτέλεσμα είναι ένα δέντρο με κόμβους αποφάσεων και κόμβους φύλλων. Ο κορυφαίος κόμβος απόφασης σε ένα δέντρο, αντιστοιχεί στον καλύτερο προγνωστικό παράγοντα και ονομάζεται root node. Τα δέντρα αποφάσεων μπορούν να χειριστούν, τόσο κατηγορικά, όσο και αριθμητικά δεδομένα.
- **JRip** εφαρμόζει έναν μαθητευόμενο κανόνα προτάσεων, εκτελεί επαναλαμβανόμενο κλιμακωτό κλάδεμα, για παραγωγή μείωσης σφάλματος (RIPPER), το οποίο προτάθηκε από τον William W. Cohen, ως βελτιστοποιημένη έκδοση του IREP (Αγαχ, 2017). JRip είναι μια μέθοδος, από κάτω προς τα πάνω, που μαθαίνει κανόνες με την αντιμετώπιση συγκεκριμένης κρίσης των παραδειγμάτων στα δεδομένα εκπαίδευσης, ως τάξη και εύρεση του συνόλου κανόνων, που καλύπτουν όλα τα μέλη της τάξης.

4.3.1.2.1.2 Ρυθμίσεις δεδομένων δοκίμων

Στην υποενότητα αυτή θα δούμε πως μπορούμε να «εκπαιδεύσουμε» τα δεδομένα μας, ώστε να έχουμε το καλύτερο ποσοστό επιτυχίας. Για να βρούμε το καλύτερο ποσοστό επιτυχίας, θα πρέπει να κάνουμε αρκετές δοκιμές, με διαφορετικούς ταξινομητές και συντελεστές, όπου θα αναφερθούν παρακάτω, αλλά και στο επόμενο κεφάλαιο. Οι επιλογές είναι:

- Training set
- Supplied test set
- Cross-validation
- Percentage split

Για τις επιλογές **Training set** και **Supplied test set**, θα πρέπει να έχουμε υπό την κατοχή μας ένα εκπαιδευτικό σετ. Οπότε εμείς θα ασχοληθούμε με τα **Cross-validation** και **Percentage split**.

Στην **Cross-validation**, μπορείτε να ορίσετε τον αριθμό, που παράγει σετ ίσου μεγέθους, στις οποίες θα χωρίζονται και θα χρησιμοποιούνται ολόκληρα δεδομένα, κατά τη διάρκεια κάθε επανάληψης της προπόνησης. Για μεγαλύτερη κατανόηση, σπάμε το αρχικό δείγμα μας (100) σε μικρότερα, όπου κάθε σύνολο χωρίζεται σε δύο ομάδες:

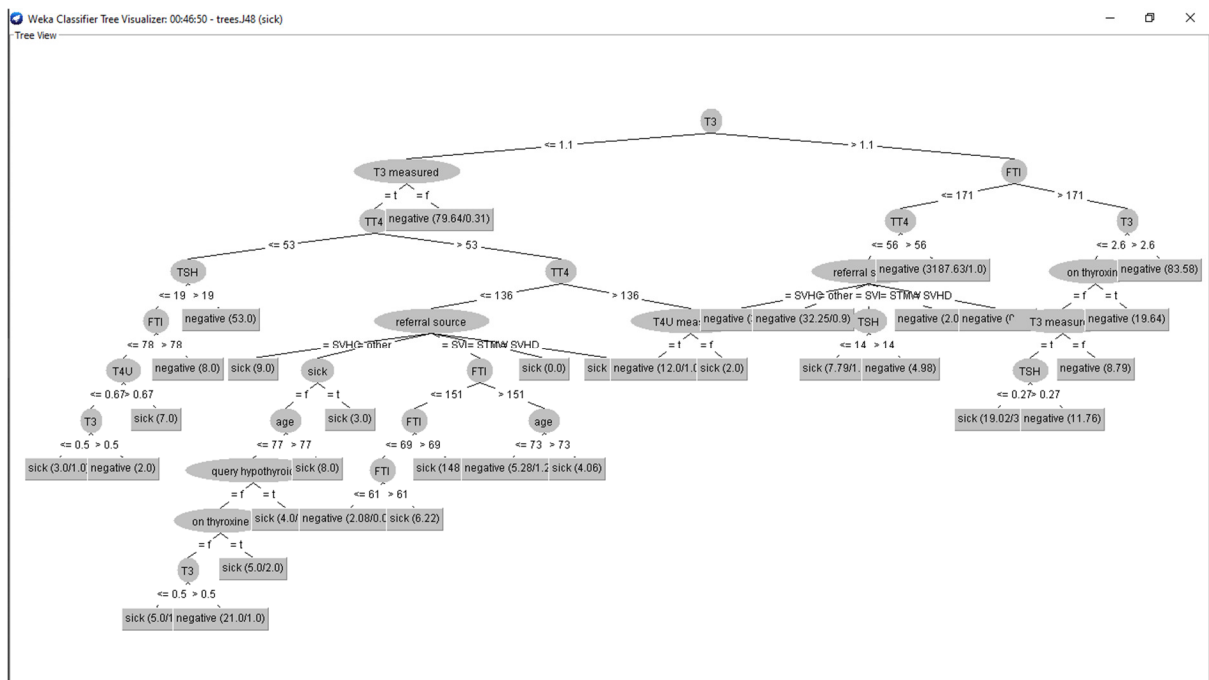
- Α ομάδα: 90 δεδομένα με ετικέτα χρησιμοποιούνται για εκπαίδευση και
- Β ομάδα: 10 δεδομένα με ετικέτα χρησιμοποιούνται για δοκιμή.

Παράγει έναν ταξινομητή με έναν αλγόριθμο από 90 δεδομένα με ετικέτα και το εφαρμόζει στα 10 δεδομένα δοκιμών για το «σύνολο 1». Κάνει το ίδιο πράγμα για το «σύνολο 2», συνεχίζει έως το «σύνολο 10» και παράγει 9 ακόμη ταξινομητές. Μέσος όρος της απόδοσης των 10 ταξινομητών που παράγονται από 10 σύνολα ίσου μεγέθους (90 προπόνηση και 10 δοκιμές). Οι ιδανικές τιμές είναι από 5 έως 12.

Στο **Percentage split**, θα διαχωρίσετε τα δεδομένα μεταξύ εκπαίδευσης και δοκιμών, χρησιμοποιώντας το καθορισμένο ποσοστό διαχωρισμού το οποίο εισάγουμε. Τα ιδανικά ποσοστά είναι από 55% έως 85%.

4.3.1.2.2 Οπτικοποίηση αποτελεσμάτων

Για να δείτε την οπτική αναπαράσταση των αποτελεσμάτων, κάντε δεξί κλικ στο αποτέλεσμα στο πλαίσιο Visualize tree (μπορείτε να το δείτε οπτικοποιημένο στην Εικόνα 24: Πίνακας Ταξινόμησης σελ.37).



Εικόνα 25: Visualize tree

5 ΕΞΟΡΥΞΗ ΓΝΩΣΗΣ ΜΕ ΤΗΝ ΕΦΑΡΜΟΓΗ WEKA ΑΠΟ ΙΑΤΡΙΚΑ ΔΕΔΟΜΕΝΑ ΤΗΣ ΝΟΣΟΥ “ΘΥΡΕΟΕΙΔΗ”

Τα ιατρικά δεδομένα που θα χρησιμοποιηθούν στο 5^ο Κεφάλαιο, για την υλοποίηση της παρούσας πτυχιακής εργασίας, σχετίζονται με τη νόσο του θυρεοειδή και έχουν παρθεί από τα αρχεία που παρέχονται στο Garavan Institute and J. Ross, σχετικά με τους ασθενείς αυτής της νόσου. Η συλλογή των δεδομένων έγινε στην Αυστραλία και συγκεκριμένα στις πολιτείες Quinlan, New South Wales το 1987. Υπάρχουν στους ιστότοπους [UCI Machine Learning Repository](#) και [UC Irvine Machine Learning Repository \(beta version\)](#). Αυτό το σύνολο δεδομένων διαθέτει άδεια χρήσης από το [Creative Commons Attribution 4.0 International](#)(CC BY 4.0). Συνεπώς, αυτό επιτρέπει την κοινή χρήση και προσαρμογή των συνόλων δεδομένων, για οποιονδήποτε σκοπό. Σε περίπτωση όπου θέλετε να δείτε την βάση δεδομένων, παρακαλώ επισκεφτείτε τον [σύνδεσμο](#). Η βάση ονομάζεται sick.data και από αρχείο .bata το μετατρέψαμε σε αρχείο .arff⁴.

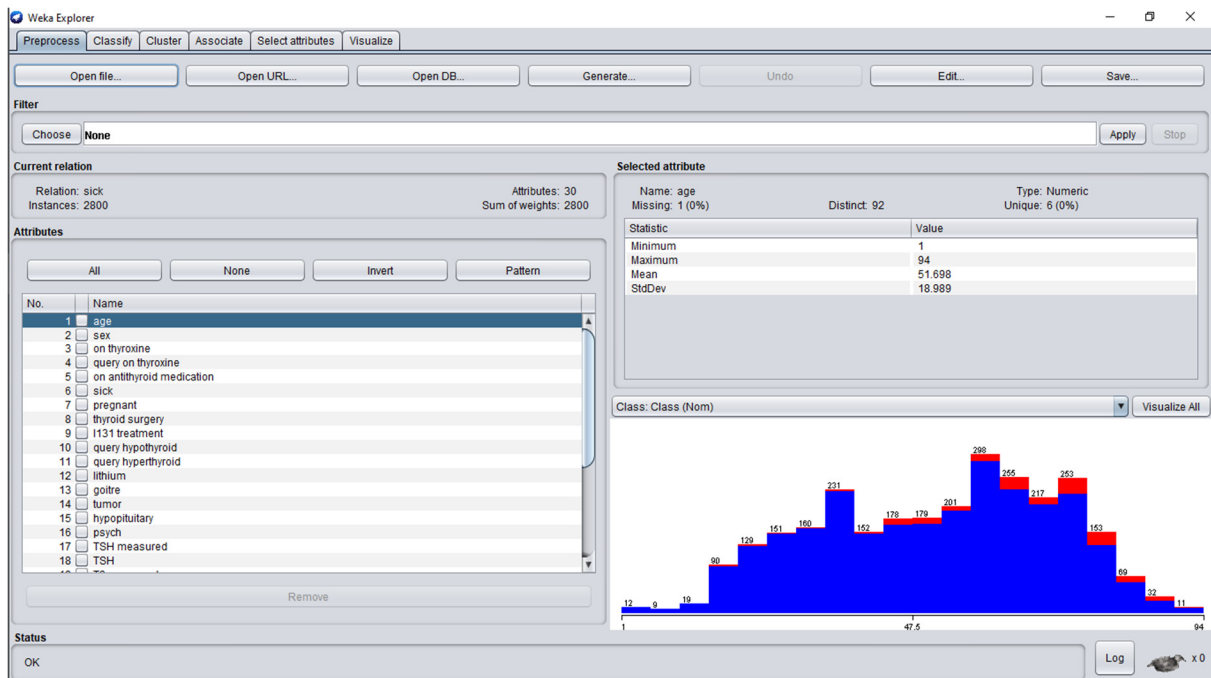
5.1 Ανάλυση των ιατρικών δεδομένων

Παρακάτω πραγματοποιείται στατιστική ανάλυση στα χαρακτηριστικά της έρευνας μας. Συγκεκριμένα αναλύουμε τα τριάντα (30) χαρακτηριστικά από τους δύο χιλιάδες οκτακόσιους (2.800) εξεταζόμενους.

⁴ ARFF: είναι απλά αρχεία κειμένου, όπου οι τιμές διαχωρίζονται με κόμμα. Το αρχείο περιέχει μια επικεφαλίδα, στην οποία ορίζονται το όνομα της σχέσης και τα πεδία. Στην αρχή του αρχείου αναφέρεται η λέξη “@relation” και ακολουθεί το όνομα του πίνακα δεδομένων. Στη συνέχεια γίνεται η δήλωση των πεδίων. Για κάθε πεδίο χρειάζεται μια γραμμή, στην αρχή της οποίας υπάρχει η λέξη “@attribute”, ακολουθεί το όνομα του πεδίου και κατόπιν δηλώνεται ο τύπος του πεδίου. Αν το πεδίο είναι αριθμητικό, χρησιμοποιείται η λέξη “numeric”. Αν το πεδίο είναι ονομαστικό, δηλώνονται οι δυνατές τιμές μέσα σε αγκύλες. Μετά τη δήλωση των πεδίων ακολουθούν τα δεδομένα. Πριν από τα κάθε αυτό δεδομένα υπάρχει μια γραμμή με τη λέξη “@data”. Τα δεδομένα είναι τιμές, οι οποίες χωρίζονται με κόμμα.

5.1.1 Χαρακτηριστικό «age»

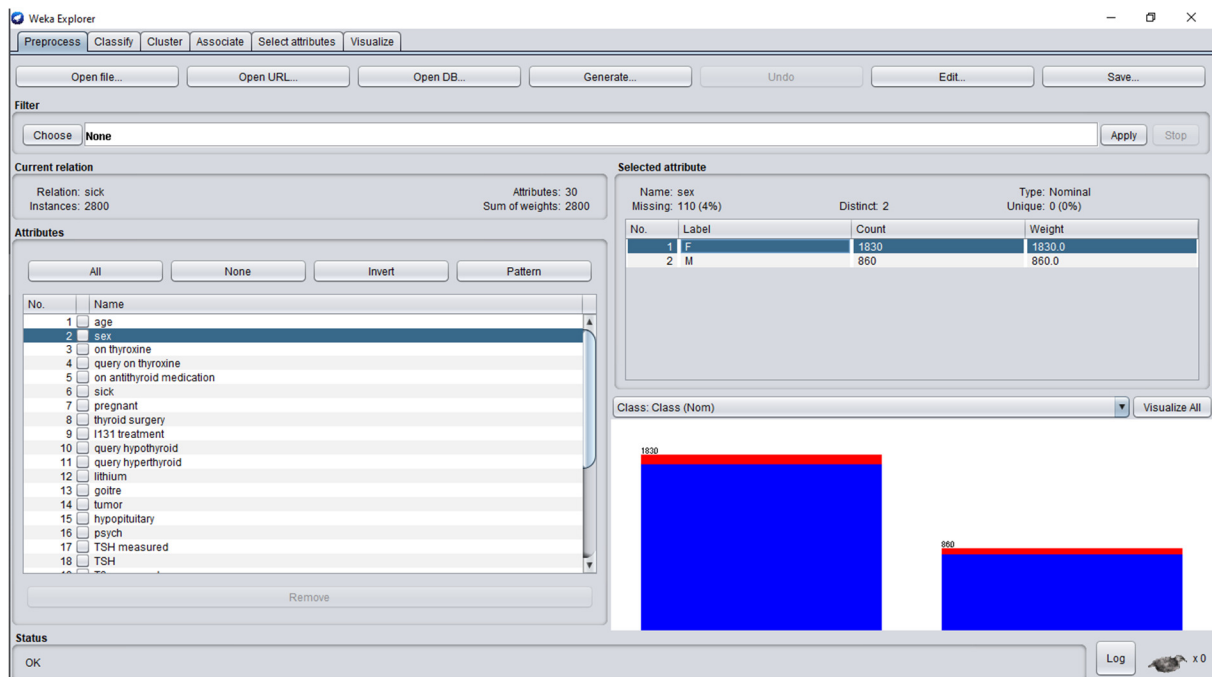
Από την στατιστική ανάλυση της μεταβλητής age (ηλικία εξεταζόμενου), προέκυψε ότι, στο σύνολο των 2.800 εξεταζόμενων, για ενδεχόμενη πάθηση της νόσο του θυρεοειδή, η μέση ηλικία των εξεταζόμενων είναι 52 ετών, με τυπική απόκλιση 18,989. Η διαφορά ανάμεσα στη μεγαλύτερη ηλικία και την μικρότερη ηλικία είναι 93 χρόνια. Συγκεκριμένα, ο νεότερος είναι 1 έτους, ενώ ο γηραιότερος είναι 94 ετών. Ένας εξεταζόμενος δεν απάντησε την ηλικία του (missing 1). Από το σύνολο των εξεταζόμενων, 6 άτομα έχουν μοναδική ηλικία (unique 6). Το πλήθος των ηλικιών που καταγράφηκαν είναι 92. Βλέπουμε πως οι ηλικίες όπου ήταν οι περισσότεροι που έκαναν την εξέταση για την νόσο του θυρεοειδή, ήταν από 15 έως 75 ετών.



Εικόνα 26: Στατιστική ανάλυση του χαρακτηριστικού «age»

5.1.2 Χαρακτηριστικό «sex»

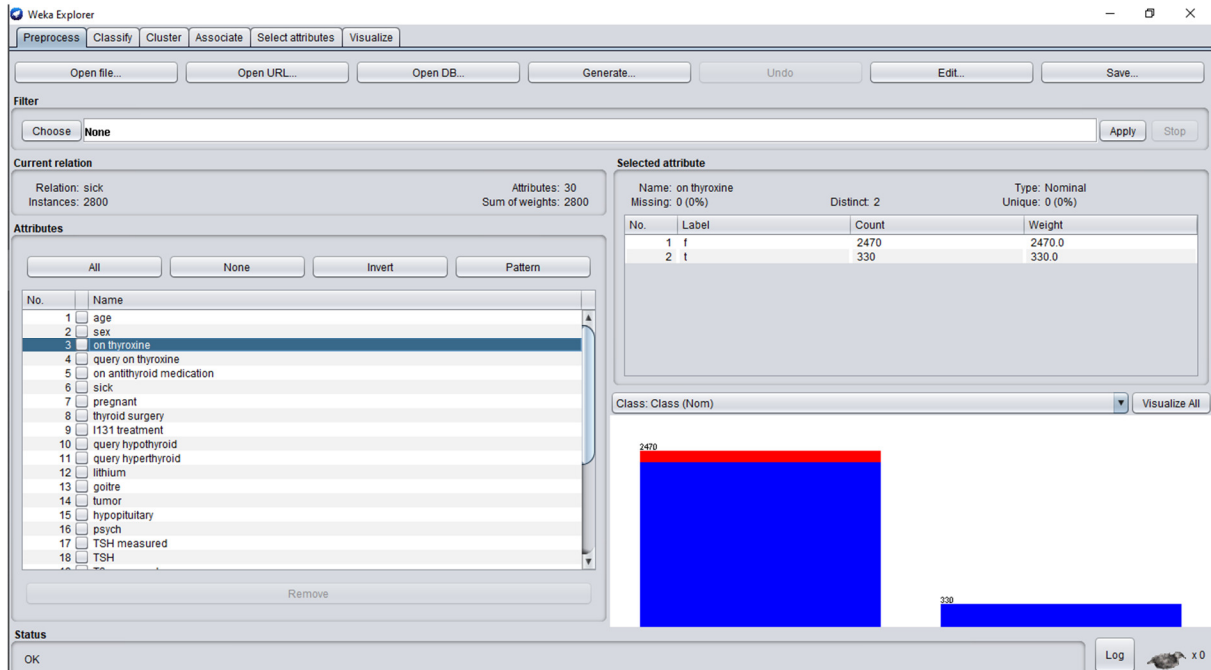
Από την στατιστική ανάλυση της μεταβλητής sex (φύλο εξεταζόμενου), προέκυψε ότι, στο σύνολο των 2.800 εξεταζόμενων, για ενδεχόμενη πάθηση της νόσο του θυρεοειδή, τα 1.830 άτομα είναι γυναίκες, ενώ τα 860 άτομα είναι άνδρες. Οι εξεταζόμενοι που απάντησαν το φύλο τους ήταν 2.690 άτομα, ενώ 110 άτομα δεν απάντησαν (missing 110).



Εικόνα 27: Στατιστική ανάλυση του χαρακτηριστικού «sex»

5.1.3 Χαρακτηριστικό «on thyroxine»

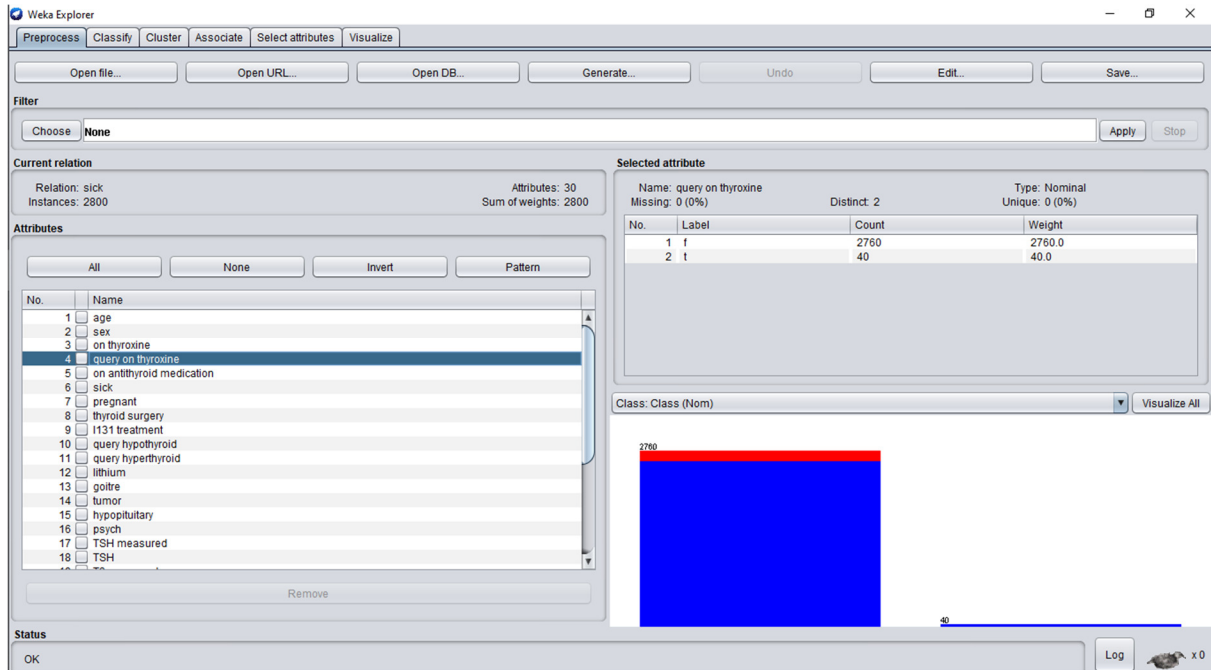
Από την στατιστική ανάλυση της μεταβλητής on thyroxine [εξέταση για θυροξίνη(T4)], προέκυψε ότι, στο σύνολο των 2.800 εξεταζόμενων, για ενδεχόμενη πάθηση της νόσο του θυρεοειδή, τα 330 άτομα ήταν θετικά στην εξέταση για θυροξίνη, ενώ τα 2.470 άτομα ήταν αρνητικά στην εξέταση για θυροξίνη. Όλοι οι εξεταζόμενοι υποβλήθηκαν στην εξέταση για την θυροξίνη (missing 0).



Εικόνα 28: Στατιστική ανάλυση του χαρακτηριστικού «on thyroxine»

5.1.4 Χαρακτηριστικό «query on thyroxine»

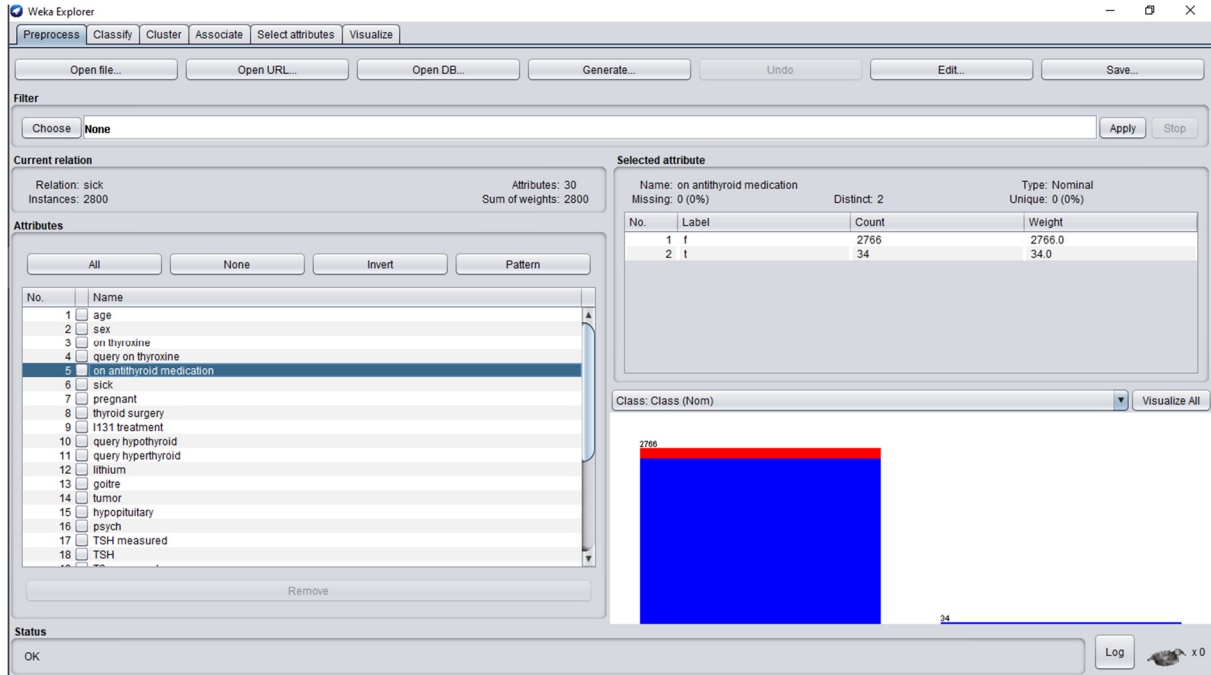
Από την στατιστική ανάλυση της μεταβλητής query on thyroxine (ασθενείς όπου είχαν ερωτηθεί αν είχαν θυροξίνη), προέκυψε ότι, στο σύνολο των 2.800 εξεταζόμενων, για ενδεχόμενη πάθηση της νόσο του θυρεοειδή, τα 2.760 άτομα δεν γνώριζαν αν έχουν ή δεν έχουν θυροξίνη, ενώ τα 40 άτομα είχαν θυροξίνη κατά την διάρκεια της εξέτασης. Όλοι οι εξεταζόμενοι απαντήσαν για το αν έχουν ή όχι θυροξίνη (missing 0).



Εικόνα 29: Στατιστική ανάλυση του χαρακτηριστικού «query on thyroxine»

5.1.5 Χαρακτηριστικό «on antithyroid medication»

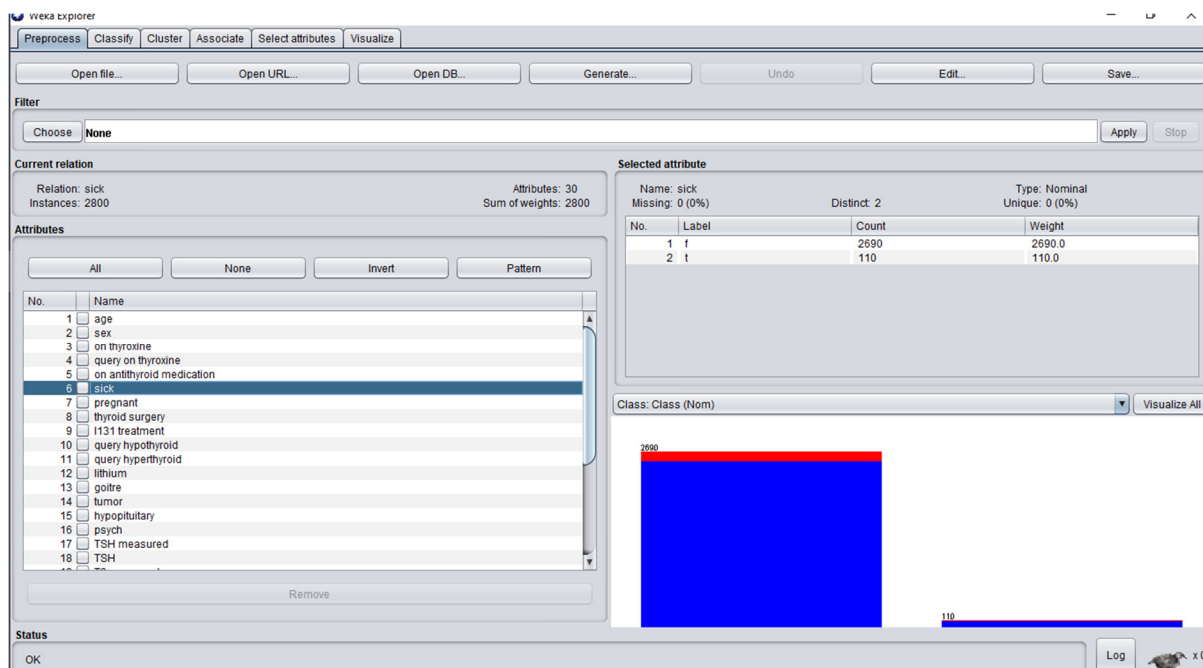
Από την στατιστική ανάλυση της μεταβλητής on antithyroid medication (αντιβιοτικά για αντιμετώπιση του θυρεοειδή), προέκυψε ότι, στο σύνολο των 2.800 εξεταζόμενων, για ενδεχόμενη πάθηση της νόσο του θυρεοειδή, τα 2.766 άτομα δεν έλαβαν αντιβιοτικά, ενώ τα 34 άτομα έλαβαν αντιβιοτικά. Όλοι οι εξεταζόμενοι απαντήσαν για την χορήγηση αντιβιοτικών για την καταπολέμηση του θυρεοειδή (missing 0).



Εικόνα 30: Στατιστική ανάλυση του χαρακτηριστικού «on antithyroid medication»

5.1.6 Χαρακτηριστικό «sick»

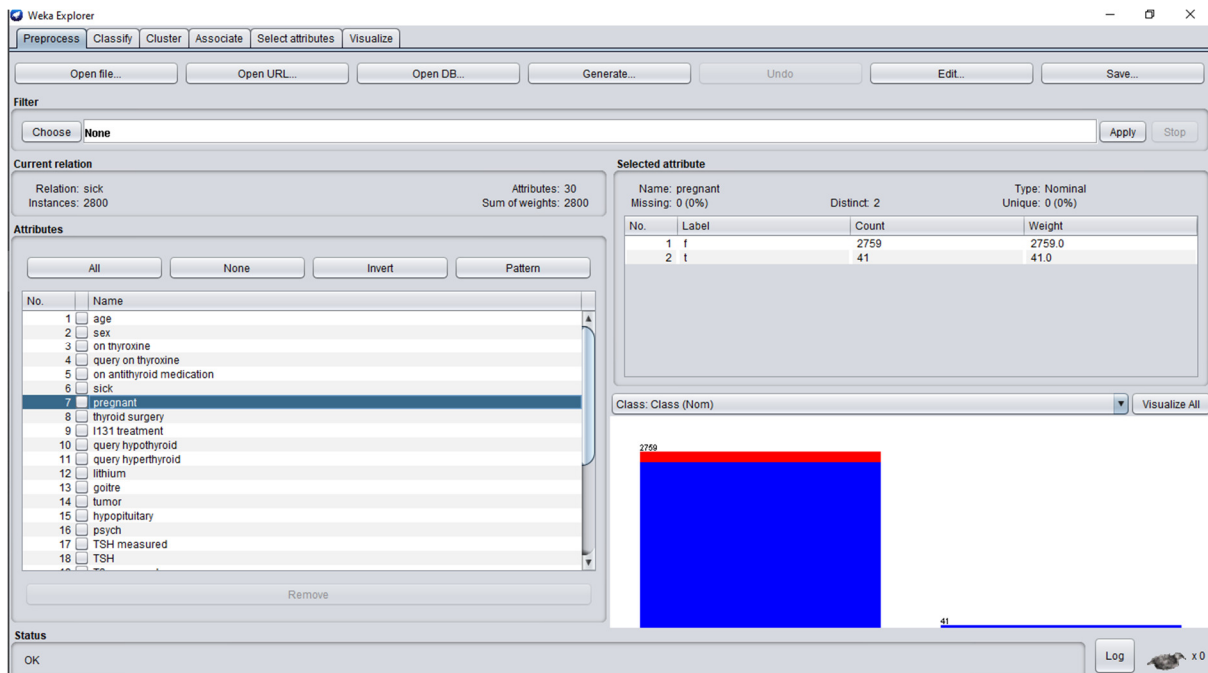
Από την στατιστική ανάλυση της μεταβλητής sick (σθενείς όπου ήταν άρρωστοι κατά την εξέταση), προέκυψε ότι, στο σύνολο των 2.800 εξεταζόμενων, για ενδεχόμενη πάθηση της νόσο του θυρεοειδή, τα 2.690 άτομα δεν ήταν άρρωστα κατά την διάρκεια της εξέτασης, ενώ τα 110 άτομα ήταν άρρωστα κατά την διάρκεια της εξέτασης. Όλοι οι εξεταζόμενοι απαντήσαν για το αν είναι άρρωστη (missing 0).



Εικόνα 31: Στατιστική ανάλυση του χαρακτηριστικού «sick»

5.1.7 Χαρακτηριστικό «pregnant»

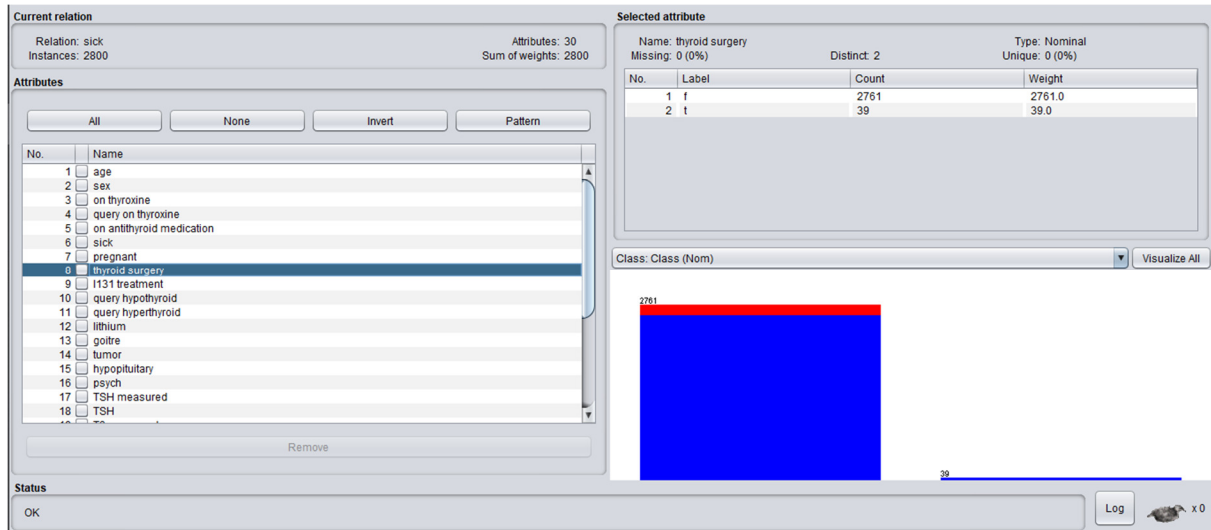
Από την στατιστική ανάλυση της μεταβλητής pregnant (ασθενείς όπου ήταν έγκυοι κατά την εξέταση), προέκυψε ότι, στο σύνολο των 2.800 εξεταζόμενων, για ενδεχόμενη πάθηση της νόσο του θυρεοειδή, τα 2.759 άτομα δεν κυοφορούσαν ή ήταν άνδρες, ενώ τα 41 γυναίκες κυοφορούσαν κατά την διάρκεια της εξέτασης. Όλοι οι εξεταζόμενοι απαντήσαν για το αν είναι έγκυοι (missing 0).



Εικόνα 32: Στατιστική ανάλυση του χαρακτηριστικού «pregnant»

5.1.8 Χαρακτηριστικό «thyroid surgery»

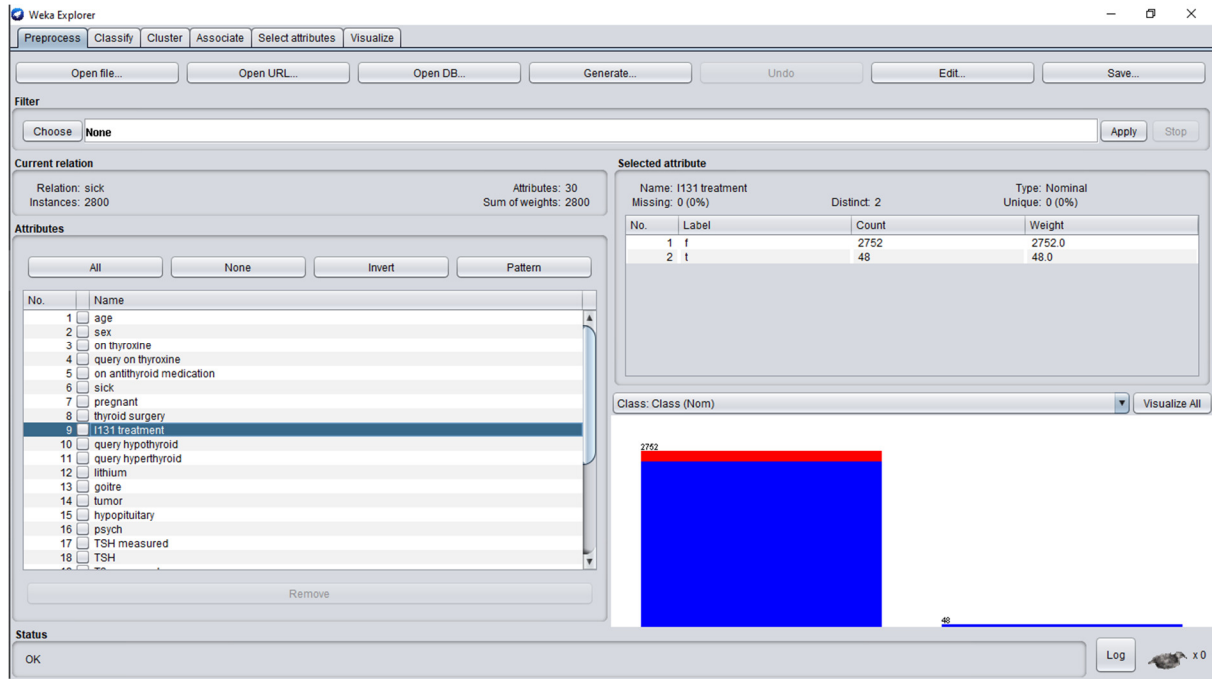
Από την στατιστική ανάλυση της μεταβλητής thyroid surgery (ασθενείς όπου είχαν υποβληθεί σε χειρουργική επέμβαση θυρεοειδούς πριν από την εξέταση), προέκυψε ότι, στο σύνολο των 2.800 εξεταζόμενων, για ενδεχόμενη πάθηση της νόσο του θυρεοειδή, τα 2.761 άτομα δεν είχαν υποβληθεί σε χειρουργική επέμβαση θυρεοειδή, ενώ τα 39 άτομα είχαν υποβληθεί σε χειρουργική επέμβαση θυρεοειδή κατά την διάρκεια της εξέτασης. Όλοι οι εξεταζόμενοι απαντήσαν για το αν είχαν υποβληθεί σε χειρουργική επέμβαση θυρεοειδη (missing 0).



Εικόνα 33: Στατιστική ανάλυση του χαρακτηριστικού «thyroid surgery»

5.1.9 Χαρακτηριστικό «I131 treatment»

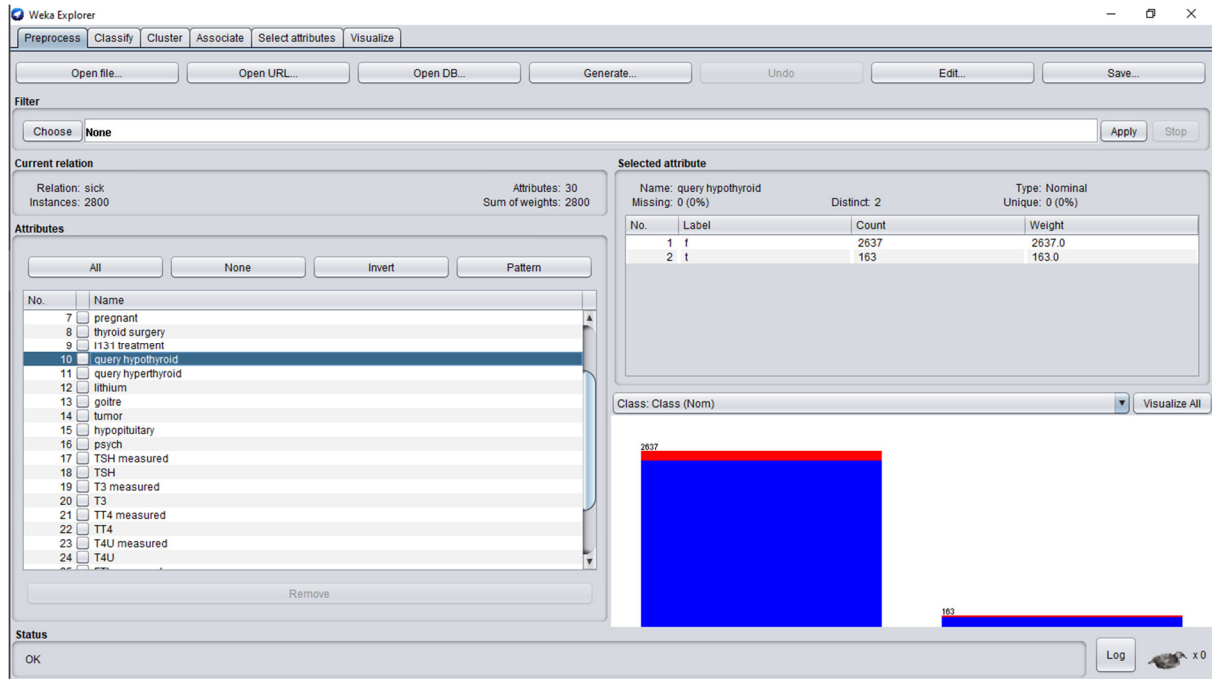
Από την στατιστική ανάλυση της μεταβλητής I131 treatment (ασθενείς όπου είχαν υποβληθεί σε Θεραπεία με Ιώδιο-131 πριν από την εξέταση), προέκυψε ότι, στο σύνολο των 2.800 εξεταζόμενων, για ενδεχόμενη πάθηση της νόσο του θυρεοειδή, τα 2.752 άτομα δεν είχαν υποβληθεί σε Θεραπεία με Ιώδιο-131, ενώ τα 48 άτομα είχαν υποβληθεί σε Θεραπεία με Ιώδιο-131 κατά την διάρκεια της εξέτασης. Όλοι οι εξεταζόμενοι απαντήσαν για το αν είχαν υποβληθεί σε Θεραπεία με Ιώδιο-131 (missing 0).



Εικόνα 34: Στατιστική ανάλυση του χαρακτηριστικού « I131 treatment »

5.1.10 Χαρακτηριστικό «query hypothyroid»

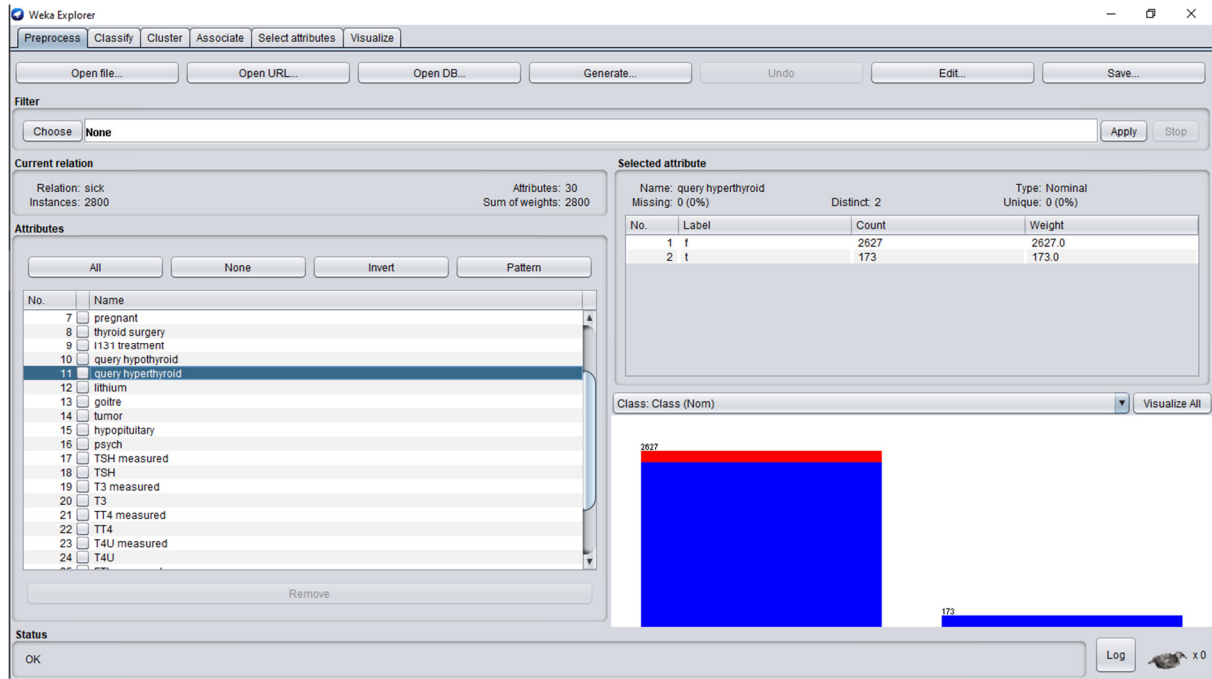
Από την στατιστική ανάλυση της μεταβλητής query hypothyroid (ασθενείς όπου είχαν ερωτηθεί αν είχαν υποθυρεοειδισμό), προέκυψε ότι, στο σύνολο των 2.800 εξεταζόμενων, για ενδεχόμενη πάθηση της νόσο του θυρεοειδή, τα 2.637 άτομα δεν γνώριζαν αν έχουν ή δεν είχαν υποθυρεοειδισμό, ενώ τα 163 άτομα είχαν υποθυρεοειδισμό κατά την διάρκεια της εξέτασης. Όλοι οι εξεταζόμενοι απαντήσαν για το αν είχαν ή όχι υποθυρεοειδισμό (missing 0).



Εικόνα 35: Στατιστική ανάλυση του χαρακτηριστικού « query hypothyroid »

5.1.11 Χαρακτηριστικό «query hyperthyroid»

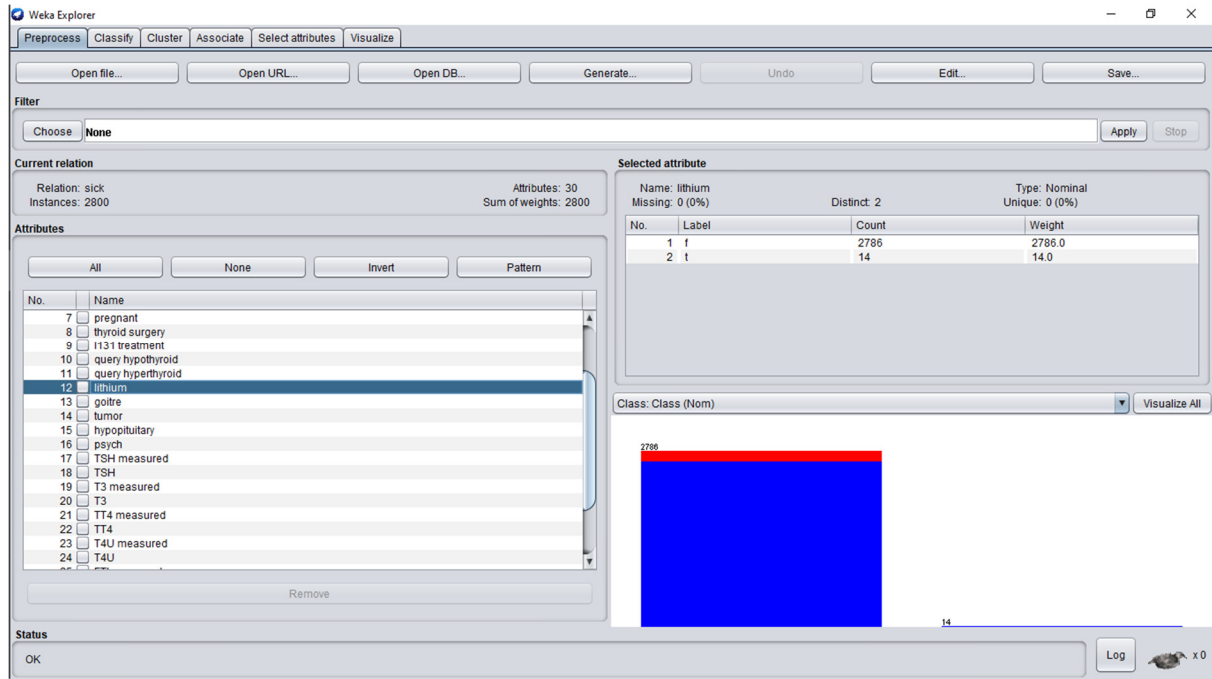
Από την στατιστική ανάλυση της μεταβλητής query hyperthyroid (ασθενείς όπου είχαν ερωτηθεί αν είχαν υπερθυρεοειδισμό), προέκυψε ότι, στο σύνολο των 2.800 εξεταζόμενων, για ενδεχόμενη πάθηση της νόσο του θυρεοειδή, τα 2.627 άτομα δεν γνώριζαν αν έχουν ή δεν είχαν υπερθυρεοειδισμό, ενώ τα 173 άτομα είχαν υπερθυρεοειδισμό κατά την διάρκεια της εξέτασης. Όλοι οι εξεταζόμενοι απαντήσαν για το αν είχαν ή όχι υπερθυρεοειδισμό (missing 0).



Εικόνα 36: Στατιστική ανάλυση του χαρακτηριστικού « query hyperthyroid »

5.1.12 Χαρακτηριστικό «lithium»

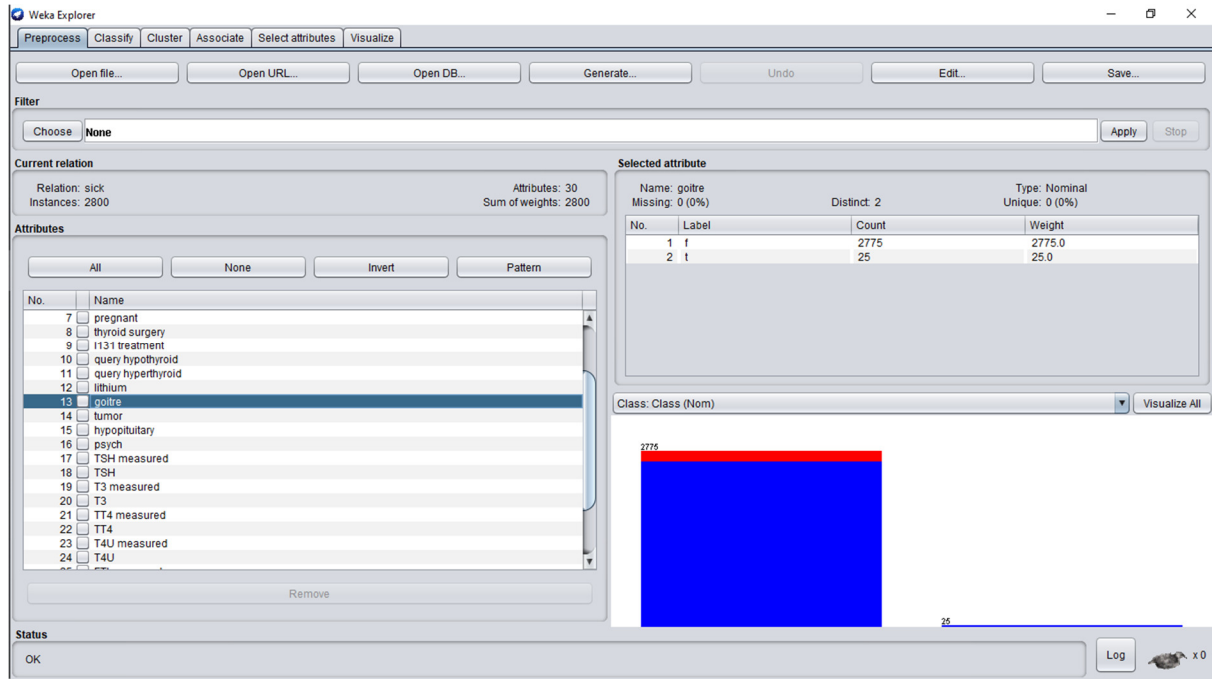
Από την στατιστική ανάλυση της μεταβλητής lithium (ασθενείς όπου είχαν υποβληθεί σε Θεραπεία με λίθιο πριν από την εξέταση), προέκυψε ότι, στο σύνολο των 2.800 εξεταζόμενων, για ενδεχόμενη πάθηση της νόσο του θυρεοειδή, τα 2.786 άτομα δεν είχαν υποβληθεί σε Θεραπεία με λίθιο, ενώ τα 14 άτομα είχαν υποβληθεί σε Θεραπεία με λίθιο κατά την διάρκεια της εξέτασης. Όλοι οι εξεταζόμενοι απαντήσαν για το αν είχαν υποβληθεί σε Θεραπεία με λίθιο (missing 0).



Εικόνα 37: Στατιστική ανάλυση του χαρακτηριστικού « lithium »

5.1.13 Χαρακτηριστικό «goitre»

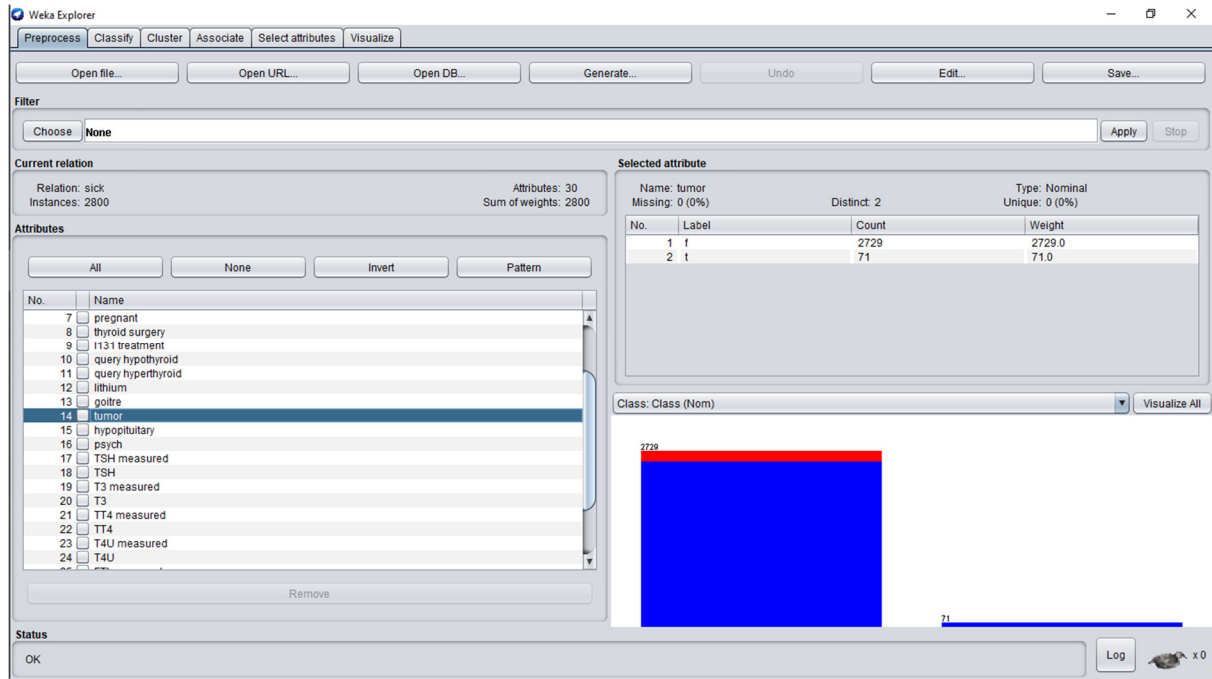
Από την στατιστική ανάλυση της μεταβλητής goiter (ασθενείς όπου είχαν βρογχοκήλη στην εξέταση), προέκυψε ότι, στο σύνολο των 2.800 εξεταζόμενων, για ενδεχόμενη πάθηση της νόσο του θυρεοειδή, τα 2.775 άτομα δεν είχαν βρογχοκήλη, ενώ τα 25 άτομα είχαν βρογχοκήλη κατά την διάρκεια της εξέτασης. Όλοι οι εξεταζόμενοι απαντήσαν για το αν είχαν βρογχοκήλη (missing 0).



Εικόνα 38: Στατιστική ανάλυση του χαρακτηριστικού « goitre »

5.1.14 Χαρακτηριστικό «tumor»

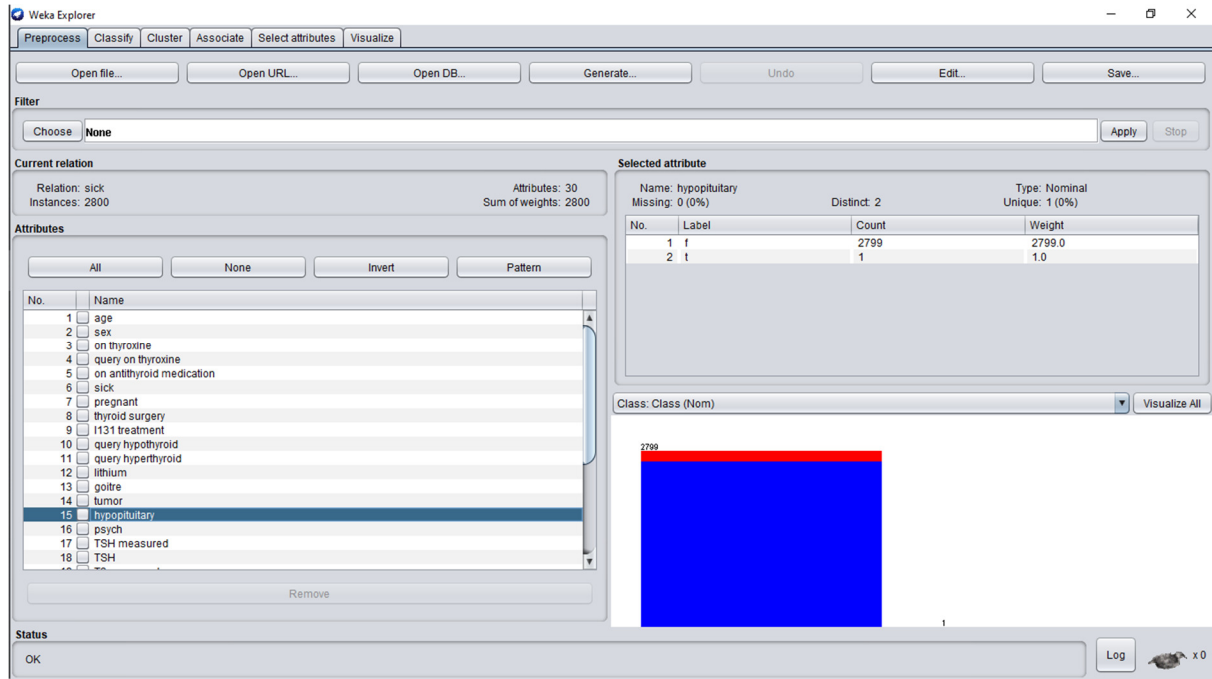
Από την στατιστική ανάλυση της μεταβλητής tumor [ασθενείς όπου είχαν Όζοι (όγκος) κατά την διάρκεια της εξέτασης], προέκυψε ότι, στο σύνολο των 2.800 εξεταζόμενων, για ενδεχόμενη πάθηση της νόσο του θυρεοειδή, τα 2.729 άτομα δεν είχαν Όζους, ενώ τα 25 άτομα είχαν Όζους κατά την διάρκεια της εξέτασης. Όλοι οι εξεταζόμενοι απαντήσαν για το αν είχαν Όζους (missing 0).



