

ΤΕΧΝΟΛΟΓΙΚΟ ΕΚΠΑΙΔΕΥΤΙΚΟ ΙΔΡΥΜΑ ΔΥΤΙΚΗΣ ΕΛΛΑΔΑΣ

**ΣΧΟΛΗ ΔΙΟΙΚΗΣΗ ΚΑΙ ΟΙΚΟΝΟΜΙΑΣ
ΤΜΗΜΑ ΔΙΟΙΚΗΣΗΣ ΕΠΙΧΕΙΡΗΣΕΩΝ**

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

**ΔΕΝΤΡΑ ΑΠΟΦΑΣΗΣ ΣΤΗΝ ΕΞΟΡΥΞΗ
ΔΕΔΟΜΕΝΩΝ
ΘΕΩΡΙΑ, ΑΛΓΟΡΙΘΜΟΙ ΚΑΙ ΕΦΑΡΜΟΓΕΣ**

ΑΘΑΝΑΣΙΟΥ ΚΩΝΣΤΑΝΤΙΝΟΣ

ΜΠΕΝΕΤΣΗΣ ΝΙΚΟΛΑΟΣ

ΤΣΑΝΟΥΛΑΣ ΚΩΝΣΤΑΝΤΙΝΟΣ

ΕΙΣΗΓΗΤΗΣ: ΣΤΑΜΟΣ ΚΩΝΣΤΑΝΤΙΝΟΣ

ΠΑΤΡΑ 2016

ΠΡΟΛΟΓΟΣ

Η ΠΑΡΟΥΣΑ ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ ΕΧΕΙ ΩΣ ΣΚΟΠΟ ΝΑ ΠΑΡΟΥΣΙΑΣΕΙ ΤΗΝ ΔΙΑΔΙΚΑΣΙΑ ΤΗΣ ΕΞΟΡΥΞΗΣ ΔΕΔΟΜΕΝΩΝ ΜΕ ΤΗΝ ΧΡΗΣΗ ΤΩΝ ΔΕΝΔΡΩΝ ΑΠΟΦΑΣΕΩΝ. ΜΕΤΑ ΤΗ ΔΗΜΙΟΥΡΓΙΑ ΕΝΟΣ ΘΕΩΡΗΤΙΚΟΥ ΠΛΑΙΣΙΟΥ, ΣΧΕΤΙΚΟΥ ΜΕ ΤΗΝ ΕΞΟΡΥΞΗ ΔΕΔΟΜΕΝΩΝ ΕΠΙΧΕΙΡΕΙΤΑΙ Η ΠΑΡΟΥΣΙΑΣΗ ΤΩΝ ΔΕΝΔΡΩΝ ΑΠΟΦΑΣΗΣ ΣΑΝ ΜΕΘΟΔΟΛΟΓΙΑ. ΣΤΗΝ ΣΥΝΕΧΕΙΑ ΕΠΙΔΕΙΚΝΥΟΝΤΑΙ ΟΙ ΣΗΜΑΝΤΙΚΟΤΕΡΟΙ ΣΧΕΤΙΚΟΙ ΑΛΓΟΡΙΘΜΟΙ ΚΑΙ ΑΞΙΟΛΟΓΟΥΝΤΑΙ ΓΙΑ ΤΗΝ ΑΠΟΤΕΛΕΣΜΑΤΙΚΟΤΗΤΑ ΤΟΥΣ. Η ΔΙΑΔΙΚΑΣΙΑ ΑΥΤΗ ΣΥΝΟΔΕΥΕΤΑΙ ΚΑΙ ΑΠΟ ΠΡΑΚΤΙΚΗ ΕΦΑΡΜΟΓΗ ΜΕ ΚΑΤΑΛΛΗΛΟ ΛΟΓΙΣΜΙΚΟ.

ΠΕΡΙΛΗΨΗ

ΟΙ ΓΡΗΓΟΡΟΙ ΡΥΘΜΟΙ ΑΝΑΠΤΥΞΗΣ ΣΤΙΣ ΤΕΧΝΟΛΟΓΙΕΣ ΠΑΡΑΓΩΓΗΣ, ΔΙΑΘΕΣΗΣ ΚΑΙ ΜΕΤΑΔΟΣΗΣ ΔΕΔΟΜΕΝΩΝ ΑΝΕΔΕΙΞΕ ΤΗΝ ΑΝΑΓΚΗ ΑΠΟΜΟΝΩΣΗΣ ΤΗΣ ΑΠΟΔΟΤΙΚΗΣ ΠΛΗΡΟΦΟΡΙΑΣ ΑΠΟ ΜΕΓΑΛΑ ΣΥΝΟΛΑ ΔΕΔΟΜΕΝΩΝ. ΈΧΟΥΝ ΠΡΟΤΑΘΕΙ ΑΡΚΕΤΕΣ ΜΕΘΟΔΟΙ ΓΙΑ ΤΟΝ ΠΡΟΣΔΙΟΡΙΣΜΟ ΤΗΣ ΟΙ ΟΠΟΙΕΣ ΒΑΣΙΖΟΝΤΑΙ ΣΕ ΜΕΓΑΛΟ ΒΑΘΜΟ ΣΕ ΕΝΝΟΙΕΣ ΚΑΙ ΔΙΑΔΙΚΑΣΙΕΣ ΤΟΥ ΠΡΑΓΜΑΤΙΚΟΥ ΚΟΣΜΟΥ. ΣΚΟΠΟΣ ΤΩΝ ΤΕΧΝΙΚΩΝ ΑΥΤΩΝ ΔΕΝ ΕΙΝΑΙ ΚΥΡΙΩΣ ΝΑ ΕΝΤΟΠΙΣΟΥΝ ΤΗΝ ΙΔΙΑ ΤΗΝ ΩΦΕΛΙΜΗ ΠΛΗΡΟΦΟΡΙΑ ΑΛΛΑ ΝΑ ΑΝΙΧΝΕΥΣΟΥΝ ΑΠΟ ΤΗΝ ΥΠΑΡΞΗ ΤΟΥΣ ΚΑΙ ΤΗΝ ΜΕΤΑΞΥ ΤΟΥΣ ΣΥΣΧΕΤΙΣΗ ΤΙΣ ΤΑΣΕΙΣ ΤΩΝ ΚΑΤΑΝΑΛΩΤΩΝ. ΟΙ ΤΑΣΕΙΣ ΑΥΤΕΣ ΧΡΗΣΙΜΟΠΟΙΟΥΝΤΑΙ ΓΙΑ ΝΑ ΠΑΡΑΞΟΥΝ ΓΝΩΣΗ ΧΡΗΣΙΜΗ ΓΙΑ ΤΗΝ ΕΚΤΙΜΗΣΗ ΜΕΛΛΟΝΤΙΚΩΝ ΚΑΤΑΣΤΑΣΕΩΝ. ΜΙΑ ΑΠΟ ΤΙΣ ΤΕΧΝΙΚΕΣ ΑΥΤΕΣ ΕΙΝΑΙ ΚΑΙ ΤΑ ΔΕΝΔΡΑ ΑΠΟΦΑΣΕΩΝ ΜΕ ΤΑ ΟΠΟΙΑ ΑΝΑΖΗΤΕΙΤΑΙ Η ΑΠΟΔΟΤΙΚΟΤΕΡΗ ΛΥΣΗ ΣΕ ΕΝΑ ΖΗΤΗΜΑ ΜΕΣΑ ΑΠΟ ΤΟΝ ΕΛΕΓΧΟ ΔΙΑΔΟΧΙΚΩΝ ΑΠΟΦΑΣΕΩΝ. Ο ΕΛΕΓΧΟΣ ΜΠΟΡΕΙ ΝΑ ΓΙΝΕΙ ΜΕ ΤΗΝ ΕΦΑΡΜΟΓΗ ΑΛΓΟΡΙΘΜΩΝ ΚΑΤΑΣΚΕΥΗΣ ΚΑΙ ΑΝΑΓΝΩΣΗΣ ΤΟΥΣ ΚΑΙ ΟΙ ΟΠΟΙΟΙ ΜΠΟΡΟΥΝ ΕΥΚΟΛΑ ΝΑ ΧΡΗΣΙΜΟΠΟΙΗΘΟΥΝ ΣΕ ΕΦΑΡΜΟΓΕΣ ΗΛΕΚΤΡΟΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ. ΜΙΑ ΤΕΤΟΙΑ ΕΦΑΡΜΟΓΗΣ ΕΙΝΑΙ ΚΑΙ ΤΟ WΕΚΑΤΟ ΟΠΟΙΟ ΜΠΟΡΕΙ ΜΕ ΕΙΣΟΔΟ ΕΝΟΣ ΜΕΓΑΛΟΥ ΣΥΝΟΛΟΥ ΔΕΔΟΜΕΝΩΝ ΝΑ ΠΑΡΟΥΣΙΑΣΕΙ ΤΑ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΓΝΩΡΙΣΜΑΤΑ ΔΕΝΔΡΩΝ ΑΠΟΦΑΣΕΩΝ.