Εικόνα 39: Στατιστική ανάλυση του χαρακτηριστικού « tumor »

5.1.15 Χαρακτηριστικό «hypopituitary»

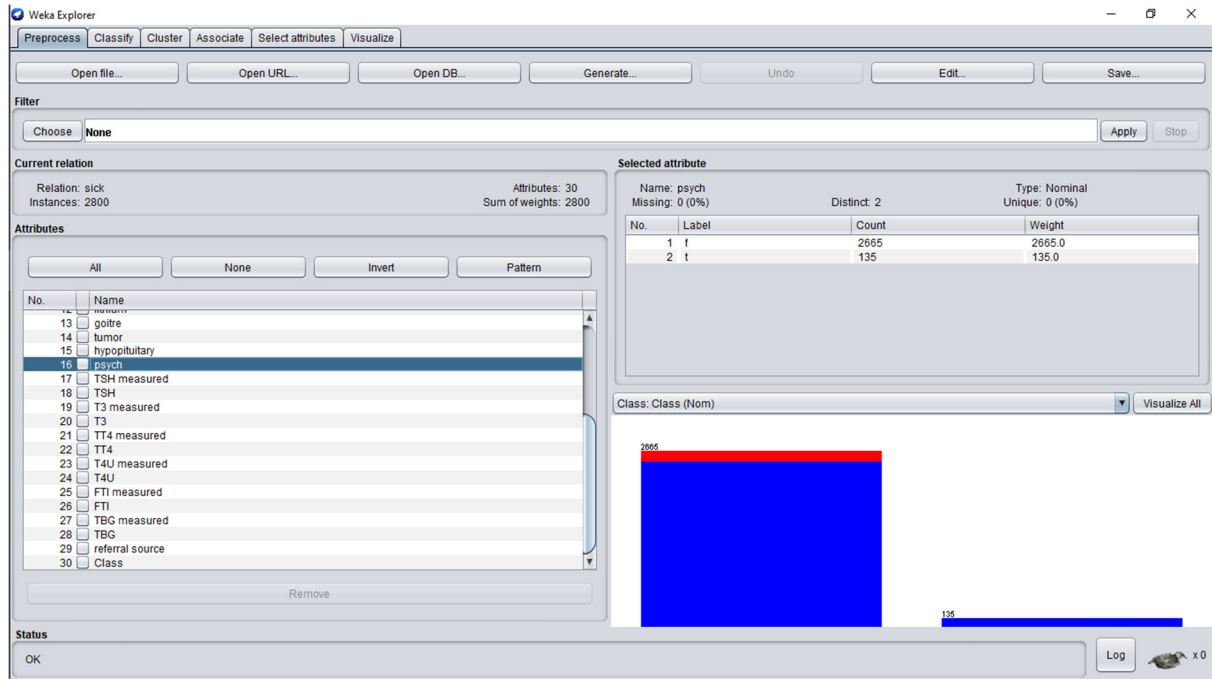
Από την στατιστική ανάλυση της μεταβλητής hypopituitary (ασθενείς όπου είχαν υποφυσιακή αποπληξία κατά την διάρκεια της εξέτασης), προέκυψε ότι, στο σύνολο των 2.800 εξεταζόμενων, για ενδεχόμενη πάθηση της νόσο του θυρεοειδή, τα 2.799 άτομα δεν είχαν υποφυσιακή αποπληξία ενώ 1 άτομο είχε υποφυσιακή αποπληξία κατά την διάρκεια της εξέτασης (Από το σύνολο των εξεταζόμενων 1 άτομο είχε υποφυσιακή αποπληξία (unique 1)). Όλοι οι εξεταζόμενοι απαντήσαν για το αν είχαν υποφυσιακή αποπληξία (missing 0).



Εικόνα 40: Στατιστική ανάλυση του χαρακτηριστικού « hypopituitary »

5.1.16 Χαρακτηριστικό «psych»

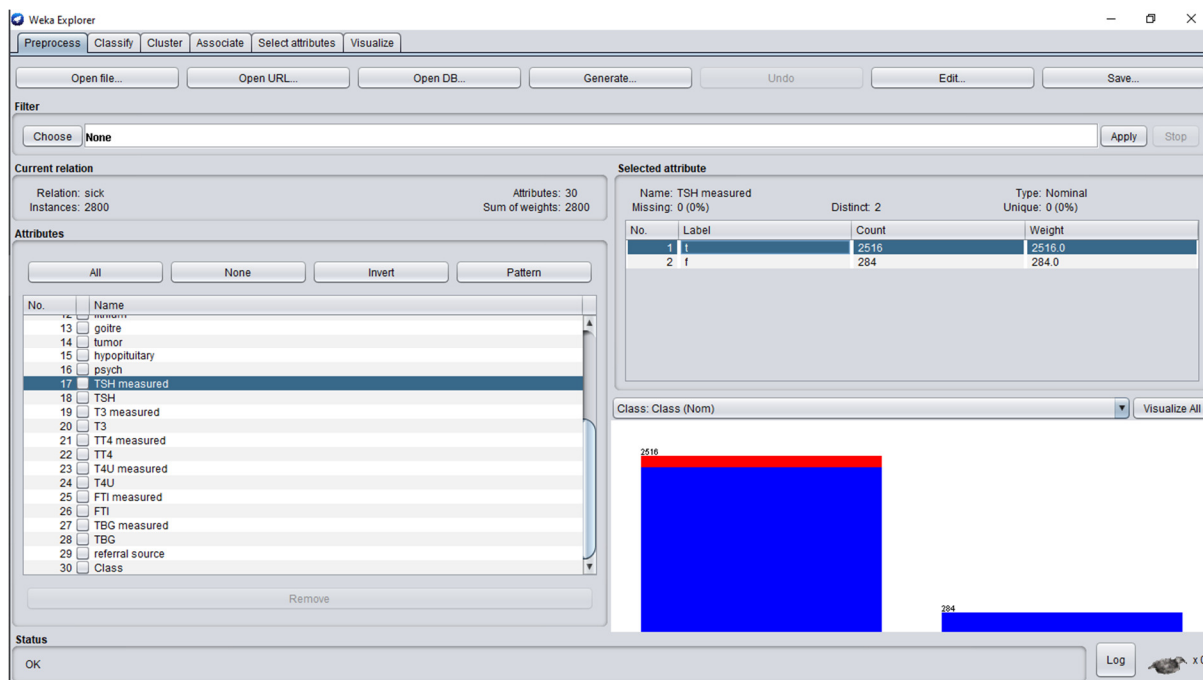
Από την στατιστική ανάλυση της μεταβλητής psych: Ασθενείς όπου διαγνώστηκαν με κάποια ψυχολογική διαταραχή κατά την διάρκεια της εξέτασης, προέκυψαν τα έξις: Στο σύνολο των 2.800 εξεταζόμενων για ενδεχόμενο της νόσο του θυρεοειδούς που μελετήσαμε, τα 2.665 άτομα δεν είχαν διαγνωστεί με κάποια ψυχολογική διαταραχή ενώ τα 135 άτομα είχαν διαγνωστεί με κάποια ψυχολογική διαταραχή κατά την διάρκεια της εξέτασης. Όλοι οι εξεταζόμενοι απάντησαν για το αν είχαν διαγνωστεί με κάποια ψυχολογική διαταραχή (missing 0).



Εικόνα 41: Στατιστική ανάλυση του χαρακτηριστικού « psych »

5.1.17 Χαρακτηριστικό «TSH measured»

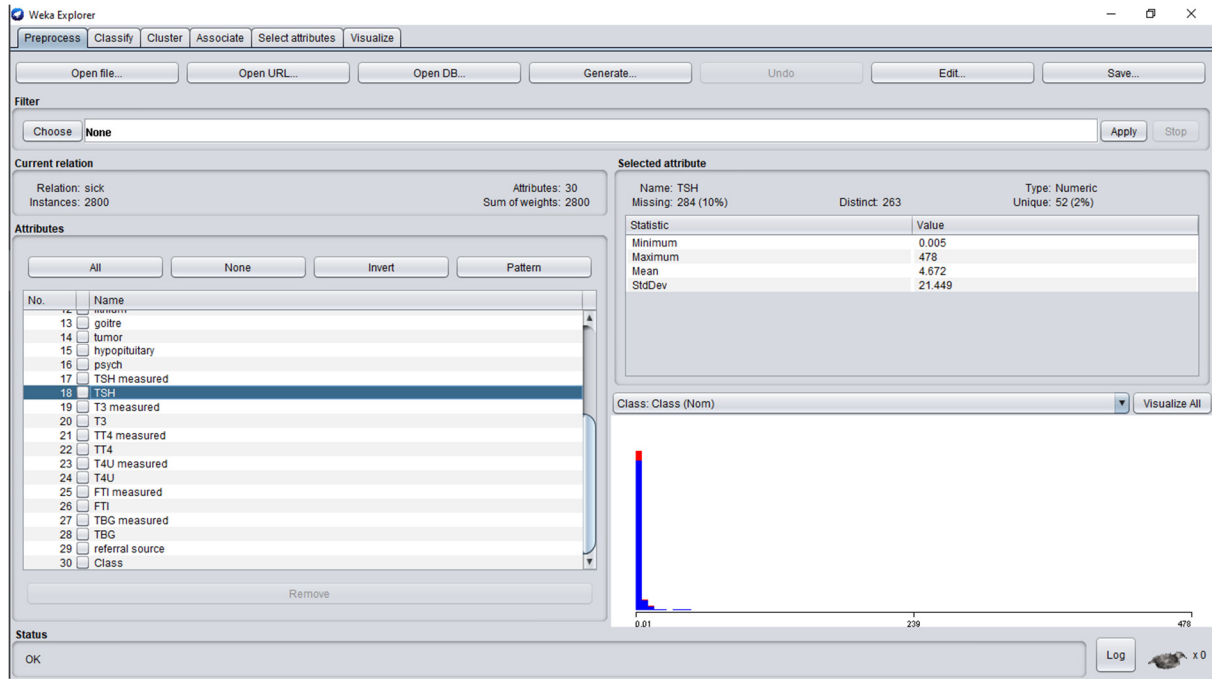
Από την στατιστική ανάλυση της μεταβλητής TSH measured (μέτρηση της ορμόνης διέγερσης του θυρεοειδούς κατά την διάρκεια της εξέτασης), προέκυψε ότι, στο σύνολο των 2.800 εξεταζόμενων, για ενδεχόμενη πάθηση της νόσο του θυρεοειδή, τα 2.516 άτομα είχαν μέτρηση την ορμόνη διέγερσης του θυρεοειδούς, ενώ τα 284 άτομα δεν είχαν μέτρηση την ορμόνη διέγερσης του θυρεοειδούς κατά την διάρκεια της εξέτασης. Όλοι οι εξεταζόμενοι απαντήσαν για το αν είχαν μέτρηση την ορμόνη διέγερσης του θυρεοειδούς (missing 0).



Εικόνα 42: Στατιστική ανάλυση του χαρακτηριστικού « TSH measured »

5.1.18 Χαρακτηριστικό «TSH»

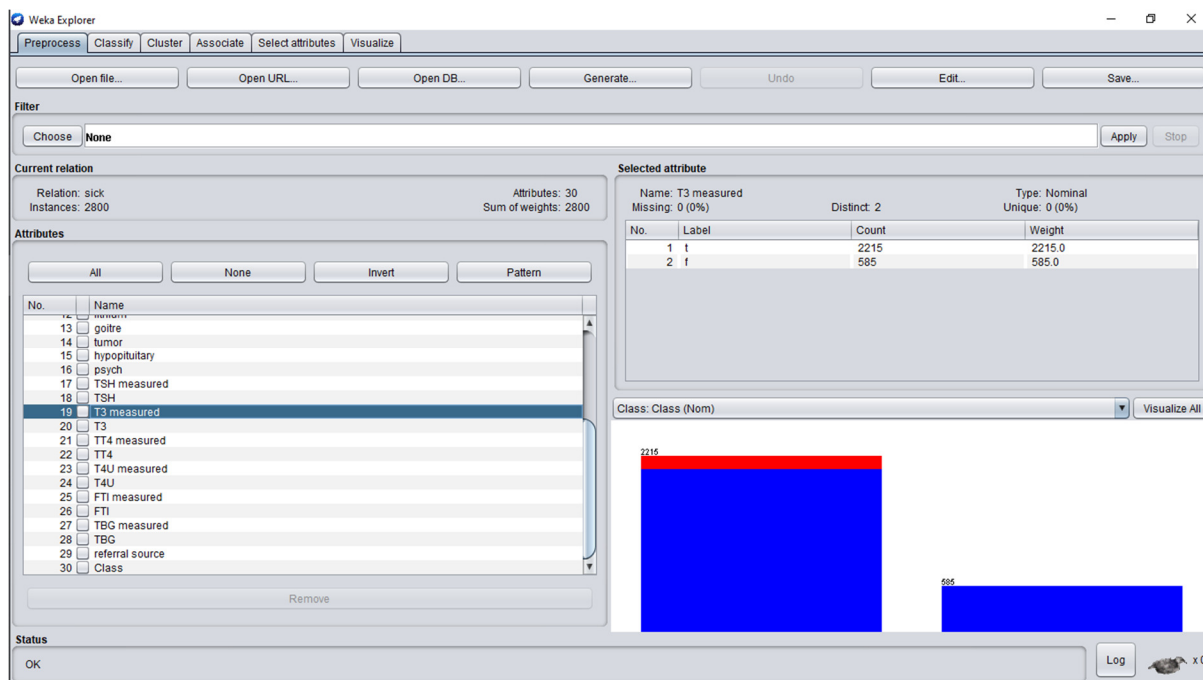
Από την στατιστική ανάλυση της μεταβλητής TSH (ορμόνη διέγερσης του θυρεοειδούς), προέκυψε ότι, στο σύνολο των 2.800 εξεταζόμενων, για ενδεχόμενη πάθηση της νόσο του θυρεοειδή, η μέση τιμή των εξεταζόμενων για την ορμόνη διέγερσης του θυρεοειδούς είναι 4,672 με τυπική απόκλιση 21,449. Η διαφορά ανάμεσα στη μεγαλύτερη και την μικρότερη ορμόνη είναι 4,667. Συγκεκριμένα ο εξεταζόμενος με την μικρότερη ορμόνη διέγερσης του θυρεοειδούς είναι 0,005, ενώ ο εξεταζόμενος με την μεγαλύτερη ορμόνη διέγερσης του θυρεοειδούς είναι 478. Από το σύνολο, 284 εξεταζόμενοι δεν απάντησαν ποια είναι η τιμή της ορμόνης διέγερσης του θυρεοειδούς (missing 284). Επίσης, από το σύνολο των εξεταζόμενων, 52 άτομα έχουν μοναδική τιμή της ορμόνης διέγερσης του θυρεοειδή (unique 52). Το πλήθος των τιμών της ορμόνης διέγερσης του θυρεοειδή οι οποίες καταγραφήκαν, είναι 263. Βλέπουμε πως οι περισσότεροι ήταν ανάμεσα στις τιμές από 0,005 έως 5,316 της ορμόνης διέγερσης του θυρεοειδή. Το φυσιολογικό εύρος δοκιμών είναι μεταξύ 0,4 και 4,0 εκατομμυρίων διεθνών μονάδων ορμόνης, ανά λίτρο αίματος (mIU/L).



Εικόνα 43: Στατιστική ανάλυση του χαρακτηριστικού « TSH »

5.1.19 Χαρακτηριστικό «T3 measured»

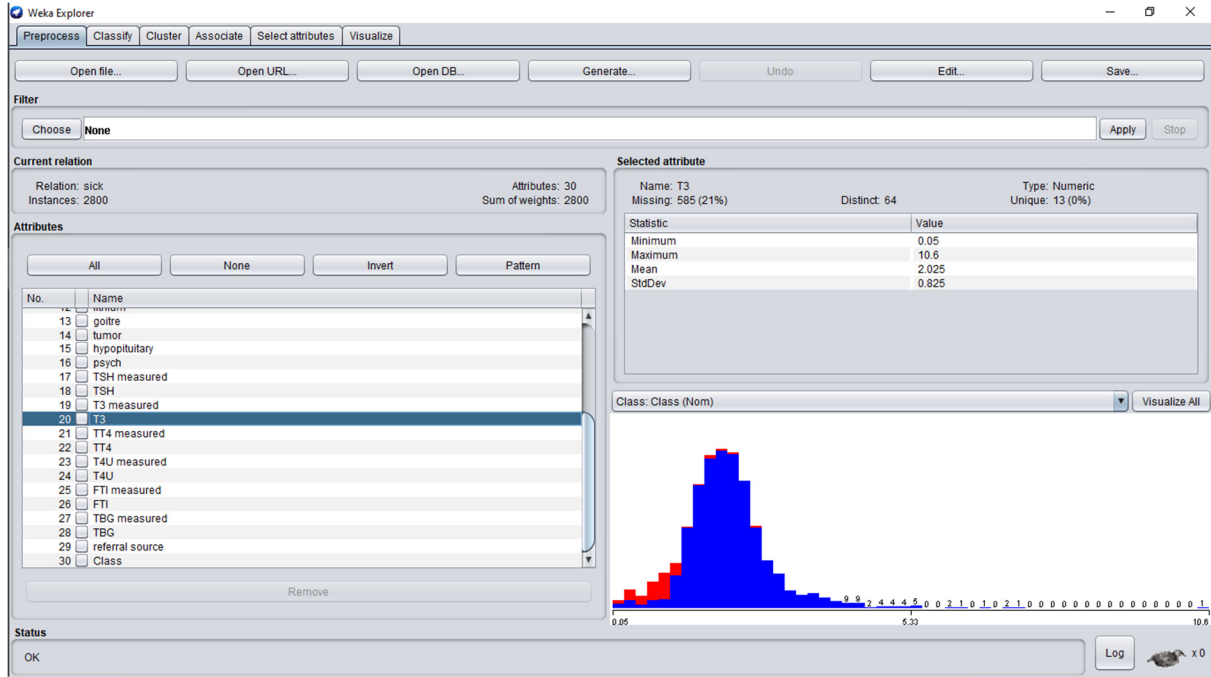
Από την στατιστική ανάλυση της μεταβλητής T3 measured (μέτρηση ορμόνης τριϊωδοθυρονίνης κατά την διάρκεια της εξέτασης), προέκυψε ότι, στο σύνολο των 2.800 εξεταζόμενων, για ενδεχόμενη πάθηση της νόσο του θυρεοειδή, τα 2.215 άτομα είχαν μέτρηση την ορμόνη τριϊωδοθυρονίνη, ενώ τα 585 άτομα δεν είχαν μέτρηση την ορμόνη τριϊωδοθυρονίνη κατά την διάρκεια της εξέτασης. Όλοι οι εξεταζόμενοι απαντήσαν, για το αν είχαν μέτρηση την ορμόνη τριϊωδοθυρονίνη (missing 0).



Εικόνα 44: Στατιστική ανάλυση του χαρακτηριστικού « T3 measured »

5.1.20 Χαρακτηριστικό «T3»

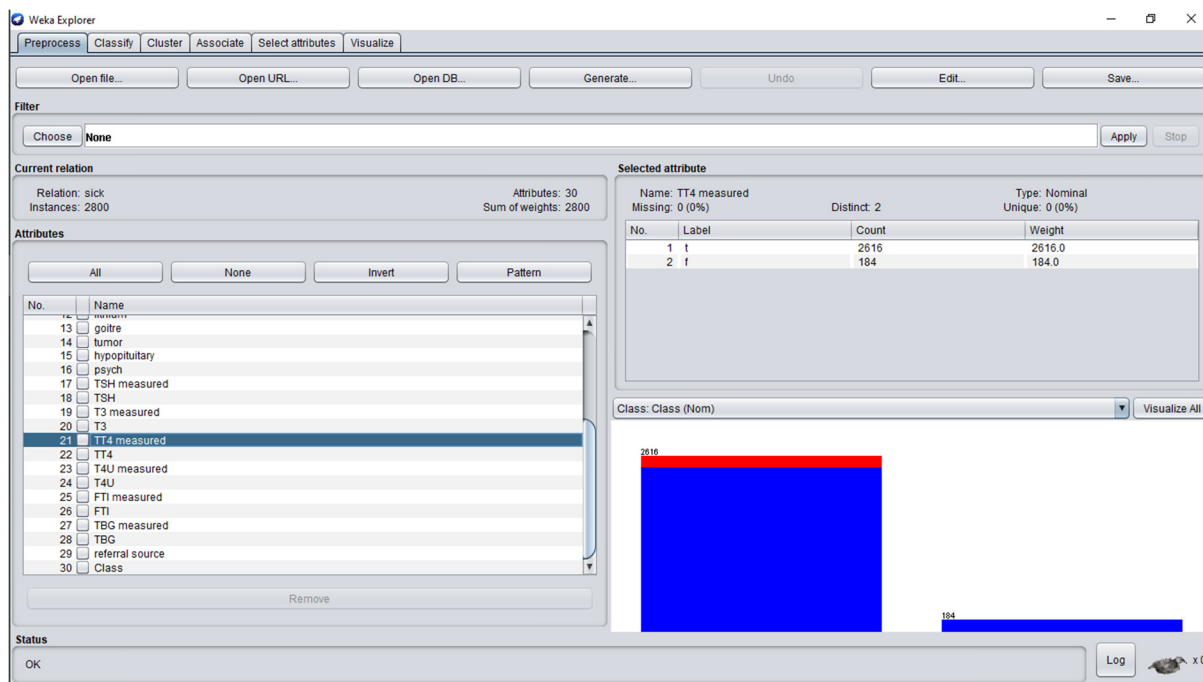
Από την στατιστική ανάλυση της μεταβλητής T3 (ορμόνη τριϊωδοθυρονίνης), προέκυψε ότι, στο σύνολο των 2.800 εξεταζόμενων, για ενδεχόμενη πάθηση της νόσο του θυρεοειδή, η μέση τιμή των εξεταζόμενων για την ορμόνη τριϊωδοθυρονίνη είναι 2,025 με τυπική απόκλιση 0,825. Η διαφορά ανάμεσα στη μεγαλύτερη ορμόνη και την μικρότερη ορμόνη είναι 10,55. Συγκεκριμένα ο εξεταζόμενος με την μικρότερη ορμόνη τριϊωδοθυρονίνη είναι 0,005, ενώ ο εξεταζόμενος με την μεγαλύτερη ορμόνη τριϊωδοθυρονίνη είναι 10,6. Από το σύνολο, 585 εξεταζόμενοι δεν απάντησαν ποια είναι η τιμή της ορμόνης τριϊωδοθυρονίνης (missing 585). Επίσης, από το σύνολο των εξεταζόμενων, 13 άτομα έχουν μοναδική τιμή της ορμόνης τριϊωδοθυρονίνης (unique 13). Το πλήθος των τιμών της ορμόνης τριϊωδοθυρονίνης που καταγραφήκαν είναι 64. Βλέπουμε πως οι τιμές της ορμόνης τριϊωδοθυρονίνης, οπου ήταν οι περισσότεροι, ήταν από 1,673 έως 2,282. Το φυσιολογικό εύρος δοκιμών είναι μεταξύ 0,1 έως 0,2 ορμόνης ανά λίτρο αίματος (μg/L).



Εικόνα 45: Στατιστική ανάλυση του χαρακτηριστικού « T3 »

5.1.21 Χαρακτηριστικό « TT4 measured »

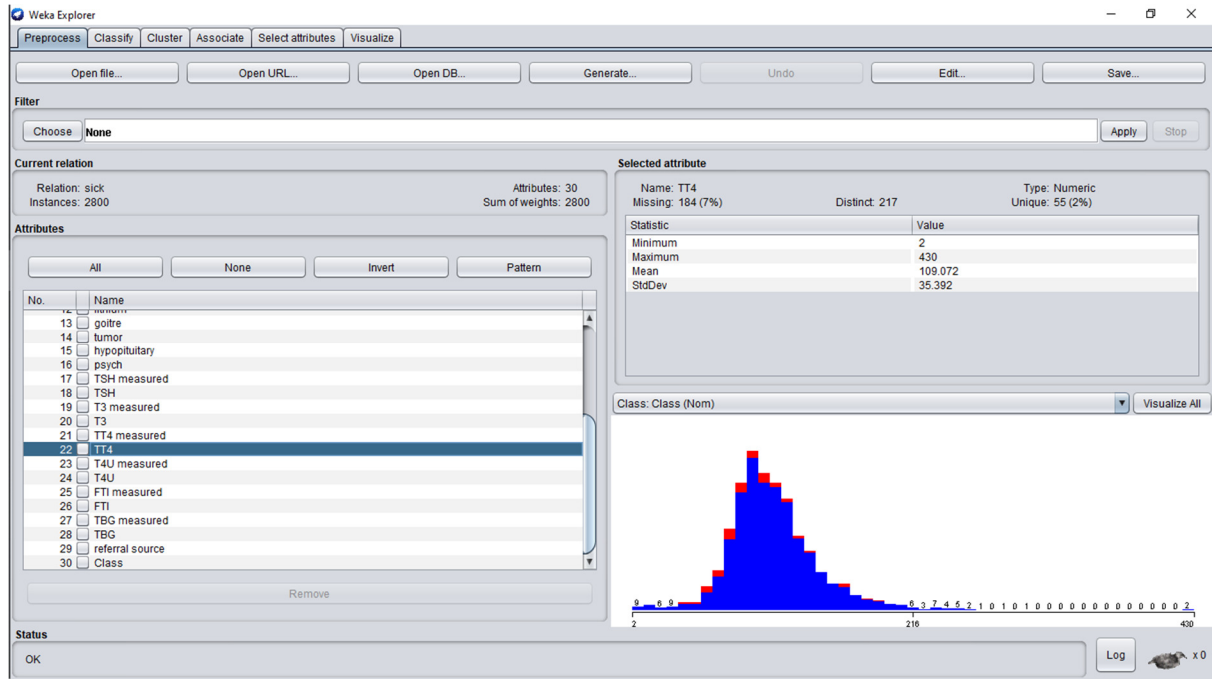
Από την στατιστική ανάλυση της μεταβλητής TT4 measured (μέτρηση θυροξίνης ολικής κατά την διάρκεια της εξέτασης), προέκυψε ότι, στο σύνολο των 2.800 εξεταζόμενων, για ενδεχόμενη πάθηση της νόσο του θυρεοειδή, τα 2.616 άτομα είχαν μέτρηση την θυροξίνη ολική, ενώ τα 184 άτομα δεν είχαν μέτρηση την θυροξίνη ολική κατά την διάρκεια της εξέτασης. Όλοι οι εξεταζόμενοι απαντήσαν για το αν είχαν μέτρηση την θυροξίνη ολική (missing 0).



Εικόνα 46: Στατιστική ανάλυση του χαρακτηριστικού « TT4 measured »

5.1.22 Χαρακτηριστικό « TT4 »

Από την στατιστική ανάλυση της μεταβλητής TT4 (θυροξίνη ολική), προέκυψε ότι, στο σύνολο των 2.800 εξεταζόμενων, για ενδεχόμενη πάθηση της νόσο του θυρεοειδή, η μέση τιμή των εξεταζόμενων για την ολική θυροξίνη είναι 109,072, με τυπική απόκλιση 32,392. Η διαφορά ανάμεσα στη μεγαλύτερη και την μικρότερη ολική θυροξίνη είναι 128. Συγκεκριμένα, ο εξεταζόμενος με την μικρότερη ολική θυροξίνη είναι 2, ενώ ο εξεταζόμενος με την μεγαλύτερη ολική θυροξίνη είναι 430. Από το σύνολο, οι 184 εξεταζόμενοι δεν απάντησαν ποια είναι η τιμή της ολικής θυροξίνης (missing 184). Επίσης, από το σύνολο των εξεταζόμενων, 55 άτομα έχουν μοναδική τιμή της ολικής θυροξίνης (unique 55). Το πλήθος των τιμών της ολικής θυροξίνης που καταγραφήκαν είναι 217. Βλέπουμε πως, οι τιμές της ολικής θυροξίνης που ήταν οι περισσότεροι ήταν από 80,612 έως τα 124,286. Το φυσιολογικό εύρος δοκιμών μεταξύ 50 και 110 ορμόνης ανά λίτρο αίματος.

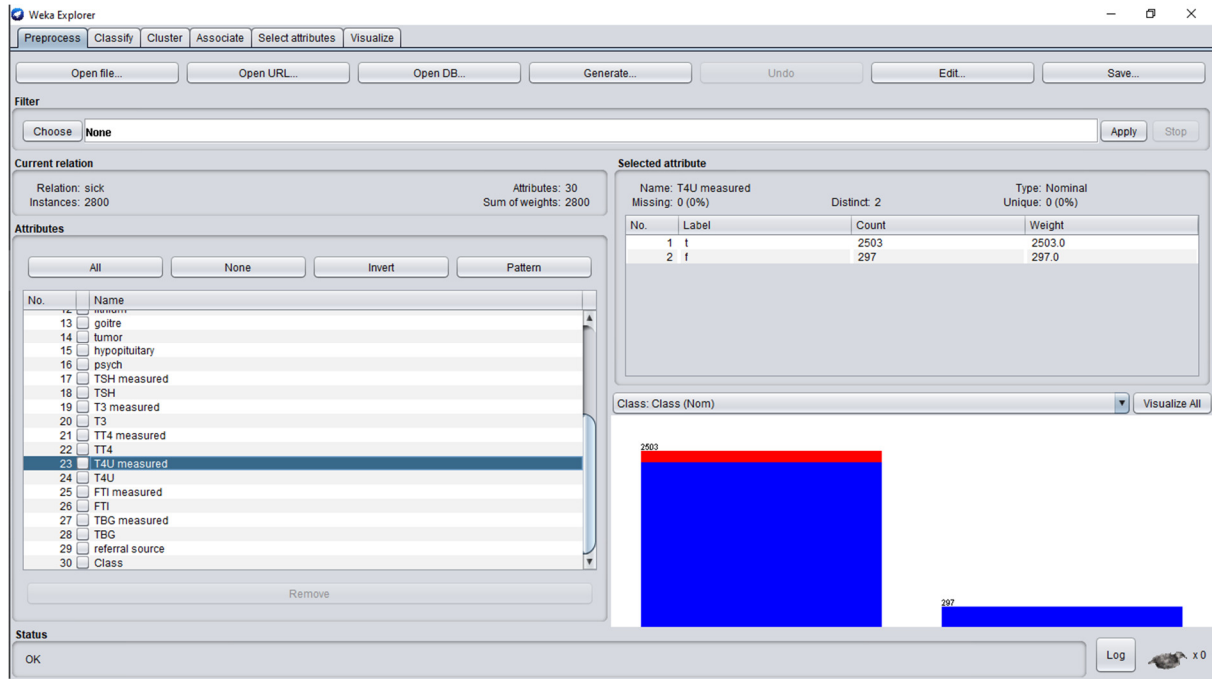


Εικόνα 47: Στατιστική ανάλυση του χαρακτηριστικού « TT4 »

5.1.23 Χαρακτηριστικό « T4U measured »

Από την στατιστική ανάλυση της μεταβλητής T4U⁵ measured (μέτρηση θυροξίνης κατά την διάρκεια της εξέτασης), προέκυψε ότι, στο σύνολο των 2.800 εξεταζόμενων, για ενδεχόμενη πάθηση της νόσο του θυρεοειδή, τα 2.503 άτομα είχαν μέτρηση την θυροξίνη, ενώ τα 297 άτομα δεν είχαν μέτρηση την θυροξίνη κατά την διάρκεια της εξέτασης. Όλοι οι εξεταζόμενοι απαντήσαν, σχετικά με το αν είχαν μέτρηση της θυροξίνης (missing 0).

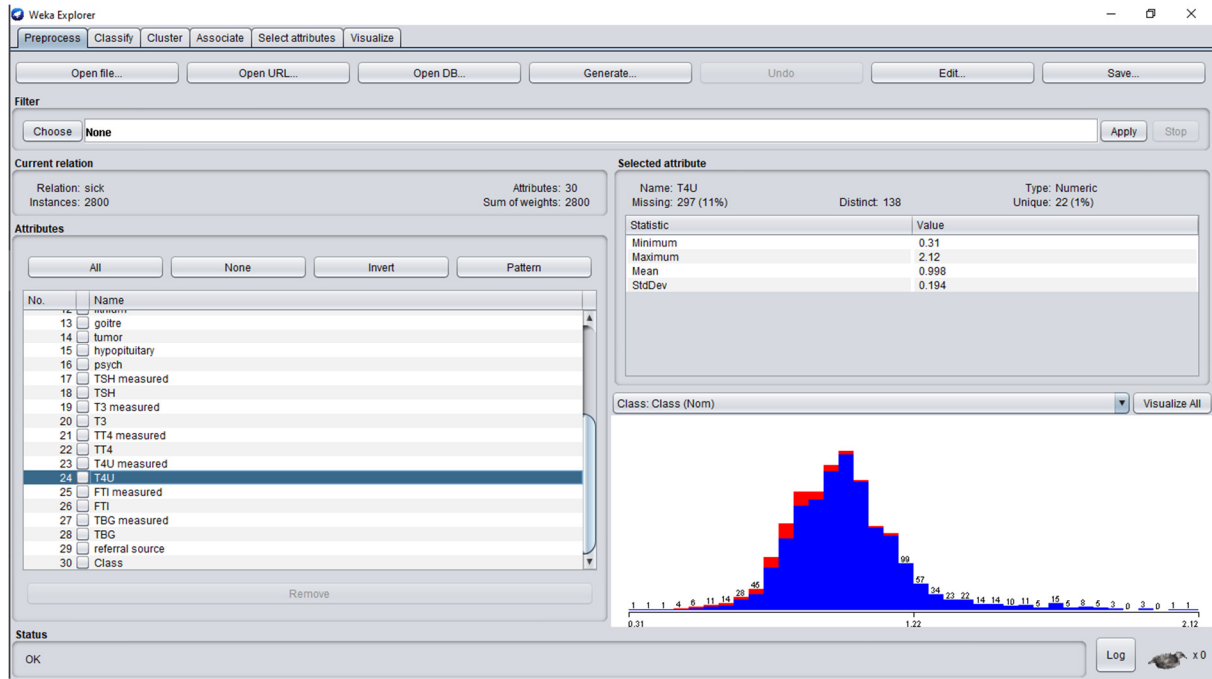
⁵ T4U: είναι ένας παλιός τρόπος μέτρησης της Θυροξίνης (T4)



Εικόνα 48: Στατιστική ανάλυση του χαρακτηριστικού « T4U measured »

5.1.24 Χαρακτηριστικό « T4U »

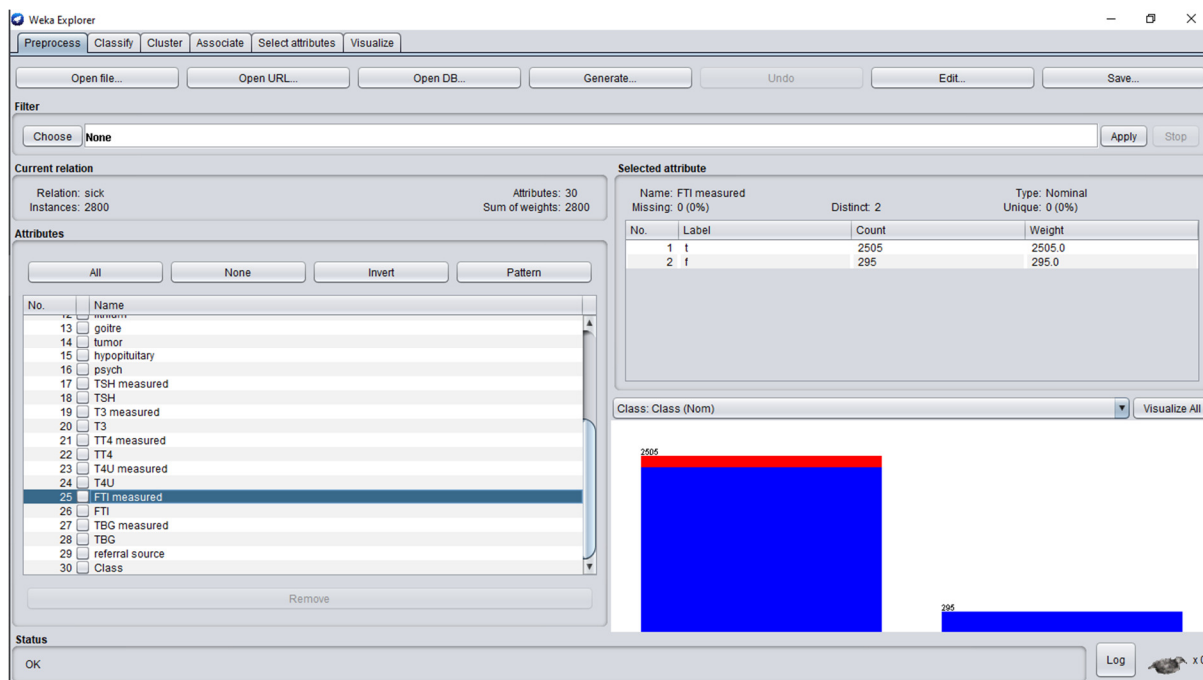
Από την στατιστική ανάλυση της μεταβλητής T4U (θυροξίνη), προέκυψαν προέκυψε ότι, στο σύνολο των 2.800 εξεταζόμενων, για ενδεχόμενη πάθηση της νόσο του θυρεοειδή, η μέση τιμή των εξεταζόμενων για την θυροξίνη είναι 0,998, με τυπική απόκλιση 0,194. Η διαφορά ανάμεσα στη μεγαλύτερη θυροξίνη και την μικρότερη θυροξίνη είναι 1,81. Συγκεκριμένα, ο εξεταζόμενος με την μικρότερη θυροξίνη είναι 0,31, ενώ ο εξεταζόμενος με την μεγαλύτερη θυροξίνη είναι 2,12. Από το σύνολο, οι 297 εξεταζόμενοι δεν απάντησαν ποια είναι η τιμή της θυροξίνης (missing 297). Επίσης, από το σύνολο των εξεταζόμενων, 22 άτομα έχουν μοναδική τιμή της θυροξίνης (unique 22). Το πλήθος των τιμών της θυροξίνης που καταγραφήκαν είναι 138. Βλέπουμε πως οι τιμές της θυροξίνης, όπου ήταν οι περισσότεροι, ήταν από 0,834 έως 1,072.



Εικόνα 49: Στατιστική ανάλυση του χαρακτηριστικού « T4U »

5.1.25 Χαρακτηριστικό « FTI measured »

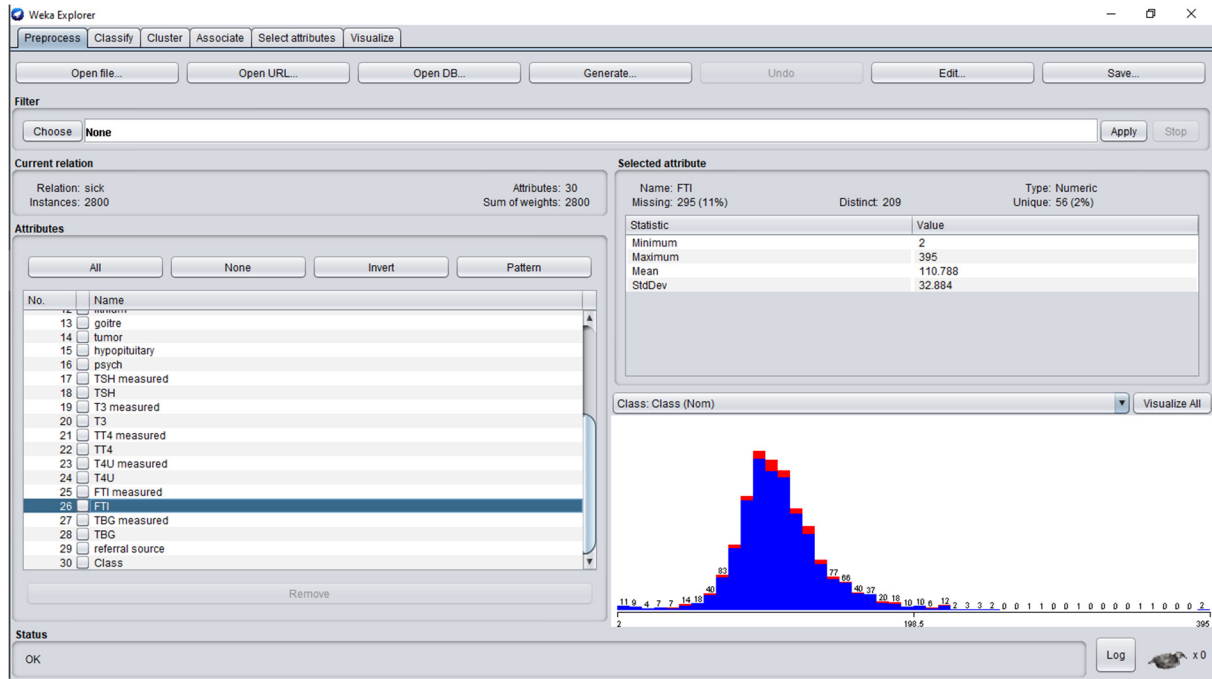
Από την στατιστική ανάλυση της μεταβλητής FTI measured (μέτρηση Δείκτη Ελεύθερης Θυροξίνης κατά την διάρκεια της εξέτασης), προέκυψε ότι, στο σύνολο των 2.800 εξεταζόμενων, για ενδεχόμενη πάθηση της νόσο του θυρεοειδή, τα 2.505 άτομα είχαν μέτρηση τον δείκτη ελεύθερης θυροξίνης, ενώ τα 295 άτομα δεν είχαν μέτρηση τον δείκτη ελεύθερης θυροξίνης, κατά την διάρκεια της εξέτασης. Όλοι οι εξεταζόμενοι απαντήσαν για το αν είχαν πραγματοποιήσει μέτρηση του δείκτη ελεύθερης θυροξίνης τους (missing 0).



Εικόνα 50: Στατιστική ανάλυση του χαρακτηριστικού « FTI measured »

5.1.26 Χαρακτηριστικό « FTI »

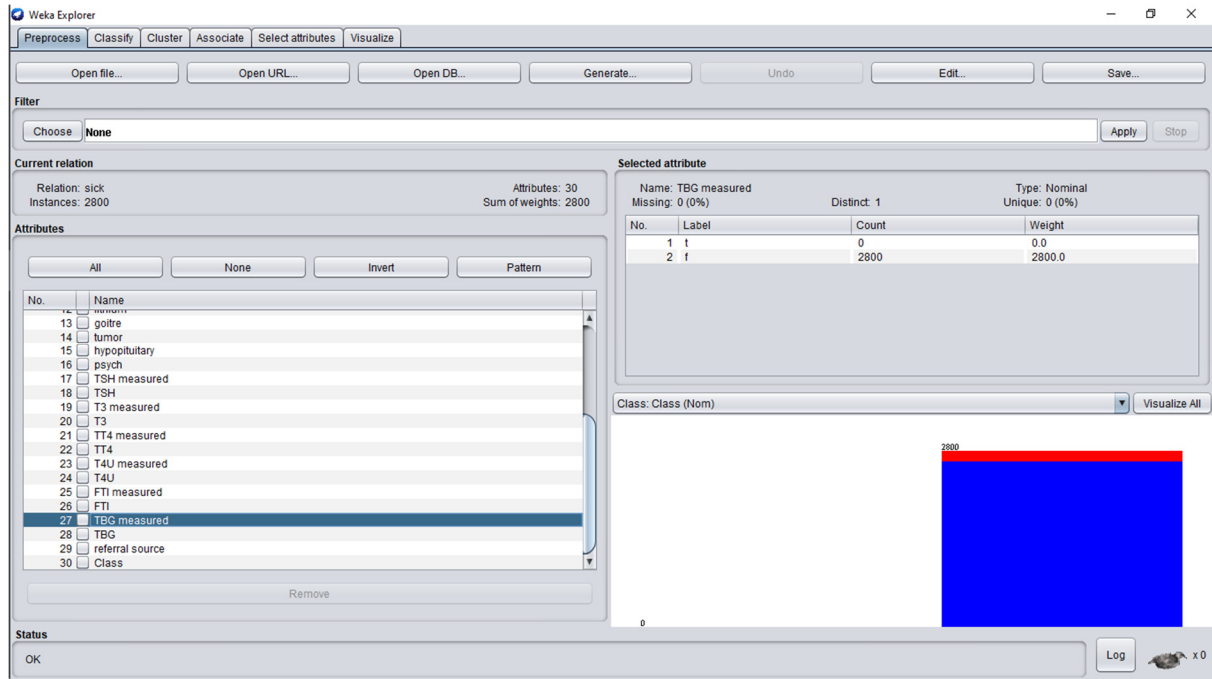
Από την στατιστική ανάλυση της μεταβλητής FTI (δείκτη ελεύθερης θυροξίνης), προέκυψε ότι, στο σύνολο των 2.800 εξεταζόμενων, για ενδεχόμενη πάθηση της νόσου του θυρεοειδή, ο μέσος δείκτης ελεύθερης θυροξίνης των εξεταζόμενων είναι 110,788, με τυπική απόκλιση 32,884. Η διαφορά ανάμεσα στον μεγαλύτερο και τον μικρότερο δείκτη ελεύθερης θυροξίνης είναι 393. Συγκεκριμένα, ο εξεταζόμενος με τον μικρότερο δείκτη ελεύθερης θυροξίνης είναι 2, ενώ ο εξεταζόμενος με τον μεγαλύτερο δείκτη ελεύθερης θυροξίνης είναι 395. Από το σύνολο, οι 295 εξεταζόμενοι δεν απάντησαν ποια είναι η τιμή του δείκτη ελεύθερης θυροξίνης (missing 295). Επίσης, από το σύνολο των εξεταζόμενων, 56 άτομα έχουν μοναδική τιμή της ολικής θυροξίνης (unique 56). Το πλήθος των τιμών των δεικτών ελεύθερης θυροξίνης που καταγραφήκαν, είναι 209. Βλέπουμε πως οι τιμές των δεικτών ελεύθερης θυροξίνης, οπου ήταν οι περισσότεροι, ήταν από 83,875 τα έως τα 116,625.



Εικόνα 51: Στατιστική ανάλυση του χαρακτηριστικού « FTI »

5.1.27 Χαρακτηριστικό « TBG measured »

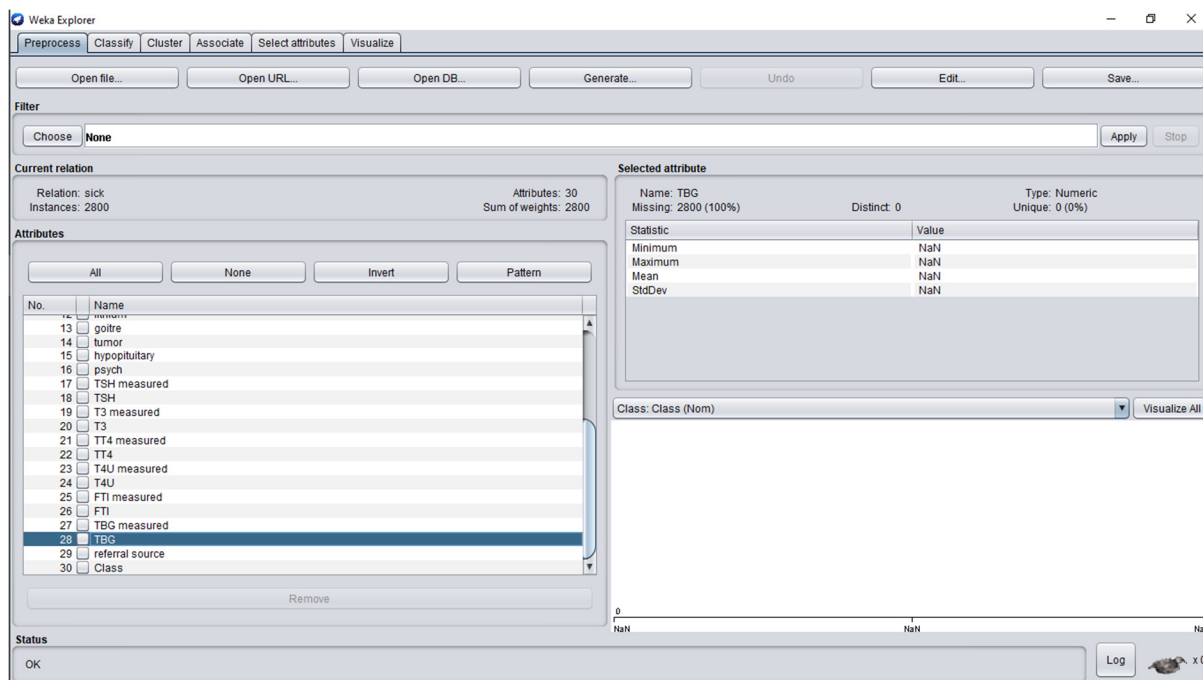
Από την στατιστική ανάλυση της μεταβλητής TBG measured (μέτρηση Σφαιρίνη Δεσμεύουσα Θυροξίνη κατά την διάρκεια της εξέτασης), προέκυψε ότι, στο σύνολο των 2.800 εξεταζόμενων, για ενδεχόμενη πάθηση της νόσο του θυρεοειδή, κανένα άτομο δεν είχε μέτρηση της σφαιρίνης δεσμεύουσας θυροξίνης. Όλοι οι εξεταζόμενοι απαντήσαν για το αν είχαν πραγματοποιήσει την μέτρηση της σφαιρίνης δεσμεύουσας θυροξίνης τους (missing 0).



Εικόνα 52: Στατιστική ανάλυση του χαρακτηριστικού « TBG measured »

5.1.28 Χαρακτηριστικό « TBG »

Από την στατιστική ανάλυση της μεταβλητής TBG (σφαιρίνη Δεσμεύουσα Θυροξίνη), προέκυψε ότι, στο σύνολο των 2.800 εξεταζόμενων, για ενδεχόμενη πάθηση της νόσο του θυρεοειδή, όλοι οι εξεταζόμενοι δεν απάντησαν για το ποια είναι η τιμή της σφαιρίνης δεσμεύουσας θυροξίνης τους (missing 100%).

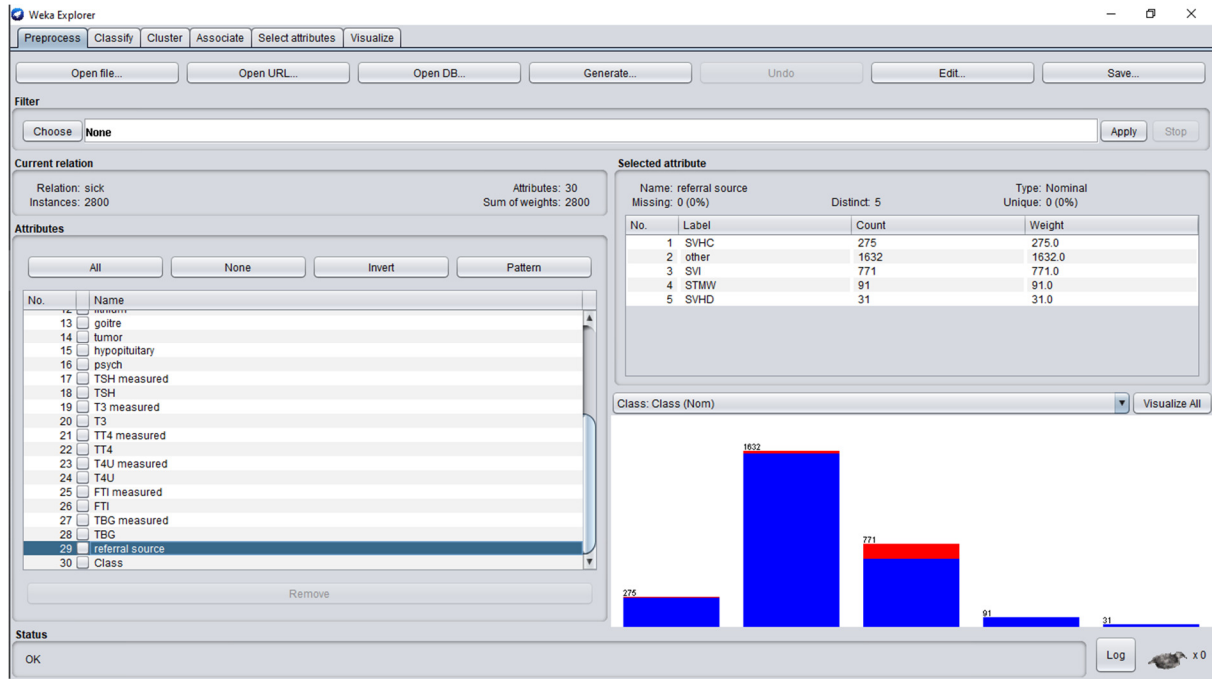


Εικόνα 53: Στατιστική ανάλυση του χαρακτηριστικού « TBG »

5.1.29 Χαρακτηριστικό « referral source »

Από την στατιστική ανάλυση της μεταβλητής referral source⁶ (Φυσιολογικές τιμές), προέκυψε ότι, στο σύνολο των 2.800 εξεταζόμενων, για ενδεχόμενη πάθηση της νόσο του θυρεοειδή, τα 275 άτομα είχαν SVHC, τα 1.632 άτομα ήταν other, τα 771 άτομα είχαν SVI, τα 91 άτομα είχαν STMW και 31 άτομα είχαν SVHD.

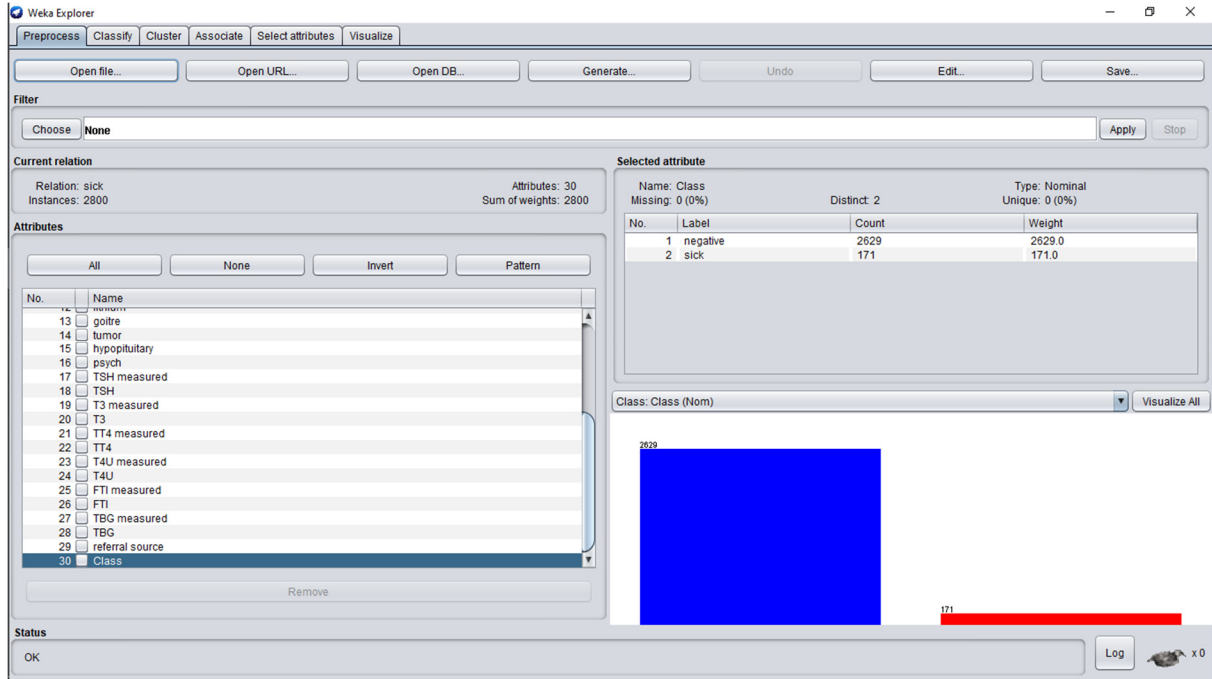
⁶ referral source: οι τιμές SVHC, SVI, STMW, SVHD ή ΑΛΛΟ αντιστοιχούν σε τύπους αντιδράσεων δεν είναι ασθένειες.



Εικόνα 54: Στατιστική ανάλυση του χαρακτηριστικού « referral source »

5.1.30 Χαρακτηριστικό « Class »

Από την στατιστική ανάλυση της μεταβλητής Class (ασθενείς όπου είχαν θυρεοειδή ή όχι), προέκυψε ότι, στο σύνολο των 2.800 εξεταζόμενων, για ενδεχόμενη πάθηση της νόσο του θυρεοειδή, τα 2.629 άτομα δεν είχαν την νόσο του θυρεοειδή, ενώ τα 171 άτομα είχαν την νόσο του θυρεοειδή.



Εικόνα 55: Στατιστική ανάλυση του χαρακτηριστικού « Class »

5.2 Ανάλυση αλγορίθμων που χρησιμοποιήθηκαν

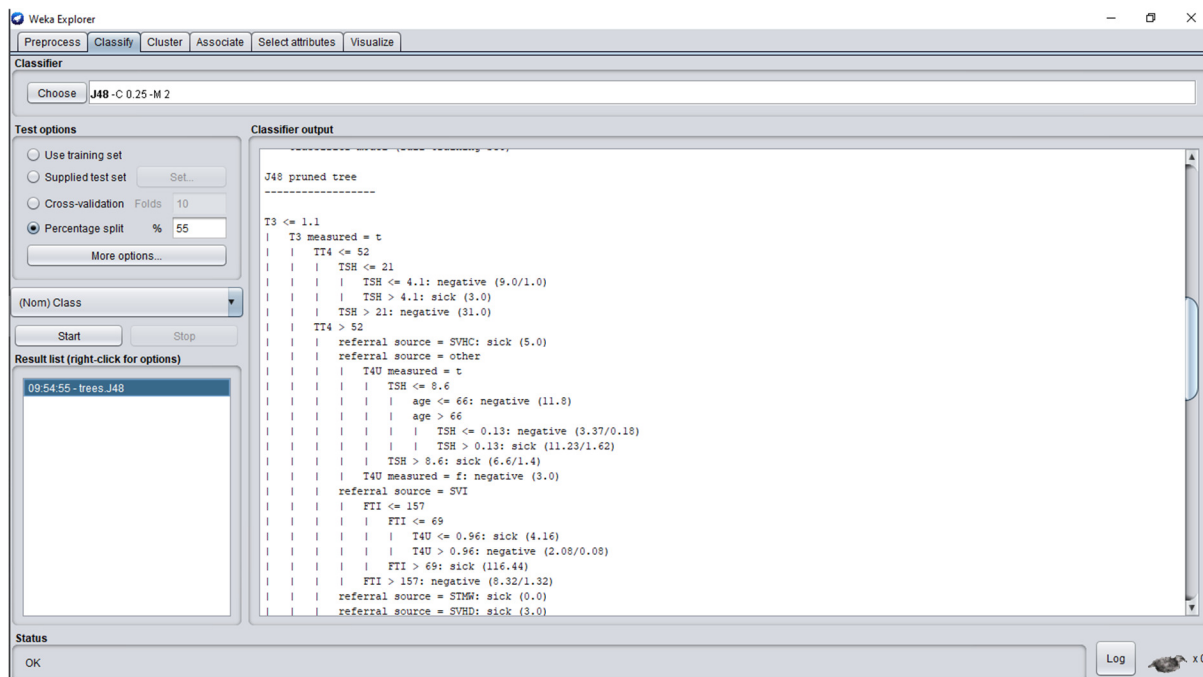
Η αναλυτική ανάλυση των αλγορίθμων είναι στο «ΔΙΕΥΚΡΙΝΙΣΕΙΣ» σελ. 139

5.2.1 Δέντρο απόφασης αλγόριθμου J48 με διακριτοποίηση τιμών

5.2.1.1 Αλγόριθμος J48 -C 0,25 -M 2 με διακριτοποίηση 55%

Εκπαίδευσάμε το 55% των δεδομένων μας με την μέθοδο της διακριτοποίησης τιμών, με τον αλγόριθμο J48 -C 0,25 -M 2 και πήραμε τα εξής αποτελέσματα: από το σύνολο των μεταβλητών που αναλύσαμε στο κεφάλαιο 5.1

Το συνολικό μέγεθος του δέντρου είναι 49 κόμβοι. Συνολικά από το μοντέλο που προκύπτει(μη εκπαιδευμένα δέντρα) βλέπουμε πως δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς 1.172 άτομα από τα οποία χάθηκαν τα 14 άτομα όπου έχουν διαγνωστεί με την νόσο του θυρεοειδούς ενώ έχουν διαγνωστεί με την νόσο του θυρεοειδούς 60 άτομα από τα οποία χάθηκαν τα 14 άτομα όπου δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς. Σε ποσοστό ακρίβειας των δεδομένων είναι 97,7778% σύμφωνα με το πρότυπο, ενώ το 2,2222% δεν είναι σύμφωνα με το πρότυπο.

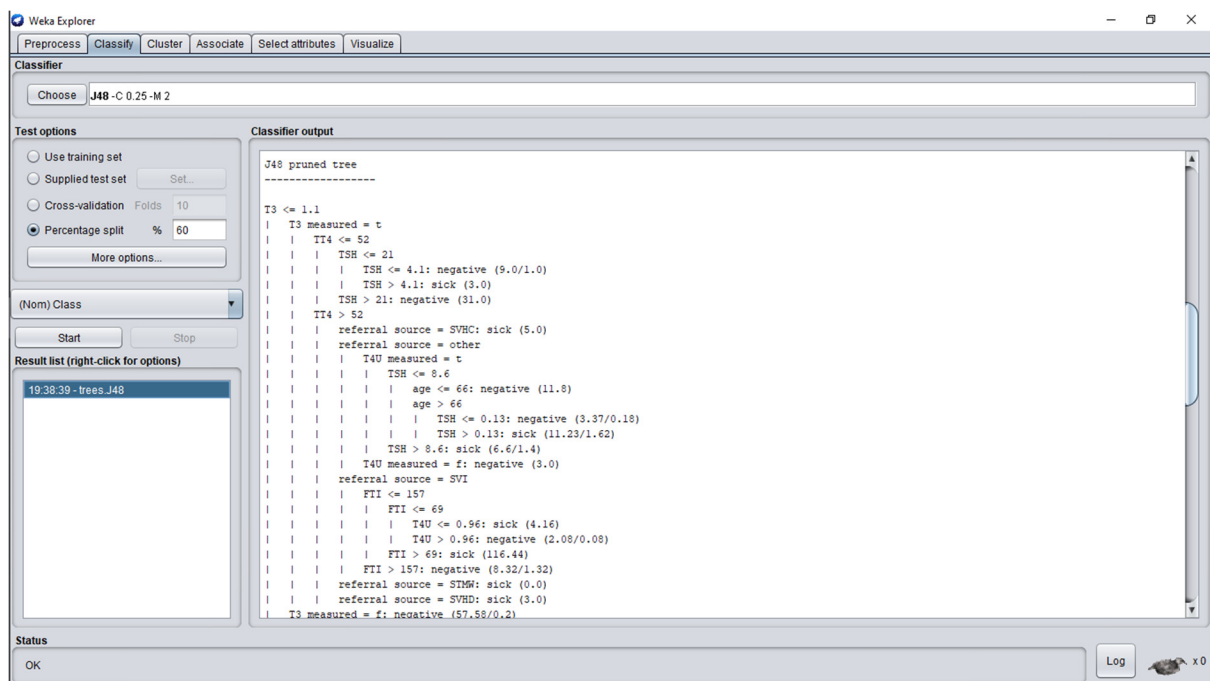


Εικόνα 56: Αλγόριθμος J48 -C 0,25 -M 2 με διακριτοποίηση 55%

5.2.1.2 Αλγόριθμος J48 -C 0,25 –M 2 με διακριτοποίηση 60%

Εκπαιδύσαμε το 60% των δεδομένων μας με την μέθοδο της διακριτοποίησης τιμών, με τον αλγόριθμο J48 -C 0,25 –M 2 και πήραμε τα εξής αποτελέσματα: από το σύνολο των μεταβλητών που αναλύσαμε στο κεφάλαιο 5.1

Το συνολικό μέγεθος του δέντρου είναι 49 κόμβοι. Συνολικά από το μοντέλο που προκύπτει(μη εκπαιδευμένα δέντρα) βλέπουμε πως δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς 1.041 άτομα από τα οποία χάθηκαν τα 17 άτομα όπου έχουν διαγνωστεί με την νόσο του θυρεοειδούς ενώ έχουν διαγνωστεί με την νόσο του θυρεοειδούς 51 άτομα από τα οποία χάθηκαν τα 11 άτομα όπου δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς. Σε ποσοστό ακρίβειας των δεδομένων είναι 97,5% σύμφωνα με το πρότυπο, ενώ το 2,5% δεν είναι σύμφωνα με το πρότυπο.

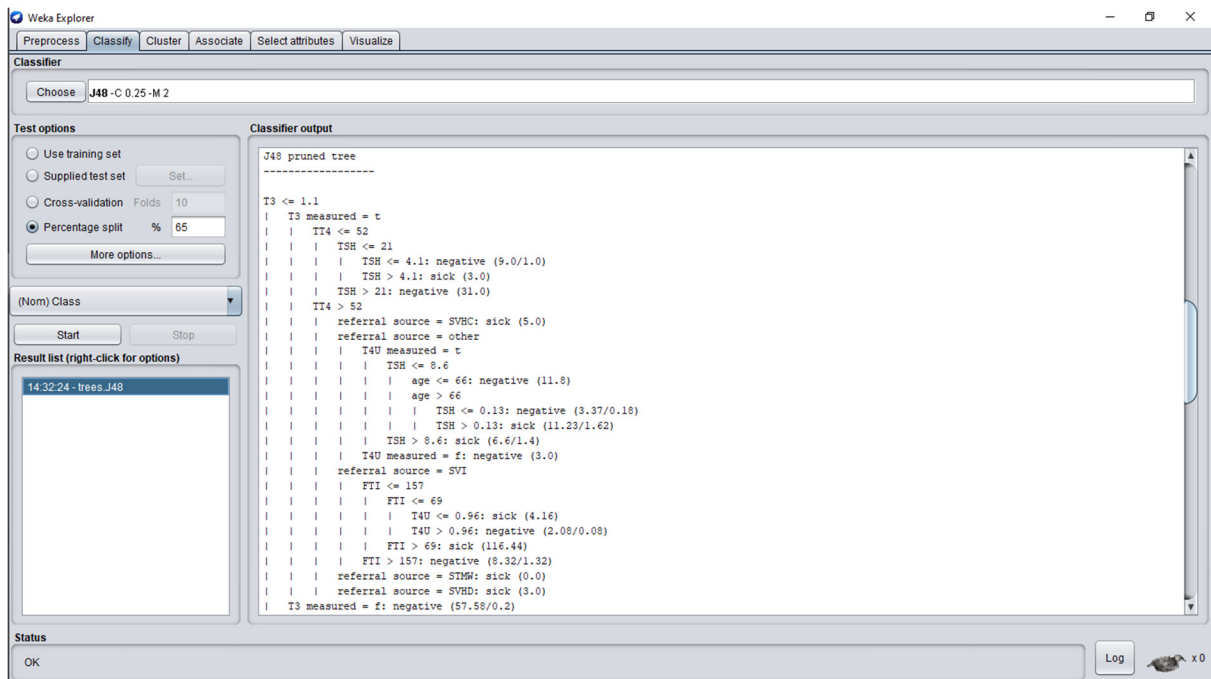


Εικόνα 57: Αλγόριθμος J48 -C 0,25 –M 2 με διακριτοποίηση 60%

5.2.1.3 Αλγόριθμος J48 -C 0,25 –M 2 με διακριτοποίηση 65%

Εκπαιδύσαμε το 65% των δεδομένων μας με την μέθοδο της διακριτοποίησης τιμών, με τον αλγόριθμο J48 -C 0,25 –M 2 και πήραμε τα εξής αποτελέσματα: από το σύνολο των μεταβλητών που αναλύσαμε στο κεφάλαιο 5.1

Το συνολικό μέγεθος του δέντρου είναι 49 κόμβοι. Συνολικά από το μοντέλο που προκύπτει(μη εκπαιδευμένα δέντρα) βλέπουμε πως δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς 920 άτομα από τα οποία χάθηκαν τα 5 άτομα όπου έχουν διαγνωστεί με την νόσο του θυρεοειδούς ενώ έχουν διαγνωστεί με την νόσο του θυρεοειδούς 43 άτομα από τα οποία χάθηκαν τα 12 άτομα όπου δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς. Σε ποσοστό ακρίβειας των δεδομένων είναι 98,2653% σύμφωνα με το πρότυπο, ενώ το 1,7347% δεν είναι σύμφωνα με το πρότυπο.



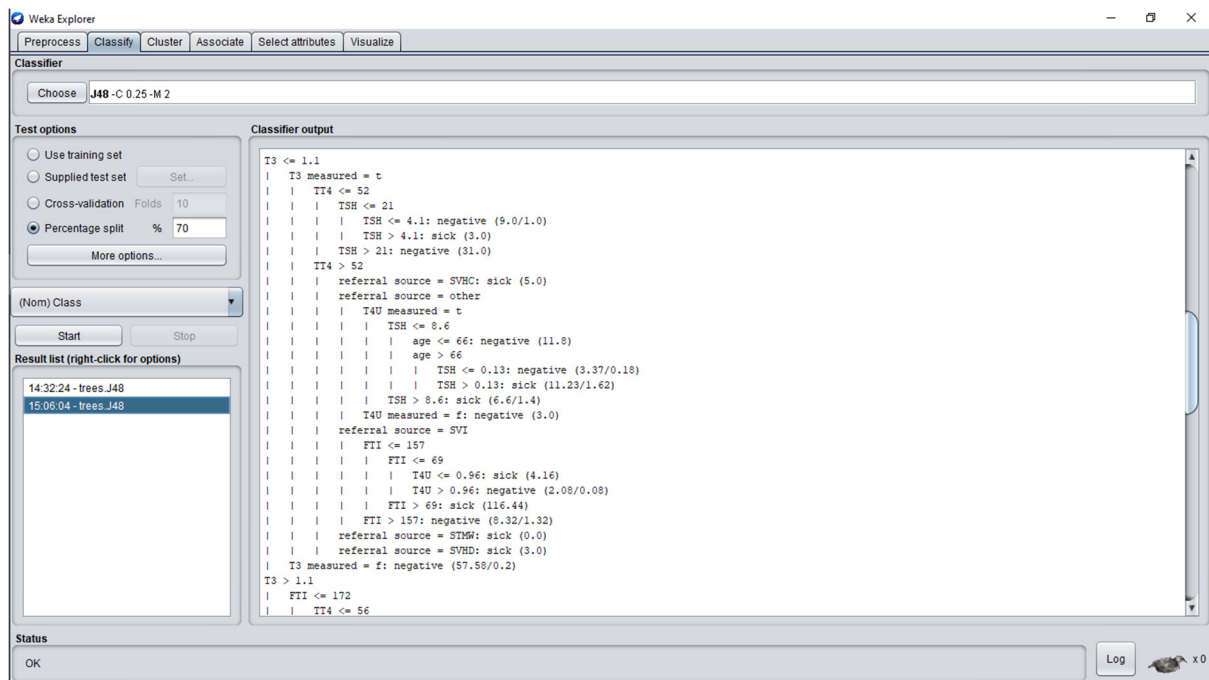
Εικόνα 58: Αλγόριθμος J48 -C 0,25 –M 2 με διακριτοποίηση 65%

5.2.1.4 Αλγόριθμος J48 -C 0,25 –M 2 με διακριτοποίηση 70%

Εκπαίδευσάμε το 70% των δεδομένων μας με την μέθοδο της διακριτοποίησης τιμών, με τον αλγόριθμο J48 -C 0,25 –M 2 και πήραμε τα εξής αποτελέσματα: από το σύνολο των μεταβλητών που αναλύσαμε στο κεφάλαιο 5.1

Το συνολικό μέγεθος του δέντρου είναι 49 κόμβοι. Συνολικά από το μοντέλο που προκύπτει(μη εκπαιδευμένα δέντρα) βλέπουμε πως δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς 789 άτομα από τα οποία χάθηκαν τα 6 άτομα όπου έχουν διαγνωστεί με την νόσο του θυρεοειδούς ενώ έχουν διαγνωστεί με την νόσο του θυρεοειδούς 38 άτομα από τα οποία

χάθηκαν τα 7 άτομα όπου δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς. Σε ποσοστό ακρίβειας των δεδομένων είναι 98,4524% σύμφωνα με το πρότυπο, ενώ το 1,5476% δεν είναι σύμφωνα με το πρότυπο.

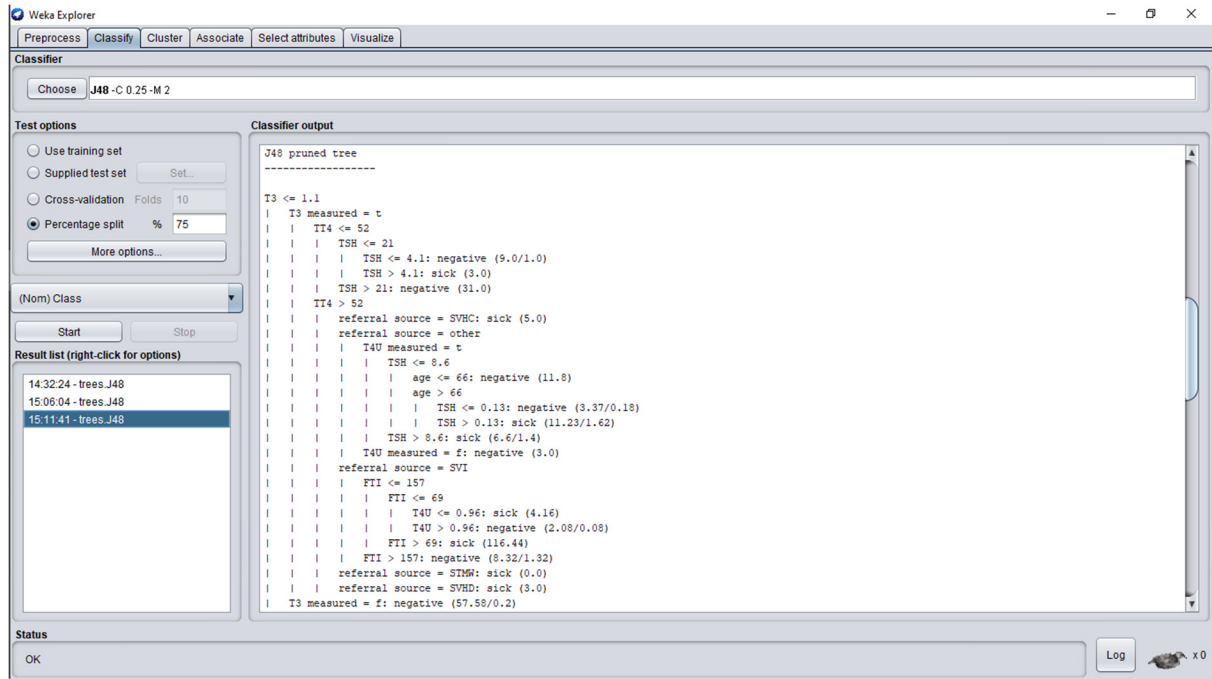


Εικόνα 59: Αλγόριθμος J48 -C 0,25 –M 2 με διακριτοποίηση 70%

5.2.1.5 Αλγόριθμος J48 -C 0,25 –M 2 με διακριτοποίηση 75%

Εκπαιδύσαμε το 75% των δεδομένων μας με την μέθοδο της διακριτοποίησης τιμών, με τον αλγόριθμο J48 -C 0,25 –M 2 και πήραμε τα εξής αποτελέσματα: από το σύνολο των μεταβλητών που αναλύσαμε στο κεφάλαιο 5.1

Το συνολικό μέγεθος του δέντρου είναι 49 κόμβοι. Συνολικά από το μοντέλο που προκύπτει(μη εκπαιδευμένα δέντρα) βλέπουμε πως δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς 663 άτομα από τα οποία χάθηκαν τα 4 άτομα όπου έχουν διαγνωστεί με την νόσο του θυρεοειδούς ενώ έχουν διαγνωστεί με την νόσο του θυρεοειδούς 29 άτομα από τα οποία χάθηκαν τα 4 άτομα όπου δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς. Σε ποσοστό ακρίβειας των δεδομένων είναι 98,8571% σύμφωνα με το πρότυπο, ενώ το 1,1429% δεν είναι σύμφωνα με το πρότυπο.

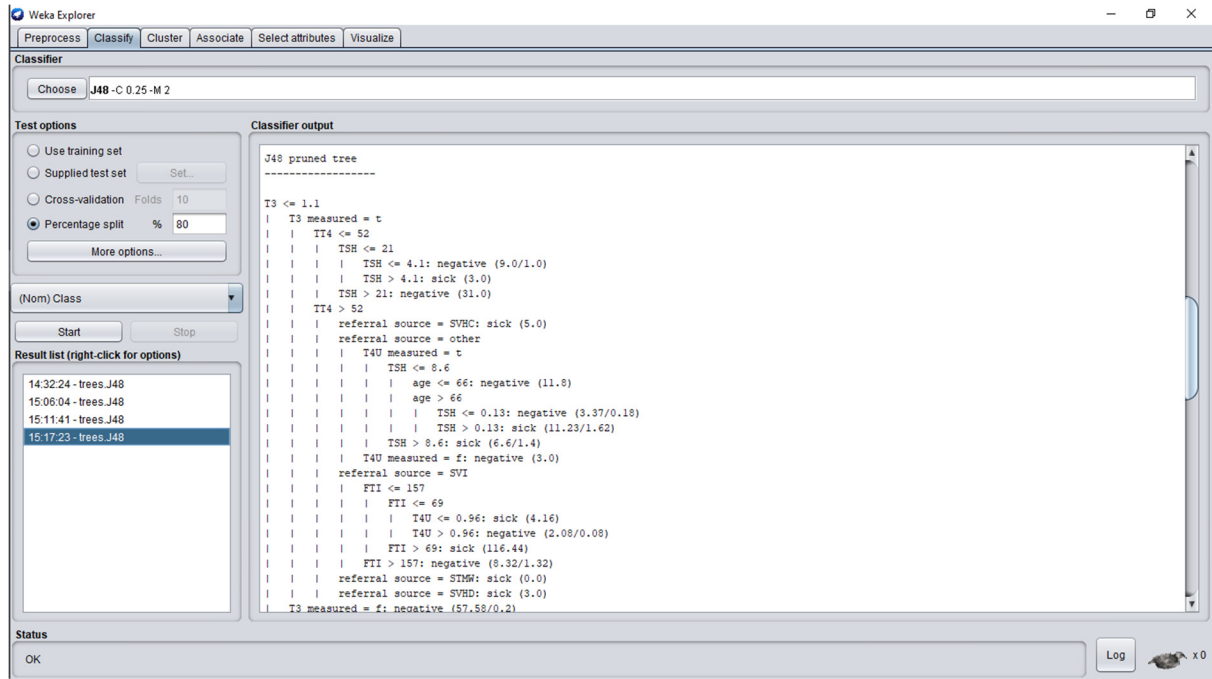


Εικόνα 60: Αλγόριθμος J48 -C 0,25 –M 2 με διακριτοποίηση 75%

5.2.1.6 Αλγόριθμος J48 -C 0,25 –M 2 με διακριτοποίηση 80%

Εκπαιδύσαμε το 80% των δεδομένων μας με την μέθοδο της διακριτοποίησης τιμών, με τον αλγόριθμο J48 -C 0,25 –M 2 και πήραμε τα εξής αποτελέσματα: από το σύνολο των μεταβλητών που αναλύσαμε στο κεφάλαιο 5.1

Το συνολικό μέγεθος του δέντρου είναι 49 κόμβοι. Συνολικά από το μοντέλο που προκύπτει(μη εκπαιδευμένα δέντρα) βλέπουμε πως δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς 530 άτομα από τα οποία χάθηκαν τα 7 άτομα όπου έχουν διαγνωστεί με την νόσο του θυρεοειδούς ενώ έχουν διαγνωστεί με την νόσο του θυρεοειδούς 21 άτομα από τα οποία χάθηκαν τα 2 άτομα όπου δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς. Σε ποσοστό ακρίβειας των δεδομένων είναι 98,3929% σύμφωνα με το πρότυπο, ενώ το 1,6071% δεν είναι σύμφωνα με το πρότυπο.

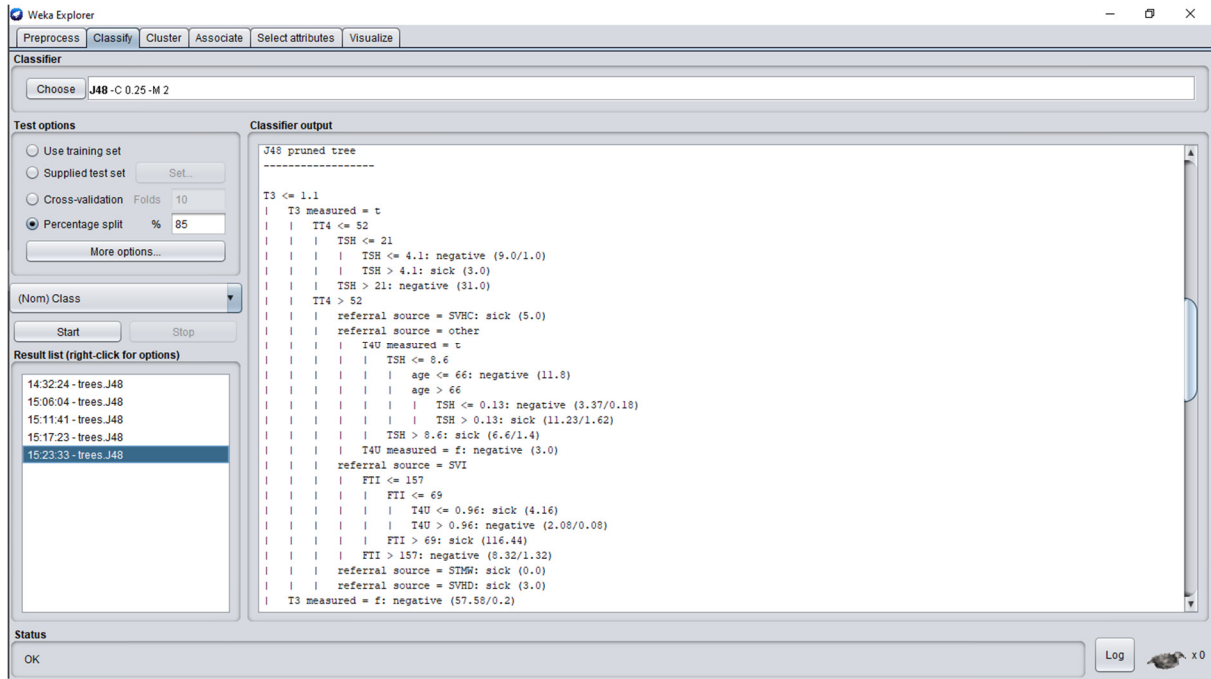


Εικόνα 61: Αλγόριθμος J48 -C 0,25 –M 2 με διακριτοποίηση 80%

5.2.1.7 Αλγόριθμος J48 -C 0,25 –M 2 με διακριτοποίηση 85%

Εκπαιδύσαμε το 85% των δεδομένων μας με την μέθοδο της διακριτοποίησης τιμών, με τον αλγόριθμο J48 -C 0,25 –M 2 και πήραμε τα εξής αποτελέσματα: από το σύνολο των μεταβλητών που αναλύσαμε στο κεφάλαιο 5.1

Το συνολικό μέγεθος του δέντρου είναι 49 κόμβοι. Συνολικά από το μοντέλο που προκύπτει(μη εκπαιδευμένα δέντρα) βλέπουμε πως δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς 530 άτομα από τα οποία χάθηκαν τα 7 άτομα όπου έχουν διαγνωστεί με την νόσο του θυρεοειδούς ενώ έχουν διαγνωστεί με την νόσο του θυρεοειδούς 21 άτομα από τα οποία χάθηκαν τα 2 άτομα όπου δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς. Σε ποσοστό ακρίβειας των δεδομένων είναι 97,8571% σύμφωνα με το πρότυπο, ενώ το 2,1429% δεν είναι σύμφωνα με το πρότυπο.



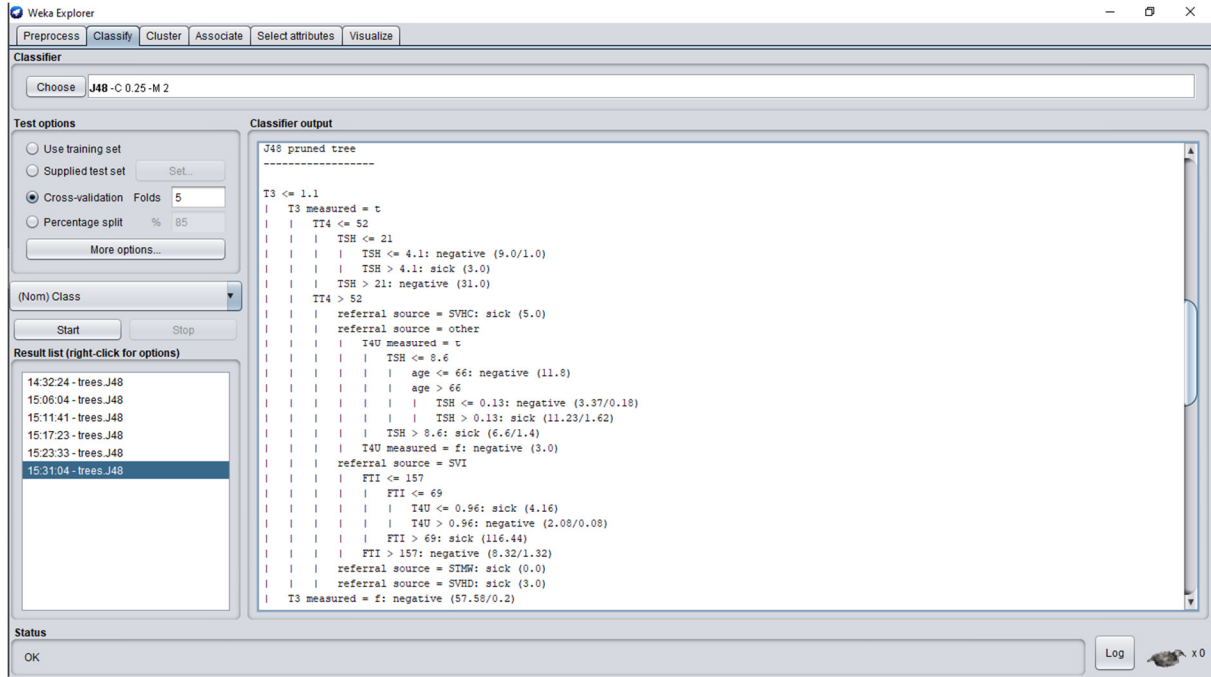
Εικόνα 62: Αλγόριθμος J48 -C 0,25 –M 2 με διακριτοποίηση 85%

5.2.2 Δέντρο απόφασης αλγόριθμου J48 με διακριτοποίηση τιμών και επικύρωση

5.2.2.1 Αλγόριθμος J48 -C 0,25 –M 2 με επικύρωση 5

Εκπαιδύσαμε τα δεδομένα μας με την μέθοδο της διακριτοποίησης τιμών με επικύρωση 5 (fold cross), με τον αλγόριθμο J48 -C 0,25 –M 2 και πήραμε τα εξής αποτελέσματα: από το σύνολο των μεταβλητών που αναλύσαμε στο κεφάλαιο 5.1

Το συνολικό μέγεθος του δέντρου είναι 49 κόμβοι. Συνολικά από το μοντέλο που προκύπτει (μη εκπαιδευμένα δέντρα) βλέπουμε πως δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς 2614 άτομα από τα οποία χάθηκαν τα 15 άτομα όπου έχουν διαγνωστεί με την νόσο του θυρεοειδούς ενώ έχουν διαγνωστεί με την νόσο του θυρεοειδούς 147 άτομα από τα οποία χάθηκαν τα 24 άτομα όπου δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς. Σε ποσοστό ακρίβειας των δεδομένων είναι 98,6071% σύμφωνα με το πρότυπο, ενώ το 1,3929% δεν είναι σύμφωνα με το πρότυπο.

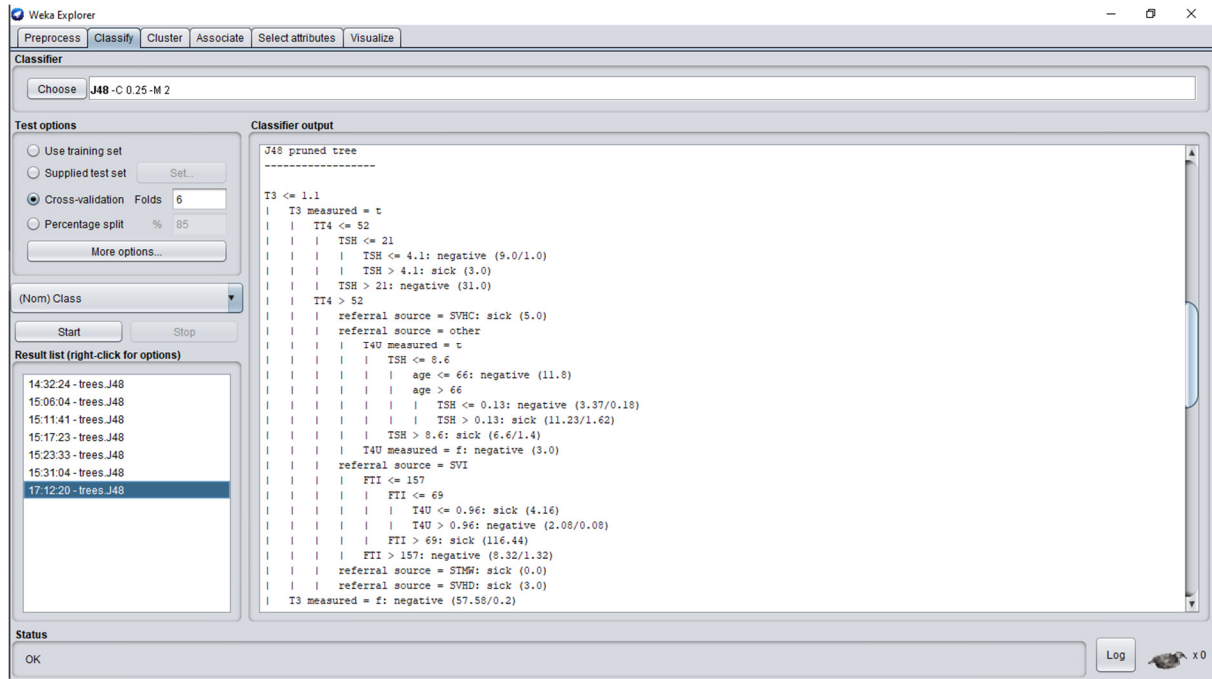


Εικόνα 63: Αλγόριθμος J48 -C 0,25 –M 2 με επικύρωση 6

5.2.2.2 Αλγόριθμος J48 -C 0,25 –M 2 με επικύρωση 6

Εκπαιδύσαμε τα δεδομένα μας με την μέθοδο της διακριτοποίησης τιμών με επικύρωση 6 (fold cross), με τον αλγόριθμο J48 -C 0,25 –M 2 και πήραμε τα εξής αποτελέσματα: από το σύνολο των μεταβλητών που αναλύσαμε στο κεφάλαιο 5.1

Το συνολικό μέγεθος του δέντρου είναι 49 κόμβοι. Συνολικά από το μοντέλο που προκύπτει βλέπουμε πως δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς 2615 άτομα από τα οποία χάθηκαν τα 14 άτομα όπου έχουν διαγνωστεί με την νόσο του θυρεοειδούς ενώ έχουν διαγνωστεί με την νόσο του θυρεοειδούς 148 άτομα από τα οποία χάθηκαν τα 23 άτομα όπου δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς. Σε ποσοστό ακρίβειας των δεδομένων είναι 98,6786% σύμφωνα με το πρότυπο, ενώ το 1,3214% δεν είναι σύμφωνα με το πρότυπο.

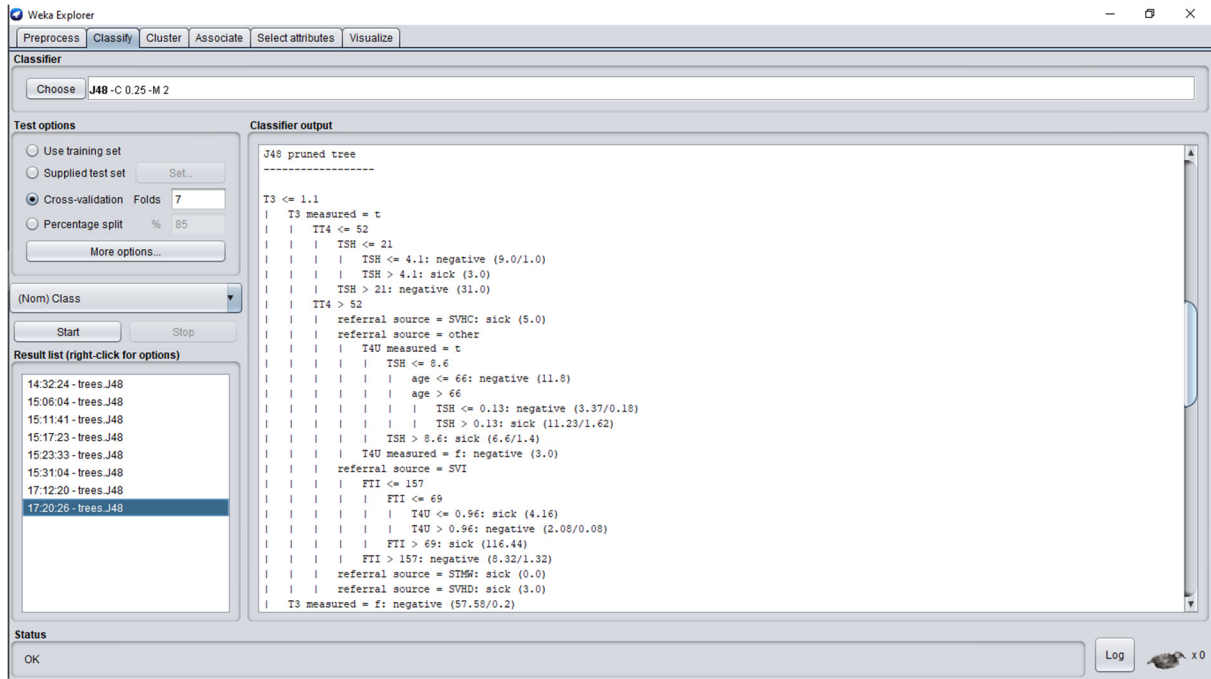


Εικόνα 64: Αλγόριθμος J48 -C 0,25 –M 2 με επικύρωση 6

5.2.2.3 Αλγόριθμος J48 -C 0,25 –M 2 με επικύρωση 7

Εκπαιδύσαμε τα δεδομένα μας με την μέθοδο της διακριτοποίησης τιμών με επικύρωση 7 (fold cross), με τον αλγόριθμο J48 -C 0,25 –M 2 και πήραμε τα εξής αποτελέσματα: από το σύνολο των μεταβλητών που αναλύσαμε στο κεφάλαιο 5.1

Το συνολικό μέγεθος του δέντρου είναι 49 κόμβοι. Συνολικά από το μοντέλο που προκύπτει βλέπουμε πως δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς 2.614 άτομα από τα οποία χάθηκαν τα 15 άτομα όπου έχουν διαγνωστεί με την νόσο του θυρεοειδούς ενώ έχουν διαγνωστεί με την νόσο του θυρεοειδούς 152 άτομα από τα οποία χάθηκαν τα 19 άτομα όπου δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς. Σε ποσοστό ακρίβειας των δεδομένων είναι 98,7857% σύμφωνα με το πρότυπο, ενώ το 1,2143% δεν είναι σύμφωνα με το πρότυπο.

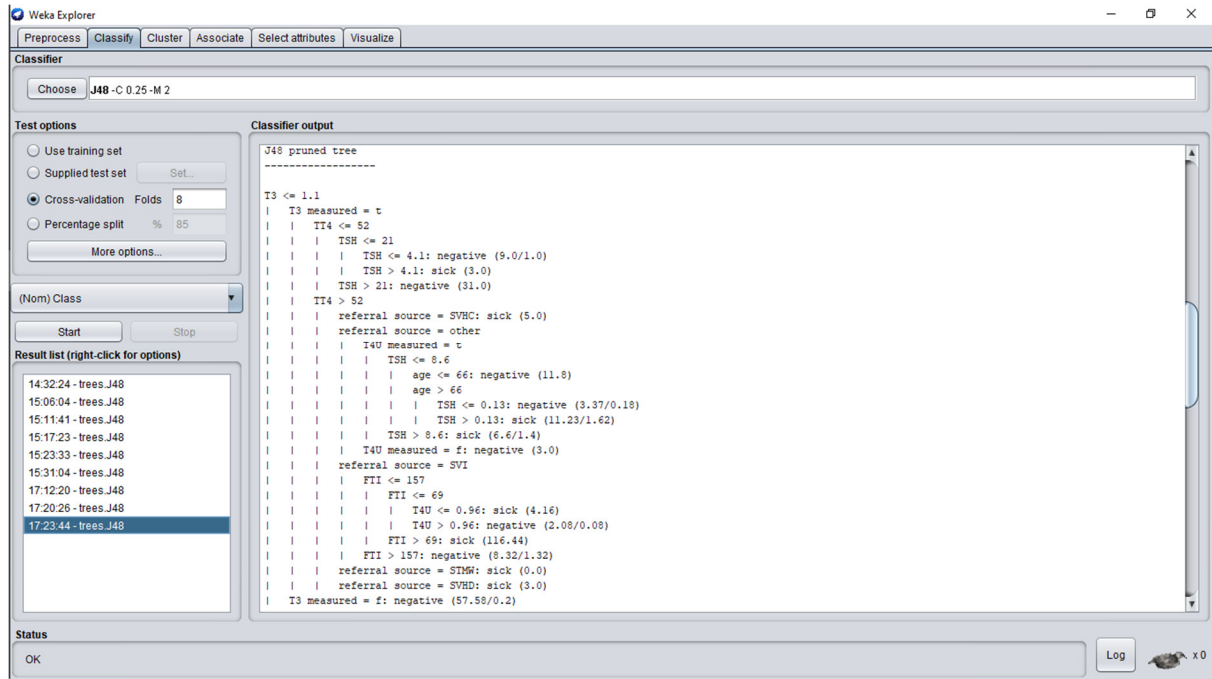


Εικόνα 65: Αλγόριθμος J48 -C 0,25 –M 2 με επικύρωση 7

5.2.2.4 Αλγόριθμος J48 -C 0,25 –M 2 με επικύρωση 8

Εκπαιδεύσαμε τα δεδομένα μας με την μέθοδο της διακριτοποίησης τιμών με επικύρωση 8 (fold cross), με τον αλγόριθμο J48 -C 0,25 –M 2 και πήραμε τα εξής αποτελέσματα: από το σύνολο των μεταβλητών που αναλύσαμε στο κεφάλαιο 5.1

Το συνολικό μέγεθος του δέντρου είναι 49 κόμβοι. Συνολικά από το μοντέλο που προκύπτει βλέπουμε πως δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς 2.617 άτομα από τα οποία χάθηκαν τα 12 άτομα όπου έχουν διαγνωστεί με την νόσο του θυρεοειδούς ενώ έχουν διαγνωστεί με την νόσο του θυρεοειδούς 149 άτομα από τα οποία χάθηκαν τα 22 άτομα όπου δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς. Σε ποσοστό ακρίβειας των δεδομένων είναι 98,7857% σύμφωνα με το πρότυπο, ενώ το 1,2143% δεν είναι σύμφωνα με το πρότυπο.

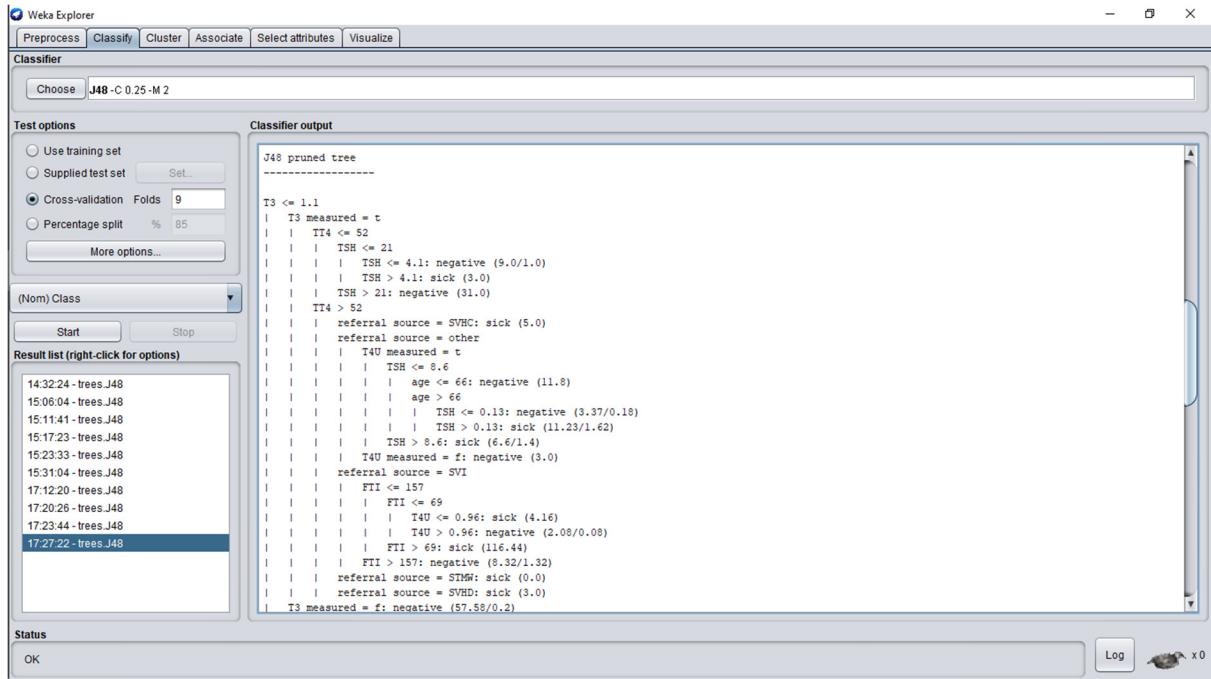


Εικόνα 66: Αλγόριθμος J48 -C 0,25 –M 2 με επικύρωση 8

5.2.2.5 Αλγόριθμος J48 -C 0,25 –M 2 με επικύρωση 9

Εκπαιδύσαμε τα δεδομένα μας με την μέθοδο της διακριτοποίησης τιμών με επικύρωση 9 (fold cross), με τον αλγόριθμο J48 -C 0,25 –M 2 και πήραμε τα εξής αποτελέσματα: από το σύνολο των μεταβλητών που αναλύσαμε στο κεφάλαιο 5.1

Το συνολικό μέγεθος του δέντρου είναι 49 κόμβοι. Συνολικά από το μοντέλο που προκύπτει βλέπουμε πως δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς 2.616 άτομα από τα οποία χάθηκαν τα 13 άτομα όπου έχουν διαγνωστεί με την νόσο του θυρεοειδούς ενώ έχουν διαγνωστεί με την νόσο του θυρεοειδούς 152 άτομα από τα οποία χάθηκαν τα 19 άτομα όπου δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς. Σε ποσοστό ακρίβειας των δεδομένων είναι 98,8571% σύμφωνα με το πρότυπο, ενώ το 1,1429% δεν είναι σύμφωνα με το πρότυπο.

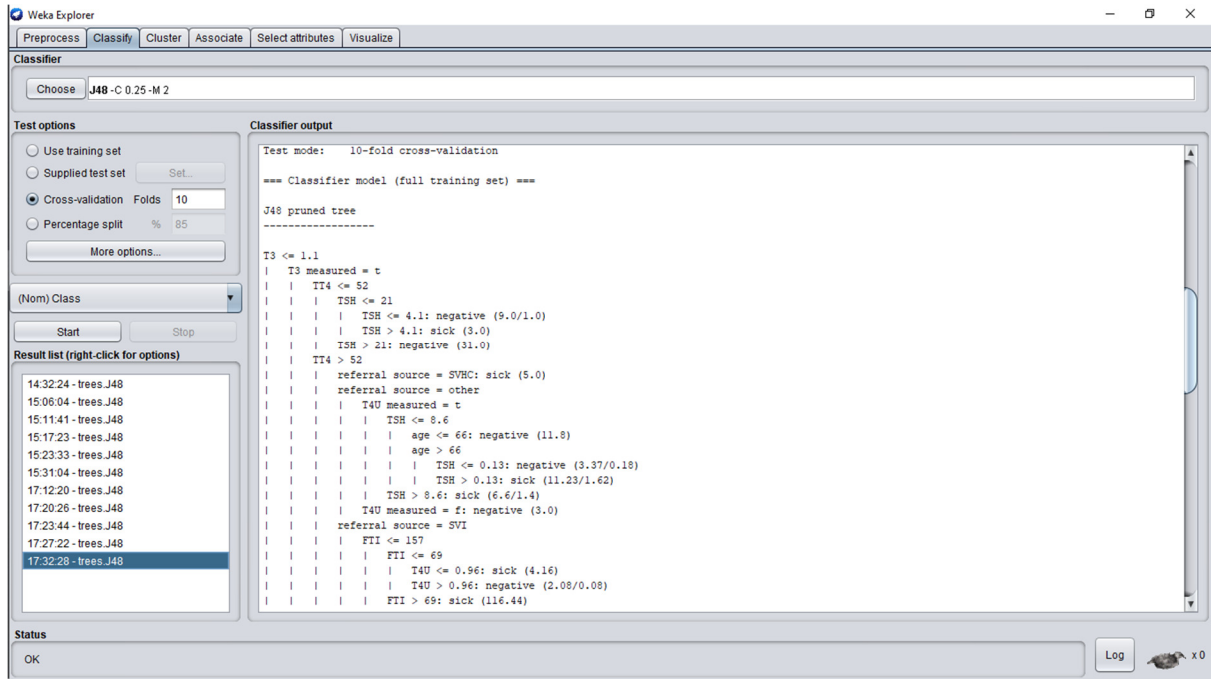


Εικόνα 67: Αλγόριθμος J48 -C 0,25 –M 2 με επικύρωση 9

5.2.2.6 Αλγόριθμος J48 -C 0,25 –M 2 με επικύρωση 10

Εκπαιδύσαμε τα δεδομένα μας με την μέθοδο της διακριτοποίησης τιμών με επικύρωση 10 (fold cross), με τον αλγόριθμο J48 -C 0,25 –M 2 και πήραμε τα εξής αποτελέσματα: από το σύνολο των μεταβλητών που αναλύσαμε στο κεφάλαιο 5.1

Το συνολικό μέγεθος του δέντρου είναι 49 κόμβοι. Συνολικά από το μοντέλο που προκύπτει βλέπουμε πως δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς 2.619 άτομα από τα οποία χάθηκαν τα 10 άτομα όπου έχουν διαγνωστεί με την νόσο του θυρεοειδούς ενώ έχουν διαγνωστεί με την νόσο του θυρεοειδούς 151 άτομα από τα οποία χάθηκαν τα 20 άτομα όπου δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς. Σε ποσοστό ακρίβειας των δεδομένων είναι 98,9286% σύμφωνα με το πρότυπο, ενώ το 1,0714% δεν είναι σύμφωνα με το πρότυπο.

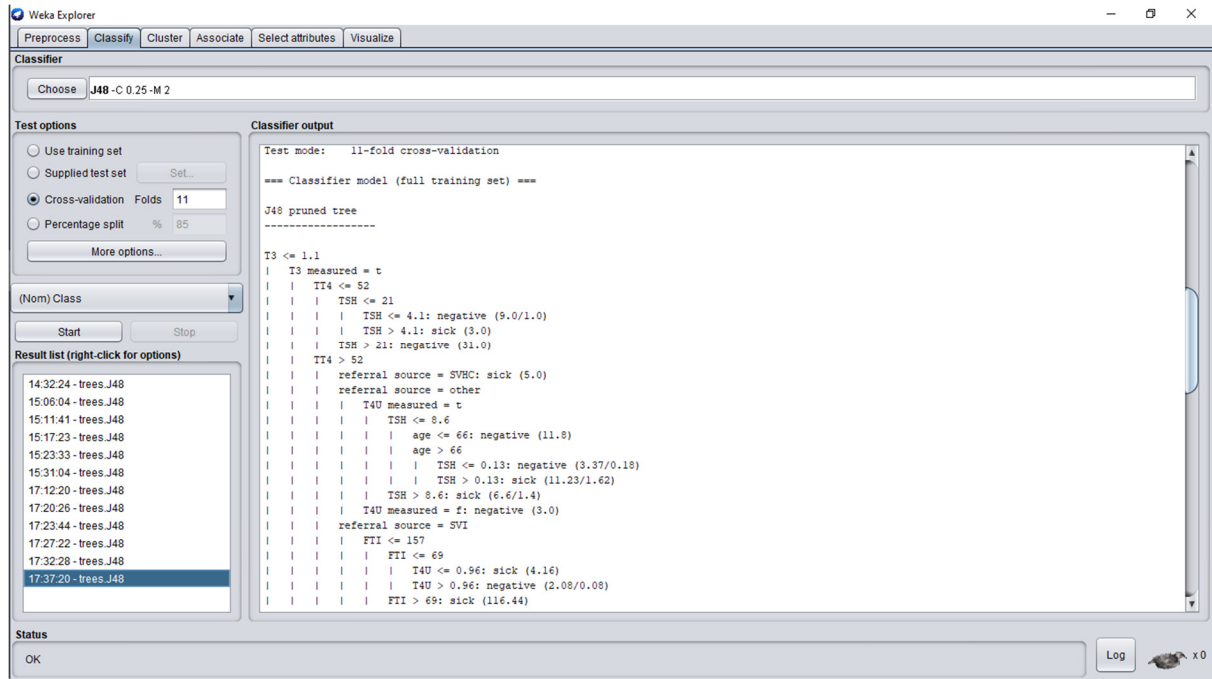


Εικόνα 68: Αλγόριθμος J48 -C 0,25 –M 2 με επικύρωση 10

5.2.2.7 Αλγόριθμος J48 -C 0,25 –M 2 με επικύρωση 11

Εκπαιδύσαμε τα δεδομένα μας με την μέθοδο της διακριτοποίησης τιμών με επικύρωση 11 (fold cross), με τον αλγόριθμο J48 -C 0,25 –M 2 και πήραμε τα εξής αποτελέσματα: από το σύνολο των μεταβλητών που αναλύσαμε στο κεφάλαιο 5.1

Το συνολικό μέγεθος του δέντρου είναι 49 κόμβοι. Συνολικά από το μοντέλο που προκύπτει βλέπουμε πως δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς 2.614 άτομα από τα οποία χάθηκαν τα 15 άτομα όπου έχουν διαγνωστεί με την νόσο του θυρεοειδούς ενώ έχουν διαγνωστεί με την νόσο του θυρεοειδούς 152 άτομα από τα οποία χάθηκαν τα 19 άτομα όπου δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς. Σε ποσοστό ακρίβειας των δεδομένων είναι 98,7857% σύμφωνα με το πρότυπο, ενώ το 1,2143% δεν είναι σύμφωνα με το πρότυπο.

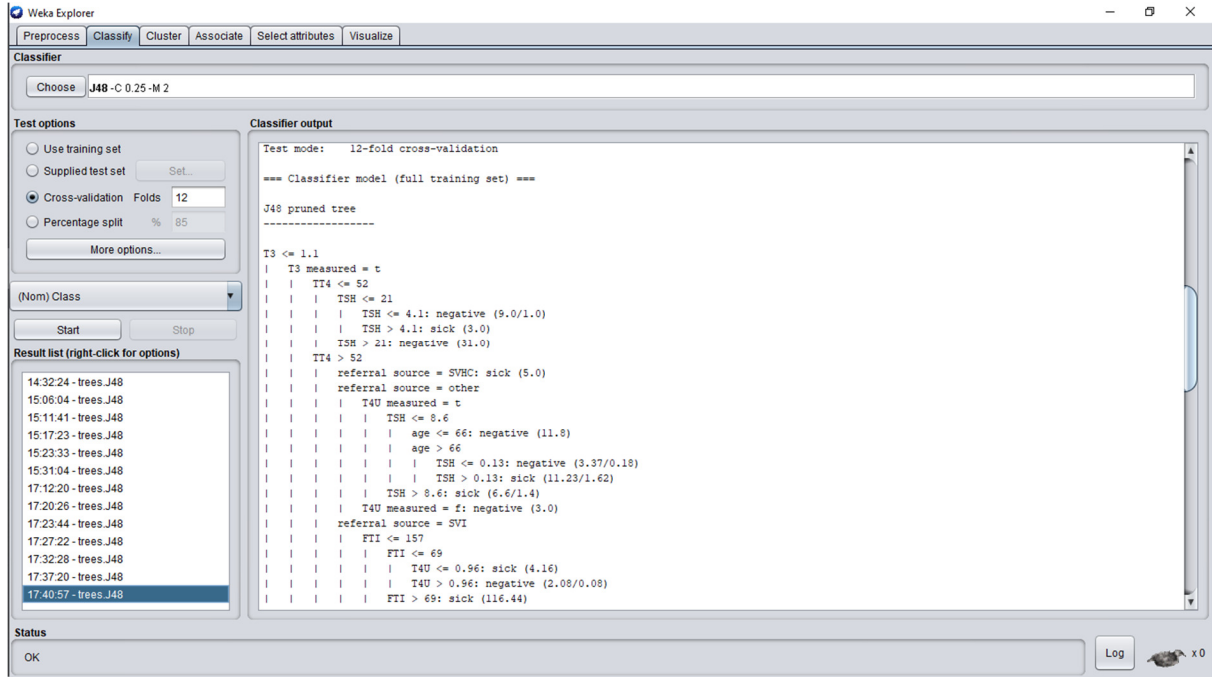


Εικόνα 69: Αλγόριθμος J48 -C 0,25 –M 2 με επικύρωση 11

5.2.2.8 Αλγόριθμος J48 -C 0,25 –M 2 με επικύρωση 12

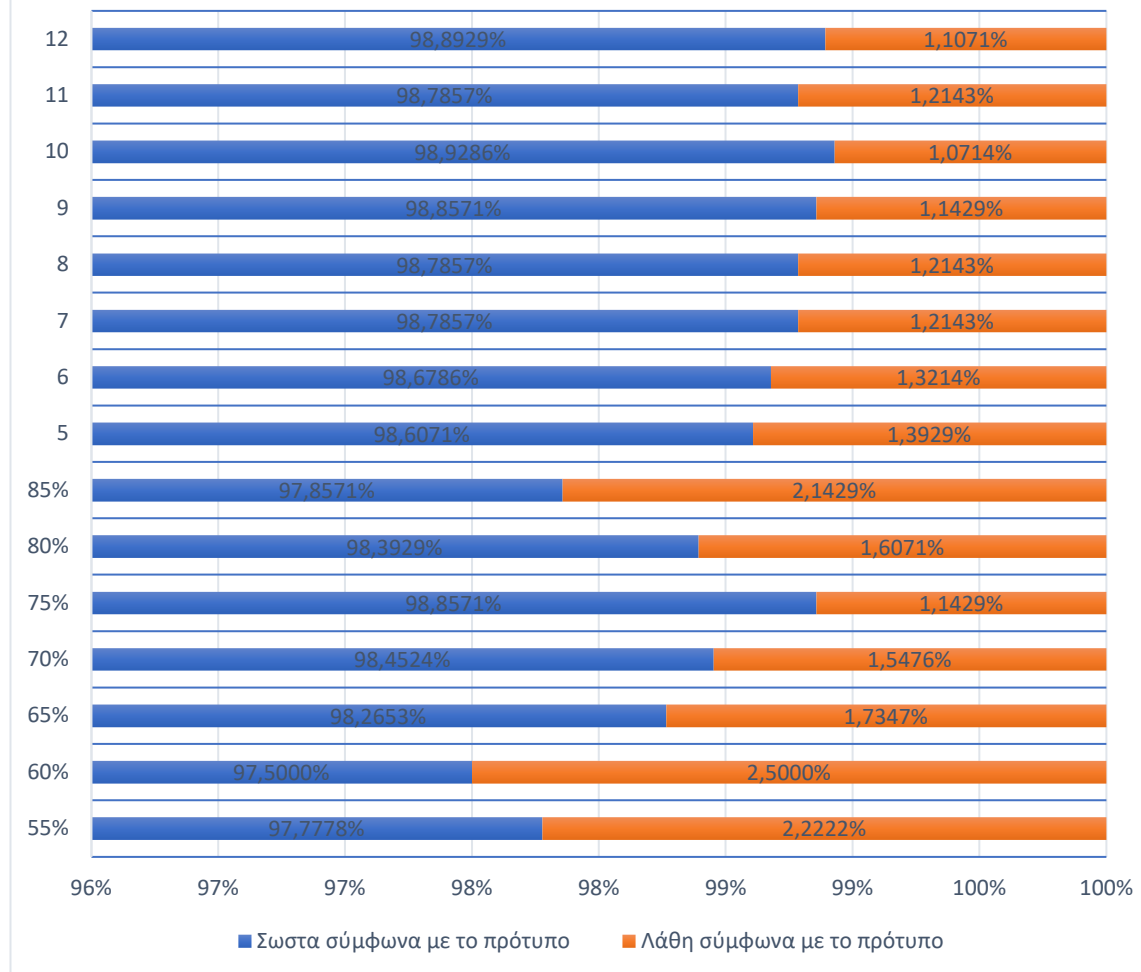
Εκπαιδύσαμε τα δεδομένα μας με την μέθοδο της διακριτοποίησης τιμών με επικύρωση 12 (fold cross), με τον αλγόριθμο J48 -C 0,25 –M 2 και πήραμε τα εξής αποτελέσματα: από το σύνολο των μεταβλητών που αναλύσαμε στο κεφάλαιο 5.1

Το συνολικό μέγεθος του δέντρου είναι 49 κόμβοι. Συνολικά από το μοντέλο που προκύπτει βλέπουμε πως δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς 2.618 άτομα από τα οποία χάθηκαν τα 11 άτομα όπου έχουν διαγνωστεί με την νόσο του θυρεοειδούς ενώ έχουν διαγνωστεί με την νόσο του θυρεοειδούς 151 άτομα από τα οποία χάθηκαν τα 20 άτομα όπου δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς. Σε ποσοστό ακρίβειας των δεδομένων είναι 98,8929% σύμφωνα με το πρότυπο, ενώ το 1,1071% δεν είναι σύμφωνα με το πρότυπο.