SUMMARY

THE VAST TECHNOLOGY GROWTH IN PRODUCING TECHNOLOGYS, DISPOSAL AND DATA TRANSMITING CREATED THE NEED TO ISOLATE THE EFFICIENT INFORMATION AMONG HUGE DATA BASIS.THERE HAVE BEEN PROPOSED A LOT OF METHODS FOR ITS DEFINITION WHICH ARE MAINLY BASED INTO REALITYS CONCEPTS AND PROCEDURES.THE MAIN PURPOSE FOR THOSE TECHNIQUES IS NOT TO FIND THAT SPECIFIC INFORMATION BUT TO SCAN THE CONSUMERS TRENDS THROUGH THEIR EXISTANCE AND CORRELATION.THOSE TRENDS ARE USED TO PRODUCE KNOWLEDGE USEFULL FOR FUTURE PURPOSES.ONE OF THOSE TECHNIQUES IS DECISION TREES WHICH ARE USED TO FIND THE MOST EFFICIENT SOLUTION TO AN ISSUE THROUGH CONTROLLING SUCCESSIVE DECISIONS.THE RESEARCH CAN BE ACCOMPLISHED WITH CONSTRUCTION AND READING ALGORITHMS WHICH CAN BE EASILY USED TO COMPUTER APPLICATIONS.ONE APPLICATION OF THESE IS "WEKA" WHICH CAN PROVIDE THE CHARACTERISTICS OF DECESION TREES AFTER RUNNING THE DATA BASE WE PROVIDED.

ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ

ΕΙΣΑΓΩΓΗ.....	7
1. ΕΞΟΥΣΙΑ ΔΕΔΟΜΕΝΩΝ.....	10
1.1 ΟΡΙΣΜΟΣ.....	10
1.2 ΠΕΡΙΓΡΑΦΗ.....	16
1.3 ΜΕΘΟΔΟΛΟΓΙΕΣ.....	24
ΤΕΧΝΗΤΑΝΕΥΡΩΝΙΚΑΔΙΚΤΥΑ (ARTIFICIAL NEURAL NETWORKS).....	24
ΔΕΝΤΡΑ ΑΠΟΦΑΣΕΩΝ (DECISION TREES).....	31
ΓΕΝΕΤΙΚΟΙ ΑΛΓΟΡΙΘΜΟΙ.....	31
ΕΠΑΓΩΓΗ ΚΑΝΟΝΩΝ (RULE INDUCTION).....	34
ΜΕΘΟΔΟΣ ΤΟΥ ΚΟΝΤΙΝΟΤΕΡΟΥ ΓΕΙΤΟΝΑ (NEAREST NEIGHBOR METHOD).....	34
ΓΡΑΦΙΚΗ ΑΝΑΛΥΣΗ (DATA VISUALIZATION).....	37
1.4 ΜΟΝΤΕΛΑ ΠΛΗΡΟΦΟΡΙΩΝ.....	41
1.5 ΕΦΑΡΜΟΓΕΣ.....	45
1.6 ΕΞΟΥΣΙΑ ΔΕΔΟΜΕΝΩΝ ΚΑΙ ΙΔΙΩΤΙΚΟΤΗΤΑ.....	48
2. ΔΕΝΔΡΑ ΑΠΟΦΑΣΕΩΝ.....	53
2.1 ΠΕΡΙΓΡΑΦΗ.....	53
2.2 ΑΛΓΟΡΙΘΜΟΙ.....	59
ID3.....	61
C4.5.....	69
C5.0.....	73
3. ΧΡΗΣΗ ΤΟΥ WEKA.....	74
3.1 ΠΕΡΙΓΡΑΦΗ ΤΟΥ ΕΡΓΑΛΕΙΟΥ.....	74
3.2 ΧΡΗΣΗ.....	75
3.3 ΔΟΜΗ ΑΡΧΕΙΩΝ ARFF.....	77
3.4 ΜΕΛΕΤΗ ΠΕΡΙΠΤΩΣΗΣ.....	79
4. ΣΥΜΠΕΡΑΣΜΑΤΑ.....	91
ΑΝΑΦΟΡΕΣ.....	95

ΠΙΝΑΚΑΣ ΕΙΚΟΝΩΝ

ΕΙΚΟΝΑ 1: Η ΤΟΠΟΘΕΤΗΣΗ ΤΗΣ ΕΞΟΡΥΞΗΣ ΔΕΔΟΜΕΝΩΝ ΩΣ ΠΡΟΣ ΑΛΛΕΣ ΜΕΘΟΔΟΛΟΓΙΕΣ ΚΑΙ ΤΕΧΝΙΚΕΣ	14
ΕΙΚΟΝΑ 2: ΔΙΑΔΙΚΑΣΙΑ ΕΞΟΡΥΞΗΣ ΓΝΩΣΗΣ.....	23
ΕΙΚΟΝΑ 3: ΑΝΤΙΣΤΟΙΧΙΑ ΒΙΟΛΟΓΙΚΩΝ ΚΑΙ ΤΕΧΝΗΤΩΝ ΝΕΥΡΩΝΙΚΩΝ ΔΙΚΤΥΩΝ.....	25
ΕΙΚΟΝΑ 4: ΜΟΡΦΗ ΤΕΧΝΗΤΟΥ ΝΕΥΡΩΝΑ	26
ΕΙΚΟΝΑ 5: ΜΟΡΦΗ ΝΕΥΡΩΝΙΚΟΥ ΔΙΚΤΥΟΥ	28
ΕΙΚΟΝΑ 6: ΚΑΤΗΓΟΡΙΟΠΟΙΗΣΗ ΜΕ ΤΗΝ ΜΕΘΟΔΟ ΤΟΥ ΚΟΝΤΙΝΟΤΕΡΟΥ ΓΕΙΤΟΝΑ.....	36
ΕΙΚΟΝΑ 7: ΠΑΡΑΔΕΙΓΜΑ ΙΣΤΟΓΡΑΜΜΑΤΟΣ	38
ΕΙΚΟΝΑ 8: ΔΙΑΓΡΑΜΜΑ ΚΑΤΑΝΟΜΗΣ ΣΥΧΝΟΤΗΤΩΝ.....	38
ΕΙΚΟΝΑ 9: ΠΑΡΑΔΕΙΓΜΑ ΔΙΑΓΡΑΜΜΑΤΟΣ ΔΙΑΣΠΟΡΑΣ.....	40
ΕΙΚΟΝΑ 10: ΜΟΡΦΗ ΠΙΝΑΚΑ ΣΥΝΑΦΕΙΑΣ	41
ΕΙΚΟΝΑ 11: ΚΟΜΒΟΣ ΑΠΟΦΑΣΗΣ ΣΕ ΔΕΝΔΡΟ ΑΠΟΦΑΣΗΣ.....	55
ΕΙΚΟΝΑ 12: ΚΟΜΒΟΣ ΤΥΧΗΣ ΣΕ ΔΕΝΔΡΟ ΑΠΟΦΑΣΗΣ	56
ΕΙΚΟΝΑ 13: ΜΟΡΦΗ ΔΕΝΔΡΟΥ ΑΠΟΦΑΣΗΣ.....	57
ΕΙΚΟΝΑ 14: ΠΑΡΑΔΕΙΓΜΑ ARFF ΑΡΧΕΙΟΥ	79
ΕΙΚΟΝΑ 15: ΑΡΧΙΚΗ ΟΘΟΝΗ WEKA.....	80
ΕΙΚΟΝΑ 16: ΕΠΙΛΟΓΗ ΣΥΝΟΛΟΥ ΔΕΔΟΜΕΝΩΝ ΑΠΟ ΤΟ ΔΙΑΔΙΚΤΥΟ.....	80
ΕΙΚΟΝΑ 17: ΑΝΑΚΤΗΣΗ ΣΤΟΙΧΕΙΩΝ ΤΟΥ ΣΥΝΟΛΟΥ ΔΕΔΟΜΕΝΩΝ	81
ΕΙΚΟΝΑ 18: ΕΠΙΛΟΓΗ ΦΙΛΤΡΩΝ ΣΤΟ WEKA	82
ΕΙΚΟΝΑ 21: ΓΡΑΦΙΚΗ ΑΠΕΙΚΟΝΙΣΗ ΑΠΟΤΕΛΕΣΜΑΤΩΝ	90
ΕΙΚΟΝΑ 22: ΛΕΠΤΟΜΕΡΕΙΑ ΓΡΑΦΙΚΗΣ ΑΠΕΙΚΟΝΙΣΗΣ	90