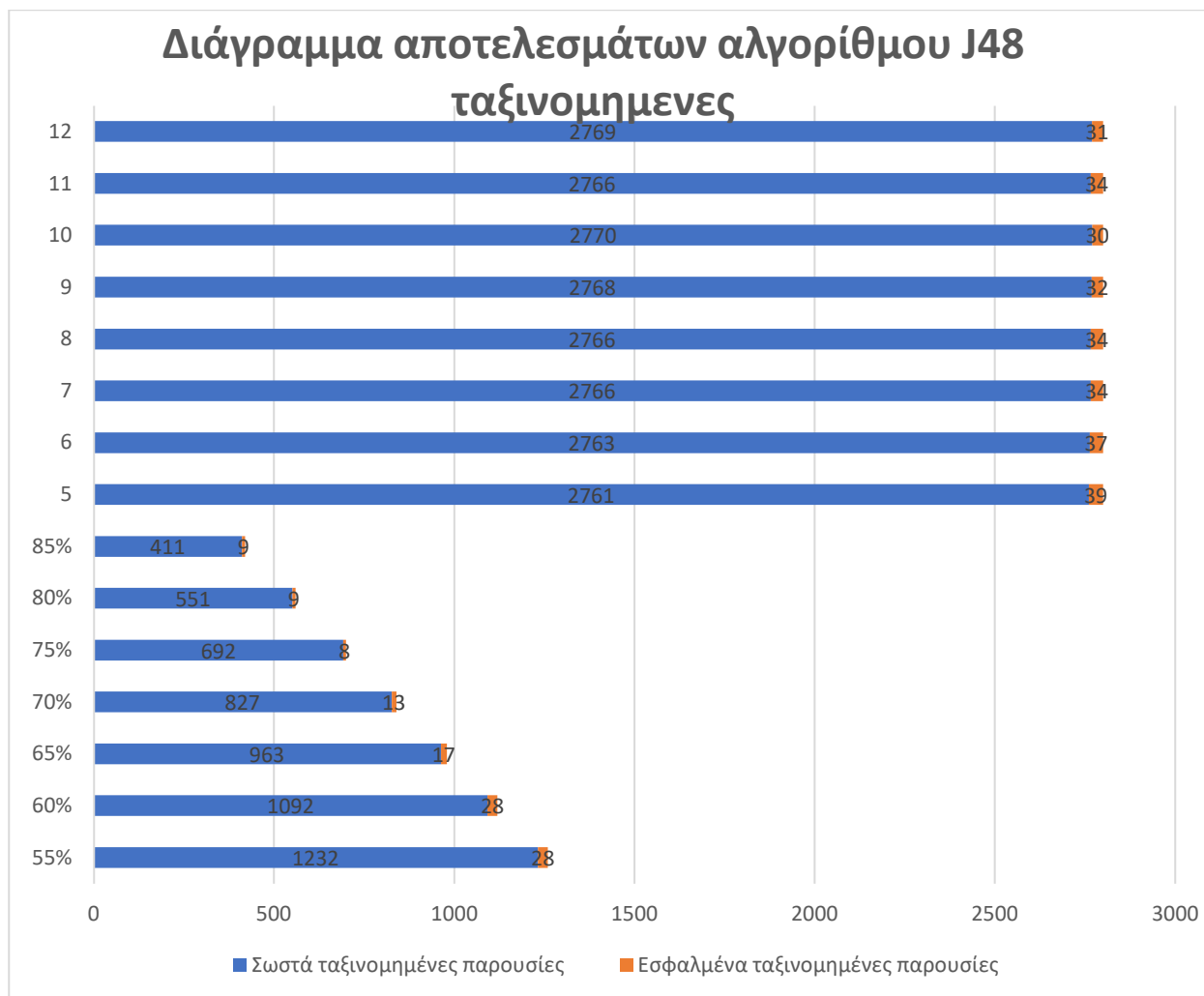


Εικόνα 70: Αλγόριθμος J48 -C 0,25 –M 2 με επικύρωση 12

Διάγραμμα αποτελεσμάτων αλγορίθμου J48



Διάγραμμα 1: Διάγραμμα αποτελεσμάτων αλγορίθμου J48



Διάγραμμα 2: Διάγραμμα αποτελεσμάτων αλγορίθμου J48 ταξινομημένες

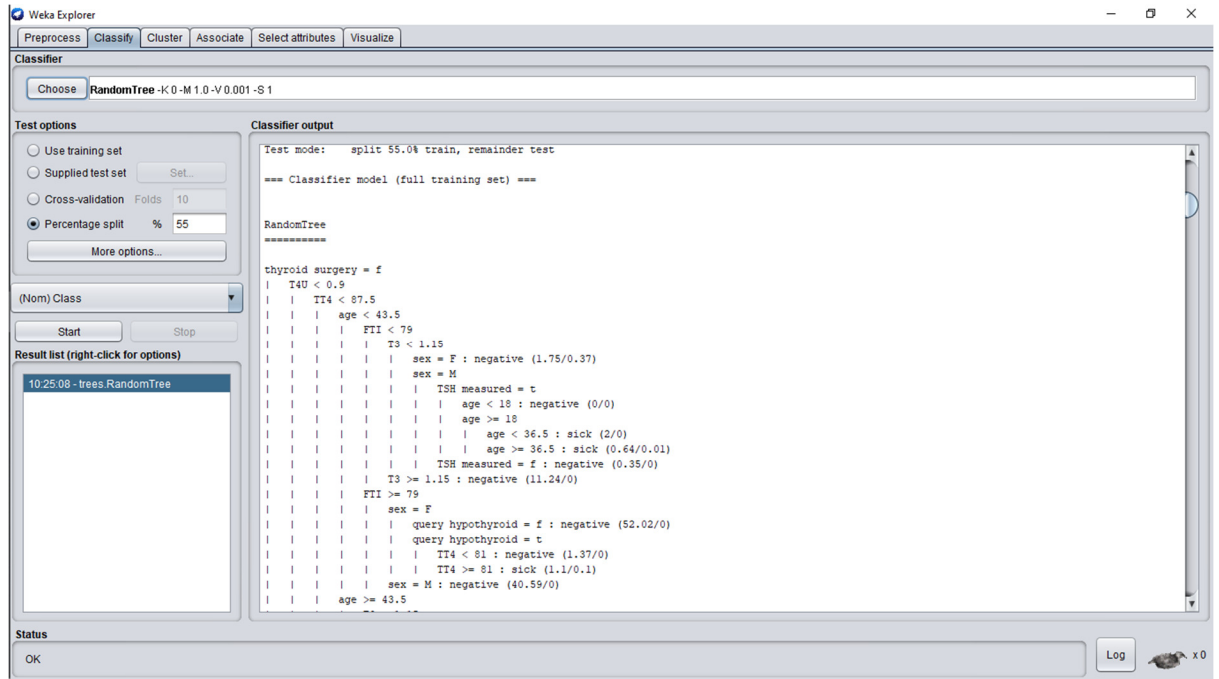
5.2.3 Δέντρο απόφασης αλγορίθμου Random Tree με διακριτοποίηση τιμών

5.2.3.1 Αλγόριθμος Random Tree -K 0 -M 1,0 -V 0,001 -S 1 με διακριτοποίηση 55%

Εκπαιδεύσαμε το 55% των δεδομένων μας με την μέθοδο της διακριτοποίησης τιμών, με τον αλγόριθμο Random Tree -K 0 -M 1,0 -V 0,001 -S 1 και πήραμε τα εξής αποτελέσματα: από το σύνολο των μεταβλητών που αναλύσαμε στο κεφάλαιο 5.1

Το συνολικό μέγεθος του δέντρου είναι 420 κόμβοι. Συνολικά από το μοντέλο που προκύπτει(μη εκπαιδευμένα δέντρα) βλέπουμε πως δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς 1.159 άτομα από τα οποία χάθηκαν τα 27 άτομα όπου έχουν διαγνωστεί με την νόσο του θυρεοειδούς ενώ έχουν διαγνωστεί με την νόσο του θυρεοειδούς 63 άτομα από τα οποία χάθηκαν τα 11 άτομα όπου δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς. Σε

ποσοστό ακρίβειας των δεδομένων είναι 96,9841% σύμφωνα με το πρότυπο, ενώ το 3,0159% δεν είναι σύμφωνα με το πρότυπο.



Εικόνα 71: Αλγόριθμος Random Tree -K 0 –M 1,0 -V 0,001 -S 1 με διακριτοποίηση 55%

5.2.3.2 Αλγόριθμος Random Tree -K 0 –M 1,0 -V 0,001 -S 1 με διακριτοποίηση 60%

Εκπαιδύσαμε το 60% των δεδομένων μας με την μέθοδο της διακριτοποίησης τιμών, με τον αλγόριθμο Random Tree -K 0 –M 1,0 -V 0,001 -S 1 και πήραμε τα έξι αποτελέσματα: από το σύνολο των μεταβλητών που αναλύσαμε στο κεφάλαιο 5.1

Το συνολικό μέγεθος του δέντρου είναι 420 κόμβοι. Συνολικά από το μοντέλο που προκύπτει(μη εκπαιδευμένα δέντρα) βλέπουμε πως δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς 1.048 άτομα από τα οποία χάθηκαν τα 10 άτομα όπου έχουν διαγνωστεί με την νόσο του θυρεοειδούς ενώ έχουν διαγνωστεί με την νόσο του θυρεοειδούς 45 άτομα από τα οποία χάθηκαν τα 17 άτομα όπου δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς. Σε ποσοστό ακρίβειας των δεδομένων είναι 97,5893% σύμφωνα με το πρότυπο, ενώ το 2,4107% δεν είναι σύμφωνα με το πρότυπο.

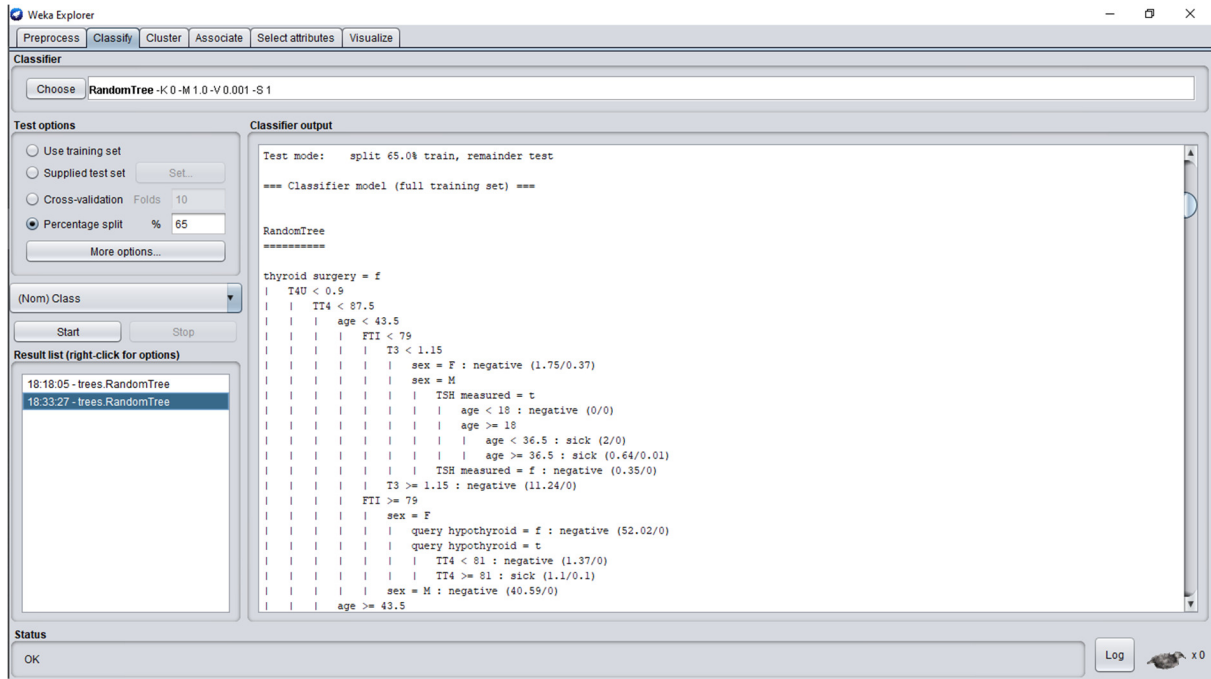


Εικόνα 72: Αλγόριθμος Random Tree -K 0 –M 1,0 -V 0,001 -S 1 με διακριτοποίηση 60%

5.2.3.3 Αλγόριθμος Random Tree -K 0 –M 1,0 -V 0,001 -S 1 με διακριτοποίηση 65%

Εκπαιδύσαμε το 65% των δεδομένων μας με την μέθοδο της διακριτοποίησης τιμών, με τον αλγόριθμο Random Tree -K 0 –M 1,0 -V 0,001 -S 1 και πήραμε τα έξι αποτελέσματα: από το σύνολο των μεταβλητών που αναλύσαμε στο κεφάλαιο 5.1

Το συνολικό μέγεθος του δέντρου είναι 420 κόμβοι. Συνολικά από το μοντέλο που προκύπτει(μη εκπαιδευμένα δέντρα) βλέπουμε πως δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς 914 άτομα από τα οποία χάθηκαν τα 11 άτομα όπου έχουν διαγνωστεί με την νόσο του θυρεοειδούς ενώ έχουν διαγνωστεί με την νόσο του θυρεοειδούς 44 άτομα από τα οποία χάθηκαν τα 11 άτομα όπου δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς. Σε ποσοστό ακρίβειας των δεδομένων είναι 97,7551% σύμφωνα με το πρότυπο, ενώ το 2,2449% δεν είναι σύμφωνα με το πρότυπο.

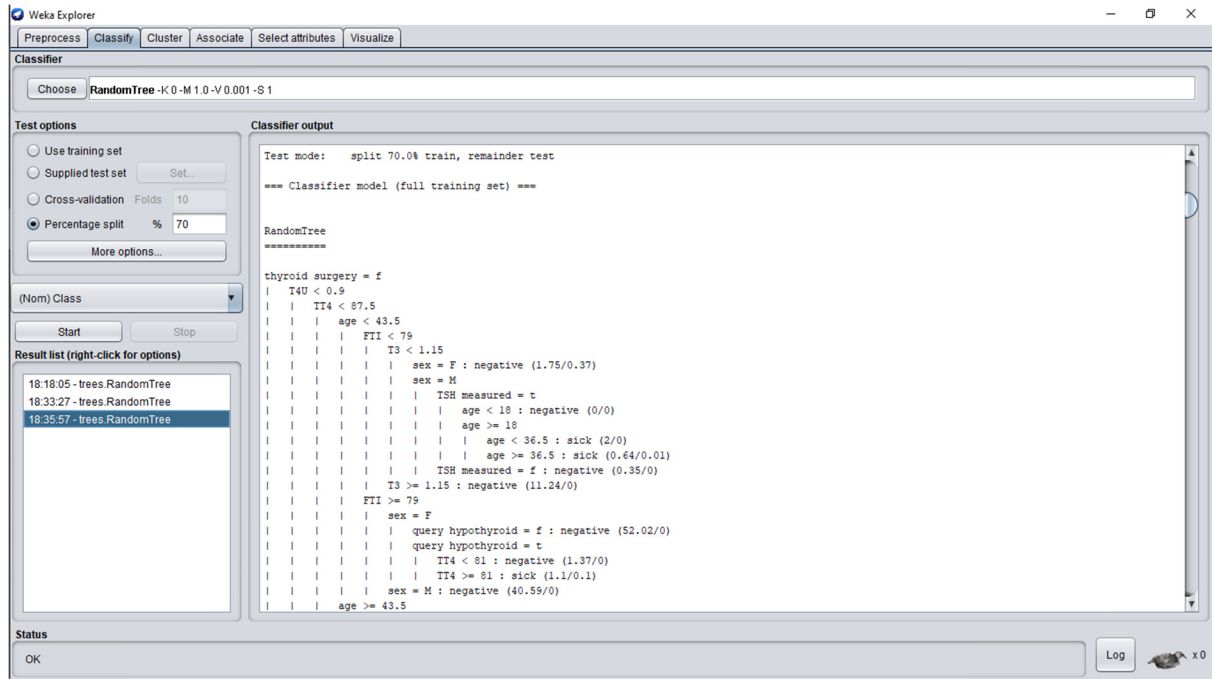


Εικόνα 73: Αλγόριθμος Random Tree -K 0 –M 1,0 -V 0,001 -S 1 με διακριτοποίηση 65%

5.2.3.4 Αλγόριθμος Random Tree -K 0 –M 1,0 -V 0,001 -S 1 με διακριτοποίηση 70%

Εκπαιδύσαμε το 70% των δεδομένων μας με την μέθοδο της διακριτοποίησης τιμών, με τον αλγόριθμο Random Tree -K 0 –M 1,0 -V 0,001 -S 1 και πήραμε τα έξι αποτελέσματα: από το σύνολο των μεταβλητών που αναλύσαμε στο κεφάλαιο 5.1

Το συνολικό μέγεθος του δέντρου είναι 420 κόμβοι. Συνολικά από το μοντέλο που προκύπτει(μη εκπαιδευμένα δέντρα) βλέπουμε πως δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς 787 άτομα από τα οποία χάθηκαν τα 8 άτομα όπου έχουν διαγνωστεί με την νόσο του θυρεοειδούς ενώ έχουν διαγνωστεί με την νόσο του θυρεοειδούς 38 άτομα από τα οποία χάθηκαν τα 7 άτομα όπου δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς. Σε ποσοστό ακρίβειας των δεδομένων είναι 98,2143% σύμφωνα με το πρότυπο, ενώ το 1,7857% δεν είναι σύμφωνα με το πρότυπο.

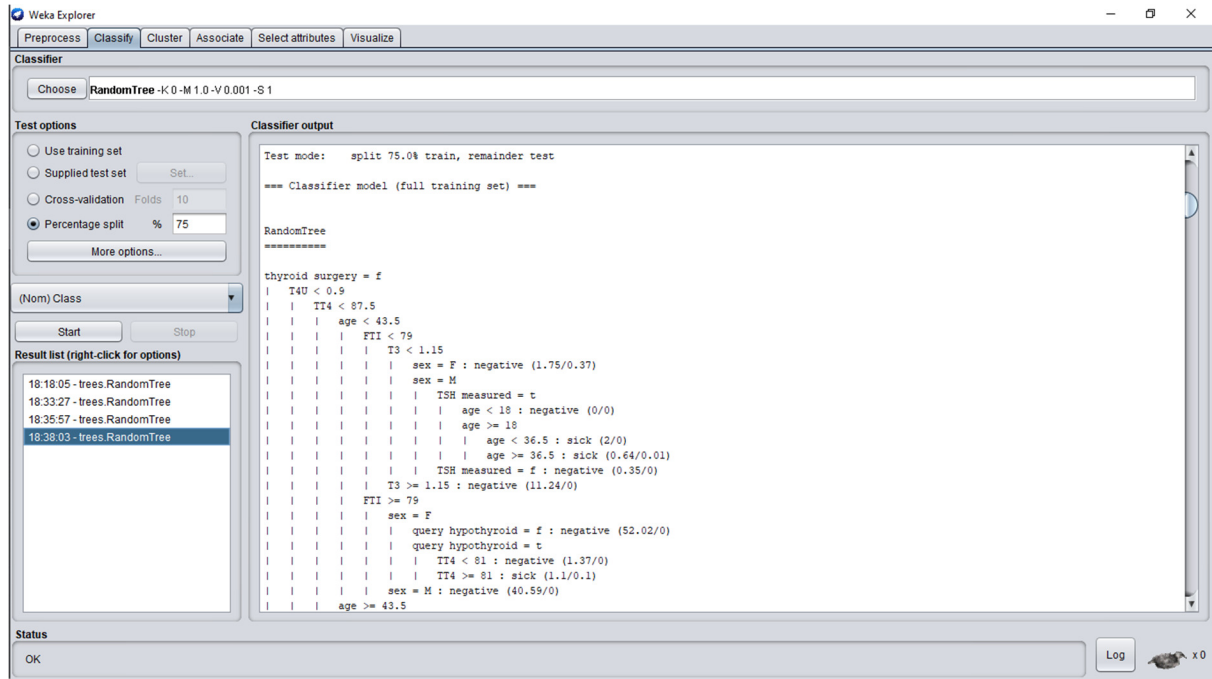


Εικόνα 74: Αλγόριθμος Random Tree -K 0 –M 1,0 -V 0,001 -S 1 με διακριτοποίηση 70%

5.2.3.5 Αλγόριθμος Random Tree -K 0 –M 1,0 -V 0,001 -S 1 με διακριτοποίηση 75%

Εκπαιδύσαμε το 75% των δεδομένων μας με την μέθοδο της διακριτοποίησης τιμών, με τον αλγόριθμο Random Tree -K 0 –M 1,0 -V 0,001 -S 1 και πήραμε τα έξι αποτελέσματα: από το σύνολο των μεταβλητών που αναλύσαμε στο κεφάλαιο 5.1

Το συνολικό μέγεθος του δέντρου είναι 420 κόμβοι. Συνολικά από το μοντέλο που προκύπτει(μη εκπαιδευμένα δέντρα) βλέπουμε πως δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς 649 άτομα από τα οποία χάθηκαν τα 18 άτομα όπου έχουν διαγνωστεί με την νόσο του θυρεοειδούς ενώ έχουν διαγνωστεί με την νόσο του θυρεοειδούς 23 άτομα από τα οποία χάθηκαν τα 10 άτομα όπου δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς. Σε ποσοστό ακρίβειας των δεδομένων είναι 96% σύμφωνα με το πρότυπο, ενώ το 4% δεν είναι σύμφωνα με το πρότυπο.

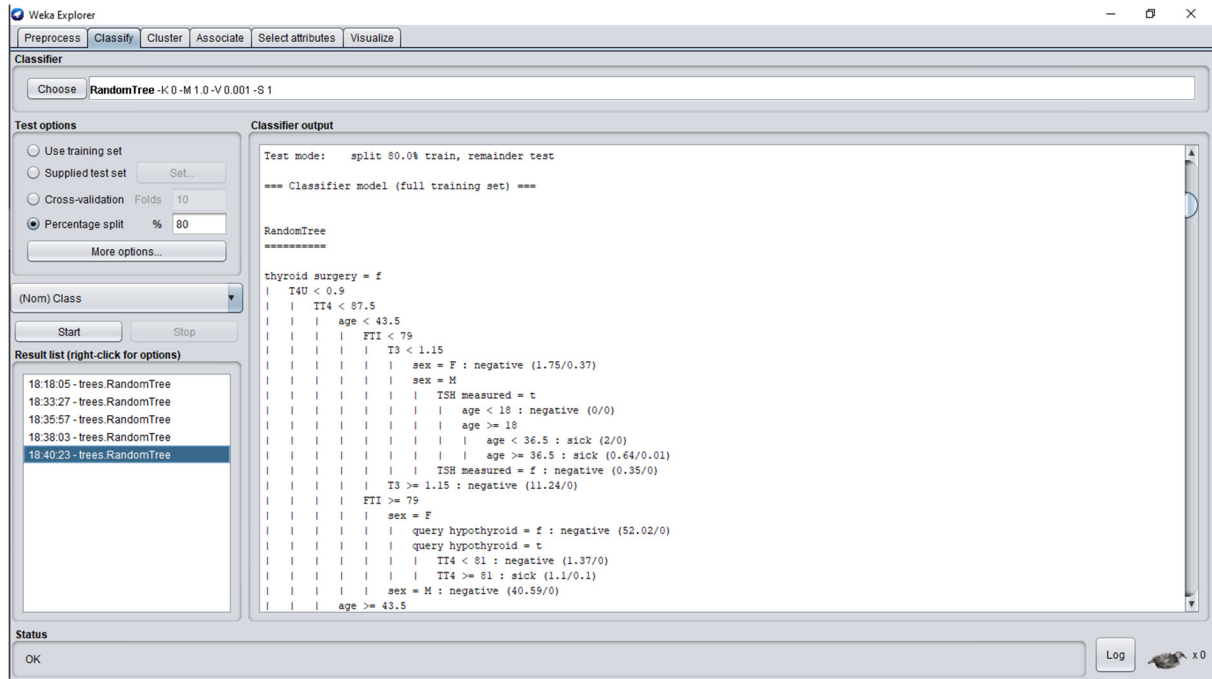


Εικόνα 75: Αλγόριθμος Random Tree -K 0 –M 1,0 -V 0,001 -S 1 με διακριτοποίηση 75%

5.2.3.6 Αλγόριθμος Random Tree -K 0 –M 1,0 -V 0,001 -S 1 με διακριτοποίηση 80%

Εκπαιδύσαμε το 80% των δεδομένων μας με την μέθοδο της διακριτοποίησης τιμών, με τον αλγόριθμο Random Tree -K 0 –M 1,0 -V 0,001 -S 1 και πήραμε τα έξι αποτελέσματα: από το σύνολο των μεταβλητών που αναλύσαμε στο κεφάλαιο 5.1

Το συνολικό μέγεθος του δέντρου είναι 420 κόμβοι. Συνολικά από το μοντέλο που προκύπτει(μη εκπαιδευμένα δέντρα) βλέπουμε πως δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς 527 άτομα από τα οποία χάθηκαν τα 10 άτομα όπου έχουν διαγνωστεί με την νόσο του θυρεοειδούς ενώ έχουν διαγνωστεί με την νόσο του θυρεοειδούς 19 άτομα από τα οποία χάθηκαν τα 4 άτομα όπου δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς. Σε ποσοστό ακρίβειας των δεδομένων είναι 97,5% σύμφωνα με το πρότυπο, ενώ το 2,5% δεν είναι σύμφωνα με το πρότυπο.

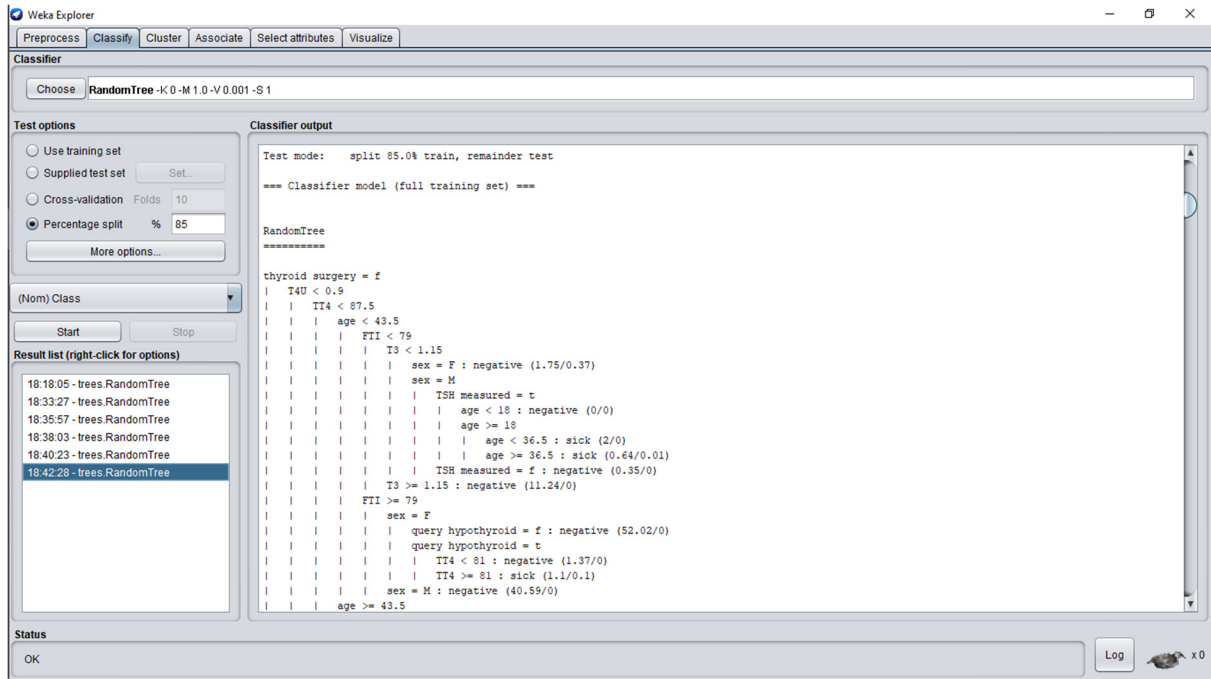


Εικόνα 76: Αλγόριθμος Random Tree -K 0 –M 1,0 -V 0,001 -S 1 με διακριτοποίηση 80%

5.2.3.7 Αλγόριθμος Random Tree -K 0 –M 1,0 -V 0,001 -S 1 με διακριτοποίηση 85%

Εκπαιδύσαμε το 85% των δεδομένων μας με την μέθοδο της διακριτοποίησης τιμών, με τον αλγόριθμο Random Tree -K 0 –M 1,0 -V 0,001 -S 1 και πήραμε τα έξι αποτελέσματα: από το σύνολο των μεταβλητών που αναλύσαμε στο κεφάλαιο 5.1

Το συνολικό μέγεθος του δέντρου είναι 420 κόμβοι. Συνολικά από το μοντέλο που προκύπτει(μη εκπαιδευμένα δέντρα) βλέπουμε πως δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς 393 άτομα από τα οποία χάθηκαν τα 7 άτομα όπου έχουν διαγνωστεί με την νόσο του θυρεοειδούς ενώ έχουν διαγνωστεί με την νόσο του θυρεοειδούς 17 άτομα από τα οποία χάθηκαν τα 3 άτομα όπου δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς. Σε ποσοστό ακρίβειας των δεδομένων είναι 97,619% σύμφωνα με το πρότυπο, ενώ το 2,381% δεν είναι σύμφωνα με το πρότυπο.



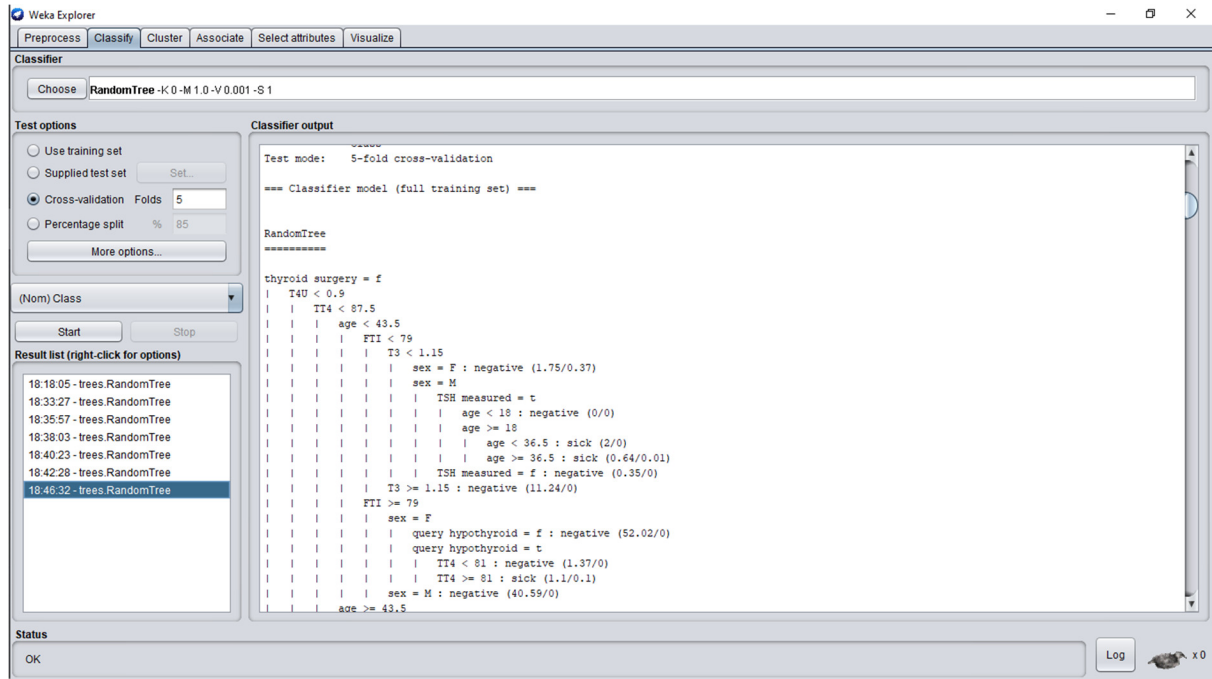
Εικόνα 77: Αλγόριθμος Random Tree -K 0 –M 1,0 -V 0,001 -S 1 με διακριτοποίηση 85%

5.2.4 Δέντρο απόφασης αλγορίθμου Random Tree με διακριτοποίηση τιμών και επικύρωση

5.2.4.1 Αλγόριθμος Random Tree -K 0 –M 1,0 -V 0,001 -S 1 με επικύρωση 5

Εκπαίδευσάμε τα δεδομένα μας με την μέθοδο της διακριτοποίησης τιμών με επικύρωση 5 (fold cross), με τον αλγόριθμο Random Tree -K 0 –M 1,0 -V 0,001 -S 1 και πήραμε τα έξι αποτελέσματα: από το σύνολο των μεταβλητών που αναλύσαμε στο κεφάλαιο 5.1

Συνολικά από το μοντέλο που προκύπτει(μη εκπαιδευμένα δέντρα) βλέπουμε πως δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς 2.587 άτομα από τα οποία χάθηκαν τα 42 άτομα όπου έχουν διαγνωστεί με την νόσο του θυρεοειδούς ενώ έχουν διαγνωστεί με την νόσο του θυρεοειδούς 123 άτομα από τα οποία χάθηκαν τα 48 άτομα όπου δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς. Σε ποσοστό ακρίβειας των δεδομένων είναι 96,7857% σύμφωνα με το πρότυπο, ενώ το 3,2143% δεν είναι σύμφωνα με το πρότυπο.

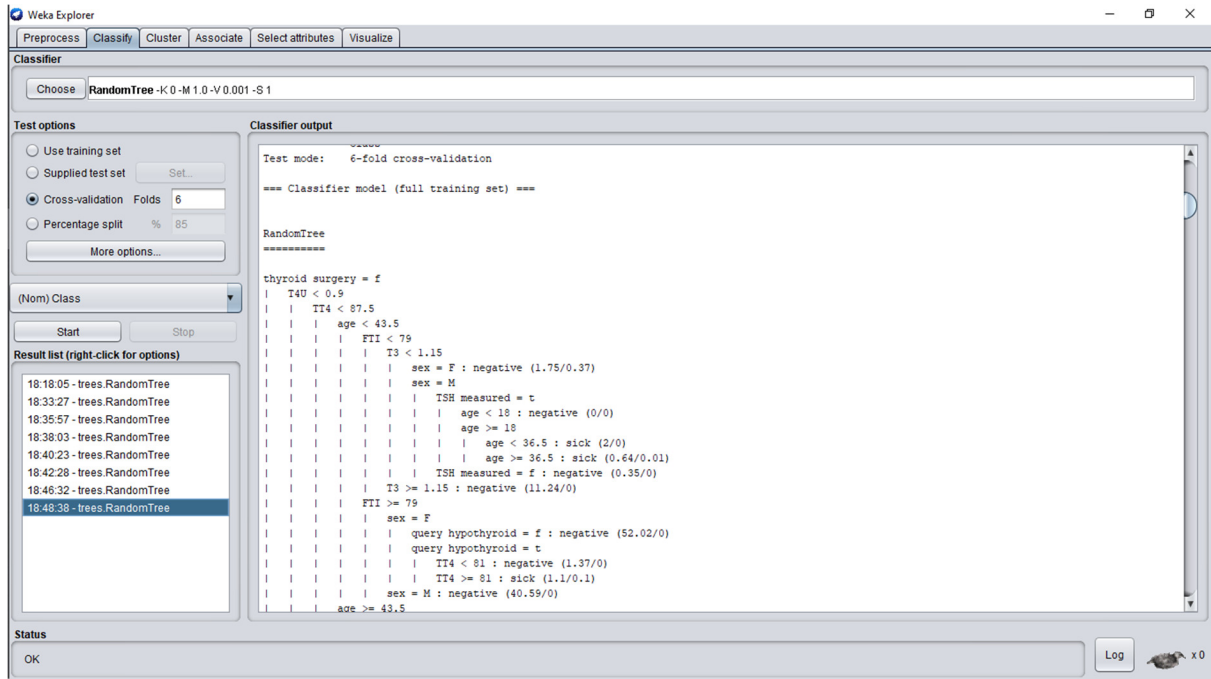


Εικόνα 78: Αλγόριθμος Random Tree -K 0 –M 1,0 -V 0,001 -S 1 με επικύρωση 5

5.2.4.2 Αλγόριθμος Random Tree -K 0 –M 1,0 -V 0,001 -S 1 με επικύρωση 6

Εκπαιδύσαμε τα δεδομένα μας με την μέθοδο της διακριτοποίησης τιμών με επικύρωση 6 (fold cross), με τον αλγόριθμο Random Tree -K 0 –M 1,0 -V 0,001 -S 1 και πήραμε τα έξι αποτελέσματα: από το σύνολο των μεταβλητών που αναλύσαμε στο κεφάλαιο 5.1

Συνολικά από το μοντέλο που προκύπτει(μη εκπαιδευμένα δέντρα) βλέπουμε πως δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς 2.590 άτομα από τα οποία χάθηκαν τα 39 άτομα όπου έχουν διαγνωστεί με την νόσο του θυρεοειδούς ενώ έχουν διαγνωστεί με την νόσο του θυρεοειδούς 116 άτομα από τα οποία χάθηκαν τα 55 άτομα όπου δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς. Σε ποσοστό ακρίβειας των δεδομένων είναι 96,6429% σύμφωνα με το πρότυπο, ενώ το 3,3571% δεν είναι σύμφωνα με το πρότυπο.



Εικόνα 79: Αλγόριθμος Random Tree -K 0 –M 1,0 -V 0,001 -S 1 με επικύρωση 6

5.2.4.3 Αλγόριθμος Random Tree -K 0 –M 1,0 -V 0,001 -S 1 με επικύρωση 7

Εκπαιδύσαμε τα δεδομένα μας με την μέθοδο της διακριτοποίησης τιμών με επικύρωση 7 (fold cross), με τον αλγόριθμο Random Tree -K 0 –M 1,0 -V 0,001 -S 1 και πήραμε τα έξι αποτελέσματα: από το σύνολο των μεταβλητών που αναλύσαμε στο κεφάλαιο 5.1

Συνολικά από το μοντέλο που προκύπτει(μη εκπαιδευμένα δέντρα) βλέπουμε πως δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς 2.594 άτομα από τα οποία χάθηκαν τα 35 άτομα όπου έχουν διαγνωστεί με την νόσο του θυρεοειδούς ενώ έχουν διαγνωστεί με την νόσο του θυρεοειδούς 129 άτομα από τα οποία χάθηκαν τα 42 άτομα όπου δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς. Σε ποσοστό ακρίβειας των δεδομένων είναι 97,25% σύμφωνα με το πρότυπο, ενώ το 2,75% δεν είναι σύμφωνα με το πρότυπο.

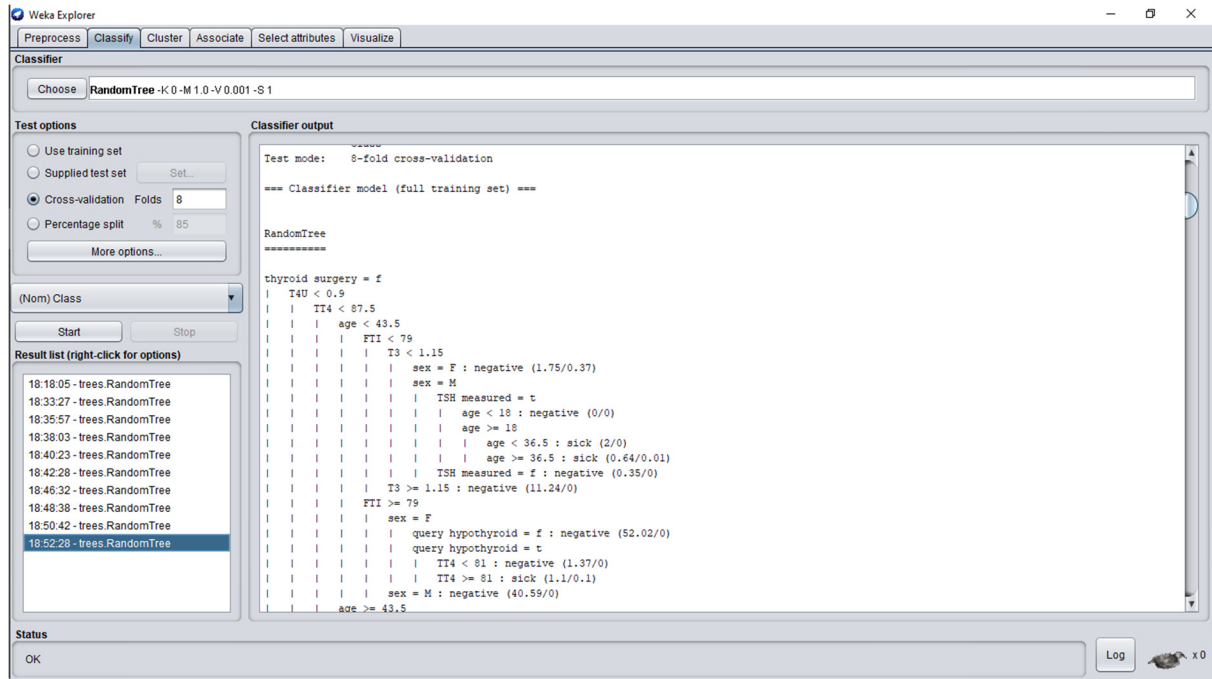


Εικόνα 80: Αλγόριθμος Random Tree -K 0 –M 1,0 -V 0,001 -S 1 με επικύρωση 7

5.2.4.4 Αλγόριθμος Random Tree -K 0 –M 1,0 -V 0,001 -S 1 με επικύρωση 8

Εκπαιδύσαμε τα δεδομένα μας με την μέθοδο της διακριτοποίησης τιμών με επικύρωση 8 (fold cross), με τον αλγόριθμο Random Tree -K 0 –M 1,0 -V 0,001 -S 1 και πήραμε τα έξι αποτελέσματα: από το σύνολο των μεταβλητών που αναλύσαμε στο κεφάλαιο 5.1

Συνολικά από το μοντέλο που προκύπτει(μη εκπαιδευμένα δέντρα) βλέπουμε πως δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς 2.600 άτομα από τα οποία χάθηκαν τα 29 άτομα όπου έχουν διαγνωστεί με την νόσο του θυρεοειδούς ενώ έχουν διαγνωστεί με την νόσο του θυρεοειδούς 112 άτομα από τα οποία χάθηκαν τα 59 άτομα όπου δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς. Σε ποσοστό ακρίβειας των δεδομένων είναι 96,8571% σύμφωνα με το πρότυπο, ενώ το 3,1429% δεν είναι σύμφωνα με το πρότυπο.

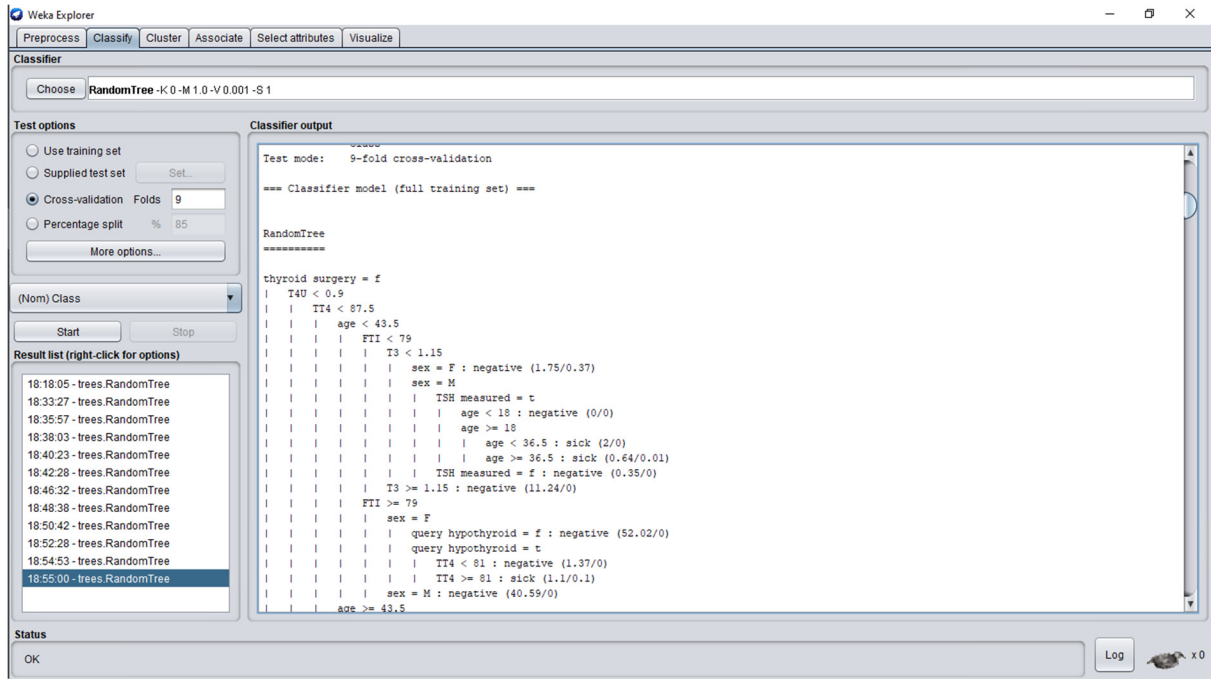


Εικόνα 81: Αλγόριθμος Random Tree -K 0 –M 1,0 -V 0,001 -S 1 με επικύρωση 8

5.2.4.5 Αλγόριθμος Random Tree -K 0 –M 1,0 -V 0,001 -S 1 με επικύρωση 9

Εκπαιδύσαμε τα δεδομένα μας με την μέθοδο της διακριτοποίησης τιμών με επικύρωση 9 (fold cross), με τον αλγόριθμο Random Tree -K 0 –M 1,0 -V 0,001 -S 1 και πήραμε τα έξι αποτελέσματα: από το σύνολο των μεταβλητών που αναλύσαμε στο κεφάλαιο 5.1

Συνολικά από το μοντέλο που προκύπτει (μη εκπαιδευμένα δέντρα) βλέπουμε πως δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς 2.584 άτομα από τα οποία χάθηκαν τα 45 άτομα όπου έχουν διαγνωστεί με την νόσο του θυρεοειδούς ενώ έχουν διαγνωστεί με την νόσο του θυρεοειδούς 125 άτομα από τα οποία χάθηκαν τα 46 άτομα όπου δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς. Σε ποσοστό ακρίβειας των δεδομένων είναι 96,75% σύμφωνα με το πρότυπο, ενώ το 3,25% δεν είναι σύμφωνα με το πρότυπο.

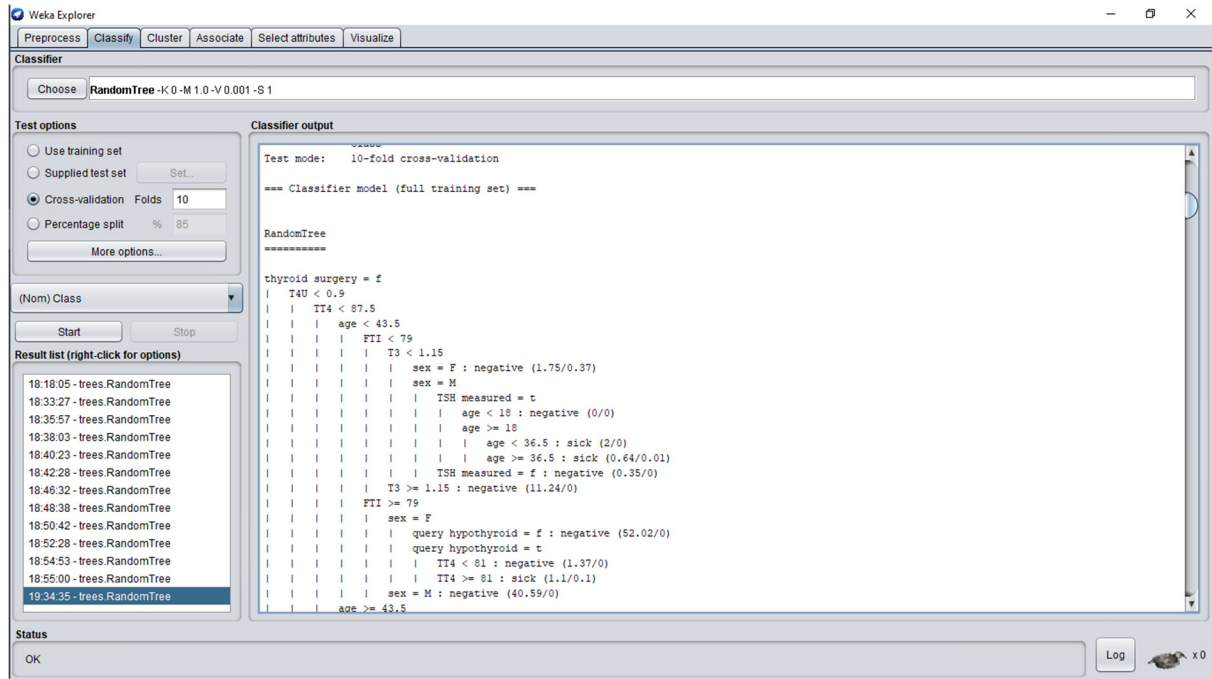


Εικόνα 82: Αλγόριθμος Random Tree -K 0 –M 1,0 -V 0,001 -S 1 με επικύρωση 9

5.2.4.6 Αλγόριθμος Random Tree -K 0 –M 1,0 -V 0,001 -S 1 με επικύρωση 10

Εκπαιδύσαμε τα δεδομένα μας με την μέθοδο της διακριτοποίησης τιμών με επικύρωση 10 (fold cross), με τον αλγόριθμο Random Tree -K 0 –M 1,0 -V 0,001 -S 1 και πήραμε τα έξι αποτελέσματα: από το σύνολο των μεταβλητών που αναλύσαμε στο κεφάλαιο 5.1

Συνολικά από το μοντέλο που προκύπτει(μη εκπαιδευμένα δέντρα) βλέπουμε πως δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς 2.594 άτομα από τα οποία χάθηκαν τα 35 άτομα όπου έχουν διαγνωστεί με την νόσο του θυρεοειδούς ενώ έχουν διαγνωστεί με την νόσο του θυρεοειδούς 131 άτομα από τα οποία χάθηκαν τα 40 άτομα όπου δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς. Σε ποσοστό ακρίβειας των δεδομένων είναι 97,3214% σύμφωνα με το πρότυπο, ενώ το 2,6786% δεν είναι σύμφωνα με το πρότυπο.

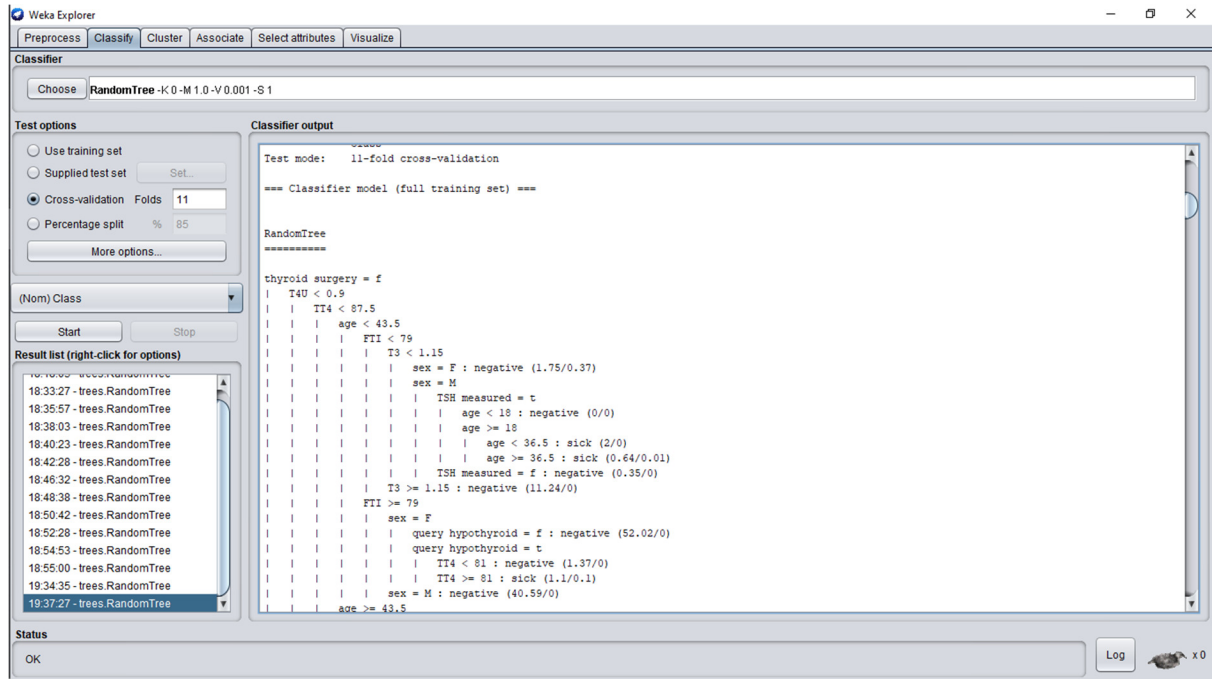


Εικόνα 83: Αλγόριθμος Random Tree -K 0 –M 1,0 -V 0,001 -S 1 με επικύρωση 10

5.2.4.7 Αλγόριθμος Random Tree -K 0 –M 1,0 -V 0,001 -S 1 με επικύρωση 11

Εκπαιδύσαμε τα δεδομένα μας με την μέθοδο της διακριτοποίησης τιμών με επικύρωση 11 (fold cross), με τον αλγόριθμο Random Tree -K 0 –M 1,0 -V 0,001 -S 1 και πήραμε τα έξι αποτελέσματα: από το σύνολο των μεταβλητών που αναλύσαμε στο κεφάλαιο 5.1

Συνολικά από το μοντέλο που προκύπτει(μη εκπαιδευμένα δέντρα) βλέπουμε πως δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς 2.584 άτομα από τα οποία χάθηκαν τα 45 άτομα όπου έχουν διαγνωστεί με την νόσο του θυρεοειδούς ενώ έχουν διαγνωστεί με την νόσο του θυρεοειδούς 118 άτομα από τα οποία χάθηκαν τα 53 άτομα όπου δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς. Σε ποσοστό ακρίβειας των δεδομένων είναι 96,5% σύμφωνα με το πρότυπο, ενώ το 3,5% δεν είναι σύμφωνα με το πρότυπο.