ΕΙΣΑΓΩΓΗ

Η ΕΚΡΗΞΗ ΤΗΣ ΑΝΑΠΤΥΞΗΣ ΤΗΣ ΤΕΧΝΟΛΟΓΙΑΣ ΔΗΜΙΟΥΡΓΗΣΕ ΤΙΣ ΠΡΟΫΠΟΘΕΣΕΙΣ ΓΙΑ ΤΗΝ ΔΙΕΥΚΟΛΥΝΣΗ ΠΑΡΑΓΩΓΗΣ, ΔΙΑΚΙΝΗΣΗΣ ΚΑΙ ΑΠΟΘΗΚΕΥΣΗΣ ΜΕΓΑΛΟΥ ΟΓΚΟΥ ΠΛΗΡΟΦΟΡΙΩΝ. Η ΥΠΟΛΟΓΙΣΤΙΚΗ ΕΠΕΞΕΡΓΑΣΤΙΚΗ ΙΣΧΥΣ ΤΩΝ ΣΥΓΧΡΟΝΩΝ ΗΛΕΚΤΡΟΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ ΑΥΞΑΝΕΤΑΙ ΜΕ ΕΚΘΕΤΙΚΟΥΣ ΡΥΘΜΟΥΣ ΚΑΙ ΠΑΡΑΛΛΗΛΑ ΤΟ ΑΠΑΙΤΟΥΜΕΝΟ ΚΟΣΤΟΣ ΚΤΗΣΗΣ ΚΑΙ ΧΡΗΣΗΣ ΜΕΙΩΝΕΤΑΙ. ΑΠΟΤΕΛΕΣΜΑ ΑΥΤΟΥ ΕΙΝΑΙ ΠΛΕΟΝ, ΔΙΑΤΑΞΕΙΣ ΥΨΗΛΗΣ ΕΠΕΞΕΡΓΑΣΤΙΚΗΣ ΙΣΧΥΟΣ ΝΑ ΕΙΝΑΙ ΠΛΕΟΝ ΠΡΟΣΙΤΕΣ ΑΚΟΜΑ ΚΑΙ ΣΕ ΜΕΜΟΝΩΜΕΝΟΥΣ ΧΡΗΣΤΕΣ. ΤΟ ΓΕΓΟΝΟΣ ΑΥΤΟ ΔΙΕΥΚΟΛΥΝΕ ΤΗΝ ΕΠΕΞΕΡΓΑΣΙΑ ΜΕΓΑΛΩΝ ΣΥΝΟΛΩΝ ΔΕΔΟΜΕΝΩΝ ΚΑΙ ΟΔΗΓΗΣΕ ΣΤΗΝ ΑΝΑΠΤΥΞΗ ΜΕΓΑΛΩΝ ΑΠΟΘΗΚΩΝ ΔΕΔΟΜΕΝΩΝ. ΤΑΥΤΟΧΡΟΝΑ Η ΑΝΑΒΑΘΜΙΣΗ ΤΩΝ ΤΗΛΕΠΙΚΟΙΝΩΝΙΑΚΩΝ ΥΠΟΔΟΜΩΝ ΑΛΛΑ ΚΑΙ Η ΑΝΑΠΤΥΞΗ ΑΠΟΔΟΤΙΚΩΝ ΕΠΙΚΟΙΝΩΝΙΑΚΩΝ ΠΡΩΤΟΚΟΛΛΩΝ ΕΙΧΕ ΣΑΝ ΣΥΝΕΠΕΙΑ ΤΗΝ ΕΠΙΤΕΥΞΗ ΜΕΓΑΛΩΝΤΑΧΥΤΗΤΩΝ ΑΝΤΑΛΛΑΓΗΣ ΔΕΔΟΜΕΝΩΝ.

Η ΚΑΤΑΣΤΑΣΗ, ΠΟΥ ΔΙΑΜΟΡΦΩΘΗΚΕ ΣΑΝ ΣΥΝΕΠΕΙΑ ΤΩΝ ΕΞΕΛΙΞΕΩΝ ΑΥΤΩΝ, ΟΔΗΓΗΣΕ ΣΤΗΝ ΑΝΑΓΚΗ ΔΙΑΧΩΡΙΣΜΟΥ ΤΩΝ ΩΦΕΛΙΜΩΝ ΔΕΔΟΜΕΝΩΝ ΑΠΟ ΤΑ ΣΥΝΟΛΑ ΑΥΤΑ. ΈΤΣΙ ΑΝΑΖΗΤΗΘΗΚΑΝ ΤΡΟΠΟΙ ΓΙΑ ΤΗΝ ΕΞΟΡΥΞΗ ΤΗΣ ΟΥΣΙΑΣΤΙΚΗΣ ΓΝΩΣΗΣ ΜΕ ΤΗΝ ΜΟΡΦΗ ΠΡΟΤΥΠΩΝ ΠΟΥ ΕΠΙΒΕΒΑΙΩΝΟΥΝ ΤΙΣ ΜΕΜΟΝΩΜΕΝΕΣ ΜΟΝΑΔΕΣ ΔΕΔΟΜΕΝΩΝ. ΈΧΟΥΝ ΠΡΟΤΑΘΕΙ ΑΡΚΕΤΕΣ ΤΕΧΝΙΚΕΣ ΚΑΙ ΜΕΘΟΔΟΛΟΓΙΕΣ ΕΞΟΡΥΞΗΣ ΓΝΩΣΗΣ, ΕΝΩ ΑΠΑΝΤΑΤΑΙ ΚΑΙ ΣΕ ΠΛΗΘΟΣ ΔΙΑΦΟΡΕΤΙΚΩΝ ΕΦΑΡΜΟΓΩΝ. ΔΙΑΦΟΡΕΤΙΚΟΙ ΤΟΜΕΙΣ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ ΧΡΗΣΙΜΟΠΟΙΟΥΝΤΑΙ ΣΤΗΝ ΔΙΑΔΙΚΑΣΙΑ ΕΞΟΡΥΞΗΣ ΓΝΩΣΗΣ, ΟΠΩΣ ΠΛΗΡΟΦΟΡΙΚΗ (ΘΕΩΡΙΑ ΤΗΣ ΠΛΗΡΟΦΟΡΙΑΣ ΚΑΙ ΒΑΣΕΙΣ ΔΕΔΟΜΕΝΩΝ), ΤΕΧΝΙΚΕΣ ΟΠΤΙΚΟΠΟΙΗΣΗΣ, ΣΤΑΤΙΣΤΙΚΗ, ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ.

ΜΙΑ ΑΠΟ ΤΙΣ ΔΟΜΕΣ ΠΟΥ ΧΡΗΣΙΜΟΠΟΙΟΥΝΤΑΙ ΑΠΟΔΟΤΙΚΑ ΣΤΗ ΕΞΟΡΥΞΗ ΔΕΔΟΜΕΝΩΝ ΕΙΝΑΙ ΤΑ ΔΕΝΔΡΑ ΑΠΟΦΑΣΗΣ. ΠΡΟΚΕΙΤΑΙ ΓΙΑ ΜΕΘΟΔΟΛΟΓΙΑ ΠΟΥ ΧΡΗΣΙΜΟΠΟΙΕΙΤΑΙ ΣΕ ΠΕΡΙΠΤΩΣΕΙΣ ΠΟΥ ΑΠΑΙΤΕΙΤΑΙ Η ΔΙΑΔΟΧΙΚΗ ΛΗΨΗ ΑΠΟΦΑΣΕΩΝ ΓΙΑ ΤΗΝ ΕΠΙΛΥΣΗ

ΠΡΟΒΛΗΜΑΤΩΝ. ΑΠΕΙΚΟΝΙΖΕΙ ΓΡΑΦΙΚΑ ΤΗΝ ΑΛΛΗΛΟΥΧΙΑ ΤΩΝ ΠΙΘΑΝΩΝ ΑΠΟΦΑΣΕΩΝ ΠΟΥ ΟΔΗΓΟΥΝ ΣΕ ΕΝΑΛΛΑΚΤΙΚΕΣ ΕΠΙΛΥΣΕΙΣ ΤΟΥ ΙΔΙΟΥ ΠΡΟΒΛΗΜΑΤΟΣ.

ΤΟ ΥΠΟΛΟΙΠΟ ΤΗΣ ΕΡΓΑΣΙΑΣ ΕΙΝΑΙ ΔΟΜΗΜΕΝΟ ΩΣ ΕΞΗΣ:

- ΣΤΟ ΠΡΩΤΟ ΚΕΦΑΛΑΙΟ ΕΠΙΧΕΙΡΕΙΤΑΙ ΜΙΑ ΕΠΙΣΚΟΠΗΣΗ ΤΗΣ ΕΝΝΟΙΑΣ ΤΗΣ ΕΞΟΡΥΞΗΣ ΔΕΔΟΜΕΝΩΝ. ΠΕΡΙΓΡΑΦΕΤΑΙ ΤΟ ΠΩΣ ΕΞΕΛΙΣΣΕΤΑΙ ΣΑΝ ΔΙΑΔΙΚΑΣΙΑ ΚΑΙ ΣΕ ΠΟΙΟ ΜΕΡΟΣ ΤΗΣ ΛΕΙΤΟΥΡΓΙΑΣ ΤΗΣ ΔΗΜΙΟΥΡΓΙΑΣ ΓΝΩΣΗΣ ΑΝΤΙΣΤΟΙΧΕΙ. ΕΠΙΣΗΣ ΠΑΡΟΥΣΙΑΖΟΝΤΑΙ ΣΥΝΟΠΤΙΚΑ ΤΟΜΕΙΣ ΤΗΣ ΑΝΘΡΩΠΙΝΗΣ ΔΡΑΣΤΗΡΙΟΤΗΤΑΣ ΠΟΥ ΒΡΙΣΚΕΙ ΕΦΑΡΜΟΓΕΣ, ΕΝΩ ΑΝΑΦΕΡΟΝΤΑΙ ΚΑΙ ΖΗΤΗΜΑΤΑ ΠΟΥ ΠΡΟΚΥΠΤΟΥΝ ΜΕ ΤΗΝ ΙΔΙΩΤΙΚΟΤΗΤΑ ΤΩΝ ΥΠΟΚΕΙΜΕΝΩΝ ΠΟΥ ΣΥΜΜΕΤΕΧΟΥΝ ΣΕ ΔΙΑΔΙΚΑΣΙΕΣ ΕΞΟΡΥΞΗΣ ΓΝΩΣΗΣ.
- ΣΤΟ ΔΕΥΤΕΡΟ ΚΕΦΑΛΑΙΟ ΠΑΡΟΥΣΙΑΖΕΤΑΙ ΓΕΝΙΚΑ ΤΟ ΠΩΣ ΕΞΕΛΙΣΣΕΤΑΙ Η ΕΞΟΡΥΞΗ ΓΝΩΣΗΣ ΜΕ ΤΗΝ ΧΡΗΣΗ ΔΕΝΔΡΩΝ ΑΠΟΦΑΣΗΣ. ΠΑΡΟΥΣΙΑΖΟΝΤΑΙ ΕΠΙΣΗΣ ΤΑ ΒΑΣΙΚΑ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΤΩΝ ΚΥΡΙΟΤΕΡΩΝ ΣΧΕΤΙΚΩΝ ΑΛΓΟΡΙΘΜΩΝ.
- ΤΟ ΤΡΙΤΟ ΚΕΦΑΛΑΙΟ ΑΠΟΤΕΛΕΙ ΜΙΑ ΠΡΑΚΤΙΚΗ ΕΦΑΡΜΟΓΗ ΤΩΝ ΑΛΓΟΡΙΘΜΩΝ ΠΟΥ ΠΑΡΟΥΣΙΑΣΤΗΚΑΝ ΣΤΟ 2^ο ΚΕΦΑΛΑΙΟ. ΜΕΣΑ ΑΠΟ ΤΗΝ ΕΦΑΡΜΟΓΗ ΑΥΤΗ ΕΚΤΙΜΑΤΑΙ ΟΤΙ ΕΙΝΑΙ ΕΜΦΑΝΕΣΤΕΡΟΣ Ο ΤΡΟΠΟΣ ΜΕ ΤΟΝ ΟΠΟΙΟ ΕΞΕΛΙΣΣΟΝΤΑΙ.
- ΤΕΛΟΣ ΠΑΡΟΥΣΙΑΖΟΝΤΑΙ ΣΥΜΠΕΡΑΣΜΑΤΑ ΠΟΥ ΠΡΟΚΥΠΤΟΥΝ ΑΠΟ ΤΗΝ ΜΕΛΕΤΗ ΚΑΙ ΠΟΥ ΣΧΕΤΙΖΟΝΤΑΙ ΜΕ ΤΗΝ ΑΠΟΔΟΤΙΚΟΤΗΤΑ ΓΕΝΙΚΑ ΤΩΝ ΜΗΧΑΝΙΣΜΩΝ ΤΩΝ

ΔΕΝΔΡΩΝ ΑΠΟΦΑΣΗΣ ΣΤΗΝ ΕΞΟΥΣΙΑ ΓΝΩΣΗΣ ΑΛΛΑ ΚΑΙ
ΤΩΝ ΕΠΙΜΕΡΟΥΣ ΑΛΓΟΡΙΘΜΩΝ.