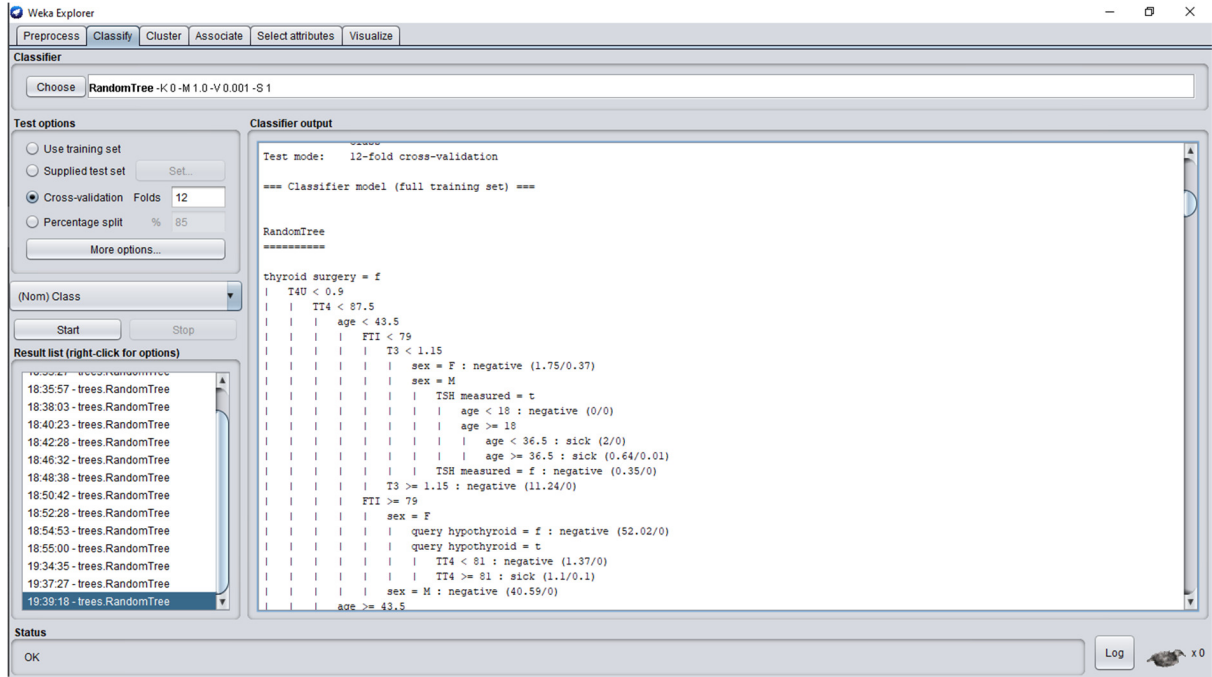


Εικόνα 84: Αλγόριθμος Random Tree -K 0 –M 1,0 -V 0,001 -S 1 με επικύρωση 11

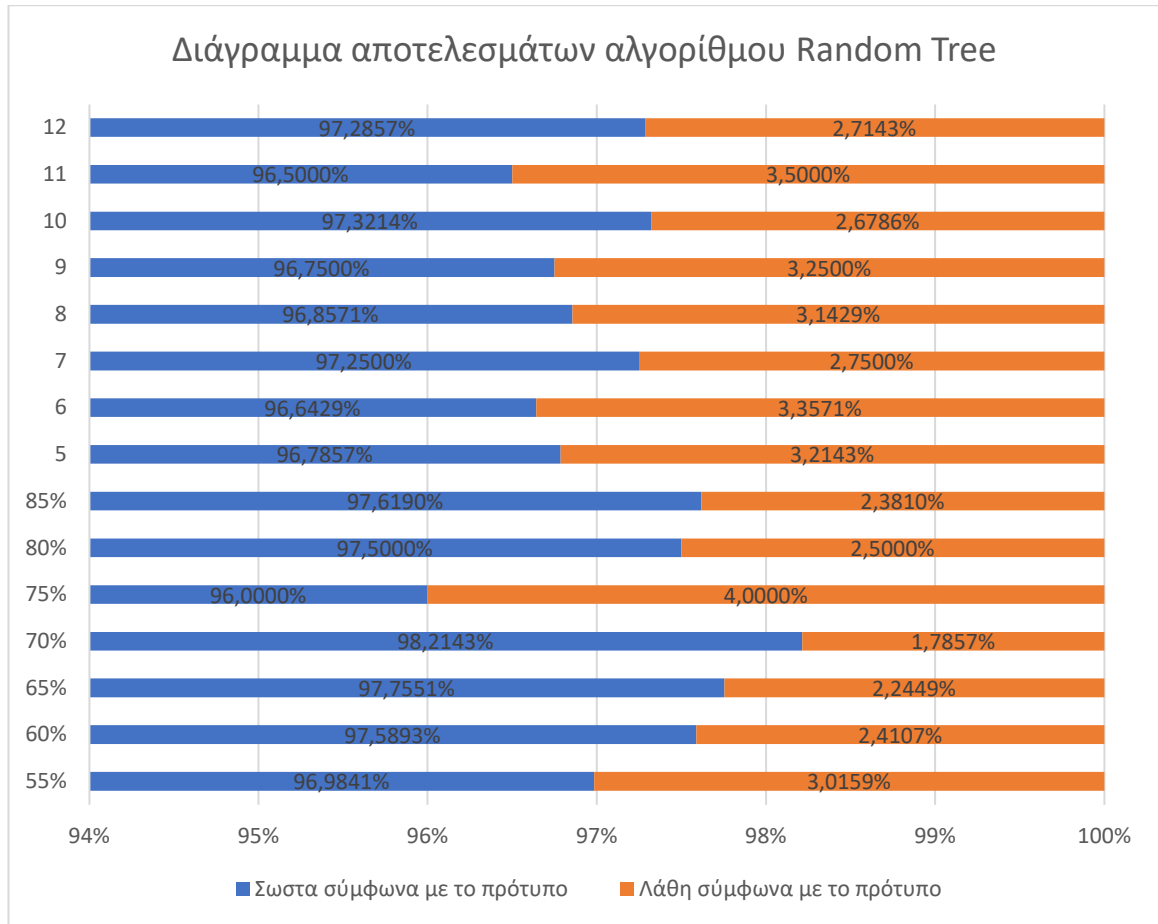
5.2.4.8 Αλγόριθμος Random Tree -K 0 –M 1,0 -V 0,001 -S 1 με επικύρωση 12

Εκπαιδύσαμε τα δεδομένα μας με την μέθοδο της διακριτοποίησης τιμών με επικύρωση 12 (fold cross), με τον αλγόριθμο Random Tree -K 0 –M 1,0 -V 0,001 -S 1 και πήραμε τα έξι αποτελέσματα: από το σύνολο των μεταβλητών που αναλύσαμε στο κεφάλαιο 5.1

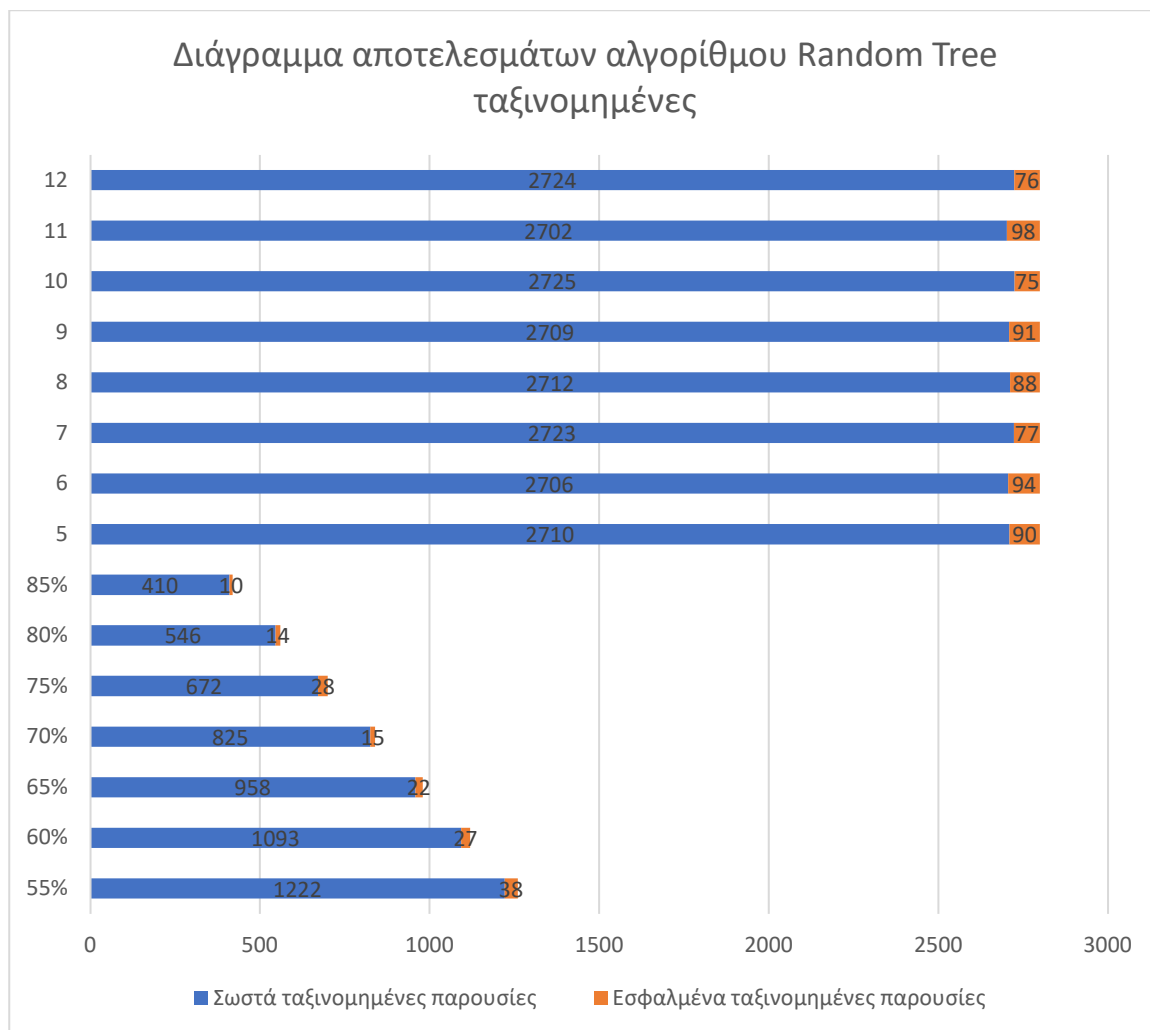
Συνολικά από το μοντέλο που προκύπτει(μη εκπαιδευμένα δέντρα) βλέπουμε πως δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς 2.593 άτομα από τα οποία χάθηκαν τα 36 άτομα όπου έχουν διαγνωστεί με την νόσο του θυρεοειδούς ενώ έχουν διαγνωστεί με την νόσο του θυρεοειδούς 131 άτομα από τα οποία χάθηκαν τα 40 άτομα όπου δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς. Σε ποσοστό ακρίβειας των δεδομένων είναι 97,2857% σύμφωνα με το πρότυπο, ενώ το 2,7143% δεν είναι σύμφωνα με το πρότυπο.



Εικόνα 85: Αλγόριθμος Random Tree -K 0 -M 1,0 -V 0,001 -S 1 με επικύρωση 12



Διάγραμμα 3: Διάγραμμα αποτελεσμάτων αλγορίθμου Random Tree



Διάγραμμα 4: Διάγραμμα αποτελεσμάτων αλγορίθμου Random Tree ταξινομημένες

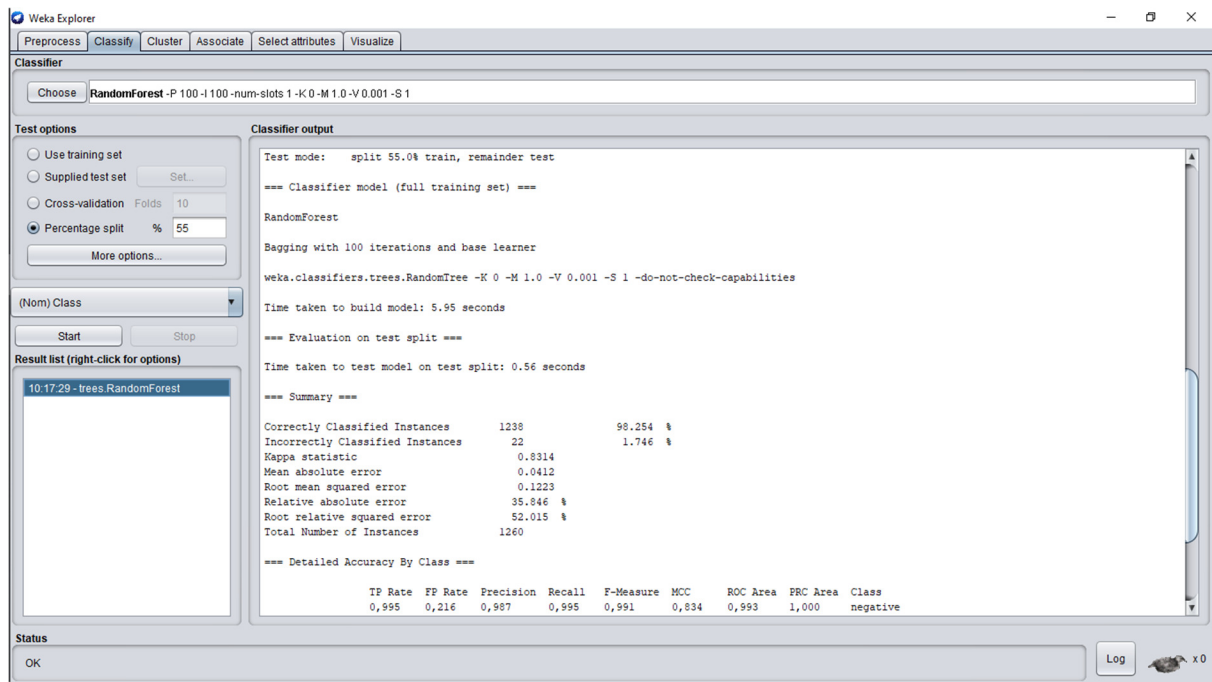
5.2.5 Δέντρο απόφασης αλγορίθμου Random Forest με διακριτοποίηση τιμών

5.2.5.1 Αλγόριθμος RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1 με διακριτοποίηση 55%

Εκπαίδευσάμε το 55% των δεδομένων μας με την μέθοδο της διακριτοποίησης τιμών, με τον αλγόριθμο RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1 και πήραμε τα εξής αποτελέσματα: από το σύνολο των μεταβλητών που αναλύσαμε στο κεφάλαιο 5.1

- Συνολικά από το μοντέλο που προκύπτει(μη εκπαιδευμένα δέντρα) βλέπουμε πως δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς 1.180 άτομα από τα οποία χάθηκαν τα 6 άτομα όπου έχουν διαγνωστεί με την νόσο του θυρεοειδούς ενώ έχουν διαγνωστεί με την νόσο του θυρεοειδούς 58 άτομα από τα οποία χάθηκαν τα 16 άτομα όπου δεν έχουν

διαγνωστεί με την νόσο του θυρεοειδούς. Σε ποσοστό ακρίβειας των δεδομένων είναι 98,254% σύμφωνα με το πρότυπο, ενώ το 1,746% δεν είναι σύμφωνα με το πρότυπο.

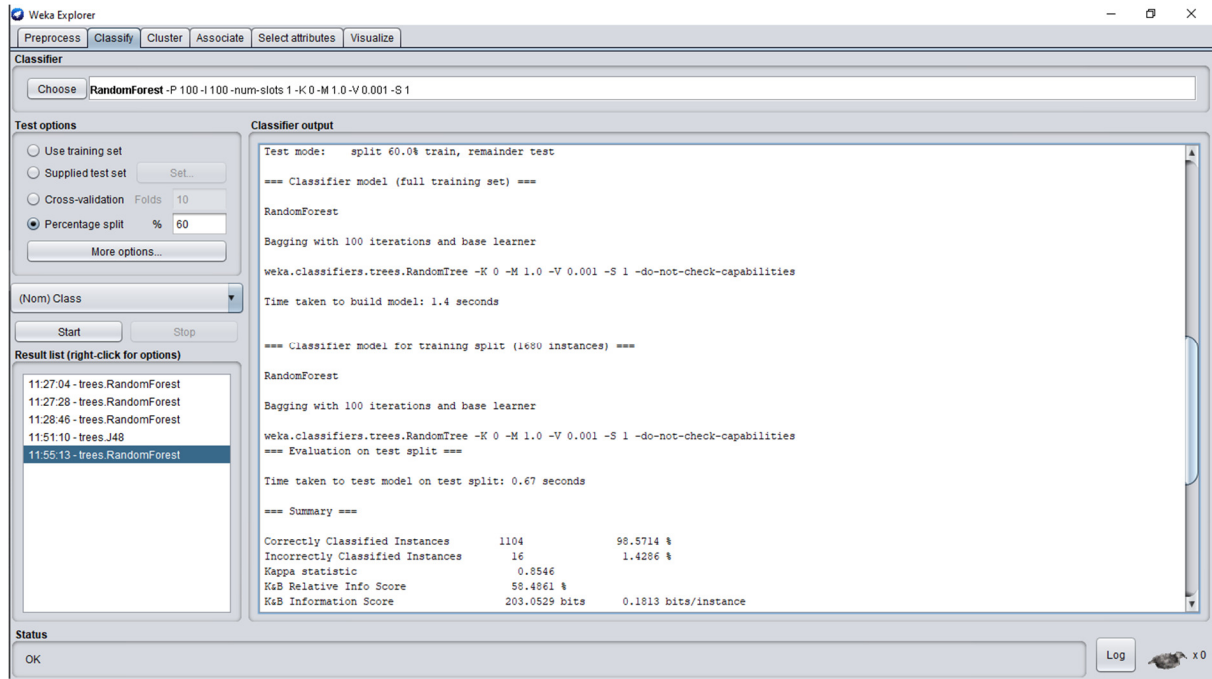


Εικόνα 86: Αλγόριθμος RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1 με διακριτοποίηση 55%

5.2.5.2 Αλγόριθμος RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1 με διακριτοποίηση 60%

Εκπαιδύσαμε το 60% των δεδομένων μας με την μέθοδο της διακριτοποίησης τιμών, με τον αλγόριθμο RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1 και πήραμε τα εξής αποτελέσματα: από το σύνολο των μεταβλητών που αναλύσαμε στο κεφάλαιο 5.1

- Συνολικά από το μοντέλο που προκύπτει(μη εκπαιδευμένα δέντρα) βλέπουμε πως δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς 1.054 άτομα από τα οποία χάθηκαν τα 4 άτομα όπου έχουν διαγνωστεί με την νόσο του θυρεοειδούς ενώ έχουν διαγνωστεί με την νόσο του θυρεοειδούς 50 άτομα από τα οποία χάθηκαν τα 12 άτομα όπου δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς. Σε ποσοστό ακρίβειας των δεδομένων είναι 98,5714% σύμφωνα με το πρότυπο, ενώ το 1,5714% δεν είναι σύμφωνα με το πρότυπο.

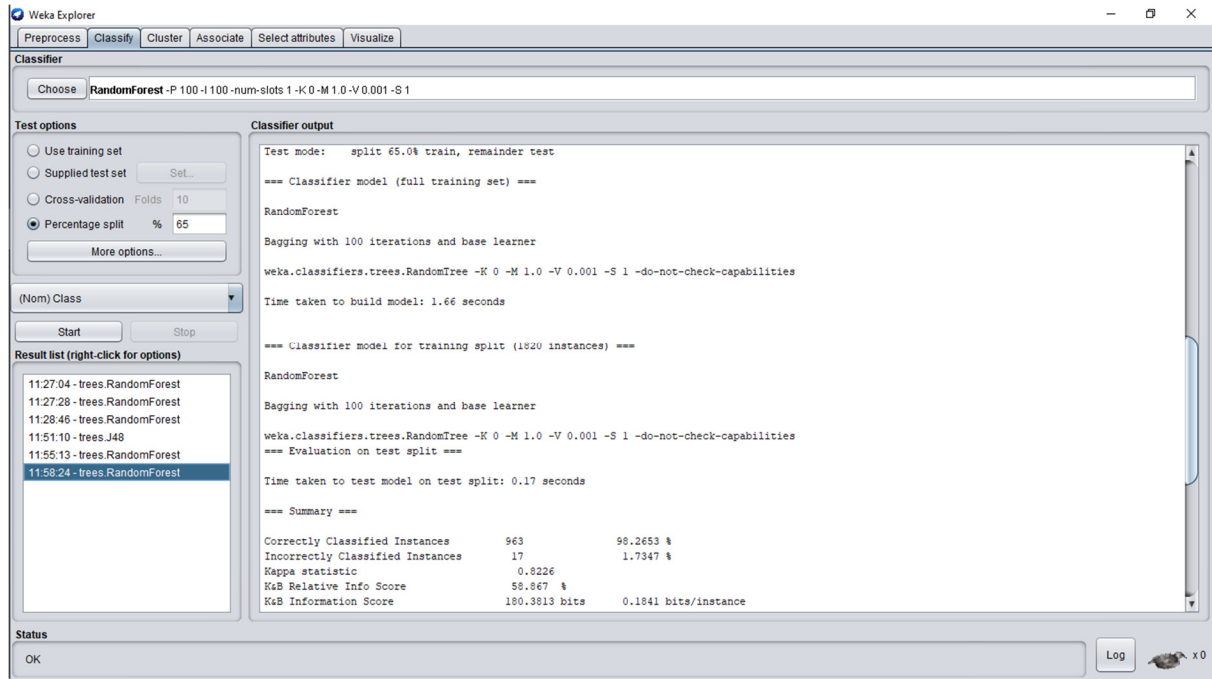


Εικόνα 87: Αλγόριθμος RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1 με διακριτοποίηση 60%

5.2.5.3 Αλγόριθμος RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1 με διακριτοποίηση 65%

Εκπαιδεύσαμε το 65% των δεδομένων μας με την μέθοδο της διακριτοποίησης τιμών, με τον αλγόριθμο RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1 και πήραμε τα εξής αποτελέσματα: από το σύνολο των μεταβλητών που αναλύσαμε στο κεφάλαιο 5.1

- Συνολικά από το μοντέλο που προκύπτει(μη εκπαιδευμένα δέντρα) βλέπουμε πως δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς 921 άτομα από τα οποία χάθηκαν τα 4 άτομα όπου έχουν διαγνωστεί με την νόσο του θυρεοειδούς ενώ έχουν διαγνωστεί με την νόσο του θυρεοειδούς 42 άτομα από τα οποία χάθηκαν τα 13 άτομα όπου δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς. Σε ποσοστό ακρίβειας των δεδομένων είναι 98,2653% σύμφωνα με το πρότυπο, ενώ το 1,7347% δεν είναι σύμφωνα με το πρότυπο.

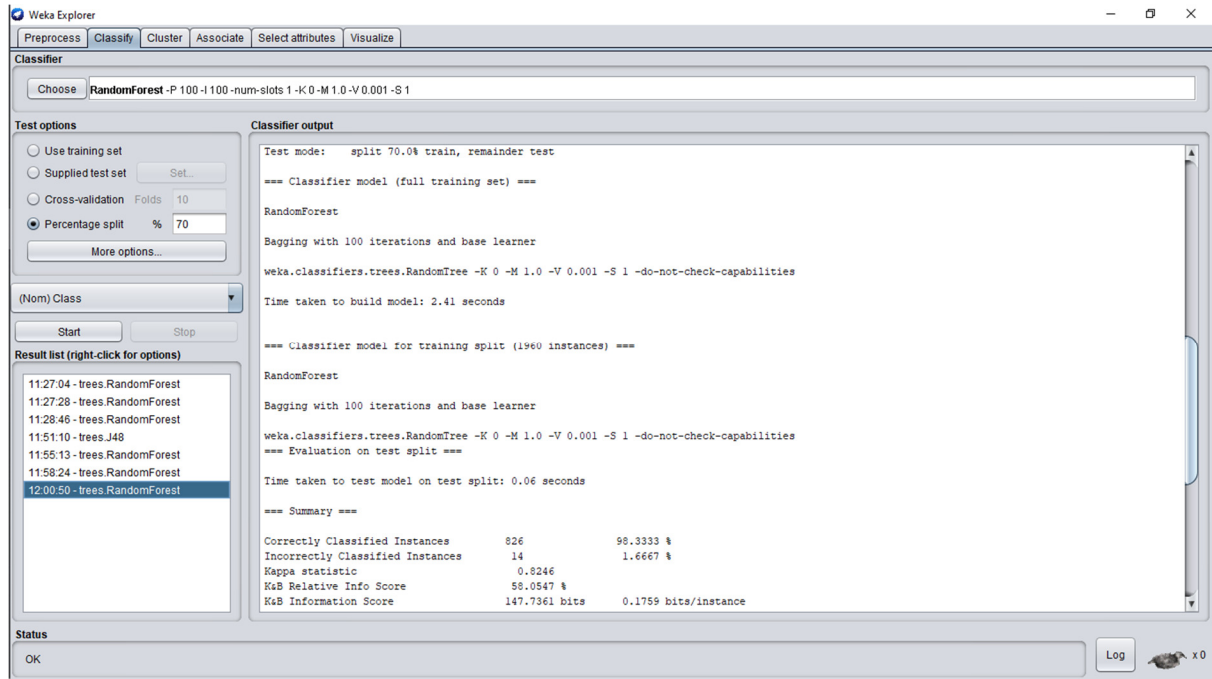


Εικόνα 88: Αλγόριθμος RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1 με διακριτοποίηση 65%

5.2.5.4 Αλγόριθμος RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1 με διακριτοποίηση 70%

Εκπαιδύσαμε το 70% των δεδομένων μας με την μέθοδο της διακριτοποίησης τιμών, με τον αλγόριθμο RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1 και πήραμε τα εξής αποτελέσματα: από το σύνολο των μεταβλητών που αναλύσαμε στο κεφάλαιο 5.1

- Συνολικά από το μοντέλο που προκύπτει(μη εκπαιδευμένα δέντρα) βλέπουμε πως δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς 791 άτομα από τα οποία χάθηκαν τα 4 άτομα όπου έχουν διαγνωστεί με την νόσο του θυρεοειδούς ενώ έχουν διαγνωστεί με την νόσο του θυρεοειδούς 35 άτομα από τα οποία χάθηκαν τα 10 άτομα όπου δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς. Σε ποσοστό ακρίβειας των δεδομένων είναι 98,3333% σύμφωνα με το πρότυπο, ενώ το 1,6667% δεν είναι σύμφωνα με το πρότυπο.

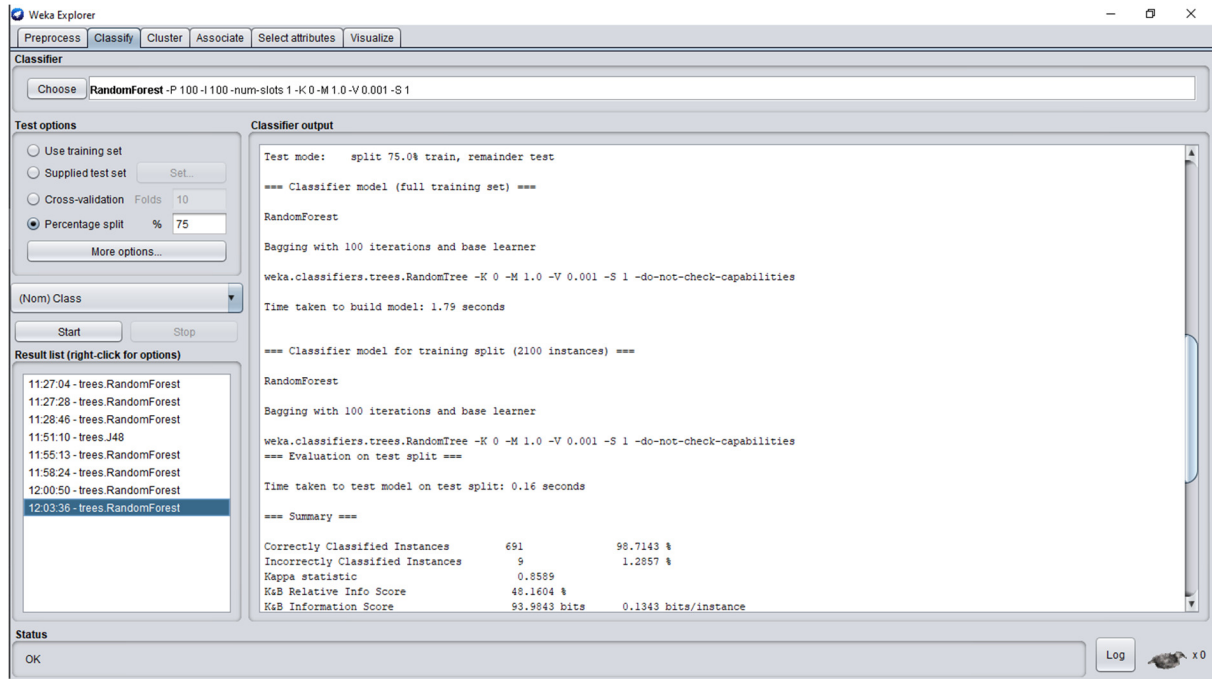


Εικόνα 89: Αλγόριθμος RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1 με διακριτοποίηση 70%

5.2.5.5 Αλγόριθμος RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1 με διακριτοποίηση 75%

Εκπαιδύσαμε το 75% των δεδομένων μας με την μέθοδο της διακριτοποίησης τιμών, με τον αλγόριθμο RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1 και πήραμε τα εξής αποτελέσματα: από το σύνολο των μεταβλητών που αναλύσαμε στο κεφάλαιο 5.1

- Συνολικά από το μοντέλο που προκύπτει(μη εκπαιδευμένα δέντρα) βλέπουμε πως δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς 662 άτομα από τα οποία χάθηκαν τα 5 άτομα όπου έχουν διαγνωστεί με την νόσο του θυρεοειδούς ενώ έχουν διαγνωστεί με την νόσο του θυρεοειδούς 29 άτομα από τα οποία χάθηκαν τα 4 άτομα όπου δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς. Σε ποσοστό ακρίβειας των δεδομένων είναι 98,7143% σύμφωνα με το πρότυπο, ενώ το 1,2857% δεν είναι σύμφωνα με το πρότυπο.

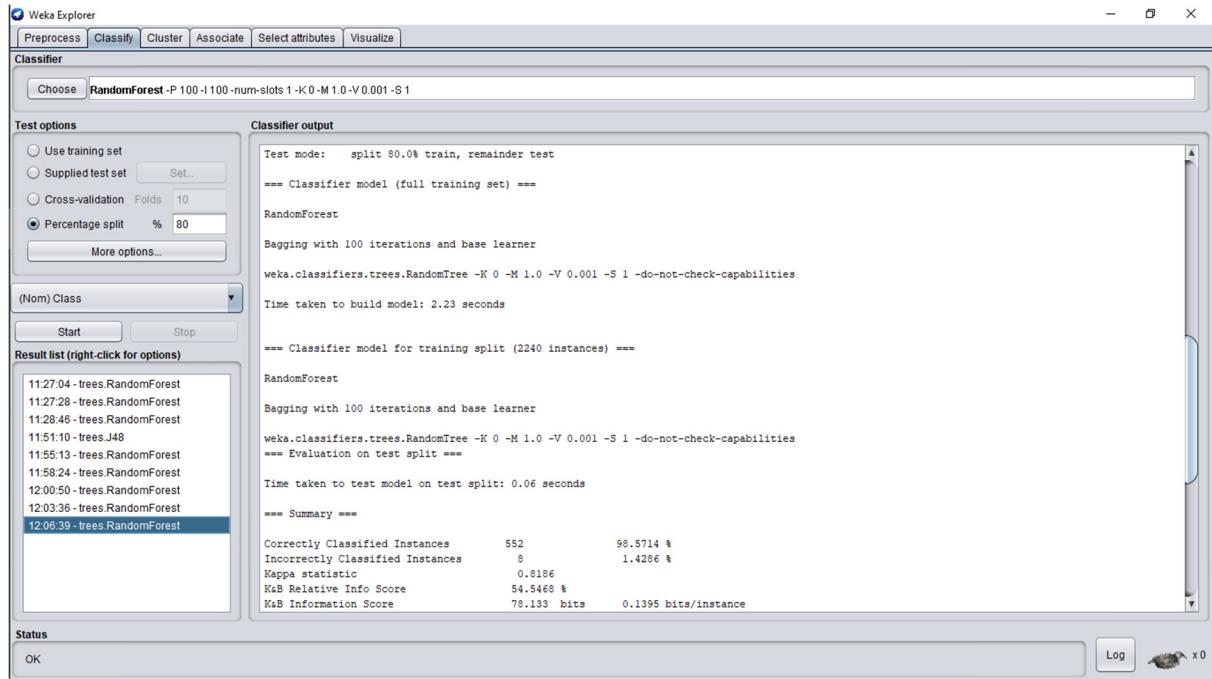


Εικόνα 90: Αλγόριθμος RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1 με διακριτοποίηση 75%

5.2.5.6 Αλγόριθμος RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1 με διακριτοποίηση 80%

Εκπαιδύσαμε το 80% των δεδομένων μας με την μέθοδο της διακριτοποίησης τιμών, με τον αλγόριθμο RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1 και πήραμε τα εξής αποτελέσματα: από το σύνολο των μεταβλητών που αναλύσαμε στο κεφάλαιο 5.1

- Συνολικά από το μοντέλο που προκύπτει(μη εκπαιδευμένα δέντρα) βλέπουμε πως δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς 533 άτομα από τα οποία χάθηκαν τα 4 άτομα όπου έχουν διαγνωστεί με την νόσο του θυρεοειδούς ενώ έχουν διαγνωστεί με την νόσο του θυρεοειδούς 19 άτομα από τα οποία χάθηκαν τα 4 άτομα όπου δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς. Σε ποσοστό ακρίβειας των δεδομένων είναι 98,5714% σύμφωνα με το πρότυπο, ενώ το 1,4286% δεν είναι σύμφωνα με το πρότυπο.

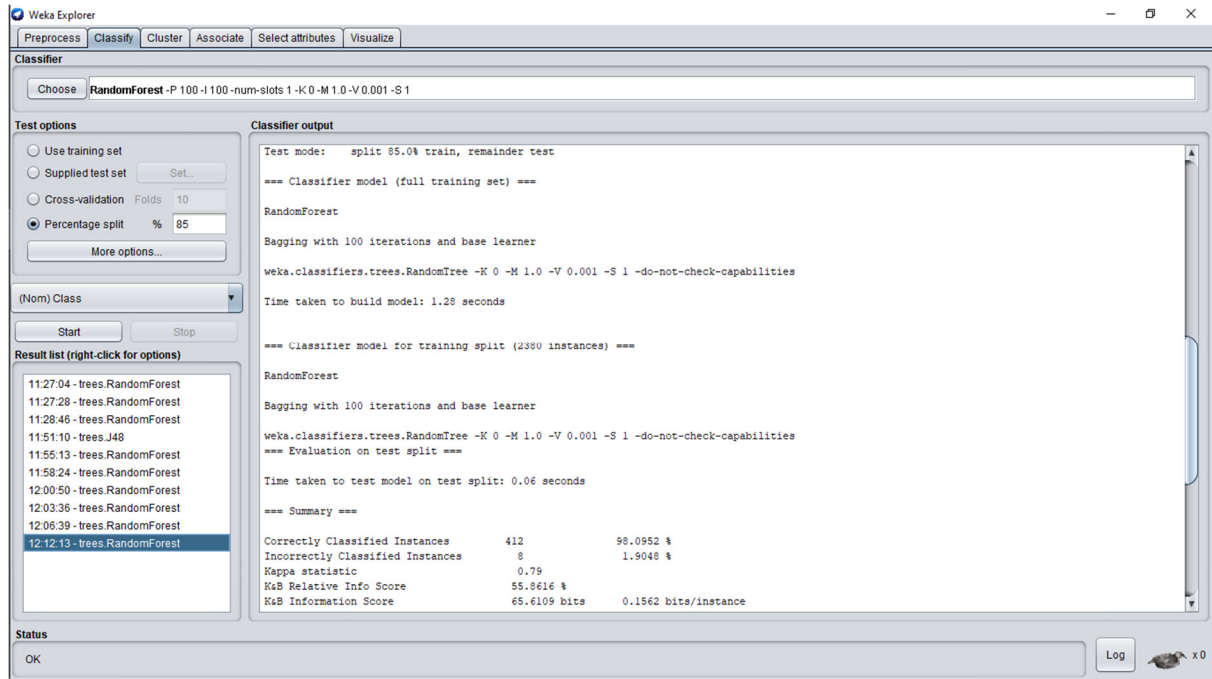


Εικόνα 91: Αλγόριθμος RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1 με διακριτοποίηση 80%

5.2.5.7 Αλγόριθμος RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1 με διακριτοποίηση 85%

Εκπαιδύσαμε το 85% των δεδομένων μας με την μέθοδο της διακριτοποίησης τιμών, με τον αλγόριθμο RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1 και πήραμε τα εξής αποτελέσματα: από το σύνολο των μεταβλητών που αναλύσαμε στο κεφάλαιο 5.1

- Συνολικά από το μοντέλο που προκύπτει(μη εκπαιδευμένα δέντρα) βλέπουμε πως δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς 396 άτομα από τα οποία χάθηκαν τα 4 άτομα όπου έχουν διαγνωστεί με την νόσο του θυρεοειδούς ενώ έχουν διαγνωστεί με την νόσο του θυρεοειδούς 16 άτομα από τα οποία χάθηκαν τα 4 άτομα όπου δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς. Σε ποσοστό ακρίβειας των δεδομένων είναι 98,0952% σύμφωνα με το πρότυπο, ενώ το 1,9048% δεν είναι σύμφωνα με το πρότυπο.



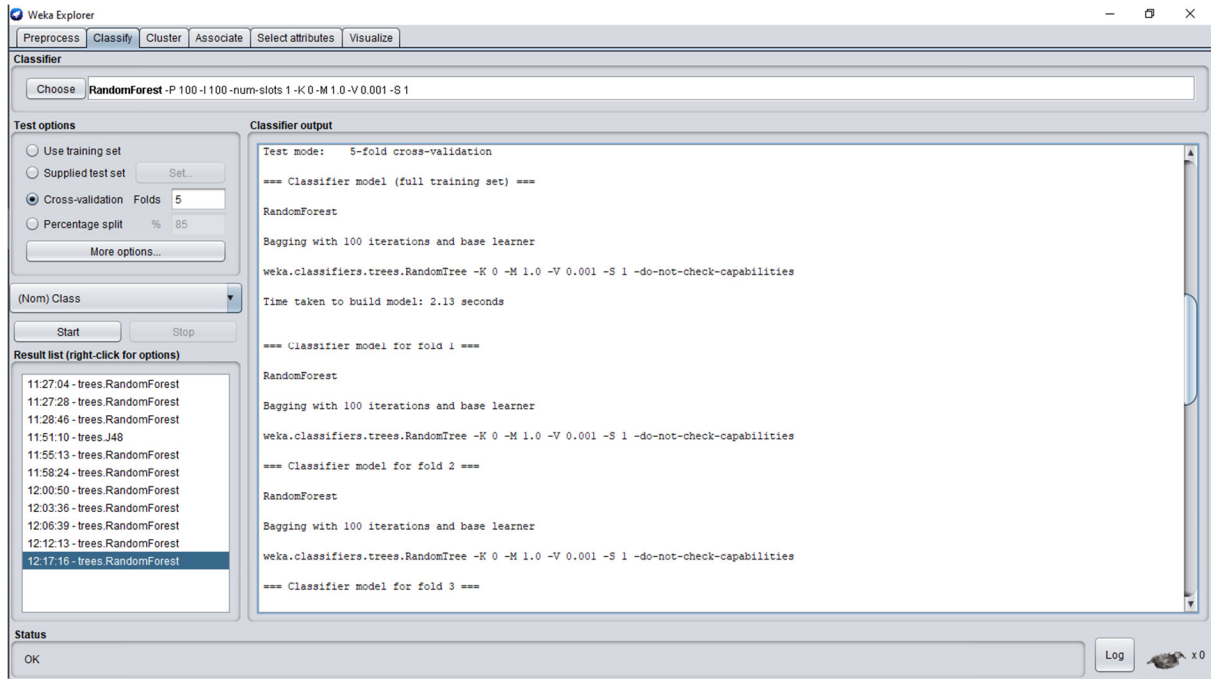
Εικόνα 92: Αλγόριθμος RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1 με διακριτοποίηση 85%

5.2.6 Δέντρο απόφασης αλγορίθμου RandomForest με διακριτοποίηση τιμών και επικύρωση

5.2.6.1 Αλγόριθμος RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1 με επικύρωση 5

Εκπαιδύσαμε τα δεδομένα μας με την μέθοδο της διακριτοποίησης τιμών με επικύρωση 5 (fold cross), με τον αλγόριθμο RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1 και πήραμε τα έξι αποτελέσματα: από το σύνολο των μεταβλητών που αναλύσαμε στο κεφάλαιο 5.1

- Συνολικά από το μοντέλο που προκύπτει(μη εκπαιδευμένα δέντρα) βλέπουμε πως δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς 2.621 άτομα από τα οποία χάθηκαν τα 8 άτομα όπου έχουν διαγνωστεί με την νόσο του θυρεοειδούς ενώ έχουν διαγνωστεί με την νόσο του θυρεοειδούς 129 άτομα από τα οποία χάθηκαν τα 42 άτομα όπου δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς. Σε ποσοστό ακρίβειας των δεδομένων είναι 98,2143% σύμφωνα με το πρότυπο, ενώ το 1,7857% δεν είναι σύμφωνα με το πρότυπο.

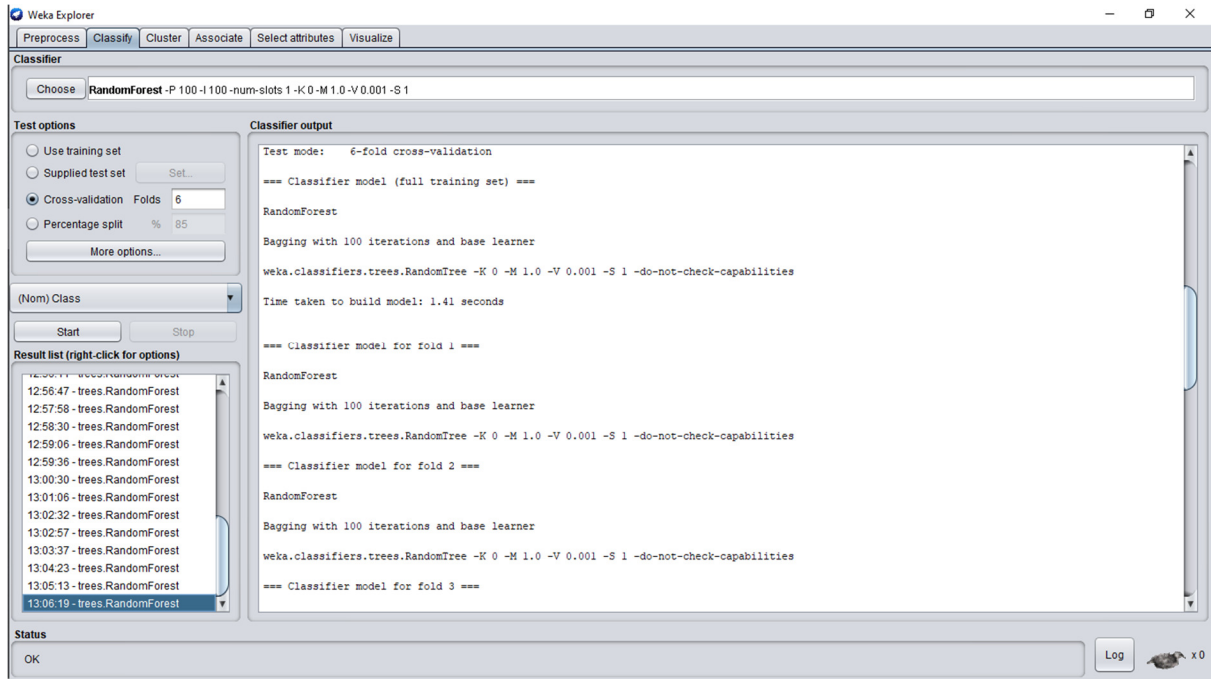


Εικόνα 93: Αλγόριθμος RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1 με επικύρωση 5

5.2.6.2 Αλγόριθμος RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1 με επικύρωση 6

Εκπαιδύσαμε τα δεδομένα μας με την μέθοδο της διακριτοποίησης τιμών με επικύρωση 6 (fold cross), με τον αλγόριθμο RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1 και πήραμε τα έξι αποτελέσματα: από το σύνολο των μεταβλητών που αναλύσαμε στο κεφάλαιο 5.1

- Συνολικά από το μοντέλο που προκύπτει (μη εκπαιδευμένα δέντρα) βλέπουμε πως δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς 2.651 άτομα από τα οποία χάθηκαν τα 12 άτομα όπου έχουν διαγνωστεί με την νόσο του θυρεοειδούς ενώ έχουν διαγνωστεί με την νόσο του θυρεοειδούς 120 άτομα από τα οποία χάθηκαν τα 51 άτομα όπου δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς. Σε ποσοστό ακρίβειας των δεδομένων είναι 97,75% σύμφωνα με το πρότυπο, ενώ το 2,25% δεν είναι σύμφωνα με το πρότυπο.

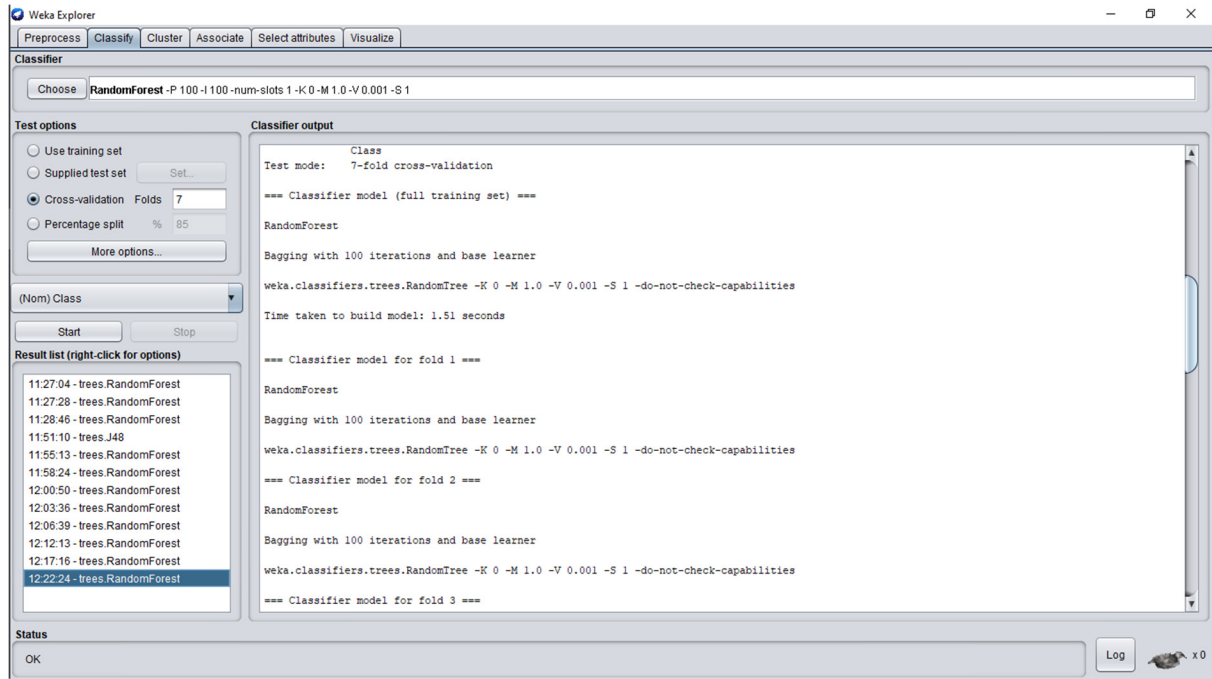


Εικόνα 94: Αλγόριθμος RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1 με επικύρωση 6

5.2.6.3 Αλγόριθμος RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1 με επικύρωση 7

Εκπαιδύσαμε τα δεδομένα μας με την μέθοδο της διακριτοποίησης τιμών με επικύρωση 7 (fold cross), με τον αλγόριθμο RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1 και πήραμε τα έξι αποτελέσματα: από το σύνολο των μεταβλητών που αναλύσαμε στο κεφάλαιο 5.1

- Συνολικά από το μοντέλο που προκύπτει (μη εκπαιδευμένα δέντρα) βλέπουμε πως δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς 2.619 άτομα από τα οποία χάθηκαν τα 10 άτομα όπου έχουν διαγνωστεί με την νόσο του θυρεοειδούς ενώ έχουν διαγνωστεί με την νόσο του θυρεοειδούς 129 άτομα από τα οποία χάθηκαν τα 42 άτομα όπου δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς. Σε ποσοστό ακρίβειας των δεδομένων είναι 98,1429% σύμφωνα με το πρότυπο, ενώ το 1,8571% δεν είναι σύμφωνα με το πρότυπο.

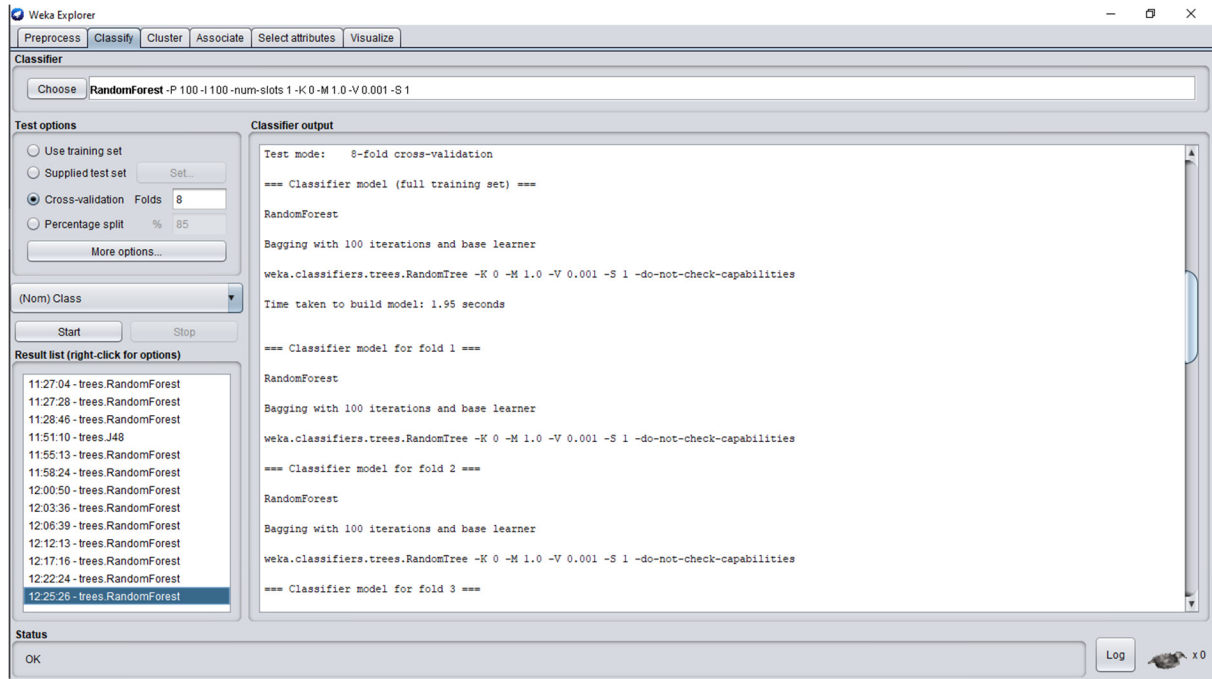


Εικόνα 95: Αλγόριθμος RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1 με επικύρωση 7

5.2.6.4 Αλγόριθμος RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1 με επικύρωση 8

Εκπαιδύσαμε τα δεδομένα μας με την μέθοδο της διακριτοποίησης τιμών με επικύρωση 8 (fold cross), με τον αλγόριθμο RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1 και πήραμε τα έξι αποτελέσματα: από το σύνολο των μεταβλητών που αναλύσαμε στο κεφάλαιο 5.1

- Συνολικά από το μοντέλο που προκύπτει (μη εκπαιδευμένα δέντρα) βλέπουμε πως δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς 2.619 άτομα από τα οποία χάθηκαν τα 10 άτομα όπου έχουν διαγνωστεί με την νόσο του θυρεοειδούς ενώ έχουν διαγνωστεί με την νόσο του θυρεοειδούς 129 άτομα από τα οποία χάθηκαν τα 42 άτομα όπου δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς. Σε ποσοστό ακρίβειας των δεδομένων είναι 98,1429% σύμφωνα με το πρότυπο, ενώ το 1,8571% δεν είναι σύμφωνα με το πρότυπο.

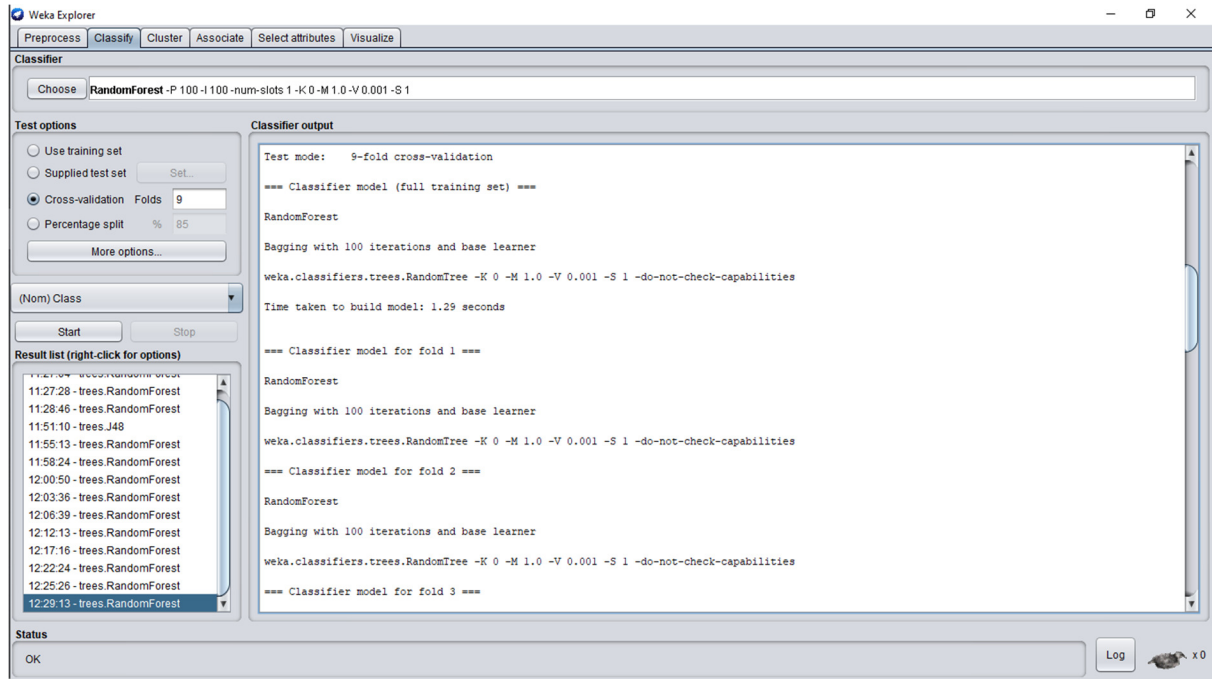


Εικόνα 96: Αλγόριθμος RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1 με επικύρωση 8

5.2.6.5 Αλγόριθμος RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1 με επικύρωση 9

Εκπαιδύσαμε τα δεδομένα μας με την μέθοδο της διακριτοποίησης τιμών με επικύρωση 9 (fold cross), με τον αλγόριθμο RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1 και πήραμε τα έξι αποτελέσματα: από το σύνολο των μεταβλητών που αναλύσαμε στο κεφάλαιο 5.1

- Συνολικά από το μοντέλο που προκύπτει (μη εκπαιδευμένα δέντρα) βλέπουμε πως δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς 2.617 άτομα από τα οποία χάθηκαν τα 12 άτομα όπου έχουν διαγνωστεί με την νόσο του θυρεοειδούς ενώ έχουν διαγνωστεί με την νόσο του θυρεοειδούς 130 άτομα από τα οποία χάθηκαν τα 41 άτομα όπου δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς. Σε ποσοστό ακρίβειας των δεδομένων είναι 98,1071% σύμφωνα με το πρότυπο, ενώ το 1,8929% δεν είναι σύμφωνα με το πρότυπο.

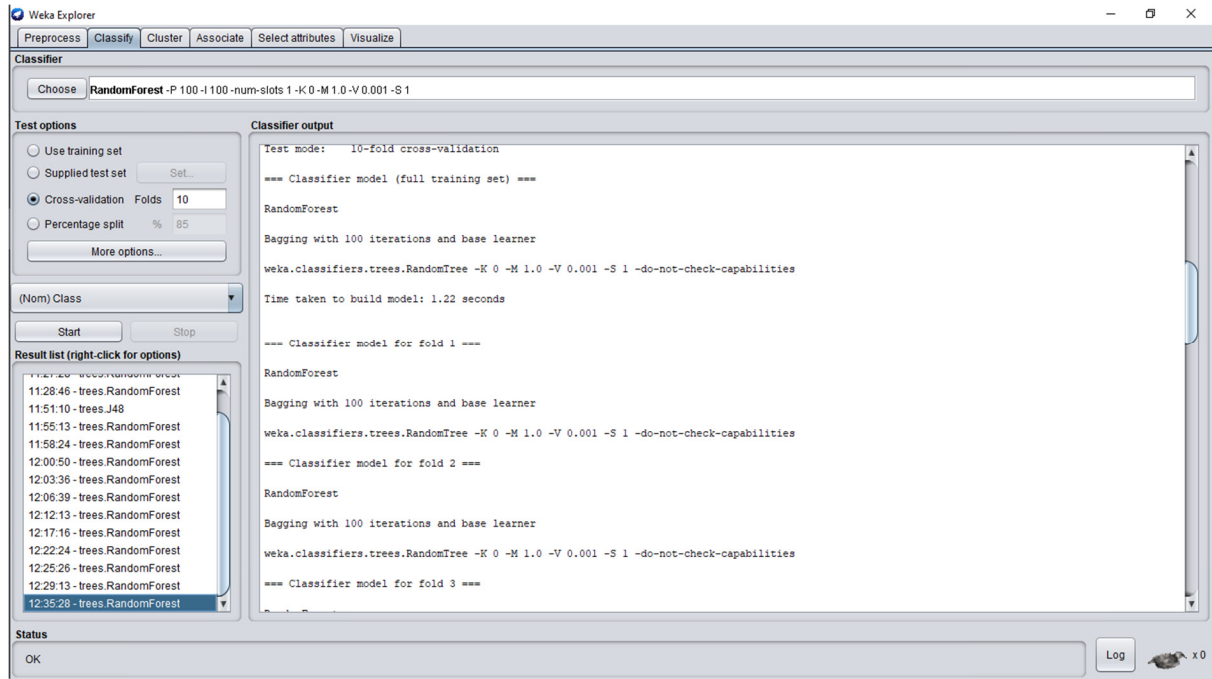


Εικόνα 97: Αλγόριθμος RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1 με επικύρωση 9

5.2.6.6 Αλγόριθμος RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1 με επικύρωση 10

Εκπαιδύσαμε τα δεδομένα μας με την μέθοδο της διακριτοποίησης τιμών με επικύρωση 10 (fold cross), με τον αλγόριθμο RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1 και πήραμε τα έξι αποτελέσματα: από το σύνολο των μεταβλητών που αναλύσαμε στο κεφάλαιο 5.1

- Συνολικά από το μοντέλο που προκύπτει (μη εκπαιδευμένα δέντρα) βλέπουμε πως δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς 2.620 άτομα από τα οποία χάθηκαν τα 9 άτομα όπου έχουν διαγνωστεί με την νόσο του θυρεοειδούς ενώ έχουν διαγνωστεί με την νόσο του θυρεοειδούς 129 άτομα από τα οποία χάθηκαν τα 42 άτομα όπου δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς. Σε ποσοστό ακρίβειας των δεδομένων είναι 98,1786% σύμφωνα με το πρότυπο, ενώ το 1,8214% δεν είναι σύμφωνα με το πρότυπο.

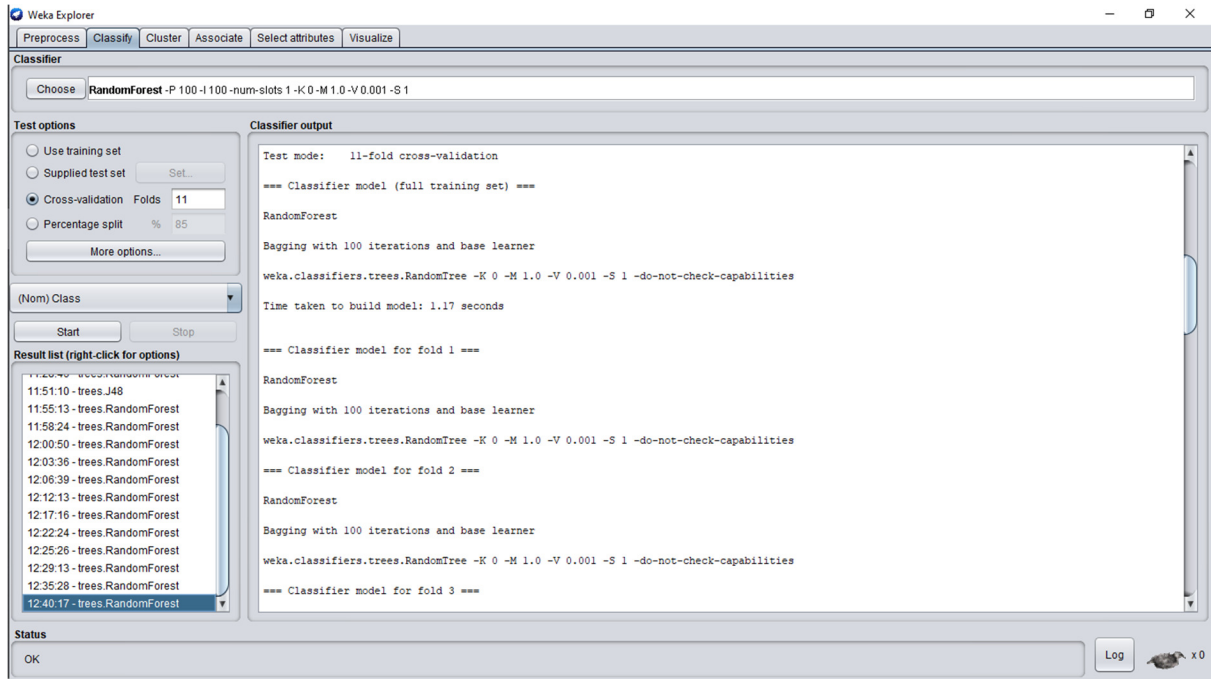


Εικόνα 98: Αλγόριθμος RandomForest -P100 -I 100 -num-slots 1 -K 0-M 1.0-V 0.001-S 1 με επικύρωση 10

5.2.6.7 Αλγόριθμος RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1 με επικύρωση 11

Εκπαιδύσαμε τα δεδομένα μας με την μέθοδο της διακριτοποίησης τιμών με επικύρωση 11 (fold cross), με τον αλγόριθμο RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1 και πήραμε τα έξι αποτελέσματα: από το σύνολο των μεταβλητών που αναλύσαμε στο κεφάλαιο 5.1

- Συνολικά από το μοντέλο που προκύπτει(μη εκπαιδευμένα δέντρα) βλέπουμε πως δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς 2.618 άτομα από τα οποία χάθηκαν τα 11 άτομα όπου έχουν διαγνωστεί με την νόσο του θυρεοειδούς ενώ έχουν διαγνωστεί με την νόσο του θυρεοειδούς 125 άτομα από τα οποία χάθηκαν τα 46 άτομα όπου δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς. Σε ποσοστό ακρίβειας των δεδομένων είναι 99,9643% σύμφωνα με το πρότυπο, ενώ το 2,0357% δεν είναι σύμφωνα με το πρότυπο.