1. ΕΞΟΥΥΞΗ ΔΕΔΟΜΕΝΩΝ

1.1 ΟΡΙΣΜΟΣ

Η ΠΡΟΟΔΟΣ ΤΗΣ ΕΠΙΣΤΗΜΗΣ ΤΗΣ ΠΛΗΡΟΦΟΡΙΚΗΣ ΕΙΧΕ ΩΣ ΑΠΟΤΕΛΕΣΜΑ ΤΗΝ ΕΚΡΗΚΤΙΚΗ ΑΝΑΠΤΥΞΗ ΤΕΧΝΟΛΟΓΙΩΝ ΚΑΙ ΜΕΘΟΔΟΛΟΓΙΩΝ ΠΟΥ ΒΡΗΚΑΝ ΕΦΑΡΜΟΓΗ ΣΤΑ ΠΛΗΡΟΦΟΡΙΑΚΑ ΣΥΣΤΗΜΑΤΑ ΤΩΝ ΣΥΓΧΡΟΝΩΝ ΟΡΓΑΝΙΣΜΩΝ ΚΑΘΕ ΚΛΙΜΑΚΑΣ. ΣΗΜΑΝΤΙΚΟ ΣΥΣΤΑΤΙΚΟ ΤΩΝ ΠΛΗΡΟΦΟΡΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ ΑΥΤΩΝ ΑΠΟΤΕΛΟΥΝ ΤΑ ΔΕΔΟΜΕΝΑ ΠΟΥ ΔΙΑΚΙΝΟΥΝ, ΑΠΟΘΗΚΕΥΟΥΝ ΚΑΙ ΕΠΕΞΕΡΓΑΖΟΝΤΑΙ. ΣΤΙΣ ΠΕΡΙΣΣΟΤΕΡΕΣ ΠΕΡΙΠΤΩΣΕΙΣ ΤΑ ΔΕΔΟΜΕΝΑ ΑΠΟΤΕΛΟΥΝ ΚΑΙ ΤΟ ΠΟΛΥΤΙΜΟΤΕΡΟ ΑΓΑΘΟ ΤΩΝ ΣΥΣΤΗΜΑΤΩΝ ΑΥΤΩΝ. Η ΕΥΧΕΡΕΙΑ ΓΙΑ ΠΡΟΣΒΑΣΗ ΣΕ ΜΕΓΑΛΟ ΠΛΗΘΟΣ ΔΙΑΦΟΡΕΤΙΚΩΝ ΠΗΓΩΝ ΔΕΔΟΜΕΝΩΝ ΔΗΜΙΟΥΡΓΗΣΕ ΥΠΕΡΠΛΗΘΩΡΑ ΔΙΑΘΕΣΙΜΗΣ ΠΛΗΡΟΦΟΡΙΑΣ. ΜΕΓΑΛΟ ΠΟΣΟΣΤΟ ΤΟΥ ΟΓΚΟΥ ΑΥΤΗΣ ΤΗΣ ΠΛΗΡΟΦΟΡΙΑΣ ΔΕΝ ΕΙΝΑΙ ΩΦΕΛΙΜΟ. ΤΟ ΜΕΙΟΝΕΚΤΗΜΑ ΤΗΣ ΥΠΕΡΠΛΗΘΩΡΑΣ ΑΥΤΗΣ ΕΙΝΑΙ ΝΑ ΣΠΑΤΑΛΑΤΑΙ ΠΟΛΥΤΙΜΟΣ ΥΠΟΛΟΓΙΣΤΙΚΟΣ ΚΑΙ ΕΠΕΞΕΡΓΑΣΤΙΚΟΣ ΧΡΟΝΟΣ ΧΩΡΙΣ ΝΑ ΥΠΑΡΧΕΙ ΑΝΤΑΠΟΔΟΤΙΚΟ ΟΦΕΛΟΣ. ΈΝΑ ΑΚΟΜΑ ΠΡΟΒΛΗΜΑ ΠΟΥ ΔΗΜΙΟΥΡΓΗΘΗΚΕ ΑΠΟ ΤΙΣ ΑΝΑΒΑΘΜΙΣΕΙΣ ΣΤΟΝ ΤΟΜΕΑ ΤΗΣ ΠΛΗΡΟΦΟΡΙΚΗΣ ΗΤΑΝ ΚΑΙ ΟΙ ΑΥΞΗΜΕΝΕΣ ΑΝΑΓΚΕΣ ΓΙΑ ΠΟΛΥΠΛΟΚΗΣ ΜΟΡΦΗΣ ΔΕΔΟΜΕΝΩΝ, ΤΑ ΟΠΟΙΑ ΕΙΝΑΙ ΧΡΗΣΙΜΑ ΚΑΘΩΣ ΤΙΣ ΠΕΡΙΣΣΟΤΕΡΕΣ ΦΟΡΕΣ ΠΕΡΙΛΑΜΒΑΝΟΥΝ ΜΕΓΑΛΗ ΑΚΡΙΒΕΙΑ ΣΤΗΝ ΠΕΡΙΓΡΑΦΗ ΤΟΥΣ, ΩΣΤΟΣΟ ΑΠΑΙΤΟΥΝ ΕΞΕΙΔΙΚΕΥΜΕΝΕΣ ΚΑΙ ΑΠΑΙΤΗΤΙΚΕΣ ΣΕ ΧΡΟΝΟ ΤΕΧΝΙΚΕΣ ΓΙΑ ΤΗΝ ΕΚΜΕΤΑΛΛΕΥΣΗ ΤΟΥΣ.

ΤΟ ΑΠΟΤΕΛΕΣΜΑ ΗΤΑΝ ΟΙ ΠΑΡΑΔΟΣΙΑΚΕΣ ΜΕΘΟΔΟΙ ΕΛΕΓΧΟΥ, ΑΞΙΟΛΟΓΗΣΗΣ ΚΑΙ ΑΝΑΛΥΣΗΣ ΔΕΔΟΜΕΝΩΝ ΝΑ ΜΗΝ ΕΙΝΑΙ ΕΠΑΡΚΕΙΣ ΓΙΑ ΝΑ ΚΑΛΥΨΟΥΝ ΤΙΣ ΑΥΞΗΜΕΝΕΣ ΑΝΑΓΚΕΣ ΠΟΥ ΔΗΜΙΟΥΡΓΗΘΗΚΑΝ. ΤΟ ΖΗΤΗΜΑ ΑΥΤΟ ΧΡΗΖΕΙ ΑΠΟΤΕΛΕΣΜΑΤΙΚΗΣ ΑΝΤΙΜΕΤΩΠΙΣΗΣ, ΑΦΟΥ ΣΤΗΝ ΠΟΛΥΠΛΟΚΟΤΗΤΑ ΤΗΣ ΑΝΑΖΗΤΗΣΗΣ ΚΑΙ ΑΝΑΚΤΗΣΗΣ ΠΛΗΡΟΦΟΡΙΑΣ, ΠΡΟΣΤΙΘΕΤΑΙ ΚΑΙ ΑΥΤΗ ΤΗΣ ΔΙΑΛΟΓΗΣ ΤΟΥΣ ΚΑΙ ΤΗΣ ΠΑΡΑΓΩΓΗΣ ΣΥΜΠΕΡΑΣΜΑΤΩΝ ΑΠΟ ΑΥΤΑ.

ΠΑΡΑΛΛΗΛΑ, Η ΠΤΩΣΗ ΤΟΥ ΟΙΚΟΝΟΜΙΚΟΥ ΚΟΣΤΟΥΣ ΑΠΟΘΗΚΕΥΣΗΣ ΤΩΝ ΔΕΔΟΜΕΝΩΝ ΕΧΕΙ ΕΠΙΤΡΕΨΕΙ ΣΤΟΥΣ ΟΡΓΑΝΙΣΜΟΥΣ ΝΑ ΔΗΜΙΟΥΡΓΟΥΝ, ΜΕΓΑΛΕΣ ΔΙΑΤΑΞΕΙΣ ΑΠΟΘΗΚΕΥΣΗΣ ΔΕΔΟΜΕΝΩΝ (ΣΥΝΗΘΩΣ ΒΑΣΕΙΣ ΔΕΔΟΜΕΝΩΝ). ΜΙΑ ΑΠΟΘΗΚΗ ΔΕΔΟΜΕΝΩΝ ΜΠΟΡΕΙ ΝΑ ΑΠΟΘΗΚΕΥΕΙ ΜΕΓΑΛΕΣ ΠΟΣΟΤΗΤΕΣ ΔΕΔΟΜΕΝΩΝ ΑΠΟ ΣΥΓΚΕΚΡΙΜΕΝΕΣ ΚΑΤΗΓΟΡΙΕΣ ΚΑΙ ΕΠΙΤΡΕΠΕΙ ΤΗΝ ΑΚΡΙΒΗ ΚΑΤΗΓΟΡΙΟΠΟΙΗΣΗ ΤΟΥΣ ΑΛΛΑ ΚΑΙ ΤΗΝ ΤΑΧΕΙΑ ΑΝΑΚΤΗΣΗ. Η ΕΥΧΕΡΕΙΑ ΑΥΤΗ ΣΤΗΝ ΑΠΟΘΗΚΕΥΣΗ ΜΕΓΑΛΩΝ ΠΟΣΟΤΗΤΩΝ ΔΕΔΟΜΕΝΩΝ ΔΗΜΙΟΥΡΓΗΣΕ ΚΑΙ ΤΗΝ ΑΝΑΓΚΗ ΓΙΑ ΤΗΝ ΔΗΜΙΟΥΡΓΙΑ ΓΝΩΣΗΣ ΑΠΟ ΑΥΤΑ.

ΠΡΟΣΑΝΑΤΟΛΙΣΜΕΝΗ ΣΤΙΣ ΑΝΑΓΚΕΣ ΑΥΤΕΣ ΕΙΝΑΙ Η ΔΙΑΔΙΚΑΣΙΑ ΤΗΣ ΕΞΟΡΥΞΗΣ ΔΕΔΟΜΕΝΩΝ. ΑΠΟΤΕΛΕΙ ΜΕΡΟΣ ΤΗΣ ΕΥΡΥΤΕΡΗΣ ΔΙΑΔΙΚΑΣΙΑΣ ΤΗΣ ΕΞΟΡΥΞΗΣ ΓΝΩΣΗΣ ΑΠΟ ΜΕΓΑΛΑ ΣΥΝΟΛΑ ΔΕΔΟΜΕΝΩΝ, ΠΟΥ ΣΤΟΧΕΥΕΙ ΚΥΡΙΩΣ ΣΤΗΝ ΥΠΟΣΤΗΡΙΞΗ ΛΗΨΗΣ ΣΤΡΑΤΗΓΙΚΩΝ ΑΠΟΦΑΣΕΩΝ. Η ΕΞΟΡΥΞΗ ΔΕΔΟΜΕΝΩΝ (DATAMINING) ΕΙΝΑΙ Η ΔΙΑΔΙΚΑΣΙΑ ΑΝΑΚΑΛΥΨΗΣ (DISCOVERY) ΝΕΩΝ ΠΡΟΤΥΠΩΝ (PATTERNS) ΠΟΥ ΤΕΚΜΗΡΙΩΝΟΝΤΑΙ ΚΑΙ ΜΠΟΡΕΙ ΝΑ ΕΙΝΑΙ ΧΡΗΣΙΜΑ ΚΑΙ ΚΑΤΑΝΟΗΤΑ ΚΑΙ ΤΩΝ ΟΠΟΙΩΝ Η ΑΝΑΛΥΣΗ ΟΔΗΓΕΙ ΣΕ ΜΗ ΑΝΑΜΕΝΟΜΕΝΕΣ ΣΧΕΣΕΙΣ ΑΝΑΜΕΣΑ ΣΤΑ ΕΞΕΤΑΖΟΜΕΝΑ ΣΥΝΟΛΑ ΔΕΔΟΜΕΝΩΝ. ΔΕΝ ΑΞΙΟΠΟΙΕΙ ΑΠΛΑ ΤΗΝ ΙΔΙΑ ΠΛΗΡΟΦΟΡΙΑ ΑΛΛΑ ΒΑΣΙΖΟΜΕΝΗ ΣΕ ΑΥΤΗΝ ΣΤΟΧΕΥΕΙ ΣΤΗΝ ΠΑΡΑΓΩΓΗ ΩΦΕΛΙΜΗΣ ΓΝΩΣΗΣ. ΠΕΡΙΛΑΜΒΑΝΕΙ ΕΠΙΣΗΣ ΤΗΝ ΕΚΦΡΑΣΗ ΤΟΥΣ ΜΕ ΝΕΟΥΣ ΤΡΟΠΟΥΣ ΠΟΥ ΕΙΝΑΙ ΚΑΤΑΝΟΗΤΟΙ ΚΑΙ ΕΥΧΡΗΣΤΟΙ ΓΙΑ ΤΟΥΣ ΧΕΙΡΙΣΤΕΣ ΤΟΥΣ. ΧΡΗΣΙΜΟΠΟΙΕΙ ΤΕΧΝΙΚΕΣ:

- ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ (MACHINELEARNING): ΘΕΩΡΕΙΤΑΙ Η ΑΝΑΖΗΤΗΣΗ ΣΕ ΕΝΑ ΣΥΝΟΛΟ ΕΝΑΛΛΑΚΤΙΚΩΝ ΥΠΟΘΕΣΕΩΝ, ΕΚΕΙΝΗΣ ΠΟΥ ΤΑΙΡΙΑΖΕΙ ΠΕΡΙΣΣΟΤΕΡΟ ΣΤΑ ΥΠΟ ΕΞΕΤΑΣΗ ΔΕΔΟΜΕΝΑ ΚΑΙ ΣΤΗΝ ΕΝΔΕΧΟΜΕΝΗ ΠΡΟΫΠΑΡΧΟΥΣΑ ΓΝΩΣΗ.
- ΕΠΑΓΩΓΙΚΗΣ ΕΞΑΓΩΓΗΣ ΣΥΜΠΕΡΑΣΜΑΤΩΝ (INDUCTINGINFERENCE): ΑΦΟΡΑ ΤΗΝ ΕΞΑΓΩΓΗ ΣΥΜΠΕΡΑΣΜΑΤΩΝ ΜΕΣΑ ΑΠΟ ΤΗΝ ΜΕΛΕΤΗ ΕΜΠΕΙΡΙΚΩΝ ΔΕΔΟΜΕΝΩΝ ΚΑΙ ΠΑΡΑΔΕΙΓΜΑΤΩΝ.

- ΔΙΑΧΕΙΡΙΣΗΣ ΜΕΓΑΛΩΝ ΚΑΙ ΕΤΕΡΟΓΕΝΩΝ ΒΑΣΕΩΝ ΔΕΔΟΜΕΝΩΝ: ΠΡΟΚΕΙΤΑΙ ΓΙΑ ΤΕΧΝΙΚΕΣ ΠΟΥ ΠΡΟΣΑΡΜΟΖΟΥΝ ΔΕΔΟΜΕΝΑ ΠΡΟΕΡΧΟΜΕΝΑ ΑΠΟ ΔΙΑΦΟΡΕΤΙΚΕΣ ΚΑΙ ΕΤΕΡΟΓΕΝΕΙΣ ΒΑΣΕΙΣ ΔΕΔΟΜΕΝΩΝ ΣΕ ΜΟΡΦΕΣ ΤΕΤΟΙΕΣ ΠΟΥ ΝΑ ΕΙΝΑΙ ΕΠΕΞΕΡΓΑΣΙΜΕΣ ΚΑΙ ΔΙΑΧΕΙΡΙΣΙΜΕΣ ΑΠΟ ΤΑ ΥΠΟΛΟΓΙΣΤΙΚΑ ΣΥΣΤΗΜΑΤΑ.