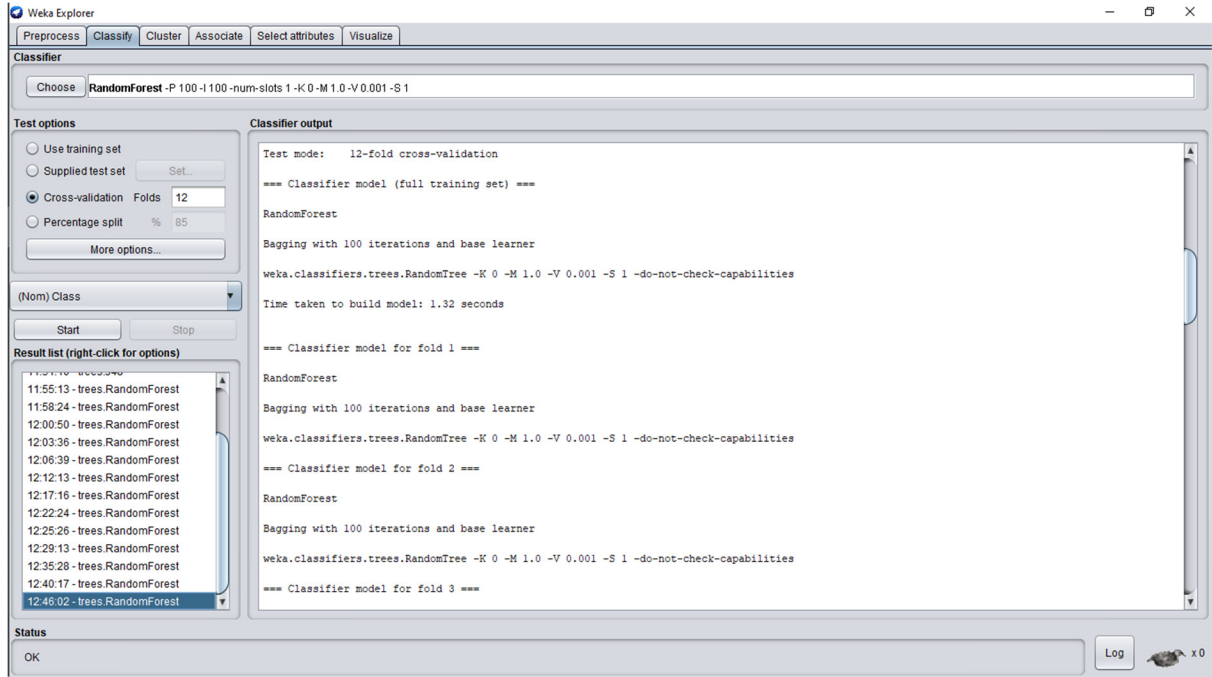


Εικόνα 99: Αλγόριθμος RandomForest -P100 -I 100 -num-slots 1 -K 0-M 1.0-V 0.001-S 1 με επικύρωση 11

5.2.6.8 Αλγόριθμος RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1 με επικύρωση 12

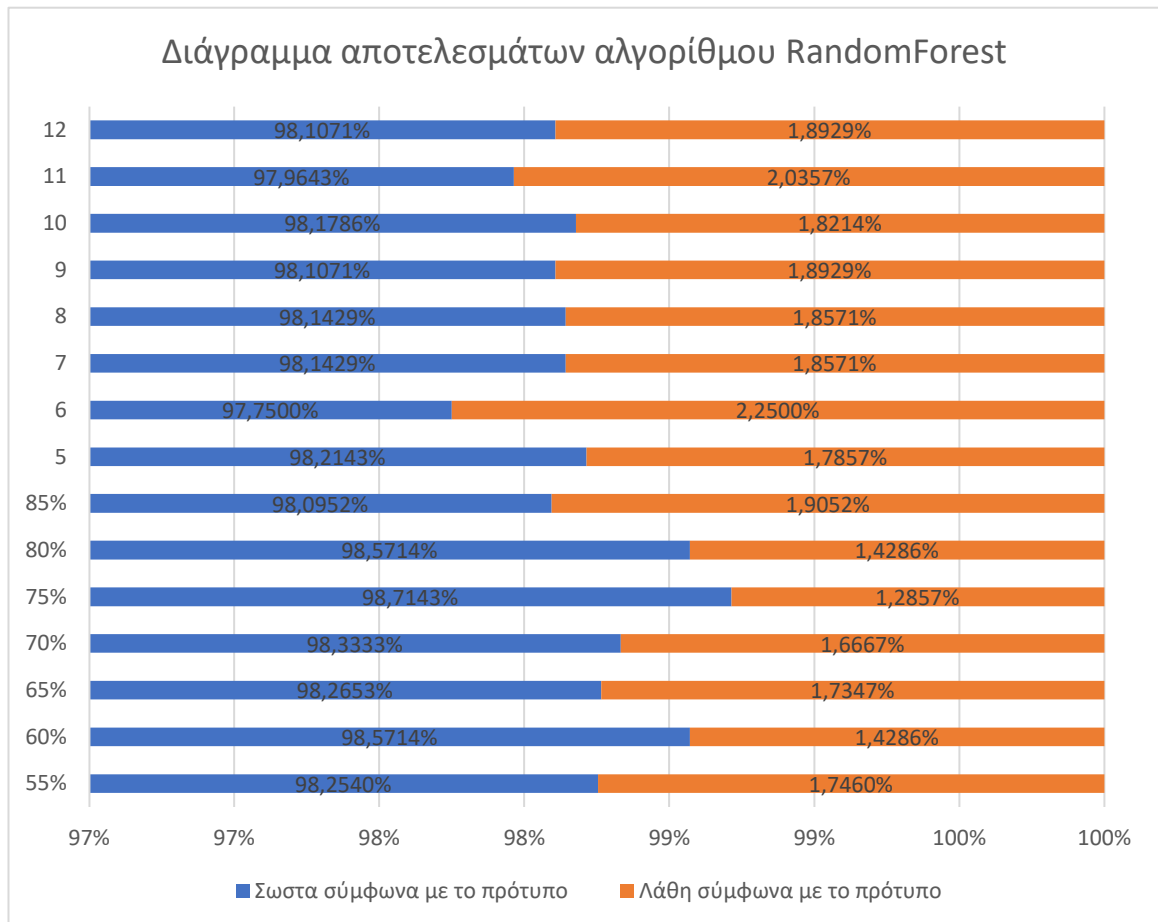
Εκπαιδύσαμε τα δεδομένα μας με την μέθοδο της διακριτοποίησης τιμών με επικύρωση 12 (fold cross), με τον αλγόριθμο RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1 και πήραμε τα έξι αποτελέσματα: από το σύνολο των μεταβλητών που αναλύσαμε στο κεφάλαιο 5.1

- Συνολικά από το μοντέλο που προκύπτει(μη εκπαιδευμένα δέντρα) βλέπουμε πως δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς 2.619 άτομα από τα οποία χάθηκαν τα 10 άτομα όπου έχουν διαγνωστεί με την νόσο του θυρεοειδούς ενώ έχουν διαγνωστεί με την νόσο του θυρεοειδούς 128 άτομα από τα οποία χάθηκαν τα 43 άτομα όπου δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς. Σε ποσοστό ακρίβειας των δεδομένων είναι 98,1071% σύμφωνα με το πρότυπο, ενώ το 1,8929% δεν είναι σύμφωνα με το πρότυπο.

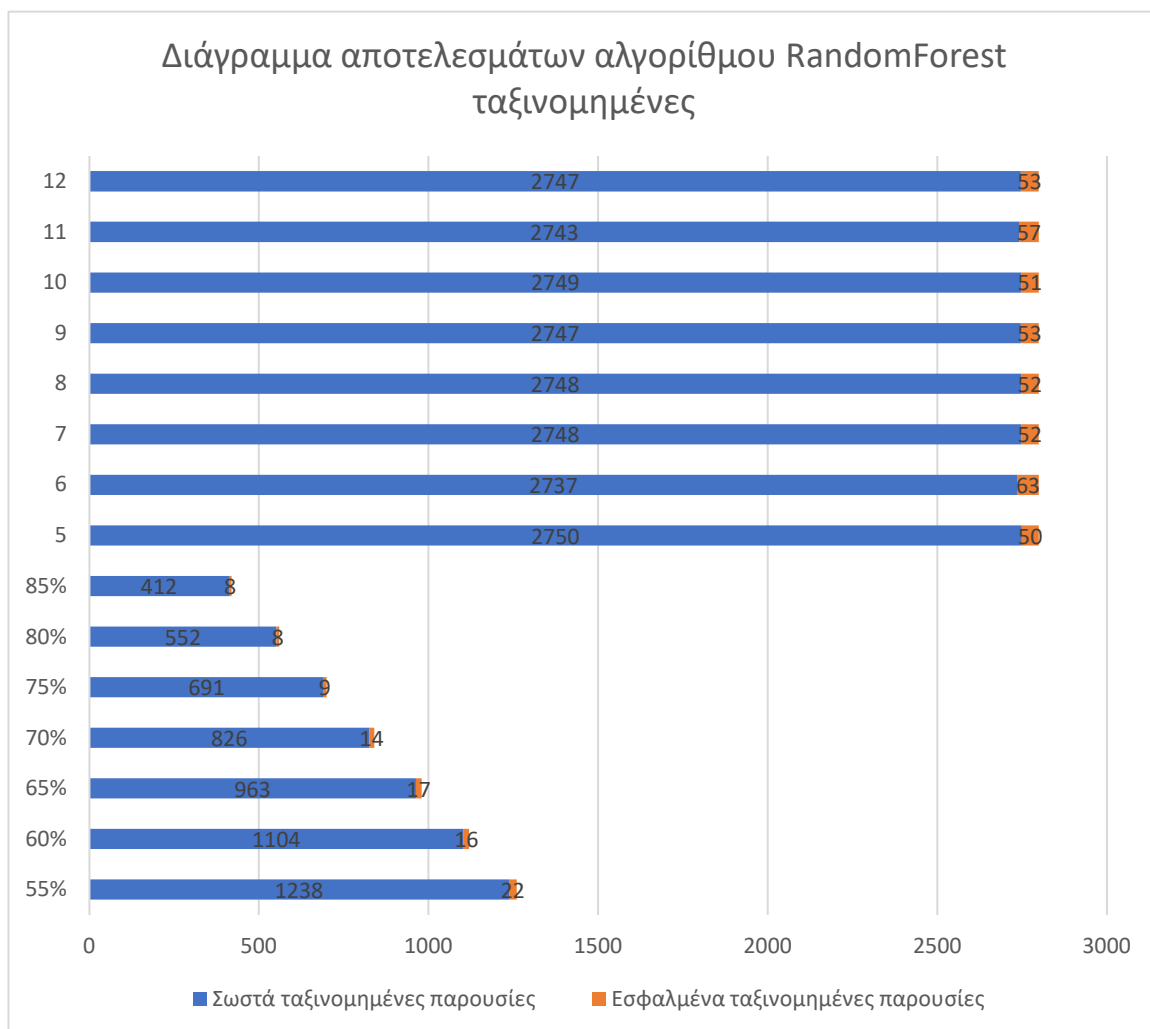


Εικόνα 100: Αλγόριθμος RandomForest -P100 -I 100 -num-slots 1 -K 0-M 1.0-V 0.001-S 1 με επικύρωση

12



Διάγραμμα 5: Διάγραμμα αποτελεσμάτων αλγορίθμου RandomForest



Διάγραμμα 6: Διάγραμμα αποτελεσμάτων αλγορίθμου RandomForest ταξινομημένες

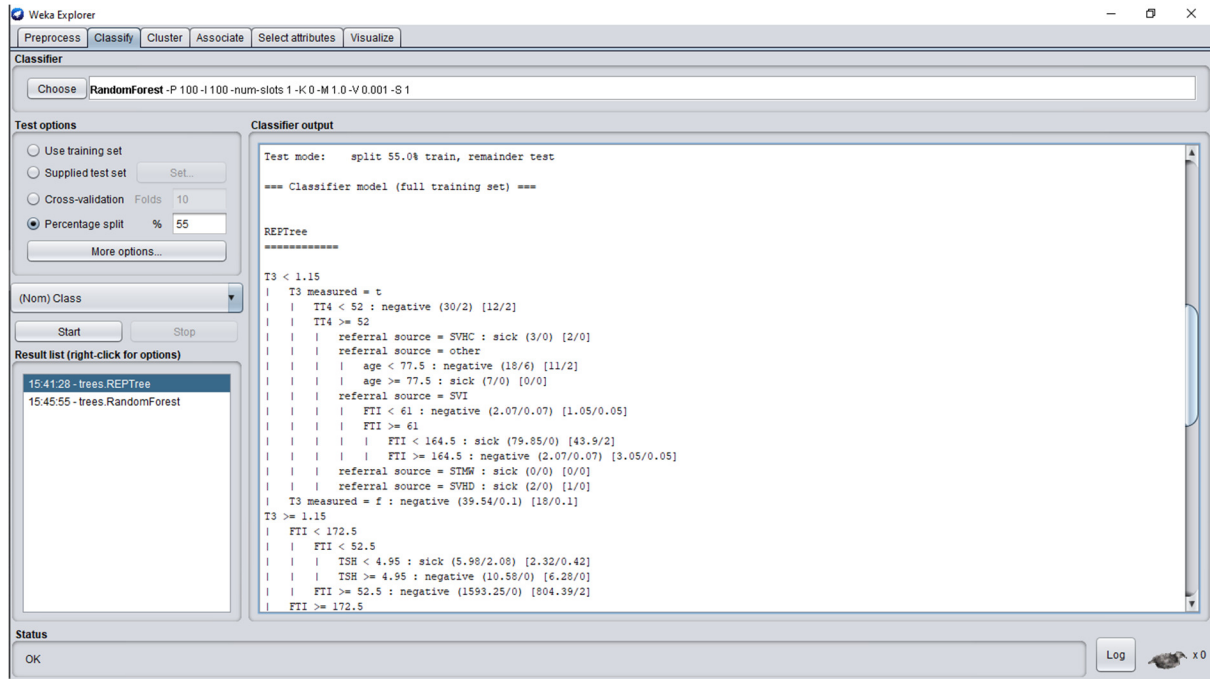
5.2.7 Δέντρο απόφασης αλγορίθμου REP Tree με διακριτοποίηση τιμών

5.2.7.1 Αλγόριθμος REPTree -M 2 -V 0,001 -N 3 -S 1 -L -1 -I 0.0 με διακριτοποίηση 55%

Εκπαιδύσαμε το 55% των δεδομένων μας με την μέθοδο της διακριτοποίησης τιμών, με τον αλγόριθμο REP Tree -M 2 -V 0,001 -N 3 -S 1 -L -1 -I 0.0 και πήραμε τα εξής αποτελέσματα: από το σύνολο των μεταβλητών που αναλύσαμε στο κεφάλαιο 5.1

Το συνολικό μέγεθος του δέντρου είναι 32 κόμβοι. Συνολικά από το μοντέλο που προκύπτει(μη εκπαιδευμένα δέντρα) βλέπουμε πως δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς 1.167 άτομα από τα οποία χάθηκαν τα 19 άτομα όπου έχουν διαγνωστεί με την νόσο του θυρεοειδούς ενώ έχουν διαγνωστεί με την νόσο του θυρεοειδούς 63 άτομα από τα

οποία χάθηκαν τα 11 άτομα όπου δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς. Σε ποσοστό ακρίβειας των δεδομένων είναι 97,619% σύμφωνα με το πρότυπο, ενώ το 2,381% δεν είναι σύμφωνα με το πρότυπο.

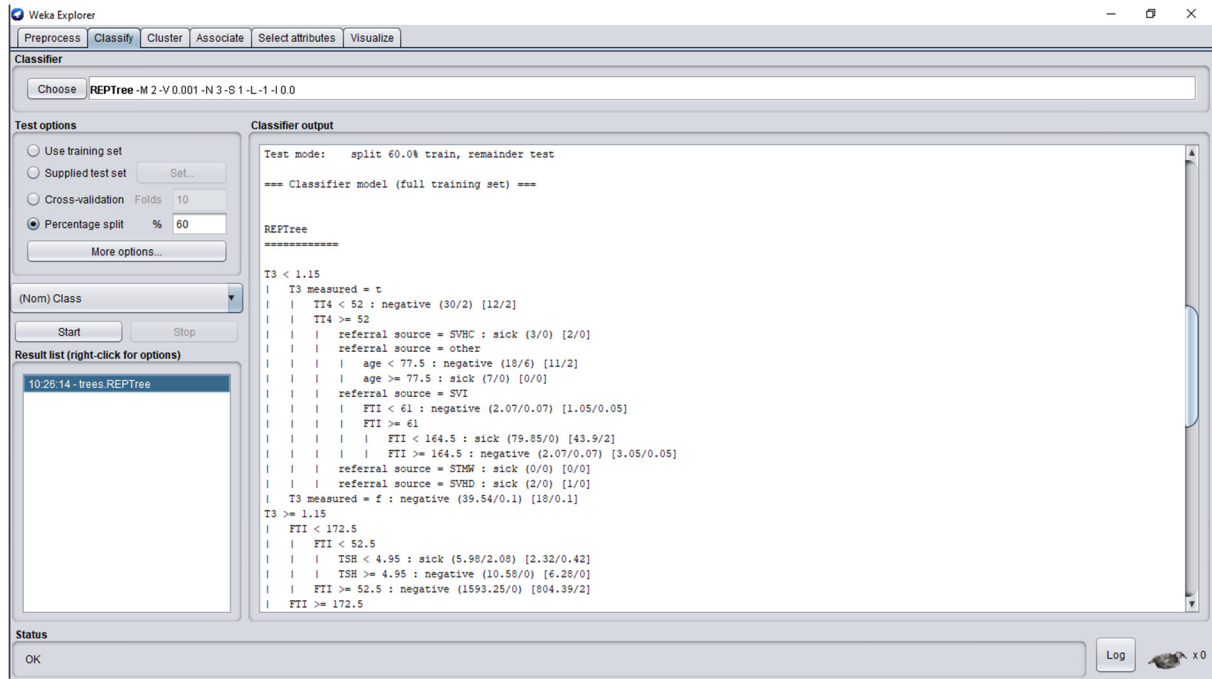


Εικόνα 101: Αλγόριθμος REPTree -M 2 -V 0,001 -N 3 -S 1 -L -1 -I 0.0 με διακριτοποίηση 55%

5.2.7.2 Αλγόριθμος REPTree -M 2 -V 0,001 -N 3 -S 1 -L -1 -I 0.0 με διακριτοποίηση 60%

Εκπαιδύσαμε το 60% των δεδομένων μας με την μέθοδο της διακριτοποίησης τιμών, με τον αλγόριθμο REPTree -M 2 -V 0,001 -N 3 -S 1 -L -1 -I 0.0 και πήραμε τα εξής αποτελέσματα: από το σύνολο των μεταβλητών που αναλύσαμε στο κεφάλαιο 5.1

Το συνολικό μέγεθος του δέντρου είναι 32 κόμβοι. Συνολικά από το μοντέλο που προκύπτει(μη εκπαιδευμένα δέντρα) βλέπουμε πως δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς 1.050 άτομα από τα οποία χάθηκαν τα 8 άτομα όπου έχουν διαγνωστεί με την νόσο του θυρεοειδούς ενώ έχουν διαγνωστεί με την νόσο του θυρεοειδούς 50 άτομα από τα οποία χάθηκαν τα 12 άτομα όπου δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς. Σε ποσοστό ακρίβειας των δεδομένων είναι 98,2143% σύμφωνα με το πρότυπο, ενώ το 1,7857% δεν είναι σύμφωνα με το πρότυπο.

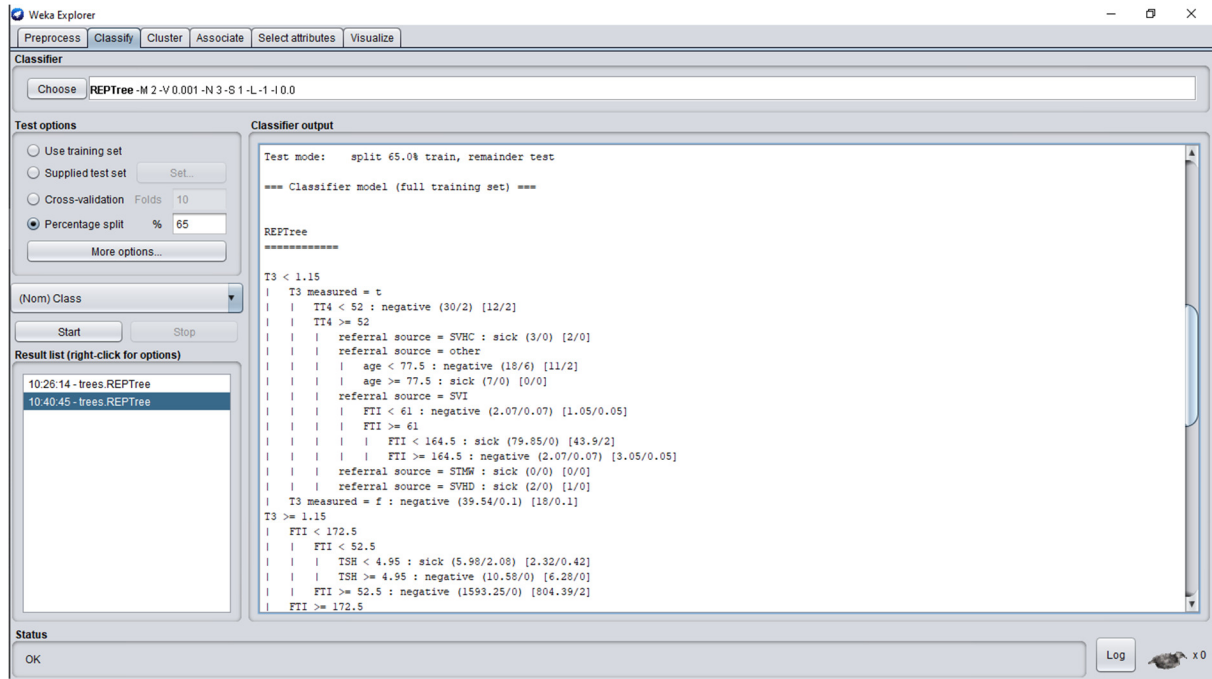


Εικόνα 102: Αλγόριθμος REPTree -M 2 -V 0,001 -N 3 -S 1 -L -1 -I 0.0 με διακριτοποίηση 60%

5.2.7.3 Αλγόριθμος REPTree -M 2 -V 0,001 -N 3 -S 1 -L -1 -I 0.0 με διακριτοποίηση 65%

Εκπαίδευσάμε το 65% των δεδομένων μας με την μέθοδο της διακριτοποίησης τιμών, με τον αλγόριθμο REPTree -M 2 -V 0,001 -N 3 -S 1 -L -1 -I 0.0 και πήραμε τα εξής αποτελέσματα: από το σύνολο των μεταβλητών που αναλύσαμε στο κεφάλαιο 5.1

Το συνολικό μέγεθος του δέντρου είναι 32 κόμβοι. Συνολικά από το μοντέλο που προκύπτει(μη εκπαιδευμένα δέντρα) βλέπουμε πως δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς 914 άτομα από τα οποία χάθηκαν τα 11 άτομα όπου έχουν διαγνωστεί με την νόσο του θυρεοειδούς ενώ έχουν διαγνωστεί με την νόσο του θυρεοειδούς 48 άτομα από τα οποία χάθηκαν τα 7 άτομα όπου δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς. Σε ποσοστό ακρίβειας των δεδομένων είναι 98,1633% σύμφωνα με το πρότυπο, ενώ το 1,8367% δεν είναι σύμφωνα με το πρότυπο.

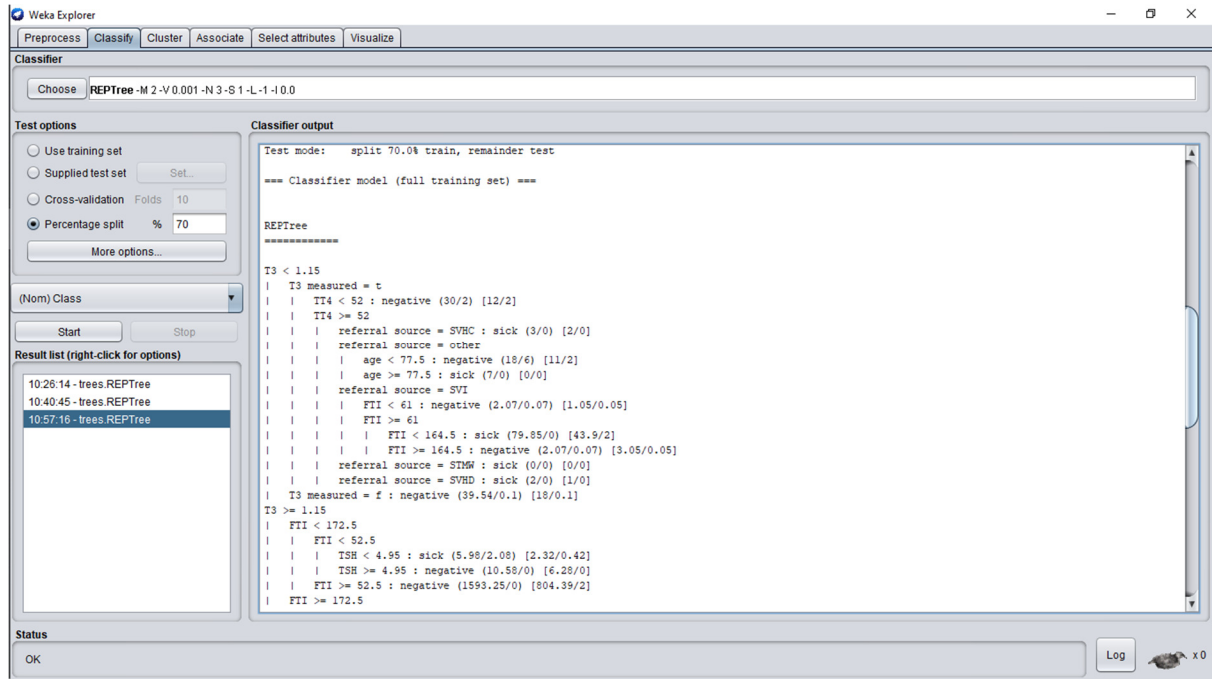


Εικόνα 103: Αλγόριθμος REPTree -M 2 -V 0,001 -N 3 -S 1 -L -1 -I 0.0 με διακριτοποίηση 65%

5.2.7.4 Αλγόριθμος REPTree -M 2 -V 0,001 -N 3 -S 1 -L -1 -I 0.0 με διακριτοποίηση 70%

Εκπαιδύσαμε το 70% των δεδομένων μας με την μέθοδο της διακριτοποίησης τιμών, με τον αλγόριθμο REPTree -M 2 -V 0,001 -N 3 -S 1 -L -1 -I 0.0 και πήραμε τα εξής αποτελέσματα: από το σύνολο των μεταβλητών που αναλύσαμε στο κεφάλαιο 5.1

Το συνολικό μέγεθος του δέντρου είναι 32 κόμβοι. Συνολικά από το μοντέλο που προκύπτει(μη εκπαιδευμένα δέντρα) βλέπουμε πως δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς 791 άτομα από τα οποία χάθηκαν τα 4 άτομα όπου έχουν διαγνωστεί με την νόσο του θυρεοειδούς ενώ έχουν διαγνωστεί με την νόσο του θυρεοειδούς 37 άτομα από τα οποία χάθηκαν τα 8 άτομα όπου δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς. Σε ποσοστό ακρίβειας των δεδομένων είναι 98,5714% σύμφωνα με το πρότυπο, ενώ το 1,4286% δεν είναι σύμφωνα με το πρότυπο.

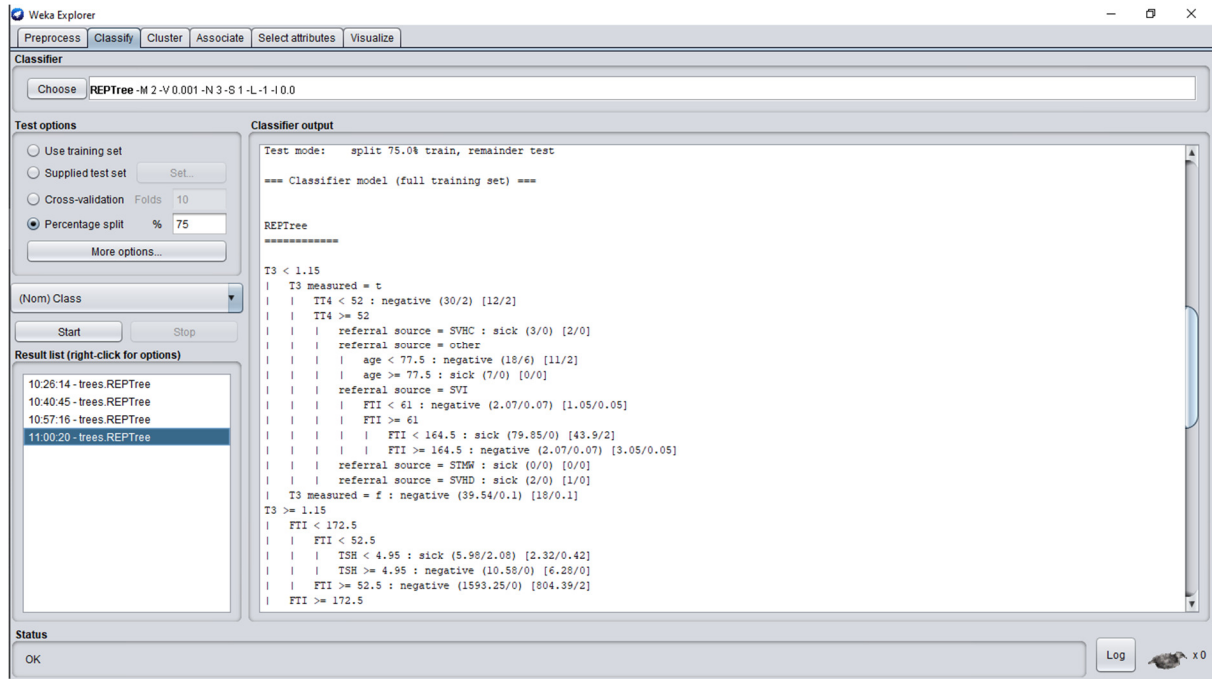


Εικόνα 104: Αλγόριθμος REPTree -M 2 -V 0,001 -N 3 -S 1 -L -1 -I 0.0 με διακριτοποίηση 70%

5.2.7.5 Αλγόριθμος REPTree -M 2 -V 0,001 -N 3 -S 1 -L -1 -I 0.0 με διακριτοποίηση 75%

Εκπαιδύσαμε το 75% των δεδομένων μας με την μέθοδο της διακριτοποίησης τιμών, με τον αλγόριθμο REPTree -M 2 -V 0,001 -N 3 -S 1 -L -1 -I 0.0 και πήραμε τα έξεις αποτελέσματα: από το σύνολο των μεταβλητών που αναλύσαμε στο κεφάλαιο 5.1

Το συνολικό μέγεθος του δέντρου είναι 32 κόμβοι. Συνολικά από το μοντέλο που προκύπτει(μη εκπαιδευμένα δέντρα) βλέπουμε πως δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς 657 άτομα από τα οποία χάθηκαν τα 10 άτομα όπου έχουν διαγνωστεί με την νόσο του θυρεοειδούς ενώ έχουν διαγνωστεί με την νόσο του θυρεοειδούς 29 άτομα από τα οποία χάθηκαν τα 4 άτομα όπου δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς. Σε ποσοστό ακρίβειας των δεδομένων είναι 98% σύμφωνα με το πρότυπο, ενώ το 2% δεν είναι σύμφωνα με το πρότυπο.

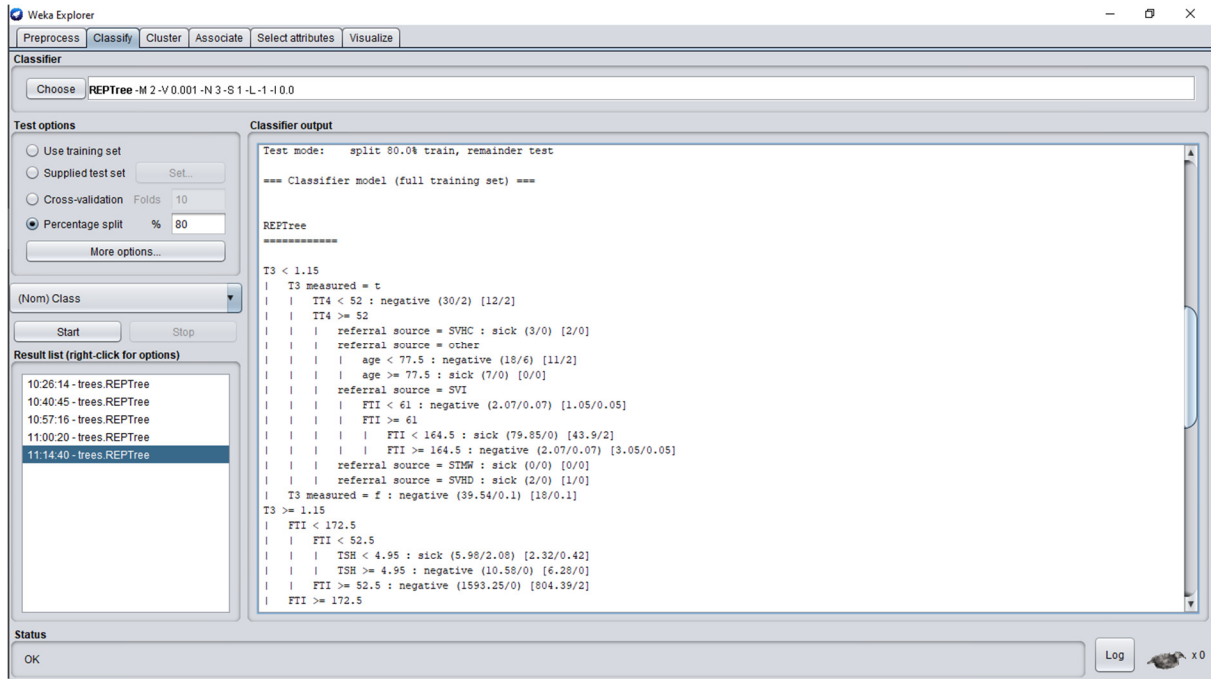


Εικόνα 105: Αλγόριθμος REPTree -M 2 -V 0,001 -N 3 -S 1 -L -1 -I 0.0 με διακριτοποίηση 75%

5.2.7.6 Αλγόριθμος REPTree -M 2 -V 0,001 -N 3 -S 1 -L -1 -I 0.0 με διακριτοποίηση 80%

Εκπαιδύσαμε το 80% των δεδομένων μας με την μέθοδο της διακριτοποίησης τιμών, με τον αλγόριθμο REPTree -M 2 -V 0,001 -N 3 -S 1 -L -1 -I 0.0 και πήραμε τα εξής αποτελέσματα: από το σύνολο των μεταβλητών που αναλύσαμε στο κεφάλαιο 5.1

Το συνολικό μέγεθος του δέντρου είναι 32 κόμβοι. Συνολικά από το μοντέλο που προκύπτει(μη εκπαιδευμένα δέντρα) βλέπουμε πως δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς 534 άτομα από τα οποία χάθηκαν τα 3 άτομα όπου έχουν διαγνωστεί με την νόσο του θυρεοειδούς ενώ έχουν διαγνωστεί με την νόσο του θυρεοειδούς 20 άτομα από τα οποία χάθηκαν τα 3 άτομα όπου δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς. Σε ποσοστό ακρίβειας των δεδομένων είναι 98,9286% σύμφωνα με το πρότυπο, ενώ το 1,0714% δεν είναι σύμφωνα με το πρότυπο.

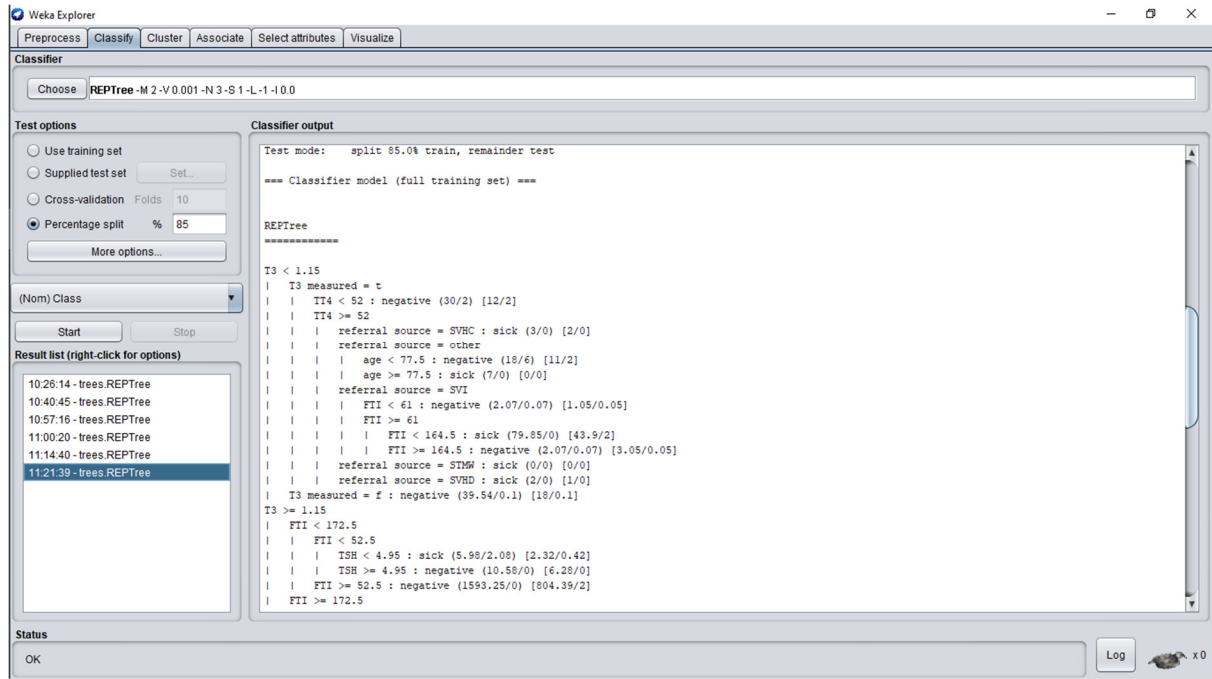


Εικόνα 106: Αλγόριθμος REPTree -M 2 -V 0,001 -N 3 -S 1 -L -1 -I 0.0 με διακριτοποίηση 80%

5.2.7.7 Αλγόριθμος REPTree -M 2 -V 0,001 -N 3 -S 1 -L -1 -I 0.0 με διακριτοποίηση 85%

Εκπαιδύσαμε το 85% των δεδομένων μας με την μέθοδο της διακριτοποίησης τιμών, με τον αλγόριθμο REPTree -M 2 -V 0,001 -N 3 -S 1 -L -1 -I 0.0 και πήραμε τα εξής αποτελέσματα: από το σύνολο των μεταβλητών που αναλύσαμε στο κεφάλαιο 5.1

Το συνολικό μέγεθος του δέντρου είναι 32 κόμβοι. Συνολικά από το μοντέλο που προκύπτει(μη εκπαιδευμένα δέντρα) βλέπουμε πως δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς 400 άτομα από τα οποία χάθηκαν τα 0 άτομα όπου έχουν διαγνωστεί με την νόσο του θυρεοειδούς ενώ έχουν διαγνωστεί με την νόσο του θυρεοειδούς 17 άτομα από τα οποία χάθηκαν τα 3 άτομα όπου δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς. Σε ποσοστό ακρίβειας των δεδομένων είναι 99,2857% σύμφωνα με το πρότυπο, ενώ το 0,7143% δεν είναι σύμφωνα με το πρότυπο.



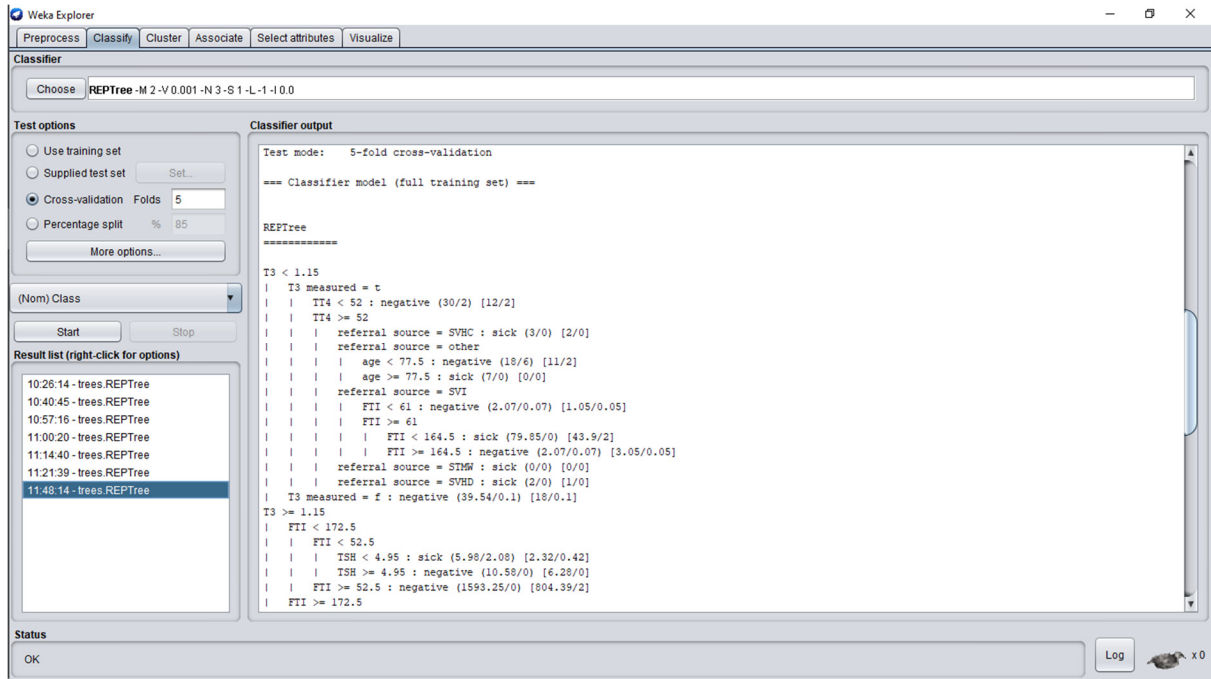
Εικόνα 107: Αλγόριθμος REPTree -M 2 -V 0,001 -N 3 -S 1 -L -1 -I 0.0 με διακριτοποίηση 85%

5.2.8 Δέντρο απόφασης αλγορίθμου REP Tree με διακριτοποίηση τιμών και επικύρωση

5.2.8.1 Αλγόριθμος REPTree -M 2 -V 0,001 -N 3 -S 1 -L -1 -I 0.0 με επικύρωση 5

Εκπαίδευσάμε τα δεδομένα μας με την μέθοδο της διακριτοποίησης τιμών με επικύρωση 5 (fold cross), με τον αλγόριθμο REPTree -M 2 -V 0,001 -N 3 -S 1 -L -1 -I 0.0 και πήραμε τα έξεις αποτελέσματα: από το σύνολο των μεταβλητών που αναλύσαμε στο κεφάλαιο 5.1

Το συνολικό μέγεθος του δέντρου είναι 32 κόμβοι. Συνολικά από το μοντέλο που προκύπτει(μη εκπαιδευμένα δέντρα) βλέπουμε πως δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς 2.613 άτομα από τα οποία χάθηκαν τα 16 άτομα όπου έχουν διαγνωστεί με την νόσο του θυρεοειδούς ενώ έχουν διαγνωστεί με την νόσο του θυρεοειδούς 137 άτομα από τα οποία χάθηκαν τα 34 άτομα όπου δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς. Σε ποσοστό ακρίβειας των δεδομένων είναι 98,2143% σύμφωνα με το πρότυπο, ενώ το 1,7857% δεν είναι σύμφωνα με το πρότυπο.

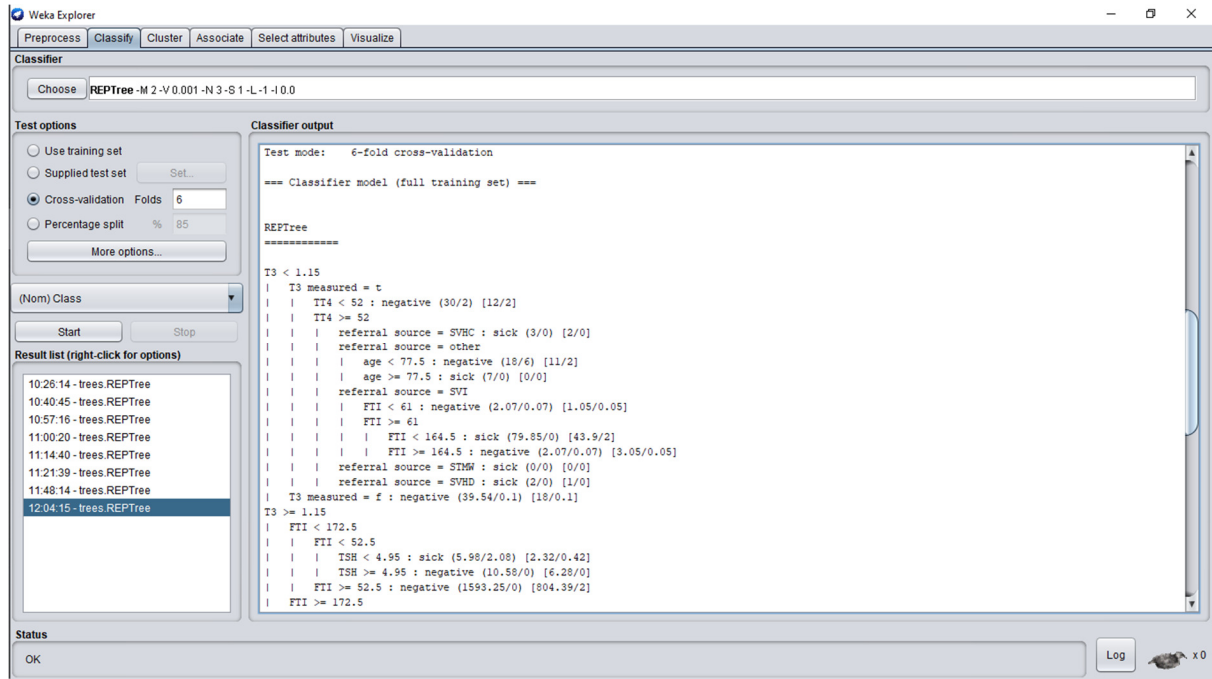


Εικόνα 108: Αλγόριθμος REPTree -M 2 -V 0,001 -N 3 -S 1 -L -1 -I 0.0 με επικύρωση 5

5.2.8.2 Αλγόριθμος REPTree -M 2 -V 0,001 -N 3 -S 1 -L -1 -I 0.0 με επικύρωση 6

Εκπαιδύσαμε τα δεδομένα μας με την μέθοδο της διακριτοποίησης τιμών με επικύρωση 6 (fold cross), με τον αλγόριθμο REPTree -M 2 -V 0,001 -N 3 -S 1 -L -1 -I 0.0 και πήραμε τα εξής αποτελέσματα: από το σύνολο των μεταβλητών που αναλύσαμε στο κεφάλαιο 5.1

Το συνολικό μέγεθος του δέντρου είναι 32 κόμβοι. Συνολικά από το μοντέλο που προκύπτει(μη εκπαιδευμένα δέντρα) βλέπουμε πως δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς 2.617 άτομα από τα οποία χάθηκαν τα 12 άτομα όπου έχουν διαγνωστεί με την νόσο του θυρεοειδούς ενώ έχουν διαγνωστεί με την νόσο του θυρεοειδούς 140 άτομα από τα οποία χάθηκαν τα 31 άτομα όπου δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς. Σε ποσοστό ακρίβειας των δεδομένων είναι 98,4643% σύμφωνα με το πρότυπο, ενώ το 1,5357% δεν είναι σύμφωνα με το πρότυπο.

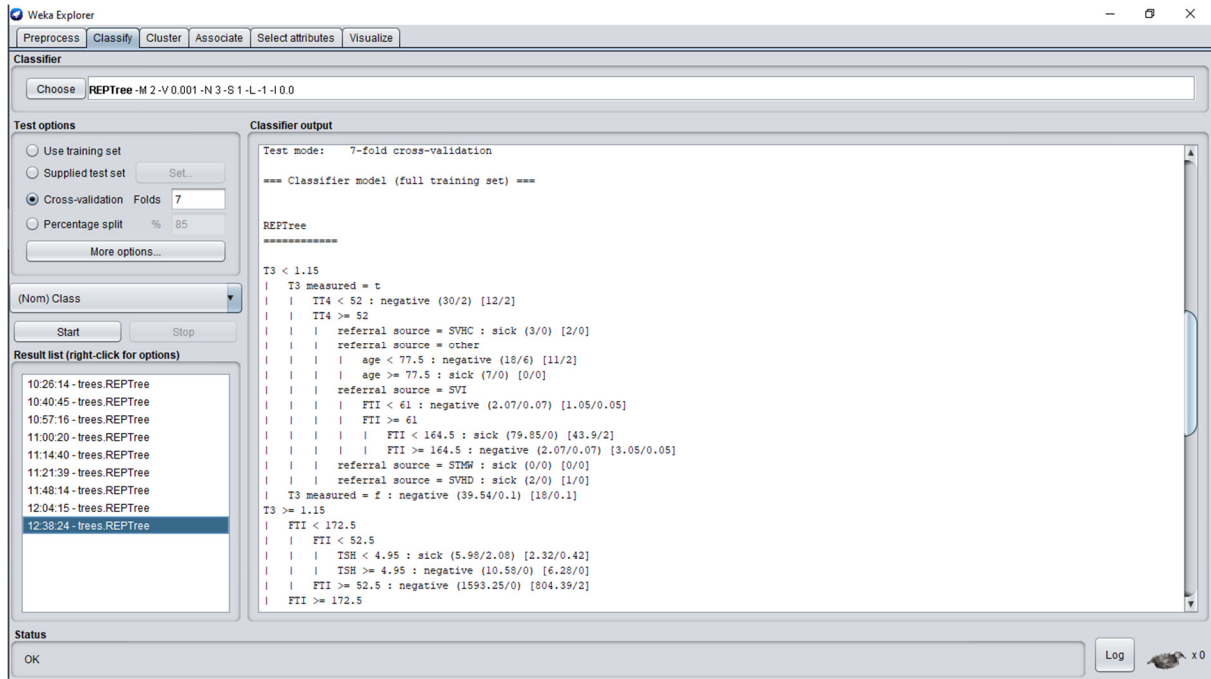


Εικόνα 109: Αλγόριθμος REPTree -M 2 -V 0,001 -N 3 -S 1 -L -1 -I 0.0 με επικύρωση 6

5.2.8.3 Αλγόριθμος REPTree -M 2 -V 0,001 -N 3 -S 1 -L -1 -I 0.0 με επικύρωση 7

Εκπαιδύσαμε τα δεδομένα μας με την μέθοδο της διακριτοποίησης τιμών με επικύρωση 7 (fold cross), με τον αλγόριθμο REPTree -M 2 -V 0,001 -N 3 -S 1 -L -1 -I 0.0 και πήραμε τα έξεις αποτελέσματα: από το σύνολο των μεταβλητών που αναλύσαμε στο κεφάλαιο 5.1

Το συνολικό μέγεθος του δέντρου είναι 32 κόμβοι. Συνολικά από το μοντέλο που προκύπτει(μη εκπαιδευμένα δέντρα) βλέπουμε πως δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς 2.616 άτομα από τα οποία χάθηκαν τα 13 άτομα όπου έχουν διαγνωστεί με την νόσο του θυρεοειδούς ενώ έχουν διαγνωστεί με την νόσο του θυρεοειδούς 138 άτομα από τα οποία χάθηκαν τα 33 άτομα όπου δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς. Σε ποσοστό ακρίβειας των δεδομένων είναι 98,3571% σύμφωνα με το πρότυπο, ενώ το 1,6429% δεν είναι σύμφωνα με το πρότυπο.

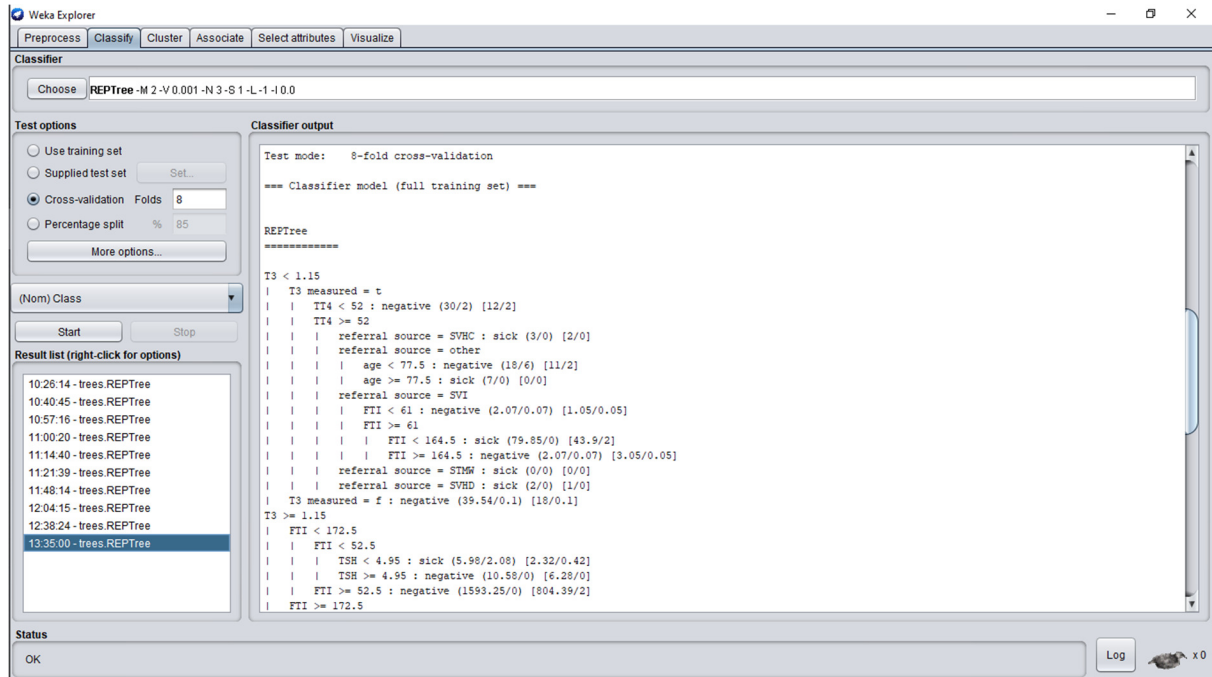


Εικόνα 110: Αλγόριθμος REPTree -M 2 -V 0,001 -N 3 -S 1 -L -1 -I 0.0 με επικύρωση 7

5.2.8.4 Αλγόριθμος REPTree -M 2 -V 0,001 -N 3 -S 1 -L -1 -I 0.0 με επικύρωση 8

Εκπαιδύσαμε τα δεδομένα μας με την μέθοδο της διακριτοποίησης τιμών με επικύρωση 8 (fold cross), με τον αλγόριθμο REPTree -M 2 -V 0,001 -N 3 -S 1 -L -1 -I 0.0 και πήραμε τα έξεις αποτελέσματα: από το σύνολο των μεταβλητών που αναλύσαμε στο κεφάλαιο 5.1

Το συνολικό μέγεθος του δέντρου είναι 32 κόμβοι. Συνολικά από το μοντέλο που προκύπτει(μη εκπαιδευμένα δέντρα) βλέπουμε πως δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς 2.617 άτομα από τα οποία χάθηκαν τα 12 άτομα όπου έχουν διαγνωστεί με την νόσο του θυρεοειδούς ενώ έχουν διαγνωστεί με την νόσο του θυρεοειδούς 149 άτομα από τα οποία χάθηκαν τα 22 άτομα όπου δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς. Σε ποσοστό ακρίβειας των δεδομένων είναι 98,7857% σύμφωνα με το πρότυπο, ενώ το 1,2143% δεν είναι σύμφωνα με το πρότυπο.