Η ΕΞΟΡΥΞΗ ΔΕΔΟΜΕΝΩΝ ΠΗΡΕ ΤΟ ΟΝΟΜΑ ΤΟΥ ΑΠΟ ΤΙΣ ΟΜΟΙΟΤΗΤΕΣ ΜΕΤΑΞΥ ΤΗΣ ΑΝΑΖΗΤΗΣΗΣ ΓΙΑ ΤΙΣ ΠΟΛΥΤΙΜΕΣ ΠΛΗΡΟΦΟΡΙΕΣ ΣΕ ΜΙΑ ΜΕΓΑΛΗ ΒΑΣΗ ΔΕΔΟΜΕΝΩΝ ΚΑΙ ΤΗΝ ΕΞΟΡΥΞΗ ΧΡΥΣΟΥ ΑΠΟ ΜΙΑ ΜΕΓΑΛΗ ΕΚΤΑΣΗ ΚΑΘΩΣ ΚΑΙ ΣΤΙΣ ΔΥΟ ΠΕΡΙΠΤΩΣΕΙΣ ΑΠΑΙΤΕΙΤΑΙ ΦΙΛΤΡΑΡΙΣΜΑ ΣΕ ΜΙΑ ΜΕΓΑΛΗ ΠΟΣΟΤΗΤΑ ΠΡΩΤΟΓΕΝΟΥΣ ΥΛΙΚΟΥ ΠΡΟΚΕΙΜΕΝΟΥ ΝΑ ΑΝΑΚΑΛΥΦΘΕΙ Η - ΜΙΚΡΗ ΣΥΝΗΘΩΣ - ΠΟΣΟΤΗΤΑ ΠΟΥ ΕΧΕΙ ΑΞΙΑ.

ΓΕΝΙΚΑ, Η ΕΞΟΡΥΞΗ ΔΕΔΟΜΕΝΩΝ ΑΝΑΦΕΡΕΤΑΙ ΣΤΙΣ ΔΙΕΡΓΑΣΙΕΣ ΓΙΑ ΤΗΝ ΑΝΑΛΥΣΗ ΔΕΔΟΜΕΝΩΝ ΑΠΟ ΔΙΑΦΟΡΕΤΙΚΕΣ ΠΛΕΥΡΕΣ ΚΑΙ Η ΑΠΟΚΟΜΙΣΗ ΧΡΗΣΙΜΩΝ ΠΛΗΡΟΦΟΡΙΩΝ ΑΠΟ ΑΥΤΑ. ΠΛΗΡΟΦΟΡΙΕΣ ΠΟΥ ΜΠΟΡΟΥΝ ΝΑ ΧΡΗΣΙΜΕΥΣΟΥΝ ΣΤΗΝ ΠΡΟΒΛΕΨΗ ΜΕΛΛΟΝΤΙΚΩΝ ΚΑΤΑΣΤΑΣΕΩΝ ΚΑΙ ΘΑ ΒΟΗΘΗΣΟΥΝ ΣΤΗ ΛΗΨΗ ΣΩΣΤΩΝ ΑΠΟΦΑΣΕΩΝ. ΟΥΣΙΑΣΤΙΚΑ Η ΕΞΟΡΥΞΗ ΔΕΔΟΜΕΝΩΝ ΕΙΝΑΙ ΕΝΑΣ ΑΡΙΘΜΟΣ ΕΡΓΑΛΕΙΩΝ ΓΙΑ ΤΗΝ ΑΝΑΛΥΣΗ ΤΩΝ ΔΕΔΟΜΕΝΩΝ ΚΑΙ ΕΠΙΤΡΕΠΕΙ ΤΗΝ ΑΝΑΛΥΣΗ ΑΥΤΗ ΑΠΟ ΠΟΛΛΕΣ ΔΙΑΦΟΡΕΤΙΚΕΣ ΓΩΝΙΕΣ. ΜΕ ΤΟΝ ΟΡΟ ΔΕΔΟΜΕΝΑ ΕΝΝΟΟΥΜΕ ΚΑΘΕ ΓΕΓΟΝΟΣ, ΑΡΙΘΜΟ Η ΚΕΙΜΕΝΟ ΠΟΥ ΜΠΟΡΕΙ ΝΑ ΑΝΑΠΑΡΑΣΤΑΘΕΙ ΚΑΙ ΝΑ ΕΠΕΞΕΡΓΑΣΤΕΙ ΗΛΕΚΤΡΟΝΙΚΟΣ ΥΠΟΛΟΓΙΣΤΗΣ. Ο ΟΡΟΣ ΠΛΗΡΟΦΟΡΙΑ ΑΠΟ ΤΗΝ ΑΛΛΗ ΣΗΜΑΙΝΕΙ, ΤΑ ΠΡΟΤΥΠΑ ΚΑΙ ΤΙΣ ΣΧΕΣΕΙΣ ΑΝΑΜΕΣΑ ΣΤΑ ΔΕΔΟΜΕΝΑ. Η ΓΝΩΣΗ ΑΠΟΚΤΙΕΤΑΙ ΜΕ ΤΙΣ ΠΛΗΡΟΦΟΡΙΕΣ ΠΟΥ ΕΧΟΥΜΕ ΣΕ ΕΝΑ ΣΥΓΚΕΚΡΙΜΕΝΟ ΧΡΟΝΙΚΟ ΔΙΑΣΤΗΜΑ ΚΑΙ ΑΦΟΡΑ ΤΗΝ ΕΥΡΕΣΗ ΠΡΟΤΥΠΩΝ ΚΑΙ ΣΧΕΣΕΩΝ ΔΙΑΜΕΣΟΥ ΤΟΥ ΧΡΟΝΟΥ.

ΠΡΟΚΕΙΜΕΝΟΥ ΝΑ ΓΙΝΕΙ ΚΑΛΥΤΕΡΑ ΑΝΤΙΛΗΠΤΗ Η ΕΞΟΡΥΞΗ ΔΕΔΟΜΕΝΩΝ ΠΑΡΑΤΙΘΕΤΑΙ ΕΝΑ ΣΥΝΗΘΕΣ ΠΑΡΑΔΕΙΓΜΑ ΑΠΟ ΤΟΝ ΠΡΑΓΜΑΤΙΚΟ ΚΟΣΜΟ. ΈΣΤΩ ΕΝΑΣ ΠΕΛΑΤΗΣ ΚΑΤΑΣΤΗΜΑΤΟΣ Χ ΠΟΥ ΣΥΝΗΘΙΖΕΙ ΝΑ ΨΩΝΙΖΕΙ ΑΠΟ ΣΥΓΚΕΚΡΙΜΕΝΟ ΚΑΤΑΣΤΗΜΑ Ψ. ΤΟ ΚΑΤΑΣΤΗΜΑ ΕΧΕΙ ΠΡΟΜΗΘΕΥΣΕΙ ΤΟΝ ΠΕΛΑΤΗ ΜΕ ΜΙΑ ΚΑΡΤΑ ΤΗΝ

ΟΠΟΙΑ ΤΟΥ ΠΡΟΤΕΙΝΕΤΑΙ ΝΑ ΧΡΗΣΙΜΟΠΟΙΕΙ ΣΤΙΣ ΑΓΟΡΕΣ ΤΟΥ ΩΣΤΕ ΑΥΤΕΣ ΝΑ ΑΝΤΙΣΤΟΙΧΙΖΟΝΤΑΙ ΜΕ ΑΥΤΟΝ. ΜΕ ΒΑΣΗ ΤΙΣ ΠΡΟΗΓΟΥΜΕΝΕΣ ΑΓΟΡΕΣ ΤΟΥ ΠΕΛΑΤΗ ΤΟ ΠΛΗΡΟΦΟΡΙΑΚΟ ΣΥΣΤΗΜΑ ΤΟΥ ΚΑΤΑΣΤΗΜΑΤΟΣ, ΧΡΗΣΙΜΟΠΟΙΩΝΤΑΣ ΤΕΧΝΙΚΕΣ ΕΞΟΡΥΞΗΣ ΔΕΔΟΜΕΝΩΝ ΚΑΙ ΕΧΟΝΤΑΣ ΣΚΙΑΓΡΑΦΗΣΕΙ ΤΟ ΑΓΟΡΑΣΤΙΚΟ ΠΡΟΦΙΛ ΤΟΥ ΠΕΛΑΤΗ, ΕΙΝΑΙ ΣΕ ΘΕΣΗ ΝΑ ΠΡΟΤΕΙΝΕΙ ΠΡΟΪΟΝΤΑ ΓΙΑ ΤΙΣ ΤΡΕΧΟΥΣΕΣ ΚΑΙ ΤΙΣ ΜΕΛΛΟΝΤΙΚΕΣ ΑΓΟΡΕΣ ΤΟΥ.