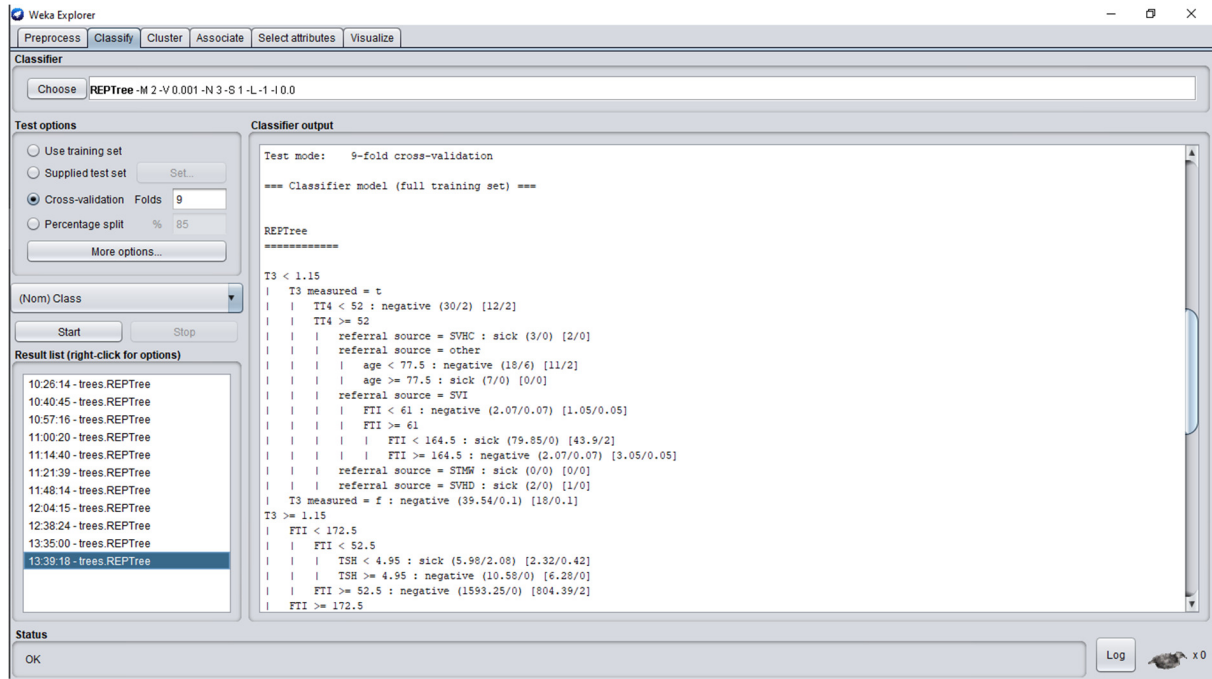


Εικόνα 111: Αλγόριθμος REPTree -M 2 -V 0,001 -N 3 -S 1 -L -1 -I 0.0 με επικύρωση 8

5.2.8.5 Αλγόριθμος REPTree -M 2 -V 0,001 -N 3 -S 1 -L -1 -I 0.0 με επικύρωση 9

Εκπαιδύσαμε τα δεδομένα μας με την μέθοδο της διακριτοποίησης τιμών με επικύρωση 9 (fold cross), με τον αλγόριθμο REPTree -M 2 -V 0,001 -N 3 -S 1 -L -1 -I 0.0 και πήραμε τα εξής αποτελέσματα: από το σύνολο των μεταβλητών που αναλύσαμε στο κεφάλαιο 5.1

Το συνολικό μέγεθος του δέντρου είναι 32 κόμβοι. Συνολικά από το μοντέλο που προκύπτει(μη εκπαιδευμένα δέντρα) βλέπουμε πως δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς 2.618 άτομα από τα οποία χάθηκαν τα 11 άτομα όπου έχουν διαγνωστεί με την νόσο του θυρεοειδούς ενώ έχουν διαγνωστεί με την νόσο του θυρεοειδούς 140 άτομα από τα οποία χάθηκαν τα 31 άτομα όπου δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς. Σε ποσοστό ακρίβειας των δεδομένων είναι 98,5% σύμφωνα με το πρότυπο, ενώ το 1,5% δεν είναι σύμφωνα με το πρότυπο.

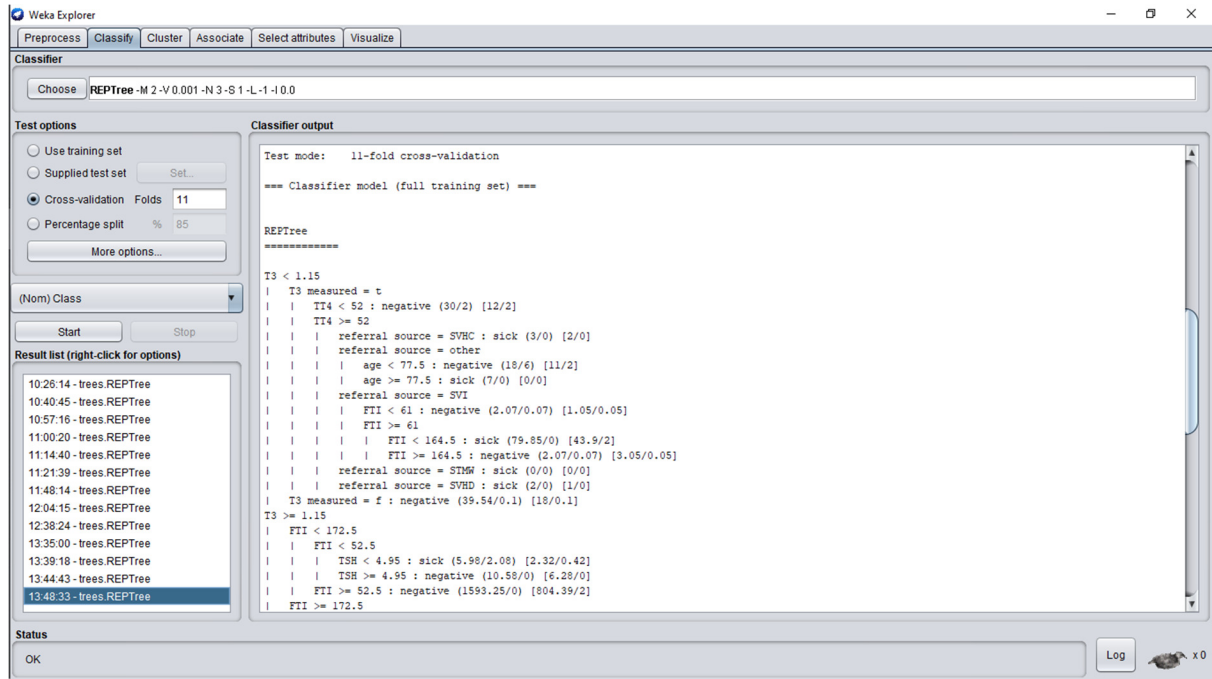


Εικόνα 112: Αλγόριθμος REPTree -M 2 -V 0,001 -N 3 -S 1 -L -1 -I 0.0 με επικύρωση 9

5.2.8.6 Αλγόριθμος REPTree -M 2 -V 0,001 -N 3 -S 1 -L -1 -I 0.0 με επικύρωση 10

Εκπαιδύσαμε τα δεδομένα μας με την μέθοδο της διακριτοποίησης τιμών με επικύρωση 10 (fold cross), με τον αλγόριθμο REPTree -M 2 -V 0,001 -N 3 -S 1 -L -1 -I 0.0 και πήραμε τα εξής αποτελέσματα: από το σύνολο των μεταβλητών που αναλύσαμε στο κεφάλαιο 5.1

Το συνολικό μέγεθος του δέντρου είναι 32 κόμβοι. Συνολικά από το μοντέλο που προκύπτει(μη εκπαιδευμένα δέντρα) βλέπουμε πως δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς 2.616 άτομα από τα οποία χάθηκαν τα 13 άτομα όπου έχουν διαγνωστεί με την νόσο του θυρεοειδούς ενώ έχουν διαγνωστεί με την νόσο του θυρεοειδούς 141 άτομα από τα οποία χάθηκαν τα 30 άτομα όπου δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς. Σε ποσοστό ακρίβειας των δεδομένων είναι 98,4643% σύμφωνα με το πρότυπο, ενώ το 1,5357% δεν είναι σύμφωνα με το πρότυπο.

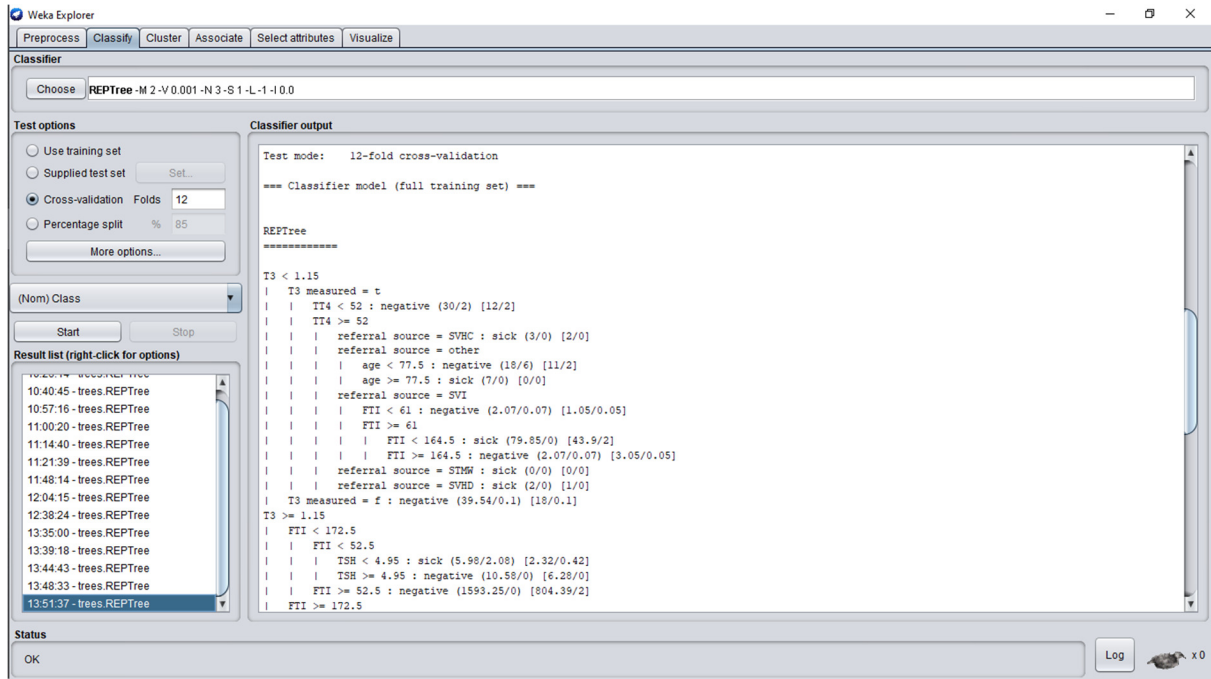


Εικόνα 113: Αλγόριθμος REPTree -M 2 -V 0,001 -N 3 -S 1 -L -1 -I 0.0 με επικύρωση 11

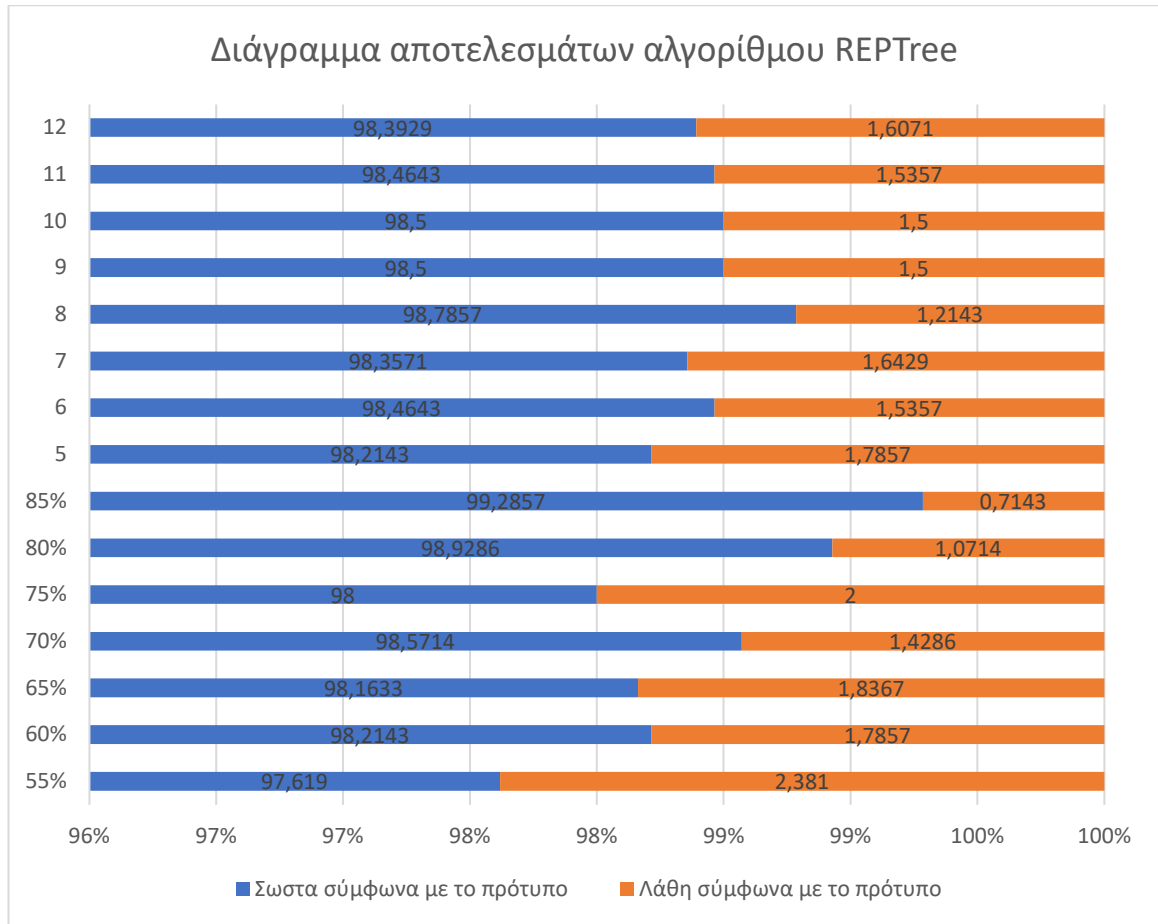
5.2.8.7 Αλγόριθμος REPTree -M 2 -V 0,001 -N 3 -S 1 -L -1 -I 0.0 με επικύρωση 12

Εκπαιδύσαμε τα δεδομένα μας με την μέθοδο της διακριτοποίησης τιμών με επικύρωση 12 (fold cross), με τον αλγόριθμο REPTree -M 2 -V 0,001 -N 3 -S 1 -L -1 -I 0.0 και πήραμε τα εξής αποτελέσματα: από το σύνολο των μεταβλητών που αναλύσαμε στο κεφάλαιο 5.1

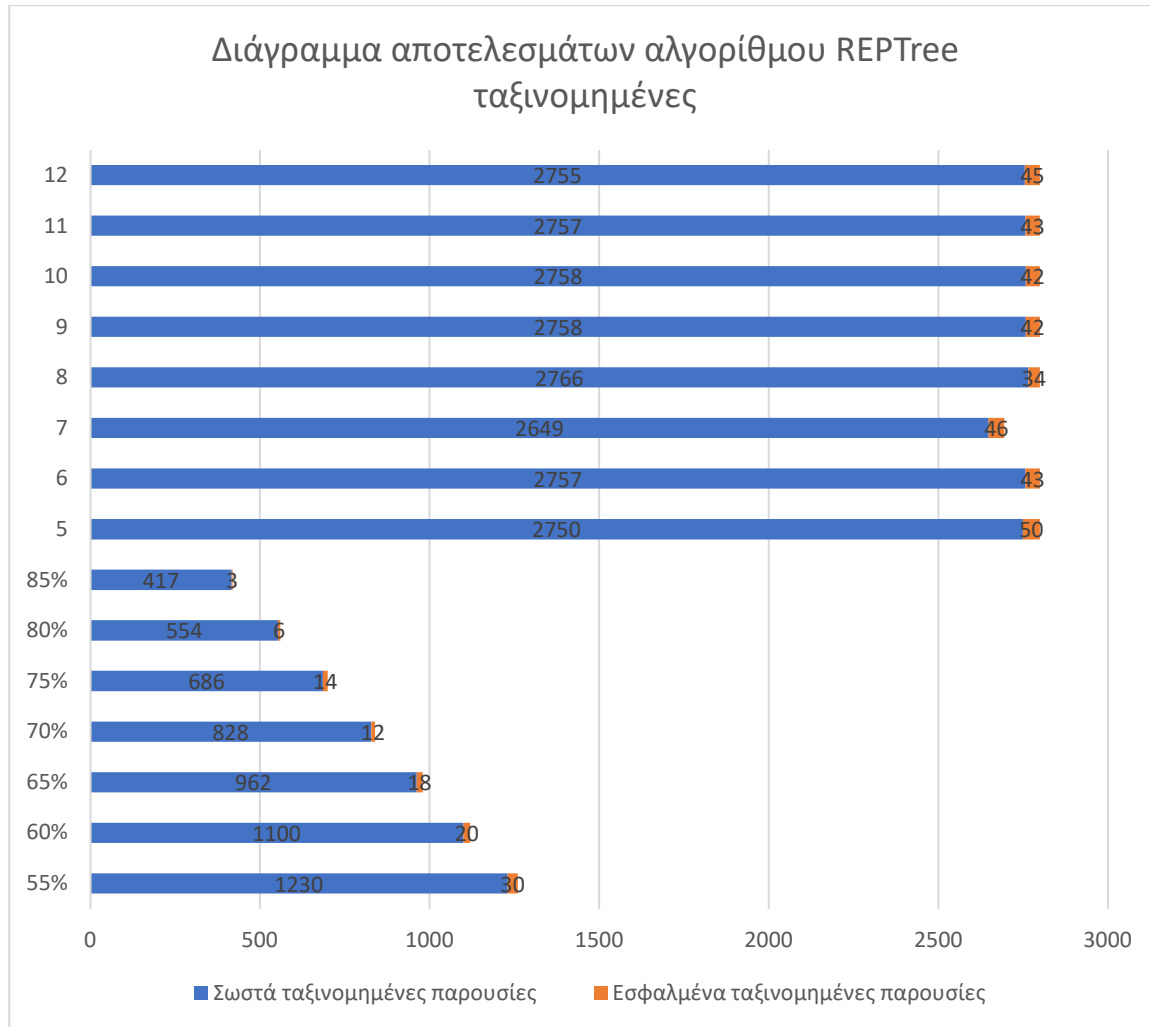
Το συνολικό μέγεθος του δέντρου είναι 32 κόμβοι. Συνολικά από το μοντέλο που προκύπτει(μη εκπαιδευμένα δέντρα) βλέπουμε πως δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς 2.620 άτομα από τα οποία χάθηκαν τα 9 άτομα όπου έχουν διαγνωστεί με την νόσο του θυρεοειδούς ενώ έχουν διαγνωστεί με την νόσο του θυρεοειδούς 135 άτομα από τα οποία χάθηκαν τα 36 άτομα όπου δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς. Σε ποσοστό ακρίβειας των δεδομένων είναι 98,3929% σύμφωνα με το πρότυπο, ενώ το 1,6071% δεν είναι σύμφωνα με το πρότυπο.



Εικόνα 114: Αλγόριθμος REPTree -M 2 -V 0,001 -N 3 -S 1 -L -1 -I 0.0 με επικύρωση 12



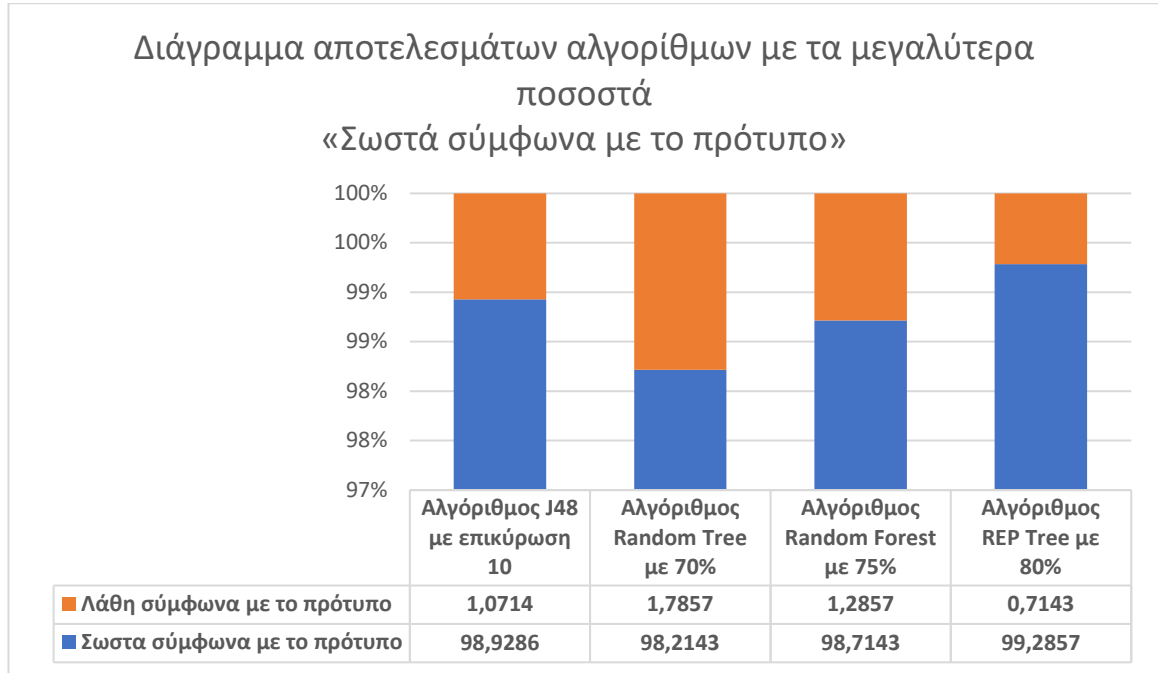
Διάγραμμα 7: Διάγραμμα αποτελεσμάτων αλγορίθμου REPTree



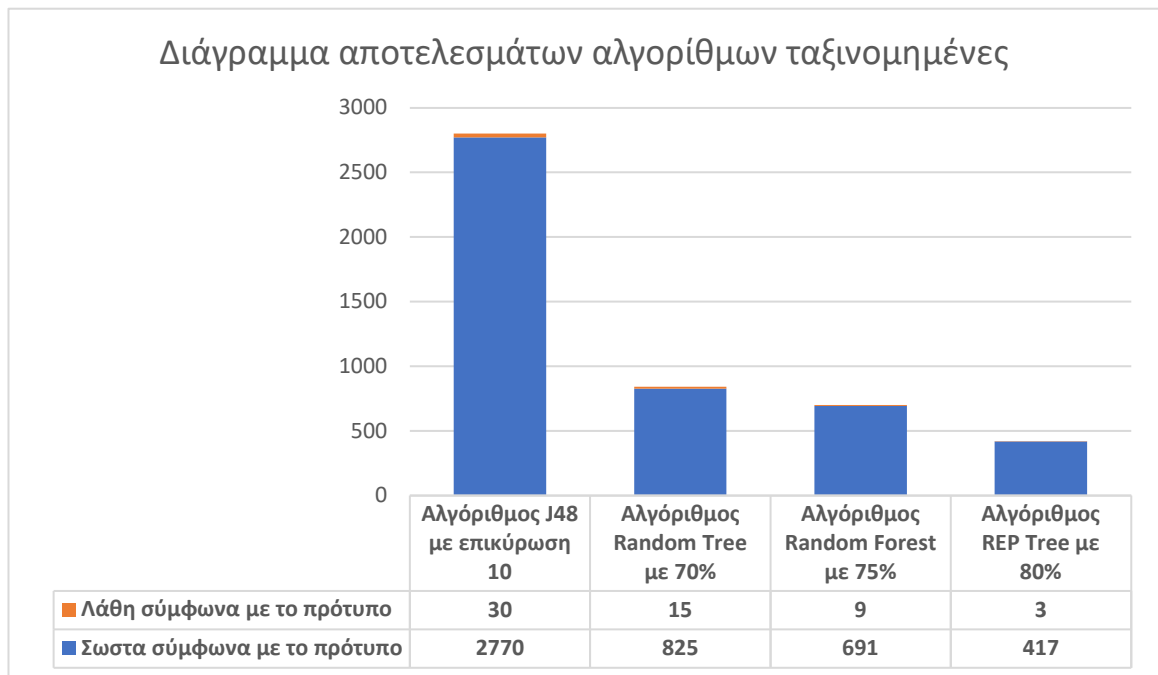
Διάγραμμα 8: Διάγραμμα αποτελεσμάτων αλγορίθμου REPTree ταξινομημένες

5.3 Συσχετίσεις Αλγορίθμων

Οι καλύτεροι αλγόριθμοι όπου ανατρέξαμε είναι J48 με επικύρωση 10, Random Tree με Percentage split 70%, Random Forest με Percentage split 75%, REP Tree με Percentage split 80%. Ο καλύτερος αλγόριθμος από το σύνολο όλων των αλγορίθμων είναι ο REP Tree με Percentage split 85% με ποσοστό επιτυχίας 99,2857% (βλέπω παρακάτω Διάγραμμα 9: Διάγραμμα αποτελεσμάτων αλγορίθμων με τα μεγαλύτερα ποσοστά «Σωστά σύμφωνα με το πρότυπο» σελ.137). Είναι το πιο ακριβές αποτέλεσμα δεδομένων που ανατρέξαμε στο πρόγραμμα weka. Ωστόσο, διότι ο αλγόριθμος REP Tree με Percentage split 85% έχει εκπαίδευση το μεγαλύτερο ποσοστό των δεδομένων (βλέπω παρακάτω Διάγραμμα 10: Διάγραμμα αποτελεσμάτων αλγορίθμων ταξινομημένες σελ. 137) και επειδή ο αλγορίθμος είναι μεγάλος (βλεπω μειονεκτήματα REP Tree σελ. 15) για αυτούς τους λόγους ως καλύτερο αλγόριθμο θεωρώ ότι είναι ο **Αλγόριθμος J48 με επικύρωση 10 με ποσοστό επιτυχίας 98,9286%**.



Διάγραμμα 9: Διάγραμμα αποτελεσμάτων αλγορίθμων με τα μεγαλύτερα ποσοστά «Σωστά σύμφωνα με το πρότυπο»



Διάγραμμα 10: Διάγραμμα αποτελεσμάτων αλγορίθμων ταξινομημένες

ΣΥΜΠΕΡΑΣΜΑΤΑ

Σε αυτό το σημείο θα θέλαμε να αναφερθούμε στα κυριότερα σημεία, που αποκτήθηκαν κατά την διάρκεια εκπόνησης της πτυχιακής μας εργασίας. Αρχικά, έγινε σχετική έρευνα για την εξόρυξη δεδομένων και την μηχανική μάθηση. Η έρευνα αυτή μας βοήθησε να αντιληφθούμε την βάση όλων των αρχών, στην οποία στηρίζεται κάθε αλγόριθμος. Στην συνέχεια της εργασίας αποτυπώσαμε την θεωρία στην πράξη. Με την βοήθεια του προγράμματος Weka, έγινε κατηγοριοποίηση σε μία βάση δεδομένων, σχετικά με τη νόσο του θυρεοειδή. Η κατηγοριοποίηση αυτή διαθέτει διάφορα στάδια, όπως ανάλυση πραγματικών στοιχείων, που αποκτήθηκαν από ερωτώμενους, τα οποία, στην πορεία της εργασίας, αναλύθηκαν με συγκεκριμένους αλγόριθμους.

ΔΙΕΥΚΡΙΝΙΣΕΙΣ

Αναλυτική ανάλυση του αλγορίθμου J48

Οι αναλυτικές αναλύσεις του αλγορίθμου J48 ισχύουν για τα υποκεφάλαια 5.2.1-5.2.2(σελ. 71~84)

- Βλέπουμε πως όσοι είχαν ορμόνη τριϊωδοθυρονίνη με τιμή 1,1 και μικρότερη και είχαν μέτρηση την ορμόνη τριϊωδοθυρονίνης και είχαν θυροξίνη ολική με τιμή 52 και μικρότερη και είχαν ορμόνη διέγερσης του θυρεοειδούς με τιμή 21 και μικρότερη και η ορμόνη διέγερσης του θυρεοειδούς είχαν τιμή 4,1 τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 9 άτομα) ενώ αν η ορμόνη διέγερσης του θυρεοειδούς είχαν τιμή 4,2 και μεγαλύτερη τότε διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 3 άτομα).
- Βλέπουμε πως όσοι είχαν ορμόνη τριϊωδοθυρονίνη με τιμή 1,1 και μικρότερη και είχαν μέτρηση την ορμόνη τριϊωδοθυρονίνης και είχαν θυροξίνη ολική με τιμή 52 και μικρότερη και είχαν ορμόνη διέγερσης του θυρεοειδούς με τιμή 22 και μεγαλύτερη τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 31 άτομα).
- Βλέπουμε πως όσοι είχαν ορμόνη τριϊωδοθυρονίνη με τιμή 1,1 και μικρότερη και είχαν μέτρηση την ορμόνη τριϊωδοθυρονίνης και είχαν θυροξίνη ολική με τιμή μεγαλύτερη από 52 και με φυσιολογικές τιμές είχαν SVHC τότε διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 5 άτομα).
- Βλέπουμε πως όσοι είχαν ορμόνη τριϊωδοθυρονίνη με τιμή 1,1 και μικρότερη και είχαν μέτρηση την ορμόνη τριϊωδοθυρονίνης και είχαν θυροξίνη ολική με τιμή μεγαλύτερη από 52 και με φυσιολογικές τιμές είχαν OTHER και είχαν μέτρηση την θυροξίνης και η ορμόνη διέγερσης του θυρεοειδούς είχαν τιμή 8,6 και μικρότερη και η ηλικία ήταν από 66 ετών και μικρότερη τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 11 άτομα) ενώ αν η ηλικία ήταν μεγαλύτερη από 66 ετών και η ορμόνη διέγερσης του θυρεοειδούς είχαν τιμή 0,13 και μικρότερη τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 3 άτομα) ενώ αν η ορμόνη διέγερσης του θυρεοειδούς είχαν τιμή μεγαλύτερη από 0,13 τότε διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 11 άτομα).

- Βλέπουμε πως όσοι είχαν ορμόνη τριϊωδοθυρονίνη με τιμή 1,1 και μικρότερη και είχαν μέτρηση την ορμόνη τριϊωδοθυρονίνης και είχαν θυροξίνη ολική με τιμή μεγαλύτερη από 52 και με φυσιολογικές τιμές είχαν OTHER και είχαν μέτρηση την θυροξίνης και η ορμόνη διέγερσης του θυρεοειδούς είχαν τιμή μεγαλύτερη από 8,6 διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 6 άτομα).
- Βλέπουμε πως όσοι είχαν ορμόνη τριϊωδοθυρονίνη με τιμή 1,1 και μικρότερη και είχαν μέτρηση την ορμόνη τριϊωδοθυρονίνης και είχαν θυροξίνη ολική με τιμή μεγαλύτερη από 52 και με φυσιολογικές τιμές είχαν OTHER και είχαν μέτρηση την θυροξίνης δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 3 άτομα).
- Βλέπουμε πως όσοι είχαν ορμόνη τριϊωδοθυρονίνη με τιμή 1,1 και μικρότερη και είχαν μέτρηση την ορμόνη τριϊωδοθυρονίνης και είχαν θυροξίνη ολική με τιμή μεγαλύτερη από 52 και με φυσιολογικές τιμές είχαν SVI και έχουν δείκτη ελεύθερης θυροξίνης με τιμή 157 και μικρότερη και έχουν δείκτη ελεύθερης θυροξίνης με τιμή 69 και μικρότερη και είχαν θυροξίνης με τιμή 0,96 και μικρότερη τότε διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 4 άτομα) ενώ είχαν θυροξίνης με τιμή μεγαλύτερη από 0,96 τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 2 άτομα).
- Βλέπουμε πως όσοι είχαν ορμόνη τριϊωδοθυρονίνη με τιμή 1,1 και μικρότερη και είχαν μέτρηση την ορμόνη τριϊωδοθυρονίνης και είχαν θυροξίνη ολική με τιμή μεγαλύτερη από 52 και με φυσιολογικές τιμές είχαν SVI και έχουν δείκτη ελεύθερης θυροξίνης με τιμή 157 και μικρότερη και έχουν δείκτη ελεύθερης θυροξίνης με τιμή μεγαλύτερη 69 τότε διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 116 άτομα).
- Βλέπουμε πως όσοι είχαν ορμόνη τριϊωδοθυρονίνη με τιμή 1,1 και μικρότερη και είχαν μέτρηση την ορμόνη τριϊωδοθυρονίνης και είχαν θυροξίνη ολική με τιμή μεγαλύτερη από 52 και με φυσιολογικές τιμές είχαν SVI και έχουν δείκτη ελεύθερης θυροξίνης με τιμή μεγαλύτερη 157 τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 8 άτομα).
- Βλέπουμε πως όσοι είχαν ορμόνη τριϊωδοθυρονίνη με τιμή 1,1 και μικρότερη και είχαν μέτρηση την ορμόνη τριϊωδοθυρονίνης και είχαν θυροξίνη ολική με τιμή μεγαλύτερη από 52 και με φυσιολογικές τιμές είχαν STMW έχουν διαγνωστεί με την νόσο του θυρεοειδούς (όπου ήταν 0 άτομα).
- Βλέπουμε πως όσοι είχαν ορμόνη τριϊωδοθυρονίνη με τιμή 1,1 και μικρότερη και είχαν μέτρηση την ορμόνη τριϊωδοθυρονίνης και είχαν θυροξίνη ολική με τιμή μεγαλύτερη

από 52 και με φυσιολογικές τιμές είχαν SVHD έχουν διαγνωστεί με την νόσο του θυρεοειδούς (όπου ήταν 3 άτομα).

- Βλέπουμε πως όσοι είχαν ορμόνη τριϊωδοθυρονίνη με τιμή 1,1 και μικρότερη και δεν είχαν μέτρηση την ορμόνη τριϊωδοθυρονίνης και δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς (όπου ήταν 57 άτομα).
- Βλέπουμε πως όσοι είχαν ορμόνη τριϊωδοθυρονίνη με τιμή μεγαλύτερη 1,1 και με δείκτη ελεύθερης θυροξίνης τιμή 172 και μικρότερη και με θυροξίνη ολική με τιμή 56 και μικρότερη και με φυσιολογικές τιμές είχαν SVHC δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς ενώ με φυσιολογικές τιμές είχαν OTHER δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς (όπου ήταν 18 άτομα).
- Βλέπουμε πως όσοι είχαν ορμόνη τριϊωδοθυρονίνη με τιμή μεγαλύτερη 1,1 και με δείκτη ελεύθερης θυροξίνης τιμή 172 και μικρότερη και με θυροξίνη ολική με τιμή 56 και μικρότερη και με φυσιολογικές τιμές είχαν SVI και με ορμόνη διέγερσης του θυρεοειδούς με τιμή 14 και μικρότερη έχουν έχουν διαγνωστεί με την νόσο του θυρεοειδούς (όπου ήταν 7 άτομα) ενώ με ορμόνη διέγερσης του θυρεοειδούς με τιμή μεγαλύτερη από 14 δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς (όπου ήταν 4 άτομα).
- Βλέπουμε πως όσοι είχαν ορμόνη τριϊωδοθυρονίνη με τιμή μεγαλύτερη 1,1 και με δείκτη ελεύθερης θυροξίνης τιμή 172 και μικρότερη και με θυροξίνη ολική με τιμή 56 και μικρότερη και με φυσιολογικές τιμές είχαν STMW δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς (όπου ήταν 2 άτομα) ενώ με φυσιολογικές τιμές είχαν SVHD δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς.
- Βλέπουμε πως όσοι είχαν ορμόνη τριϊωδοθυρονίνη με τιμή μεγαλύτερη 1,1 και με δείκτη ελεύθερης θυροξίνης τιμή 172 και μικρότερη και με θυροξίνη ολική με τιμή μεγαλύτερη 56 δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς (όπου ήταν 2.392 άτομα).
- Βλέπουμε πως όσοι είχαν ορμόνη τριϊωδοθυρονίνη με τιμή μεγαλύτερη 1,1 και με δείκτη ελεύθερης θυροξίνης τιμή μεγαλύτερη 172 και με ορμόνη τριϊωδοθυρονίνη με τιμή 2,6 και μικρότερη και δεν είχαν εξεταστή για θυροξίνη(T4) και είχαν μέτρηση την ορμόνης τριϊωδοθυρονίνης και με ορμόνη διέγερσης του θυρεοειδούς με τιμή 0,31 και μικρότερη έχουν διαγνωστεί με την νόσο του θυρεοειδούς (όπου ήταν 16 άτομα) ενώ με ορμόνη διέγερσης του θυρεοειδούς με τιμή μεγαλύτερη 0,31 δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς (όπου ήταν 7 άτομα).

- Βλέπουμε πως όσοι είχαν ορμόνη τριϊωδοθυρονίνη με τιμή μεγαλύτερη 1,1 και με δείκτη ελεύθερης θυροξίνης τιμή μεγαλύτερη 172 και με ορμόνη τριϊωδοθυρονίνη με τιμή 2,6 και μικρότερη και δεν είχαν εξεταστή για θυροξίνη(T4) και δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς (όπου ήταν 5 άτομα).
- Βλέπουμε πως όσοι είχαν ορμόνη τριϊωδοθυρονίνη με τιμή μεγαλύτερη 1,1 και με δείκτη ελεύθερης θυροξίνης τιμή μεγαλύτερη 172 και με ορμόνη τριϊωδοθυρονίνη με τιμή 2,6 και μικρότερη και είχαν εξεταστή για θυροξίνη(T4) και δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς (όπου ήταν 10 άτομα).
- Βλέπουμε πως όσοι είχαν ορμόνη τριϊωδοθυρονίνη με τιμή μεγαλύτερη 1,1 και με δείκτη ελεύθερης θυροξίνης τιμή μεγαλύτερη 172 και με ορμόνη τριϊωδοθυρονίνη με τιμή μεγαλύτερη 2,6 δεν έχουν διαγνωστεί με την νόσο του θυρεοειδούς (όπου ήταν 62 άτομα).

```

Classifier output
J48 pruned tree
-----
T3 <= 1.1
| T3 measured = t
| | TT4 <= 52
| | | TSH <= 21
| | | | TSH <= 4.1: negative (9.0/1.0)
| | | | TSH > 4.1: sick (3.0)
| | | | TSH > 21: negative (31.0)
| | | TT4 > 52
| | | | referral source = SVHC: sick (5.0)
| | | | referral source = other
| | | | | T4U measured = t
| | | | | | TSH <= 8.6
| | | | | | | age <= 66: negative (11.8)
| | | | | | | age > 66
| | | | | | | | TSH <= 0.13: negative (3.37/0.18)
| | | | | | | | TSH > 0.13: sick (11.23/1.62)
| | | | | | | | TSH > 8.6: sick (6.6/1.4)
| | | | | | | | T4U measured = f: negative (3.0)
| | | | | | | referral source = SVI
| | | | | | | | FTI <= 157
| | | | | | | | | FTI <= 69
| | | | | | | | | T4U <= 0.96: sick (4.16)
| | | | | | | | | T4U > 0.96: negative (2.08/0.08)
| | | | | | | | | FTI > 69: sick (116.44)
| | | | | | | | | FTI > 157: negative (8.32/1.32)
| | | | | | | | referral source = STMW: sick (0.0)
| | | | | | | | referral source = SVHD: sick (3.0)
| | | | | T3 measured = f: negative (57.58/0.2)

```

Εικόνα 115: Αναλυτική ανάλυση του αλγορίθμου J48

Αναλυτική ανάλυση του αλγορίθμου Random Tree

Οι αναλυτικές αναλύσεις του αλγορίθμου Random Tree ισχύουν για τα υποκεφάλαια 5.2.3-5.2.4(σελ. 87~94).

- Βλέπουμε πως όσοι δεν είχαν υποβληθεί σε χειρουργική επέμβαση θυρεοειδούς και είχαν θυροξίνη με τιμή μικρότερη 0,9 και είχαν θυροξίνη ολική με τιμή μικρότερη 87,5 και η ηλικία τους ήταν μικρότερη των 43 χρόνων και είχαν δείκτη ελεύθερης θυροξίνης με τιμή μικρότερη 79 και είχαν ορμόνη τριϊωδοθυρονίνη με τιμή μικρότερη 1,15 και ήταν γυναίκες τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 1 άτομο) ενώ αν ήταν άνδρες και είχαν μέτρηση την ορμόνη διέγερσης του θυρεοειδούς και η ηλικία τους ήταν μικρότερη των 18 χρόνων τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς ενώ αν ήταν 18 ετών και μεγαλύτερη τότε διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 2 άτομα).
- Βλέπουμε πως όσοι δεν είχαν υποβληθεί σε χειρουργική επέμβαση θυρεοειδούς και είχαν θυροξίνη με τιμή μικρότερη 0,9 και είχαν θυροξίνη ολική με τιμή μικρότερη 87,5 και η ηλικία τους ήταν μικρότερη των 43 χρόνων και είχαν δείκτη ελεύθερης θυροξίνης με τιμή μικρότερη 79 και είχαν ορμόνη τριϊωδοθυρονίνη με τιμή μικρότερη 1,15 και ήταν γυναίκες τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 1 άτομο) ενώ αν ήταν άνδρες και δεν είχαν μέτρηση την ορμόνη διέγερσης του θυρεοειδούς τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς.
- Βλέπουμε πως όσοι δεν είχαν υποβληθεί σε χειρουργική επέμβαση θυρεοειδούς και είχαν θυροξίνη με τιμή μικρότερη 0,9 και είχαν θυροξίνη ολική με τιμή μικρότερη 87,5 και η ηλικία τους ήταν μικρότερη των 43 χρόνων και είχαν δείκτη ελεύθερης θυροξίνης με τιμή μικρότερη 79 και είχαν ορμόνη τριϊωδοθυρονίνη με τιμή 1,15 και μεγαλύτερη τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 11 άτομα).
- Βλέπουμε πως όσοι δεν είχαν υποβληθεί σε χειρουργική επέμβαση θυρεοειδούς και είχαν θυροξίνη με τιμή μικρότερη 0,9 και είχαν θυροξίνη ολική με τιμή μικρότερη 87,5 και η ηλικία τους ήταν μικρότερη των 43 χρόνων και είχαν δείκτη ελεύθερης θυροξίνης με τιμή 79 μεγαλύτερη και ήταν γυναίκες και δεν είχαν υποθυρεοειδισμό τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 52 άτομα) ενώ αν είχαν υποθυρεοειδισμό και είχαν θυροξίνη ολική με τιμή μικρότερη από 81 τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 1 άτομο) ενώ αν είχαν θυροξίνη ολική με τιμή 81 και μεγαλύτερη τότε διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 1 άτομο)
- Βλέπουμε πως όσοι δεν είχαν υποβληθεί σε χειρουργική επέμβαση θυρεοειδούς και είχαν θυροξίνη με τιμή μικρότερη 0,9 και είχαν θυροξίνη ολική με τιμή μικρότερη 87,5 και η ηλικία τους ήταν μικρότερη των 43 χρόνων και είχαν δείκτη ελεύθερης θυροξίνης

με τιμή 79 μεγαλύτερη και ήταν άνδρες τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 40 άτομα).

- Βλέπουμε πως όσοι δεν είχαν υποβληθεί σε χειρουργική επέμβαση θυρεοειδούς και είχαν θυροξίνη με τιμή μικρότερη 0,9 και είχαν θυροξίνη ολική με τιμή μικρότερη 87,5 και η ηλικία τους ήταν 43 χρόνων και μικρότερη και είχαν ορμόνη τριϊωδοθυρονίνη με τιμή μικρότερη από 1,15 και δεν είχαν υπερθυρεοειδισμό και είχαν ορμόνη τριϊωδοθυρονίνη με τιμή μικρότερη από 0,55 και είχαν μέτρηση την ορμόνη τριϊωδοθυρονίνης και με φυσιολογικές τιμές είχαν SVHC τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς ενώ αν ήταν με φυσιολογικές τιμές είχαν OTHER και είχαν ορμόνη τριϊωδοθυρονίνη με τιμή μικρότερη από 0,2 τότε διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 1 άτομο) ενώ αν είχαν ορμόνη τριϊωδοθυρονίνη με τιμή 0,2 και μεγαλύτερη και είχαν δείκτη ελεύθερης θυροξίνης με τιμή μικρότερη από 46 τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 1 άτομο) ενώ αν είχαν δείκτη ελεύθερης θυροξίνης με τιμή 46 και μεγαλύτερη και είχαν ορμόνη διέγερσης του θυρεοειδούς με τιμή μικρότερη από 3,15 τότε διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 2 άτομα) ενώ αν είχαν ορμόνη διέγερσης του θυρεοειδούς με τιμή 3,15 και μεγαλύτερη τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς.
- Βλέπουμε πως όσοι δεν είχαν υποβληθεί σε χειρουργική επέμβαση θυρεοειδούς και είχαν θυροξίνη με τιμή μικρότερη 0,9 και είχαν θυροξίνη ολική με τιμή μικρότερη 87,5 και η ηλικία τους ήταν 43 χρόνων και μεγαλύτερη και είχαν ορμόνη τριϊωδοθυρονίνη με τιμή μικρότερη από 1,15 και δεν είχαν υπερθυρεοειδισμό και είχαν ορμόνη τριϊωδοθυρονίνη με τιμή μικρότερη από 0,55 και είχαν μέτρηση την ορμόνη τριϊωδοθυρονίνης και ήταν με φυσιολογικές τιμές είχαν SVI και θυροξίνη ολική με τιμή μικρότερη από 52 και δεν είχαν θυροξίνη(T4) και η ηλικία τους ήταν μεταξύ 43 και μικρότερη από 76 ετών και θυροξίνη με τιμή μικρότερη από 0,55 και με δείκτη ελεύθερης θυροξίνης με τιμή μικρότερη του 66,5 τότε διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 1 άτομο) ενώ αν ο δείκτης ελεύθερης θυροξίνης με τιμή 66,5 και μεγαλύτερη τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 2 άτομα) εάν όμως είχαν θυροξίνη με έχει τιμή 0,55 και μεγαλύτερη τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 6 άτομα).
- Βλέπουμε πως όσοι δεν είχαν υποβληθεί σε χειρουργική επέμβαση θυρεοειδούς και είχαν θυροξίνη με τιμή μικρότερη 0,9 και είχαν θυροξίνη ολική με τιμή μικρότερη 87,5

και η ηλικία τους ήταν 43 χρόνων και μεγαλύτερη και είχαν ορμόνη τριϊωδοθυρονίνη με τιμή μικρότερη από 1,15 και δεν είχαν υπερθυρεοειδισμό και είχαν ορμόνη τριϊωδοθυρονίνη με τιμή μικρότερη από 0,55 και είχαν μέτρηση την ορμόνη τριϊωδοθυρονίνης και ήταν με φυσιολογικές τιμές είχαν SVI και θυροξίνη ολική με τιμή μικρότερη από 52 και δεν είχαν θυροξίνη(T4) και η ηλικία τους ήταν 76 ετών και μεγαλύτερη και με ορμόνη διέγερσης του θυρεοειδούς με τιμή μικρότερη από 2,58 τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 1 άτομο) εάν η ορμόνη διέγερσης του θυρεοειδούς έχει τιμή 2,58 και μεγαλύτερη τότε διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 2 άτομα).

- Βλέπουμε πως όσοι δεν είχαν υποβληθεί σε χειρουργική επέμβαση θυρεοειδούς και είχαν θυροξίνη με τιμή μικρότερη 0,9 και είχαν θυροξίνη ολική με τιμή μικρότερη 87,5 και η ηλικία τους ήταν 43 χρόνων και μεγαλύτερη και είχαν ορμόνη τριϊωδοθυρονίνη με τιμή μικρότερη από 1,15 και δεν είχαν υπερθυρεοειδισμό και είχαν ορμόνη τριϊωδοθυρονίνη με τιμή μικρότερη από 0,55 και είχαν μέτρηση την ορμόνη τριϊωδοθυρονίνης και ήταν με φυσιολογικές τιμές και είχαν SVI και θυροξίνη ολική με τιμή μικρότερη από 52 και είχαν θυροξίνη(T4) τότε διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 1 άτομο).
- Βλέπουμε πως όσοι δεν είχαν υποβληθεί σε χειρουργική επέμβαση θυρεοειδούς και είχαν θυροξίνη με τιμή μικρότερη 0,9 και είχαν θυροξίνη ολική με τιμή μικρότερη 87,5 και η ηλικία τους ήταν 43 χρόνων και μεγαλύτερη και είχαν ορμόνη τριϊωδοθυρονίνη με τιμή μικρότερη από 1,15 και δεν είχαν υπερθυρεοειδισμό και είχαν ορμόνη τριϊωδοθυρονίνη με τιμή μικρότερη από 0,55 και είχαν μέτρηση την ορμόνη τριϊωδοθυρονίνης και ήταν με φυσιολογικές τιμές και είχαν STMW τότε διαγνωστήκαν με την νόσο του θυρεοειδούς ενώ αν με φυσιολογικές τιμές και είχαν SVHD τότε διαγνωστήκαν με την νόσο του θυρεοειδούς.
- Βλέπουμε πως όσοι δεν είχαν υποβληθεί σε χειρουργική επέμβαση θυρεοειδούς και είχαν θυροξίνη με τιμή μικρότερη 0,9 και είχαν θυροξίνη ολική με τιμή μικρότερη 87,5 και η ηλικία τους ήταν 43 χρόνων και μεγαλύτερη και είχαν ορμόνη τριϊωδοθυρονίνη με τιμή μικρότερη από 1,15 και δεν είχαν υπερθυρεοειδισμό και είχαν ορμόνη τριϊωδοθυρονίνη με τιμή μικρότερη από 0,55 και δεν είχαν μέτρηση την ορμόνη τριϊωδοθυρονίνης και η ηλικία τους ήταν μικρότερη των 53 χρόνων τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 1 άτομο) ενώ αν η ηλικία

τους ήταν 53 χρόνων και μεγαλύτερη τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 5 άτομα).

- Βλέπουμε πως όσοι δεν είχαν υποβληθεί σε χειρουργική επέμβαση θυρεοειδούς και είχαν θυροξίνη με τιμή μικρότερη 0,9 και είχαν θυροξίνη ολική με τιμή μικρότερη 87,5 και η ηλικία τους ήταν 43 χρόνων και μεγαλύτερη και είχαν ορμόνη τριϊωδοθυρονίνη με τιμή μικρότερη από 1,15 και δεν είχαν υπερθυρεοειδισμό και είχαν ορμόνη τριϊωδοθυρονίνη με τιμή 0,55 και μεγαλύτερη και η ηλικία τους ήταν από 43 και μικρότερη από 66 ετών και είχαν μέτρηση το δείκτης ελεύθερης θυροξίνης και είχαν θυροξίνη ολική με τιμή μικρότερη από 50 και είχαν με φυσιολογικές τιμές και είχαν SVHC τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς ενώ αν ήταν με φυσιολογικές τιμές και είχαν OTHER τότε διαγνωστήκαν με την νόσο του θυρεοειδούς ενώ αν ήταν με φυσιολογικές τιμές και είχαν SVI τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 3 άτομα) ενώ αν ήταν με φυσιολογικές τιμές και είχαν STMW τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς ενώ αν ήταν με φυσιολογικές τιμές και είχαν SVHD τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς.
- Βλέπουμε πως όσοι δεν είχαν υποβληθεί σε χειρουργική επέμβαση θυρεοειδούς και είχαν θυροξίνη με τιμή μικρότερη 0,9 και είχαν θυροξίνη ολική με τιμή μικρότερη 87,5 και η ηλικία τους ήταν 43 χρόνων και μεγαλύτερη και είχαν ορμόνη τριϊωδοθυρονίνη με τιμή μικρότερη από 1,15 και δεν είχαν υπερθυρεοειδισμό και είχαν ορμόνη τριϊωδοθυρονίνη με τιμή 0,55 και μεγαλύτερη και η ηλικία τους ήταν από 43 και μικρότερη από 66 ετών και είχαν μέτρηση το δείκτης ελεύθερης θυροξίνης και είχαν θυροξίνη ολική με τιμή 50 και μεγαλύτερη και ήταν με φυσιολογικές τιμές και είχαν SVHC και η ηλικία τους είναι μεταξύ 43 και μικρότερη από 47 ετών τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς ενώ αν η ηλικία τους είναι 47 ετών και μεγαλύτερη τότε διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 2 άτομα).
- Βλέπουμε πως όσοι δεν είχαν υποβληθεί σε χειρουργική επέμβαση θυρεοειδούς και είχαν θυροξίνη με τιμή μικρότερη 0,9 και είχαν θυροξίνη ολική με τιμή μικρότερη 87,5 και η ηλικία τους ήταν 43 χρόνων και μεγαλύτερη και είχαν ορμόνη τριϊωδοθυρονίνη με τιμή μικρότερη από 1,15 και δεν είχαν υπερθυρεοειδισμό και είχαν ορμόνη τριϊωδοθυρονίνη με τιμή 0,55 και μεγαλύτερη και η ηλικία τους ήταν από 43 και μικρότερη από 66 ετών και είχαν μέτρηση το δείκτης ελεύθερης θυροξίνης και είχαν θυροξίνη ολική με τιμή 50 και μεγαλύτερη και ήταν με φυσιολογικές τιμές και είχαν

OTHER τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 7 άτομα) ενώ αν ήταν με φυσιολογικές τιμές και είχαν SVI και είχαν δείκτη ελεύθερης θυροξίνης με τιμή μικρότερη από 115 τότε διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 14 άτομα) ενώ αν είχαν δείκτη ελεύθερης θυροξίνης με τιμή 115 και μεγαλύτερη και αν είχαν μέτρηση την ορμόνη τριϊωδοθυρονίνης τότε διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 2 άτομα) ένα αν δεν είχαν μέτρηση την ορμόνη τριϊωδοθυρονίνης τότε διαγνωστήκαν με την νόσο του θυρεοειδούς.

- Βλέπουμε πως όσοι δεν είχαν υποβληθεί σε χειρουργική επέμβαση θυρεοειδούς και είχαν θυροξίνη με τιμή μικρότερη 0,9 και είχαν θυροξίνη ολική με τιμή μικρότερη 87,5 και η ηλικία τους ήταν 43 χρόνων και μεγαλύτερη και είχαν ορμόνη τριϊωδοθυρονίνη με τιμή μικρότερη από 1,15 και δεν είχαν υπερθυρεοειδισμό και είχαν ορμόνη τριϊωδοθυρονίνη με τιμή 0,55 και μεγαλύτερη και η ηλικία τους ήταν από 43 και μικρότερη από 66 ετών και είχαν μέτρηση το δείκτης ελεύθερης θυροξίνης και είχαν θυροξίνη ολική με τιμή 50 και μεγαλύτερη και ήταν με φυσιολογικές τιμές και είχαν STMW τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς ενώ αν ήταν με φυσιολογικές τιμές και είχαν SVHD τότε διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 1 άτομο).
- Βλέπουμε πως όσοι δεν είχαν υποβληθεί σε χειρουργική επέμβαση θυρεοειδούς και είχαν θυροξίνη με τιμή μικρότερη 0,9 και είχαν θυροξίνη ολική με τιμή μικρότερη 87,5 και η ηλικία τους ήταν 43 χρόνων και μεγαλύτερη και είχαν ορμόνη τριϊωδοθυρονίνη με τιμή μικρότερη από 1,15 και δεν είχαν υπερθυρεοειδισμό και είχαν ορμόνη τριϊωδοθυρονίνη με τιμή 0,55 και μεγαλύτερη και η ηλικία τους ήταν από 43 και μικρότερη από 66 ετών και δεν είχαν μέτρηση το δείκτης ελεύθερης θυροξίνης τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 2 άτομα).
- Βλέπουμε πως όσοι δεν είχαν υποβληθεί σε χειρουργική επέμβαση θυρεοειδούς και είχαν θυροξίνη με τιμή μικρότερη 0,9 και είχαν θυροξίνη ολική με τιμή μικρότερη 87,5 και η ηλικία τους ήταν 43 χρόνων και μεγαλύτερη και είχαν ορμόνη τριϊωδοθυρονίνη με τιμή μικρότερη από 1,15 και δεν είχαν υπερθυρεοειδισμό και είχαν ορμόνη τριϊωδοθυρονίνη με τιμή 0,55 και μεγαλύτερη και η ηλικία τους ήταν 66 ετών και μεγαλύτερη και ήταν με φυσιολογικές τιμές και είχαν SVHC τότε διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 1 άτομο) ενώ αν ήταν με φυσιολογικές τιμές και είχαν OTHER και είχαν ορμόνη διέγερσης του θυρεοειδούς με τιμή μικρότερη 1,55 και δεν ήταν άρρωστοι και η ηλικία τους ήταν μεταξύ 43 και μικρότερη από 71 χρόνων

τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς ενώ αν η ηλικία τους ήταν 71 χρόνων και μεγαλύτερη και είχαν μέτρηση θυροξίνης ολικής τότε διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 2 άτομα) ενώ αν δεν είχαν μέτρηση θυροξίνης ολικής τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς. Όμως εάν ήταν άρρωστοι τότε διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 1 άτομο).

- Βλέπουμε πως όσοι δεν είχαν υποβληθεί σε χειρουργική επέμβαση θυρεοειδούς και είχαν θυροξίνη με τιμή μικρότερη 0,9 και είχαν θυροξίνη ολική με τιμή μικρότερη 87,5 και η ηλικία τους ήταν 43 χρόνων και μεγαλύτερη και είχαν ορμόνη τριϊωδοθυρονίνη με τιμή μικρότερη από 1,15 και δεν είχαν υπερθυρεοειδισμό και είχαν ορμόνη τριϊωδοθυρονίνη με τιμή 0,55 και μεγαλύτερη και η ηλικία τους ήταν 66 ετών και μεγαλύτερη και ήταν με φυσιολογικές τιμές και είχαν OTHER και δεν είχαν ορμόνη διέγερσης του θυρεοειδούς τότε διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 1 άτομο).
- Βλέπουμε πως όσοι δεν είχαν υποβληθεί σε χειρουργική επέμβαση θυρεοειδούς και είχαν θυροξίνη με τιμή μικρότερη 0,9 και είχαν θυροξίνη ολική με τιμή μικρότερη 87,5 και η ηλικία τους ήταν 43 χρόνων και μεγαλύτερη και είχαν ορμόνη τριϊωδοθυρονίνη με τιμή μικρότερη από 1,15 και δεν είχαν υπερθυρεοειδισμό και είχαν ορμόνη τριϊωδοθυρονίνη με τιμή 0,55 και μεγαλύτερη και η ηλικία τους ήταν 66 ετών και μεγαλύτερη και ήταν με φυσιολογικές τιμές και είχαν OTHER και είχαν ορμόνη διέγερσης του θυρεοειδούς με τιμή 1,55 και μεγαλύτερη τότε διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 2 άτομα).
- Βλέπουμε πως όσοι δεν είχαν υποβληθεί σε χειρουργική επέμβαση θυρεοειδούς και είχαν θυροξίνη με τιμή μικρότερη 0,9 και είχαν θυροξίνη ολική με τιμή μικρότερη 87,5 και η ηλικία τους ήταν 43 χρόνων και μεγαλύτερη και είχαν ορμόνη τριϊωδοθυρονίνη με τιμή μικρότερη από 1,15 και δεν είχαν υπερθυρεοειδισμό και είχαν ορμόνη τριϊωδοθυρονίνη με τιμή 0,55 και μεγαλύτερη και η ηλικία τους ήταν 66 ετών και μεγαλύτερη και ήταν με φυσιολογικές τιμές και είχαν SVI και είχαν μέτρηση ορμόνης τριϊωδοθυρονίνης τότε διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 21 άτομα) ενώ αν δεν είχαν μέτρηση ορμόνης τριϊωδοθυρονίνης τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς.
- Βλέπουμε πως όσοι δεν είχαν υποβληθεί σε χειρουργική επέμβαση θυρεοειδούς και είχαν θυροξίνη με τιμή μικρότερη 0,9 και είχαν θυροξίνη ολική με τιμή μικρότερη 87,5 και η ηλικία τους ήταν 43 χρόνων και μεγαλύτερη και είχαν ορμόνη τριϊωδοθυρονίνη

με τιμή μικρότερη από 1,15 και δεν είχαν υπερθυρεοειδισμό και είχαν ορμόνη τριϊωδοθυρονίνη με τιμή 0,55 και μεγαλύτερη και η ηλικία τους ήταν 66 ετών και μεγαλύτερη και ήταν με φυσιολογικές τιμές και είχαν STMW τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς ενώ αν ήταν με φυσιολογικές τιμές και είχαν SVHD τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς.

- Βλέπουμε πως όσοι δεν είχαν υποβληθεί σε χειρουργική επέμβαση θυρεοειδούς και είχαν θυροξίνη με τιμή μικρότερη 0,9 και είχαν θυροξίνη ολική με τιμή μικρότερη 87,5 και η ηλικία τους ήταν 43 χρόνων και μεγαλύτερη και είχαν ορμόνη τριϊωδοθυρονίνη με τιμή μικρότερη από 1,15 και είχαν υπερθυρεοειδισμό και δεν είχαν θυροξίνη(T4) τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 1 άτομο) ενώ αν είχαν θυροξίνη(T4) τότε διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 1 άτομο).
- Βλέπουμε πως όσοι δεν είχαν υποβληθεί σε χειρουργική επέμβαση θυρεοειδούς και είχαν θυροξίνη με τιμή μικρότερη 0,9 και είχαν θυροξίνη ολική με τιμή μικρότερη 87,5 και η ηλικία τους ήταν 43 χρόνων και μεγαλύτερη και είχαν ορμόνη τριϊωδοθυρονίνη με τιμή 1,15 και μεγαλύτερη και είχαν θυροξίνη ολική τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 136 άτομα).
- Βλέπουμε πως όσοι δεν είχαν υποβληθεί σε χειρουργική επέμβαση θυρεοειδούς και είχαν θυροξίνη με τιμή μικρότερη 0,9 και είχαν θυροξίνη ολική με τιμή 87,5 και μεγαλύτερη και είχαν δείκτη ελεύθερης θυροξίνης με τιμή μικρότερη από 121,5 και είχαν θυροξίνη ολική και δεν είχαν διαγνωστεί με κάποια ψυχολογική διαταραχή και είχαν ορμόνη διέγερσης του θυρεοειδούς με τιμή μικρότερη από 10,5 και είχαν δείκτη ελεύθερης θυροξίνης μικρότερη από 105,5 και δεν ήταν αρμοστοί κατά την διάρκεια της εξετάσεις και είχαν θυροξίνη ολική τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 20 άτομα) εάν όμως ήταν αρμοστοί κατά την διάρκεια της εξετάσεις τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 3 άτομα).
- Βλέπουμε πως όσοι δεν είχαν υποβληθεί σε χειρουργική επέμβαση θυρεοειδούς και είχαν θυροξίνη με τιμή μικρότερη 0,9 και είχαν θυροξίνη ολική με τιμή 87,5 και μεγαλύτερη και είχαν δείκτη ελεύθερης θυροξίνης με τιμή μικρότερη από 121,5 και είχαν θυροξίνη ολική και δεν είχαν διαγνωστεί με κάποια ψυχολογική διαταραχή και είχαν ορμόνη διέγερσης του θυρεοειδούς με τιμή μικρότερη από 10,5 και είχαν δείκτη ελεύθερης θυροξίνης 105,5 και μεγαλύτερη και είχαν θυροξίνη ολική με τιμή μεταξύ 87,5 και μικρότερη από 89,5 τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (

όπου ήταν 20 άτομα) ενώ αν είχαν θυροξίνη ολική με τιμή μεταξύ 89,5 και μικρότερη από 91,5 και είχαν ορμόνη τριϊωδοθυρονίνη με τιμή μικρότερη από 1 τότε διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 4 άτομα) ενώ αν δεν είχαν ορμόνη τριϊωδοθυρονίνη τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς ενώ αν είχαν ορμόνη τριϊωδοθυρονίνη με 1 και μεγαλύτερη τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 10 άτομα).

- Βλέπουμε πως όσοι δεν είχαν υποβληθεί σε χειρουργική επέμβαση θυρεοειδούς και είχαν θυροξίνη με τιμή μικρότερη 0,9 και είχαν θυροξίνη ολική με τιμή 87,5 και μεγαλύτερη και είχαν δείκτη ελεύθερης θυροξίνης με τιμή μικρότερη από 121,5 και είχαν θυροξίνη ολική και δεν είχαν διαγνωστεί με κάποια ψυχολογική διαταραχή και είχαν ορμόνη διέγερσης του θυρεοειδούς με τιμή μικρότερη από 10,5 και είχαν δείκτη ελεύθερης θυροξίνης 105,5 και μεγαλύτερη και είχαν θυροξίνη ολική με τιμή 91,5 και μεγαλύτερη και είχαν ορμόνη διέγερσης του θυρεοειδούς με τιμή μικρότερη από 1,45 και δεν είχαν εξεταστή για θυροξίνη(T4) και είχαν ορμόνη τριϊωδοθυρονίνης με τιμή μικρότερη από 1,05 και η ηλικία τους ήταν μικρότερη από 76 χρόνων τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς ενώ αν η ηλικία τους ήταν 76 χρόνων και μεγαλύτερη και ήταν γυναίκες τότε διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 1 μια γυναίκα) ενώ αν ήταν άνδρες τότε διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 1 ένας άνδρας) εάν όμως είχαν ορμόνη τριϊωδοθυρονίνης με τιμή 1,05 και μεγαλύτερη τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 48 άτομα).
- Βλέπουμε πως όσοι δεν είχαν υποβληθεί σε χειρουργική επέμβαση θυρεοειδούς και είχαν θυροξίνη με τιμή μικρότερη 0,9 και είχαν θυροξίνη ολική με τιμή 87,5 και μεγαλύτερη και είχαν δείκτη ελεύθερης θυροξίνης με τιμή μικρότερη από 121,5 και είχαν θυροξίνη ολική και δεν είχαν διαγνωστεί με κάποια ψυχολογική διαταραχή και είχαν ορμόνη διέγερσης του θυρεοειδούς με τιμή μικρότερη από 10,5 και είχαν δείκτη ελεύθερης θυροξίνης 105,5 και μεγαλύτερη και είχαν θυροξίνη ολική με τιμή 91,5 και μεγαλύτερη και είχαν ορμόνη διέγερσης του θυρεοειδούς με τιμή μικρότερη από 1,45 και είχαν εξεταστή για θυροξίνη(T4) τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 3 άτομα).
- Βλέπουμε πως όσοι δεν είχαν υποβληθεί σε χειρουργική επέμβαση θυρεοειδούς και είχαν θυροξίνη με τιμή μικρότερη 0,9 και είχαν θυροξίνη ολική με τιμή 87,5 και μεγαλύτερη και είχαν δείκτη ελεύθερης θυροξίνης με τιμή μικρότερη από 121,5 και

είχαν θυροξίνη ολική και δεν είχαν διαγνωστεί με κάποια ψυχολογική διαταραχή και είχαν ορμόνη διέγερσης του θυρεοειδούς με τιμή μικρότερη από 10,5 και είχαν δείκτη ελεύθερης θυροξίνης 105,5 και μεγαλύτερη και είχαν θυροξίνη ολική με τιμή 91,5 και μεγαλύτερη και είχαν ορμόνη διέγερσης του θυρεοειδούς με τιμή από 1,45 και μεγαλύτερη και δεν έπαιρναν αντιβιοτικά για αντιμετώπιση του θυρεοειδή και είχαν μέτρηση τον δείκτη ελεύθερης θυροξίνης τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 42 άτομα) ενώ αν δεν είχαν μέτρηση τον δείκτη ελεύθερης θυροξίνης και ασθενείς όπου δεν είχαν ερωτηθεί αν είχαν υποθυρεοειδισμό και είχαν μέτρηση η ορμόνη διέγερσης του θυρεοειδούς και ήταν με φυσιολογικές τιμές και είχαν SVHC τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς ενώ αν ήταν με φυσιολογικές τιμές και είχαν OTHER τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς ενώ αν ήταν με φυσιολογικές τιμές και είχαν SVI τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς ενώ αν ήταν με φυσιολογικές τιμές και είχαν STMW τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς ενώ αν ήταν με φυσιολογικές τιμές και είχαν SVHD τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς εάν όμως δεν είχαν μέτρηση η ορμόνη διέγερσης του θυρεοειδούς τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς.

- Βλέπουμε πως όσοι δεν είχαν υποβληθεί σε χειρουργική επέμβαση θυρεοειδούς και είχαν θυροξίνη με τιμή μικρότερη 0,9 και είχαν θυροξίνη ολική με τιμή 87,5 και μεγαλύτερη και είχαν δείκτη ελεύθερης θυροξίνης με τιμή μικρότερη από 121,5 και είχαν θυροξίνη ολική και δεν είχαν διαγνωστεί με κάποια ψυχολογική διαταραχή και είχαν ορμόνη διέγερσης του θυρεοειδούς με τιμή μικρότερη από 10,5 και είχαν δείκτη ελεύθερης θυροξίνης 105,5 και μεγαλύτερη και είχαν θυροξίνη ολική με τιμή 91,5 και μεγαλύτερη και είχαν ορμόνη διέγερσης του θυρεοειδούς με τιμή από 1,45 και μεγαλύτερη και δεν έπαιρναν αντιβιοτικά για αντιμετώπιση του θυρεοειδή και δεν είχαν μέτρηση τον δείκτη ελεύθερης θυροξίνης και ασθενείς όπου είχαν ερωτηθεί αν είχαν υποθυρεοειδισμό τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς.
- Βλέπουμε πως όσοι δεν είχαν υποβληθεί σε χειρουργική επέμβαση θυρεοειδούς και είχαν θυροξίνη με τιμή μικρότερη 0,9 και είχαν θυροξίνη ολική με τιμή 87,5 και μεγαλύτερη και είχαν δείκτη ελεύθερης θυροξίνης με τιμή μικρότερη από 121,5 και είχαν θυροξίνη ολική και δεν είχαν διαγνωστεί με κάποια ψυχολογική διαταραχή και είχαν ορμόνη διέγερσης του θυρεοειδούς με τιμή μικρότερη από 10,5 και είχαν δείκτη ελεύθερης θυροξίνης 105,5 και μεγαλύτερη και είχαν θυροξίνη ολική με τιμή 91,5 και

μεγαλύτερη και είχαν ορμόνη διέγερσης του θυρεοειδούς με τιμή από 1,45 και μεγαλύτερη και έπαιρναν αντιβιοτικά για αντιμετώπιση του θυρεοειδή τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς.

- Βλέπουμε πως όσοι δεν είχαν υποβληθεί σε χειρουργική επέμβαση θυρεοειδούς και είχαν θυροξίνη με τιμή μικρότερη 0,9 και είχαν θυροξίνη ολική με τιμή 87,5 και μεγαλύτερη και είχαν δείκτη ελεύθερης θυροξίνης με τιμή μικρότερη από 121,5 και είχαν θυροξίνη ολική και δεν είχαν διαγνωστεί με κάποια ψυχολογική διαταραχή και είχαν ορμόνη διέγερσης του θυρεοειδούς με τιμή 10,5 και μεγαλύτερη και η ηλικία τους ήταν μικρότερη των 54 χρόνων τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 1 άτομο) ενώ αν η ηλικία τους ήταν 54 χρόνων και μεγαλύτερη τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 1 άτομο).
- Βλέπουμε πως όσοι δεν είχαν υποβληθεί σε χειρουργική επέμβαση θυρεοειδούς και είχαν θυροξίνη με τιμή μικρότερη 0,9 και είχαν θυροξίνη ολική με τιμή 87,5 και μεγαλύτερη και είχαν δείκτη ελεύθερης θυροξίνης με τιμή μικρότερη από 121,5 και είχαν θυροξίνη ολική και είχαν διαγνωστεί με κάποια ψυχολογική διαταραχή τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 13 άτομα).
- Βλέπουμε πως όσοι δεν είχαν υποβληθεί σε χειρουργική επέμβαση θυρεοειδούς και είχαν θυροξίνη με τιμή μικρότερη 0,9 και είχαν θυροξίνη ολική με τιμή 87,5 και μεγαλύτερη και είχαν δείκτη ελεύθερης θυροξίνης με τιμή μικρότερη από 121,5 και δεν είχαν θυροξίνη ολική τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 11 άτομα).
- Βλέπουμε πως όσοι δεν είχαν υποβληθεί σε χειρουργική επέμβαση θυρεοειδούς και είχαν θυροξίνη με τιμή μικρότερη 0,9 και είχαν θυροξίνη ολική με τιμή 87,5 και μεγαλύτερη και είχαν δείκτη ελεύθερης θυροξίνης με τιμή από 121,5 και μεγαλύτερη και δεν είχαν εξέταση για θυροξίνη(T4) και είχαν μέτρηση ορμόνης τριϊωδοθυρονίνης και η ηλικία τους ήταν μικρότερη από 40 χρόνων και είχαν ορμόνη διέγερσης του θυρεοειδούς με τιμή μικρότερη 0,01 και είχαν θυροξίνη ολική με τιμή μικρότερη από 154,5 τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 1 άτομο) ενώ αν είχαν θυροξίνη ολική με τιμή μεταξύ 154,5 και μικρότερη από 182,5 τότε διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 1 άτομο) ενώ αν είχαν θυροξίνη ολική με τιμή 182,5 και μεγαλύτερη τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 1 άτομο).

- Βλέπουμε πως όσοι δεν είχαν υποβληθεί σε χειρουργική επέμβαση θυρεοειδούς και είχαν θυροξίνη με τιμή μικρότερη 0,9 και είχαν θυροξίνη ολική με τιμή 87,5 και μεγαλύτερη και είχαν δείκτη ελεύθερης θυροξίνης με τιμή από 121,5 και μεγαλύτερη και δεν είχαν εξέταση για θυροξίνη(T4) και είχαν μέτρηση ορμόνης τριϊωδοθυρονίνης και η ηλικία τους ήταν μικρότερη από 40 χρόνων και είχαν ορμόνη διέγερσης του θυρεοειδούς με τιμή 0,01 και μεγαλύτερη και είχαν ορμόνη τριϊωδοθυρονίνης με τιμή μικρότερη από 1,05 τότε διαγνωστήκαν με την νόσο του θυρεοειδούς ενώ αν είχαν ορμόνη τριϊωδοθυρονίνης με τιμή 1,05 και μεγαλύτερη τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 45 άτομα).
- Βλέπουμε πως όσοι δεν είχαν υποβληθεί σε χειρουργική επέμβαση θυρεοειδούς και είχαν θυροξίνη με τιμή μικρότερη 0,9 και είχαν θυροξίνη ολική με τιμή 87,5 και μεγαλύτερη και είχαν δείκτη ελεύθερης θυροξίνης με τιμή από 121,5 και μεγαλύτερη και δεν είχαν εξέταση για θυροξίνη(T4) και είχαν μέτρηση ορμόνης τριϊωδοθυρονίνης και η ηλικία τους ήταν 40 χρόνων και μεγαλύτερη και είχαν θυροξίνη με τιμή μικρότερη 0,79 και είχαν δείκτη ελεύθερης θυροξίνης με τιμή μικρότερη 164,5 και είχαν ορμόνη τριϊωδοθυρονίνη με τιμή μικρότερη 1,15 και είχαν δείκτη ελεύθερης θυροξίνης με τιμή μικρότερη από 128 τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 2 άτομα) ενώ αν δεν είχαν δείκτη ελεύθερης θυροξίνης τότε διαγνωστήκαν με την νόσο του θυρεοειδούς.
- Βλέπουμε πως όσοι δεν είχαν υποβληθεί σε χειρουργική επέμβαση θυρεοειδούς και είχαν θυροξίνη με τιμή μικρότερη 0,9 και είχαν θυροξίνη ολική με τιμή 87,5 και μεγαλύτερη και είχαν δείκτη ελεύθερης θυροξίνης με τιμή από 121,5 και μεγαλύτερη και δεν είχαν εξέταση για θυροξίνη(T4) και είχαν μέτρηση ορμόνης τριϊωδοθυρονίνης και η ηλικία τους ήταν 40 χρόνων και μεγαλύτερη και είχαν θυροξίνη με τιμή μικρότερη 0,79 και είχαν δείκτη ελεύθερης θυροξίνης με τιμή μικρότερη 164,5 και είχαν ορμόνη τριϊωδοθυρονίνη με τιμή μικρότερη 1,15 και είχαν δείκτη ελεύθερης θυροξίνης με τιμή από 128 και μεγαλύτερη και είχαν θυροξίνη ολική με τιμή μικρότερη από 94 και είχαν ορμόνη τριϊωδοθυρονίνης με τιμή με τιμή μικρότερη 0,85 τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 1 άτομο) ενώ αν είχαν ορμόνη τριϊωδοθυρονίνης με τιμή 0,85 και μεγαλύτερη τότε διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 2 άτομα) εάν όμως είχαν θυροξίνη ολική με τιμή 94 και μεγαλύτερη τότε διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 6 άτομα).

- Βλέπουμε πως όσοι δεν είχαν υποβληθεί σε χειρουργική επέμβαση θυρεοειδούς και είχαν θυροξίνη με τιμή μικρότερη 0,9 και είχαν θυροξίνη ολική με τιμή 87,5 και μεγαλύτερη και είχαν δείκτη ελεύθερης θυροξίνης με τιμή από 121,5 και μεγαλύτερη και δεν είχαν εξέταση για θυροξίνη(T4) και είχαν μέτρηση ορμόνης τριϊωδοθυρονίνης και η ηλικία τους ήταν 40 χρόνων και μεγαλύτερη και είχαν θυροξίνη με τιμή μικρότερη 0,79 και είχαν δείκτη ελεύθερης θυροξίνης με τιμή μικρότερη 164,5 και είχαν ορμόνη τριϊωδοθυρονίνη με τιμή 1,15 και μεγαλύτερη τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 25 άτομα).
- Βλέπουμε πως όσοι δεν είχαν υποβληθεί σε χειρουργική επέμβαση θυρεοειδούς και είχαν θυροξίνη με τιμή μικρότερη 0,9 και είχαν θυροξίνη ολική με τιμή 87,5 και μεγαλύτερη και είχαν δείκτη ελεύθερης θυροξίνης με τιμή από 121,5 και μεγαλύτερη και δεν είχαν εξέταση για θυροξίνη(T4) και είχαν μέτρηση ορμόνης τριϊωδοθυρονίνης και η ηλικία τους ήταν 40 χρόνων και μεγαλύτερη και είχαν θυροξίνη με τιμή μικρότερη 0,79 και είχαν δείκτη ελεύθερης θυροξίνης με τιμή 164,5 και μεγαλύτερη και ήταν με φυσιολογικές τιμές και είχαν SVHC τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς ενώ αν και ήταν με φυσιολογικές τιμές και είχαν OTHER τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 7 άτομα) ενώ αν ήταν με φυσιολογικές τιμές και είχαν SVI και η ηλικία τους ήταν μεταξύ 40 και μικρότερη από 83 χρόνων και είχαν μέτρηση τον δείκτη ελεύθερης θυροξίνης τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 4 άτομα) ενώ αν δεν είχαν μέτρηση τον δείκτη ελεύθερης θυροξίνης τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς. Εάν όμως η ηλικία τους ήταν 83 χρόνων και μεγαλύτερη τότε διαγνωστήκαν με την νόσο του θυρεοειδούς. Εάν όμως ήταν με φυσιολογικές τιμές και είχαν STMW τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς ενώ αν ήταν με φυσιολογικές τιμές και είχαν SVHD τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς.
- Βλέπουμε πως όσοι δεν είχαν υποβληθεί σε χειρουργική επέμβαση θυρεοειδούς και είχαν θυροξίνη με τιμή μικρότερη 0,9 και είχαν θυροξίνη ολική με τιμή 87,5 και μεγαλύτερη και είχαν δείκτη ελεύθερης θυροξίνης με τιμή από 121,5 και μεγαλύτερη και δεν είχαν εξέταση για θυροξίνη(T4) και είχαν μέτρηση ορμόνης τριϊωδοθυρονίνης και η ηλικία τους ήταν 40 χρόνων και μεγαλύτερη και είχαν θυροξίνη με τιμή 0,79 και μεγαλύτερη και είχαν ορμόνη τριϊωδοθυρονίνης με τιμή μικρότερη από 1,15 και η ηλικία τους ήταν μεταξύ 40 και μικρότερη από 76 χρόνων και είχαν δείκτη ελεύθερης θυροξίνης με τιμή μικρότερη από 156,5 τότε διαγνωστήκαν με την νόσο του

θυρεοειδούς (όπου ήταν 15 άτομα) ενώ αν είχαν δείκτη ελεύθερης θυροξίνης με τιμή 156,5 και μεγαλύτερη και είχαν ορμόνη τριϊωδοθυρονίνης με τιμή μικρότερη από 0,7 τότε διαγνωστήκαν με την νόσο του θυρεοειδούς ενώ αν είχαν ορμόνη τριϊωδοθυρονίνης με τιμή 0,7 και μεγαλύτερη και ήταν γυναίκες τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 2 άτομα) ενώ αν ήταν άνδρες τότε διαγνωστήκαν με την νόσο του θυρεοειδούς. Εάν όμως η ηλικία τους ήταν 76 χρόνων και μεγαλύτερη τότε διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 2 άτομα).

- Βλέπουμε πως όσοι δεν είχαν υποβληθεί σε χειρουργική επέμβαση θυρεοειδούς και είχαν θυροξίνη με τιμή μικρότερη 0,9 και είχαν θυροξίνη ολική με τιμή 87,5 και μεγαλύτερη και είχαν δείκτη ελεύθερης θυροξίνης με τιμή από 121,5 και μεγαλύτερη και δεν είχαν εξέταση για θυροξίνη(T4) και είχαν μέτρηση ορμόνης τριϊωδοθυρονίνης και η ηλικία τους ήταν 40 χρόνων και μεγαλύτερη και είχαν θυροξίνη με τιμή 0,79 και μεγαλύτερη και είχαν ορμόνη τριϊωδοθυρονίνης με τιμή 1,15 και μεγαλύτερη και είχαν δείκτη ελεύθερης θυροξίνης με τιμή μικρότερη από 175,5 τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 70 άτομα) ενώ αν είχαν δείκτη ελεύθερης θυροξίνης με τιμή 175,5 και μεγαλύτερη και ήταν με φυσιολογικές τιμές και είχαν SVHC τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς ενώ αν ήταν με φυσιολογικές τιμές και είχαν OTHER και είχαν θυροξίνη ολική με τιμή μικρότερη από 191 και είχαν ορμόνη τριϊωδοθυρονίνης με τιμή μικρότερη από 2,65 και ασθενείς όπου δεν είχαν ερωτηθεί αν είχαν υπερθυρεοειδισμό και είχαν ορμόνη τριϊωδοθυρονίνης με τιμή μικρότερη από 1,75 τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς ενώ αν είχαν ορμόνη τριϊωδοθυρονίνης με τιμή μεταξύ 1,75 μικρότερη από 2,65 και ήταν γυναίκες και είχαν μέτρηση τον δείκτη ελεύθερης θυροξίνης τότε διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 4 άτομα) ενώ αν δεν είχαν μέτρηση τον δείκτη ελεύθερης θυροξίνης τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς εάν όμως ήταν άνδρες τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς εάν όμως ήταν ασθενείς όπου είχαν ερωτηθεί αν είχαν υπερθυρεοειδισμό τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς εάν όμως είχαν ορμόνη τριϊωδοθυρονίνης με τιμή 2,65 και μεγαλύτερη τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς εάν όμως είχαν θυροξίνη ολική με τιμή 191 και μεγαλύτερη τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 4 άτομα).

- Βλέπουμε πως όσοι δεν είχαν υποβληθεί σε χειρουργική επέμβαση θυρεοειδούς και είχαν θυροξίνη με τιμή μικρότερη 0,9 και είχαν θυροξίνη ολική με τιμή 87,5 και μεγαλύτερη και είχαν δείκτη ελεύθερης θυροξίνης με τιμή από 121,5 και μεγαλύτερη και δεν είχαν εξεταστεί για θυροξίνη(T4) και είχαν μέτρηση ορμόνης τριϊωδοθυρονίνης και η ηλικία τους ήταν 40 χρόνων και μεγαλύτερη και είχαν θυροξίνη με τιμή 0,79 και μεγαλύτερη και είχαν ορμόνη τριϊωδοθυρονίνης με τιμή 1,15 και μεγαλύτερη και είχαν δείκτη ελεύθερης θυροξίνης με τιμή 175,5 και μεγαλύτερη και ήταν με φυσιολογικές τιμές και είχαν SVI και είχαν θυροξίνη ολική με τιμή μικρότερη από 166,5 και είχαν θυροξίνη τότε διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 4 άτομα) ενώ αν δεν είχαν μέτρηση την θυροξίνη τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς εάν όμως είχαν θυροξίνη ολική με τιμή μεταξύ 166,5 και μικρότερη από 173,5 τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 1 άτομο) ενώ αν είχαν θυροξίνη ολική με τιμή μεταξύ 173,5 και μεγαλύτερη τότε διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 1 άτομο) εάν όμως ήταν με φυσιολογικές τιμές και είχαν STMW τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς ενώ αν ήταν με φυσιολογικές τιμές και είχαν SVHD τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς. Εάν όμως δεν είχαν μέτρηση ορμόνης τριϊωδοθυρονίνης τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 34 άτομα) εάν όμως είχαν εξεταστεί για θυροξίνη(T4) τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 37 άτομα).
- Βλέπουμε πως όσοι δεν είχαν υποβληθεί σε χειρουργική επέμβαση θυρεοειδούς και είχαν θυροξίνη με τιμή 0,9 και μεγαλύτερη και η ηλικία τους ήταν μικρότερη των 72 χρόνων και είχαν μετρήσει ορμόνης τριϊωδοθυρονίνης και η ηλικία τους ήταν μικρότερη από 23 χρόνων και η ορμόνη τριϊωδοθυρονίνης είχε τιμή μικρότερη από 1,35 και η ορμόνη τριϊωδοθυρονίνης ήταν μικρότερη από 1,2 τότε διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 3 άτομα) ενώ αν η ορμόνη τριϊωδοθυρονίνης ήταν μεταξύ 1,2 και 1,35 και ο δείκτης ελεύθερης θυροξίνης είχε τιμή μικρότερη 65,5 τότε διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 1 άτομο) ενώ αν ο δείκτης ελεύθερης θυροξίνης είχε τιμή 65,5 και μεγαλύτερη τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 1 άτομο) εάν όμως η ορμόνη τριϊωδοθυρονίνης είχε τιμή 1,35 και μεγαλύτερη και η ηλικία τους ήταν μικρότερη από 18 χρόνων τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 39 άτομα) ενώ αν η ηλικία τους ήταν μεταξύ 18 και 23 χρόνων και είχαν θυροξίνη ολική με τιμή μικρότερη από

156,5 τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 65 άτομα) ενώ αν είχαν θυροξίνη ολική με τιμή 156,5 και μεγαλύτερη και ο δείκτης ελεύθερης θυροξίνης είχε τιμή μικρότερη 164,5 τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 7 άτομα) ενώ αν ο δείκτης ελεύθερης θυροξίνης είχε τιμή μεταξύ 164,5 και 174,5 και είχαν θυροξίνη ολική με τιμή μεταξύ 156,5 και 198,5 τότε διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 1 άτομο) ενώ αν είχαν θυροξίνη ολική με τιμή 198,5 και μεγαλύτερη τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 1 άτομο) εάν όμως ο δείκτης ελεύθερης θυροξίνης είχε τιμή 164,5 και μεγαλύτερη τότε διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 1 άτομο).

- Βλέπουμε πως όσοι δεν είχαν υποβληθεί σε χειρουργική επέμβαση θυρεοειδούς και είχαν θυροξίνη με τιμή 0,9 και μεγαλύτερη και η ηλικία τους ήταν μικρότερη των 72 χρόνων και είχαν μετρήσει ορμόνης τριϊωδοθυρονίνης και η ηλικία τους ήταν μεταξύ 23 και 72 χρόνων και η ηλικία τους ήταν μεταξύ 23 και 41 χρόνων και η ηλικία τους ήταν μεταξύ 23 και 26 χρόνων και είχαν ορμόνη τριϊωδοθυρονίνης με τιμή μικρότερη 1,25 και είχαν θυροξίνη με τιμή μικρότερη 1,11 τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 1 άτομο) ενώ αν είχαν θυροξίνη με τιμή 1,11 και μεγαλύτερη τότε διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 1 άτομο) εάν όμως είχαν ορμόνη τριϊωδοθυρονίνης με τιμή 1,25 μεγαλύτερη τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 53 άτομα) εάν όμως η ηλικία τους ήταν μεταξύ 26 και 41 χρόνων τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 342 άτομα) εάν όμως η ηλικία τους ήταν μεταξύ 41 και 72 χρόνων και ήταν με φυσιολογικές τιμές είχαν SVHC τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 87 άτομα) ενώ αν ήταν με φυσιολογικές τιμές είχαν OTHER και είχαν ορμόνη τριϊωδοθυρονίνης με τιμή μικρότερη από 0,75 και είχαν θυροξίνη ολική με τιμή μικρότερη από 43 τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 4 άτομα) ενώ αν είχαν θυροξίνη ολική με τιμή 43 και μεγαλύτερη τότε διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 1 άτομο) εάν όμως είχαν ορμόνη τριϊωδοθυρονίνης με τιμή 0,75 και μεγαλύτερη και είχαν θυροξίνη ολική με τιμή μικρότερη από 161,5 τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 360 άτομα) ενώ αν είχαν θυροξίνη ολική με τιμή 161,5 και μεγαλύτερη και είχαν θυροξίνη με τιμή μικρότερη από 1,02 και είχαν θυροξίνη ολική με τιμή μικρότερη 185 και είχαν θυροξίνη ολική με τιμή μικρότερη 163 τότε

διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 1 άτομο) ενώ αν είχαν θυροξίνη ολική με τιμή 163 και μεγαλύτερη και είχαν δείκτη ελεύθερης θυροξίνης με τιμή μικρότερη από 178,5 τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 2 άτομα) ενώ αν είχαν δείκτη ελεύθερης θυροξίνης με τιμή 178,5 και μεγαλύτερη τότε διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 1 άτομα) εάν όμως είχαν θυροξίνη ολική με τιμή 185 και μεγαλύτερη τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 1 άτομο) εάν όμως είχαν θυροξίνη με τιμή 1,02 και μεγαλύτερη τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 1 άτομο).

- Βλέπουμε πως όσοι δεν είχαν υποβληθεί σε χειρουργική επέμβαση θυρεοειδούς και είχαν θυροξίνη με τιμή 0,9 και μεγαλύτερη και η ηλικία τους ήταν μικρότερη των 72 χρόνων και είχαν μετρήσει ορμόνης τριϊωδοθυρονίνης και η ηλικία τους ήταν μεταξύ 23 και 72 χρόνων και η ηλικία τους ήταν μεταξύ 41 και 72 χρόνων και ήταν με φυσιολογικές τιμές είχαν SVI και είχαν ορμόνη τριϊωδοθυρονίνης με τιμή μικρότερη 1,15 και είχαν δείκτη ελεύθερης θυροξίνης με τιμή μικρότερη από 59,5 και δεν ήταν άρρωστοι κατά την εξέταση και είχαν ορμόνη διέγερσης του θυρεοειδούς με τιμή μικρότερη από 26,5 τότε διαγνωστήκαν με την νόσο του θυρεοειδούς ενώ αν είχαν ορμόνη διέγερσης του θυρεοειδούς με τιμή 26,5 και μεγαλύτερη τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 7 άτομα) ενώ αν ήταν άρρωστοι κατά την εξέταση τότε διαγνωστήκαν με την νόσο του θυρεοειδούς. Εάν όμως είχαν δείκτη ελεύθερης θυροξίνης με τιμή 59,5 και μεγαλύτερη και είχαν θυροξίνη με τιμή μικρότερη 1,05 και δεν είχαν ερωτηθεί αν είχαν υπερθυρεοειδισμό και δεν είχαν διαγνωστεί με κάποια ψυχολογική διαταραχή και η ηλικία τους ήταν μικρότερη 66 χρόνων τότε διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 1 άτομο) ενώ αν η ηλικία τους ήταν 66 χρόνων και μεγαλύτερη και είχαν ορμόνη τριϊωδοθυρονίνης με τιμή μικρότερη 0,6 τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς ενώ αν είχαν ορμόνη τριϊωδοθυρονίνης με τιμή 0,6 και μεγαλύτερη τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 2 άτομα) εάν όμως είχαν διαγνωστεί με κάποια ψυχολογική διαταραχή τότε διαγνωστήκαν με την νόσο του θυρεοειδούς εάν όμως είχαν ερωτηθεί αν είχαν υπερθυρεοειδισμό τότε διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 2 άτομα). Ενώ όμως είχαν θυροξίνη με τιμή 1,05 και μεγαλύτερη και δεν είχαν ερωτηθεί αν είχαν υποθυρεοειδισμό και είχαν θυροξίνη ολική με τιμή μικρότερη από 192,5 και είχαν μέτρηση τον δείκτης ελεύθερης

θυροξίνης τότε διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 2 άτομα) ενώ αν δεν είχαν μέτρηση τον δείκτης ελεύθερης θυροξίνης τότε διαγνωστήκαν με την νόσο του θυρεοειδούς εάν όμως είχαν θυροξίνη ολική με τιμή 192,5 και μεγαλύτερη τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 1 άτομο) εάν όμως είχαν ερωτηθεί αν είχαν υποθυρεοειδισμό τότε διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 1 άτομο).

- Βλέπουμε πως όσοι δεν είχαν υποβληθεί σε χειρουργική επέμβαση θυρεοειδούς και είχαν θυροξίνη με τιμή 0,9 και μεγαλύτερη και η ηλικία τους ήταν μικρότερη των 72 χρόνων και είχαν μετρήσει ορμόνης τριϊωδοθυρονίνης και η ηλικία τους ήταν μεταξύ 23 και 72 χρόνων και η ηλικία τους ήταν μεταξύ 41 και 72 χρόνων και ήταν με φυσιολογικές τιμές είχαν SVI και είχαν ορμόνη τριϊωδοθυρονίνης με τιμή 1,15 και μεγαλύτερη και είχαν ορμόνη διέγερσης του θυρεοειδούς με τιμή μικρότερη 3,25 τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 187 άτομα) ενώ αν είχαν ορμόνη διέγερσης του θυρεοειδούς με τιμή 3,25 και μεγαλύτερη και δεν είχαν βρογχοκήλη και είχαν θυροξίνη ολική με τιμή μικρότερη από 60,5 και είχαν θυροξίνη με τιμή μικρότερη 1,14 και είχαν μέτρηση τον δείκτης ελεύθερης θυροξίνης τότε διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 2 άτομα) ενώ αν δεν είχαν μέτρηση τον δείκτης ελεύθερης θυροξίνης τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς εάν όμως είχαν θυροξίνη με τιμή 1,14 και μεγαλύτερη τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 1 άτομο) εάν όμως είχαν θυροξίνη ολική με τιμή 60,5 και μεγαλύτερη τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 37 άτομα) εάν όμως είχαν βρογχοκήλη τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 1 άτομο).
- Βλέπουμε πως όσοι δεν είχαν υποβληθεί σε χειρουργική επέμβαση θυρεοειδούς και είχαν θυροξίνη με τιμή 0,9 και μεγαλύτερη και η ηλικία τους ήταν μικρότερη των 72 χρόνων και είχαν μετρήσει ορμόνης τριϊωδοθυρονίνης και η ηλικία τους ήταν μεταξύ 23 και 72 χρόνων και η ηλικία τους ήταν μεταξύ 41 και 72 χρόνων και ήταν με φυσιολογικές τιμές είχαν STMW τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 11 άτομα) ενώ αν ήταν με φυσιολογικές τιμές είχαν SVHD τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 17 άτομα).
- Βλέπουμε πως όσοι δεν είχαν υποβληθεί σε χειρουργική επέμβαση θυρεοειδούς και είχαν θυροξίνη με τιμή 0,9 και μεγαλύτερη και η ηλικία τους ήταν μικρότερη των 72 χρόνων και δεν είχαν μετρήσει την ορμόνη τριϊωδοθυρονίνης και είχαν μέτρηση τον

δείκτη ελεύθερης θυροξίνης και ήταν με φυσιολογικές τιμές είχαν SVHC τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 2 άτομα) ενώ αν ήταν με φυσιολογικές τιμές είχαν OTHER τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 263 άτομα) ενώ αν ήταν με φυσιολογικές τιμές είχαν SVI και δεν είχαν υποβληθεί σε θεραπεία με λίθιο και η ηλικία τους ήταν μικρότερη από 44 χρόνων και είχαν ορμόνη διέγερσης του θυρεοειδούς με τιμή μικρότερη από 1 τότε διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 1 άτομο) ενώ αν είχαν ορμόνη διέγερσης του θυρεοειδούς με τιμή 1 και μεγαλύτερη τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 2 άτομα) εάν όμως η ηλικία τους ήταν 44 χρόνων και μεγαλύτερη τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 19 άτομα) εάν όμως είχαν υποβληθεί σε θεραπεία με λίθιο τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 1 άτομο) εάν όμως ήταν με φυσιολογικές τιμές είχαν STMW τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 6 άτομα) εάν όμως ήταν με φυσιολογικές τιμές είχαν SVHD τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 1 άτομο) εάν όμως δεν είχαν μέτρηση τον δείκτη ελεύθερης θυροξίνης τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 106 άτομα).

- Βλέπουμε πως όσοι δεν είχαν υποβληθεί σε χειρουργική επέμβαση θυρεοειδούς και είχαν θυροξίνη με τιμή 0,9 και μεγαλύτερη και η ηλικία τους ήταν 72 χρόνων και μεγαλύτερη και ήταν γυναίκες και είχαν θυροξίνη με τιμή μεταξύ 0,9 και 1,15 και είχαν δείκτη ελεύθερης θυροξίνης με τιμή με τιμή μικρότερη από 172 και είχαν μέτρηση την ορμόνη τριϊωδοθυρονίνης και είχαν ορμόνη διέγερσης του θυρεοειδούς με τιμή μικρότερη από 88,5 και είχαν μέτρηση ορμόνη διέγερσης του θυρεοειδούς και ήταν με φυσιολογικές τιμές είχαν SVHC τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 4 άτομα) ενώ αν ήταν με φυσιολογικές τιμές είχαν OTHER και είχαν δείκτη ελεύθερης θυροξίνης με τιμή μικρότερη από 78 και δεν έπαιρναν αντιβιοτικά για αντιμετώπιση του θυρεοειδή και είχαν θυροξίνη και είχαν θυροξίνη ολική με τιμή μικρότερη από 68 τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 3 άτομα) ενώ αν είχαν θυροξίνη ολική με τιμή μεταξύ 68 και 70,5 τότε διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 1 άτομο) ενώ αν είχαν θυροξίνη ολική με τιμή από 70,5 και μεγαλύτερη τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 2 άτομα) εάν όμως δεν είχαν θυροξίνη τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς εάν όμως έπαιρναν αντιβιοτικά για αντιμετώπιση του θυρεοειδή τότε

δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 1 άτομο) εάν όμως είχαν δείκτη ελεύθερης θυροξίνης με τιμή από 78 και μεγαλύτερη τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 55 άτομα) εάν όμως ήταν με φυσιολογικές τιμές είχαν SVI και δεν είχαν ερωτηθεί αν είχαν υποθυρεοειδισμό και είχαν δείκτη ελεύθερης θυροξίνης με τιμή μικρότερη από 92,5 και είχαν θυροξίνη τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 11 άτομα) ενώ αν δεν είχαν θυροξίνη τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς ενώ αν είχαν δείκτη ελεύθερης θυροξίνης με τιμή μεταξύ 92,5 και 94,5 τότε διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 1 άτομο) ενώ αν είχαν δείκτη ελεύθερης θυροξίνης με τιμή 94,5 και μεγαλύτερα και δεν είχαν διαγνώστηκαν με κάποια ψυχολογική διαταραχή και είχαν ορμόνη διέγερσης του θυρεοειδούς με τιμή μικρότερη από 4 και είχαν θυροξίνη με τιμή μικρότερη από 1,04 και είχαν μέτρηση δείκτη ελεύθερης θυροξίνης τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 37 άτομα) ενώ αν δεν είχαν μέτρηση δείκτη ελεύθερης θυροξίνης τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς ενώ αν είχαν θυροξίνη με τιμή 1,04 και μεγαλύτερη και είχαν θυροξίνη ολική με τιμή μικρότερη 109,5 τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 4 άτομα) ενώ αν είχαν θυροξίνη ολική με τιμή 109,5 και μεγαλύτερη και είχαν ορμόνη τριϊωδοθυρονίνης με τιμή μικρότερη από 1,3 τότε διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 2 άτομα) ενώ αν είχαν ορμόνη τριϊωδοθυρονίνης με τιμή 1,3 και μεγαλύτερη τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 2 άτομα) εάν όμως είχαν ορμόνη διέγερσης του θυρεοειδούς με τιμή 4 και μεγαλύτερη τότε διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 3 άτομα) εάν όμως είχαν διαγνώστηκαν με κάποια ψυχολογική διαταραχή τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 3 άτομα) εάν όμως είχαν ερωτηθεί αν είχαν υποθυρεοειδισμό και είχαν ορμόνη τριϊωδοθυρονίνη με τιμή μικρότερη από 1,3 τότε διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 2 άτομα) ενώ αν είχαν ορμόνη τριϊωδοθυρονίνη με τιμή 1,3 και μεγαλύτερη τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 2 άτομα) εάν όμως ήταν με φυσιολογικές τιμές είχαν STMW τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς εάν όμως ήταν με φυσιολογικές τιμές είχαν SVHD τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 2 άτομα) εάν όμως δεν είχαν μετρήθηκε η ορμόνη διέγερσης του θυρεοειδούς τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 4 άτομα) εάν όμως είχαν ορμόνη διέγερσης του θυρεοειδούς με τιμή 88,5 και

μεγαλύτερη τότε διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 1 άτομο) εάν όμως δεν είχαν μέτρηση ορμόνης τριϊωδοθυρονίνης τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 13 άτομα) εάν όμως είχαν δείκτη ελεύθερης θυροξίνης με τιμή 172 και μεγαλύτερη και είχαν θυροξίνη με τιμή μικρότερη από 0,99 και είχαν μέτρηση δείκτης ελεύθερης θυροξίνης τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 3 άτομα) ενώ αν δεν είχαν μέτρηση δείκτης ελεύθερης θυροξίνης τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς ενώ αν είχαν θυροξίνη με τιμή 0,99 και μεγαλύτερη και είχαν θυροξίνη ολική με τιμή μικρότερη από 195 τότε διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 2 άτομα) ενώ αν είχαν θυροξίνη ολική με τιμή 195 και μεγαλύτερη τότε διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 1 άτομο) εάν όμως είχαν θυροξίνη με τιμή 1,15 και μεγαλύτερη και ήταν με φυσιολογικές τιμές είχαν SVHC τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 2 άτομα) ενώ αν ήταν με φυσιολογικές τιμές είχαν OTHER τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 26 άτομα) ενώ αν ήταν με φυσιολογικές τιμές είχαν SVI και είχαν μέτρηση δείκτης ελεύθερης θυροξίνης τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 9 άτομα) ενώ αν δεν είχαν μέτρηση δείκτης ελεύθερης θυροξίνης τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς ενώ αν ήταν με φυσιολογικές τιμές είχαν STMW τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς ενώ αν ήταν με φυσιολογικές τιμές είχαν SVHD τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 2 άτομα).

- Βλέπουμε πως όσοι δεν είχαν υποβληθεί σε χειρουργική επέμβαση θυρεοειδούς και είχαν θυροξίνη με τιμή 0,9 και μεγαλύτερη και η ηλικία τους ήταν 72 χρόνων και μεγαλύτερη και ήταν άνδρες και δεν έπαιρναν αντιβιοτικά για αντιμετώπιση του θυρεοειδή και είχαν δείκτη ελεύθερης θυροξίνης με τιμή μικρότερη από 51 και είχαν μέτρηση την θυροξίνη ολικής τότε διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 2 άτομα) ενώ αν δεν είχαν μέτρηση την θυροξίνη ολικής τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς ενώ αν δεν είχαν δείκτη ελεύθερης θυροξίνης με τιμή 51 και μεγαλύτερη και δεν είχαν ερωτηθεί αν είχαν υποθυρεοειδισμό και δεν ήταν έγκυος και είχαν δείκτη ελεύθερης θυροξίνης με τιμή μικρότερη από 102,5 τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 20 άτομα) ενώ αν είχαν δείκτη ελεύθερης θυροξίνης με τιμή 102,5 και μεγαλύτερη και είχαν δείκτη ελεύθερης θυροξίνης με τιμή μικρότερη από 107 και είχαν ορμόνη διέγερσης του θυρεοειδούς με

τιμή μικρότερη από 1,6 τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 2 άτομα) ενώ αν είχαν ορμόνη διέγερσης του θυρεοειδούς με τιμή 1,6 και μεγαλύτερη και είχαν θυροξίνη με τιμή μικρότερη από 0,95 τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 1 άτομο) ενώ αν είχαν θυροξίνη με τιμή 0,95 και μεγαλύτερη και είχαν ορμόνη διέγερσης του θυρεοειδούς με τιμή μικρότερη από 5,7 τότε διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 2 άτομα) ενώ αν είχαν ορμόνη διέγερσης του θυρεοειδούς με τιμή 5,7 και μεγαλύτερη τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς. Εάν όμως είχαν δείκτη ελεύθερης θυροξίνης με τιμή 107 και μεγαλύτερη και ασθενείς όπου δεν διαγνώστηκαν με κάποια ψυχολογική διαταραχή και η ηλικία τους ήταν μικρότερη από 72 χρόνων και είχαν θυροξίνη με τιμή μικρότερη από 0,99 τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 10 άτομα) ενώ αν είχαν θυροξίνη με τιμή 0,99 και μεγαλύτερη και είχαν ορμόνη τριϊωδοθυρονίνης και είχαν θυροξίνη με τιμή μικρότερη από 1,13 και είχαν θυροξίνη με τιμή μικρότερη από 1,12 και είχαν ορμόνη τριϊωδοθυρονίνης με τιμή μικρότερη 1,7 τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 1 άτομο) ενώ αν είχαν ορμόνη τριϊωδοθυρονίνης με τιμή 1,7 και μεγαλύτερη τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 1 άτομο) εάν όμως είχαν θυροξίνη με τιμή 1,12 και μεγαλύτερη τότε διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 1 άτομο) εάν όμως είχαν θυροξίνη με τιμή 1,13 και μεγαλύτερη τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 2 άτομα) εάν όμως δεν είχαν ορμόνη τριϊωδοθυρονίνης τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς εάν όμως η ηλικία τους ήταν 72 χρόνων και μεγαλύτερη τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 7 άτομα) εάν όμως οι ασθενείς διαγνώστηκαν με κάποια ψυχολογική διαταραχή τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 2 άτομα) εάν όμως ήταν έγκυοι τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς εάν όμως είχαν ερωτηθεί αν είχαν υποθυρεοειδισμό και είχαν θυροξίνη(T4) και είχαν θυροξίνη με τιμή μικρότερη από 0,91 τότε διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 1 άτομο) ενώ αν είχαν θυροξίνη με τιμή 0,91 και μεγαλύτερη και είχαν θυροξίνη ολική με τιμή μικρότερη από 101,5 τότε διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 1 άτομο) ενώ αν είχαν θυροξίνη ολική με τιμή 101,5 και μεγαλύτερη τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 2 άτομα) εάν όμως δεν είχαν θυροξίνη(T4) τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 1 άτομο) εάν όμως έπαιρναν

αντιβιοτικά για αντιμετώπιση του θυρεοειδή τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς εάν όμως είχαν υποβληθεί σε χειρουργική επέμβαση θυρεοειδούς τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 39 άτομα).

```

Classifier output
RandomTree
=====
thyroid surgery = f
| T4U < 0.9
| | TT4 < 87.5
| | | age < 43.5
| | | | FTI < 79
| | | | | T3 < 1.15
| | | | | | sex = F : negative (1.75/0.37)
| | | | | | sex = M
| | | | | | TSH measured = t
| | | | | | | age < 18 : negative (0/0)
| | | | | | | age >= 18
| | | | | | | | age < 36.5 : sick (2/0)
| | | | | | | | age >= 36.5 : sick (0.64/0.01)
| | | | | | | | TSH measured = f : negative (0.35/0)
| | | | | | T3 >= 1.15 : negative (11.24/0)
| | | | | FTI >= 79
| | | | | | sex = F
| | | | | | | query hypothyroid = f : negative (52.02/0)
| | | | | | | query hypothyroid = t
| | | | | | | TT4 < 81 : negative (1.37/0)
| | | | | | | TT4 >= 81 : sick (1.1/0.1)
| | | | | | sex = M : negative (40.59/0)
| | | | | age >= 43.5
| | | | | T3 < 1.15
| | | | | | query hyperthyroid = f
| | | | | | | T3 < 0.55
| | | | | | | T3 measured = t
| | | | | | | referral source = SVHC : negative (0/0)

```

Εικόνα 116: Αναλυτική ανάλυση του αλγορίθμου Random Tree

Αναλυτική ανάλυση του αλγορίθμου REP Tree

Οι αναλυτικές αναλύσεις του αλγορίθμου Random Tree ισχύουν για τα υποκεφάλαια 5.2.7-5.2.8(σελ. 120~127).

Βλέπουμε πως όσοι είχαν ορμόνη τριϊωδοθυρονίνη με τιμή μικρότερη από 1,15 και είχαν μέτρηση ορμόνης τριϊωδοθυρονίνης και είχαν θυροξίνη ολική με τιμή μικρότερη από 52 τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 30 άτομα) ενώ αν είχαν θυροξίνη ολική με τιμή 52 και μεγαλύτερη και ήταν με φυσιολογικές τιμές είχαν SVHC τότε διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 3 άτομα) ενώ αν ήταν με φυσιολογικές τιμές είχαν OTHER και η ηλικία τους ήταν μικρότερη από 77 χρόνων τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 18 άτομα) ενώ αν η ηλικία τους ήταν 77 χρόνων και μεγαλύτερη τότε διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 7 άτομα) εάν όμως ήταν με φυσιολογικές τιμές είχαν SVI και έχουν δείκτη ελεύθερης θυροξίνης με τιμή μικρότερη από 61 τότε δεν

διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 2 άτομα) ενώ αν είχαν δείκτη ελεύθερης θυροξίνης με τιμή 61 και μεγαλύτερη και είχαν δείκτη ελεύθερης θυροξίνης με τιμή μικρότερη από 164,5 τότε διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 7 άτομα) ενώ αν είχαν δείκτη ελεύθερης θυροξίνης με τιμή 164,5 και μεγαλύτερη τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 7 άτομα) εάν όμως ήταν με φυσιολογικές τιμές είχαν STMW τότε διαγνωστήκαν με την νόσο του θυρεοειδούς εάν όμως ήταν με φυσιολογικές τιμές είχαν SVHD τότε διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 2 άτομα) εάν όμως δεν είχαν μέτρηση ορμόνης τριϊωδοθυρονίνης τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 40 άτομα). Εάν όμως είχαν ορμόνη τριϊωδοθυρονίνη με τιμή 1,15 και μεγαλύτερη και είχαν δείκτη ελεύθερης θυροξίνης με τιμή μικρότερη από 172,5 και είχαν δείκτη ελεύθερης θυροξίνης με τιμή μικρότερη από 52,5 και είχαν ορμόνη διέγερσης του θυρεοειδούς με τιμή μικρότερη από 4,95 τότε διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 6 άτομα) ενώ αν είχαν ορμόνη διέγερσης του θυρεοειδούς με τιμή 4,95 και μεγαλύτερη τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 11 άτομα) εάν όμως είχαν δείκτη ελεύθερης θυροξίνης με τιμή 52,5 και μεγαλύτερη τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 1.593 άτομα) εάν όμως είχαν δείκτη ελεύθερης θυροξίνης με τιμή 172,5 και μεγαλύτερη και είχαν ορμόνη τριϊωδοθυρονίνη με τιμή μικρότερη από 2,55 και είχαν ορμόνη διέγερσης του θυρεοειδούς με τιμή μικρότερη από 0,21 και δεν έκαναν εξέταση για θυροξίνη(T4) και είχαν μέτρηση την ορμόνη τριϊωδοθυρονίνης τότε διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 11 άτομα) ενώ αν δεν είχαν μέτρηση την ορμόνη τριϊωδοθυρονίνης τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 3 άτομα) εάν όμως έκαναν εξέταση για θυροξίνη(T4) τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 5 άτομα) εάν όμως είχαν ορμόνη διέγερσης του θυρεοειδούς με τιμή 0,21 και μεγαλύτερη τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 8 άτομα) εάν όμως είχαν ορμόνη τριϊωδοθυρονίνη με τιμή 2,55 και μεγαλύτερη τότε δεν διαγνωστήκαν με την νόσο του θυρεοειδούς (όπου ήταν 44 άτομα).

```

Classifier output
REPTree
=====
T3 < 1.15
| T3 measured = t
| | FT4 < 52 : negative (30/2) [12/2]
| | FT4 >= 52
| | | referral source = SVHC : sick (3/0) [2/0]
| | | referral source = other
| | | | age < 77.5 : negative (18/6) [11/2]
| | | | age >= 77.5 : sick (7/0) [0/0]
| | | referral source = SVI
| | | | FTI < 61 : negative (2.07/0.07) [1.05/0.05]
| | | | FTI >= 61
| | | | | FTI < 164.5 : sick (79.85/0) [43.9/2]
| | | | | FTI >= 164.5 : negative (2.07/0.07) [3.05/0.05]
| | | | referral source = STMW : sick (0/0) [0/0]
| | | | referral source = SVHD : sick (2/0) [1/0]
| | T3 measured = f : negative (39.54/0.1) [18/0.1]
T3 >= 1.15
| FTI < 172.5
| | FTI < 52.5
| | | TSH < 4.95 : sick (5.98/2.08) [2.32/0.42]
| | | TSH >= 4.95 : negative (10.58/0) [6.28/0]
| | FTI >= 52.5 : negative (1593.25/0) [804.39/2]
| FTI >= 172.5
| | T3 < 2.55
| | | TSH < 0.21
| | | | on thyroxine = f
| | | | | T3 measured = t : sick (11.63/1.63) [1.24/0.24]
| | | | | T3 measured = f : negative (3.12/0) [0.98/0]

```

Εικόνα 117: Αναλυτική ανάλυση του αλγορίθμου REP Tree

Βιβλιογραφία

- Corporate Finance Institute. (2021, 11 20). Ανάκτηση από <https://corporatefinanceinstitute.com/resources/knowledge/other/random-forest/>
- Ian H. Witten, E. F. (2020, 11 21). *Weka 3 - Data Mining*. Ανάκτηση από <https://www.cs.waikato.ac.nz/~ml/publications/1999/99IHW-EF-LT-MH-GH-SJC-Tools-Java.pdf>
- Namratha M, P. T. (2021, 11 25). *COLLATIVE STUDY OF CLASSIFIERS IN PATTERN*. Ανάκτηση από International Journal of Computer Network and Security(IJCNS): <http://www.ijcns.com/pdf/ijcnsvol5no1-3.pdf>
- Richard J. Roiger, M. W. (2008). *Εξόρυξη πληροφορίας, Ένας εισαγωγικός οδηγός με παραδείγματα*. Κλειδάριθμος.
- Shams, R. (2012, 5 10). Ανάκτηση από <https://stackoverflow.com/questions/10437677/cross-validation-in-weka>
- The University of Waikato*. (2021, 03 20). Ανάκτηση από https://www.cs.waikato.ac.nz/~ml/weka/gui_explorer.html
- University of Regina. (2020, 12 30). Ανάκτηση από http://www2.cs.uregina.ca/~hamilton/courses/831/notes/ml/dtrees/4_dtrees1.html
- WEKA*. (2020, 11 14). Ανάκτηση από <https://www.cs.waikato.ac.nz/ml/weka/>
- weka. (2020, 12 20). *Class J48*. Ανάκτηση από <https://weka.sourceforge.io/doc.dev/weka/classifiers/trees/J48.html>
- wikipedia. (2020, 12 30). Ανάκτηση από https://el.wikipedia.org/wiki/%CE%95%CF%80%CE%B5%CE%BE%CE%B5%CF%81%CE%B3%CE%B1%CF%83%CE%AF%CE%B1_%CF%86%CF%85%CF%83%CE%B9%CE%BA%CE%AE%CF%82_%CE%B3%CE%BB%CF%8E%CF%83%CF%83%CE%B1%CF%82
- wikipedia. (2020, 11 21). *wikipedia*. Ανάκτηση από [https://el.wikipedia.org/wiki/Weka_\(%CE%BC%CE%B7%CF%87%CE%B1%CE%](https://el.wikipedia.org/wiki/Weka_(%CE%BC%CE%B7%CF%87%CE%B1%CE%)

BD%CE%B9%CE%BA%CE%AE_%CE%BC%CE%AC%CE%B8%CE%B7%CF%83%CE%B7)

wikipedia. (2020, 12 20). *wikipedia*. Ανάκτηση από https://el.wikipedia.org/wiki/%CE%95%CE%BD%CF%84%CF%81%CE%BF%CF%80%CE%AF%CE%B1_%CF%80%CE%BB%CE%B7%CF%81%CE%BF%CF%86%CE%BF%CF%81%CE%B9%CF%8E%CE%BD

Άγαχ, Α. (2017). Medical Applications of Artificial Intelligence. Στο Α. Άγαχ, *Medical Applications of Artificial Intelligence* (σ. 37).

Ελληνικά Ακαδημαϊκά Ηλεκτρονικά Συγγράμματα και Βοηθήματα. (2021, 06 10). Ανάκτηση από https://repository.kallipos.gr/bitstream/11419/1239/2/Kef._13.pdf

Πνευματικά δικαιώματα

Copyright © Πανεπιστήμιο Πατρών. Με επιφύλαξη παντός δικαιώματος. All rights reserved.

Δηλώνω ρητά ότι, σύμφωνα με το άρθρο 8 του Ν. 1599/1988 και τα άρθρα 2,4,6 παρ. 3 του Ν. 1256/1982, η παρούσα εργασία αποτελεί αποκλειστικά προϊόν προσωπικής εργασίας και δεν προσβάλλει κάθε μορφής πνευματικά δικαιώματα τρίτων και δεν είναι προϊόν μερικής ή ολικής αντιγραφής, οι πηγές δε που χρησιμοποιήθηκαν περιορίζονται στις βιβλιογραφικές αναφορές και μόνον.

Δημήτριος Καριωτάκης, Μαρία Βορύλλα, 2022