ΟΙ ΔΙΑΔΙΚΑΣΙΕΣ ΤΗΣ ΕΞΟΡΥΞΗΣ ΔΕΔΟΜΕΝΩΝ ΒΑΣΙΖΟΝΤΑΙ ΣΕ ΜΕΘΟΔΟΛΟΓΙΕΣ ΑΠΟ ΤΗΝ ΠΛΗΡΟΦΟΡΙΚΗ, ΤΗΝ ΣΤΑΤΙΣΤΙΚΗ ΚΑΙ ΤΗΝ ΤΕΧΝΗΤΗ ΝΟΗΜΟΣΥΝΗ ΤΙΣ ΤΕΧΝΟΛΟΓΙΕΣ ΤΩΝ ΒΑΣΕΩΝ ΔΕΔΟΜΕΝΩΝ. Η ΔΙΑΦΟΡΟΠΟΙΗΣΗ ΤΗΣ ΕΞΟΡΥΞΗΣ ΓΝΩΣΗΣ ΣΕ ΣΧΕΣΗ ΜΕ ΤΗΝ ΣΤΑΤΙΣΤΙΚΗ ΑΝΑΛΥΣΗ ΕΓΚΕΙΤΑΙ ΣΤΟ ΓΕΓΟΝΟΣ ΟΤΙ ΜΕΣΩ ΑΥΤΗΣ ΑΝΑΖΗΤΟΥΝΤΑΙ ΣΥΣΧΕΤΙΣΤΙΚΑ ΠΡΟΤΥΠΑ ΜΕΤΑΞΥ ΤΩΝ ΣΤΑΤΙΣΤΙΚΩΣ ΠΑΡΑΚΟΛΟΥΘΗΜΕΝΩΝ ΜΕΤΑΒΛΗΤΩΝ. ΠΑΡΑΛΛΗΛΑ ΕΠΕΚΤΕΙΝΕΙ ΤΗΝ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ ΠΡΟΣ ΤΗΝ ΚΑΤΕΥΘΥΝΣΗ ΤΗΣ ΔΗΜΙΟΥΡΓΙΑΣ ΠΡΟΤΥΠΩΝ ΙΚΑΝΩΝ ΝΑ ΠΡΟΒΛΕΨΟΥΝ ΜΕΛΛΟΝΤΙΚΕΣ ΤΑΣΕΙΣ. ΤΑΥΤΟΧΡΟΝΑ ΑΓΓΙΖΕΙ ΚΑΙ ΘΕΜΑΤΑ ΠΡΟΣΒΑΣΗΣ ΚΑΙ ΔΙΑΧΕΙΡΙΣΗΣ ΔΕΔΟΜΕΝΩΝ ΠΟΥ ΕΙΝΑΙ ΑΠΟΘΗΚΕΥΜΕΝΑ ΣΕ (ΣΥΧΝΑ ΚΑΤΑΝΕΜΗΜΕΝΑ ΚΑΙ ΑΝΟΜΟΙΟΓΕΝΗ) ΣΥΣΤΗΜΑΤΑ ΒΑΣΕΩΝ ΔΕΔΟΜΕΝΩΝ. Η ΤΕΧΝΟΛΟΓΙΚΗ ΑΝΑΠΤΥΞΗ ΠΟΥ ΠΑΡΑΤΗΡΕΙΤΑΙ ΤΙΣ ΤΕΛΕΥΤΑΙΕΣ ΔΕΚΑΕΤΙΕΣ ΣΤΙΣ ΕΠΙΣΤΗΜΕΣ ΑΥΤΕΣ ΚΑΙ Ο ΜΕΓΑΛΟΣ ΒΑΘΜΟΣ ΑΥΤΟΜΑΤΟΠΟΙΗΣΗΣ ΠΟΥ ΕΠΙΤΥΓΧΑΝΕΤΑΙ ΣΤΗΝ ΥΛΟΠΟΙΗΣΗ ΚΑΙ ΤΗΝ ΕΚΤΕΛΕΣΗ ΤΩΝ ΣΧΕΤΙΚΩΝ ΑΛΓΟΡΙΘΜΩΝ ΚΑΘΙΣΤΑ ΤΗΝ ΕΚΤΕΛΕΣΗ ΤΟΥΣ ΤΑΧΕΙΑ ΚΑΙ ΑΚΡΙΒΗ.



ΕΙΚΟΝΑ 1: Η ΤΟΠΟΘΕΤΗΣΗ ΤΗΣ ΕΞΟΡΥΞΗΣ ΔΕΔΟΜΕΝΩΝ ΩΣ ΠΡΟΣ ΑΛΛΕΣ ΜΕΘΟΔΟΛΟΓΙΕΣ ΚΑΙ ΤΕΧΝΙΚΕΣ (ΠΗΓΗ:DATA MINING, [HTTP://WWW.SAEDSAYAD.COM/DATA MINING.HTM](http://www.saedsayad.com/data_mining.htm)).

ΔΥΟ ΕΙΝΑΙ ΤΑ ΜΟΝΤΕΛΑ ΔΙΕΝΕΡΓΕΙΑΣ ΕΞΟΡΥΞΗΣ ΔΕΔΟΜΕΝΩΝ ΜΕ ΒΑΣΗ ΤΟ ΠΑΡΑΓΟΜΕΝΟ ΩΦΕΛΙΜΟ ΑΠΟΤΕΛΕΣΜΑ:

- ΜΟΝΤΕΛΑ ΠΡΟΒΛΕΨΗΣ (PREDICTIVEMODELS). Η ΕΚΤΙΜΗΣΗ ΤΩΝ ΜΕΛΛΟΝΤΙΚΩΝ ΤΙΜΩΝ ΜΕΤΑΒΛΗΤΩΝ ΒΑΣΙΖΕΤΑΙ ΣΤΙΣ ΤΡΕΧΟΥΣΕΣ ΤΙΜΕΣ ΤΟΥΣ Η ΣΕ ΤΙΜΕΣ ΑΛΛΩΝ ΜΕ ΤΙΣ ΟΠΟΙΕΣ ΕΧΟΥΝ ΚΑΠΟΙΑ ΣΧΕΣΗ ΜΕΤΑΞΥ ΤΟΥΣ. ΟΙ ΤΕΛΕΥΤΑΙΕΣ ΑΠΟΤΕΛΟΥΝ ΤΟ ΥΠΟΒΑΘΡΟ ΣΤΟ ΟΠΟΙΟ ΘΑ ΣΤΗΡΙΧΘΕΙ Η ΠΑΡΑΓΩΓΗ ΠΡΟΤΥΠΩΝ,ΠΟΥ ΟΤΑΝ ΘΑ ΕΦΑΡΜΟΖΟΝΤΑΙ ΜΕΛΛΟΝΤΙΚΑ ΔΕΔΟΜΕΝΑ, ΘΑ ΠΑΡΕΧΕΤΑΙ ΜΙΑ ΠΡΟΒΛΕΨΗ ΣΧΕΤΙΚΑ ΜΕ ΤΗΝ ΠΟΡΕΙΑ ΤΩΝ ΔΙΑΔΙΚΑΣΙΩΝ ΣΤΙΣ ΟΠΟΙΕΣ ΑΠΟΤΕΛΟΥΝ ΠΑΡΑΜΕΤΡΟΥΣ – ΕΙΣΟΔΟΥΣ. ΓΙΑ ΠΑΡΑΔΕΙΓΜΑ ΑΝ ΑΠΟ ΤΗΝ ΜΕΛΕΤΗ ΤΟΥ ΠΡΟΦΙΛ ΤΗΣ ΑΓΟΡΑΣΤΙΚΗΣ ΣΥΜΠΕΡΙΦΟΡΑΣ ΤΩΝ ΠΕΛΑΤΩΝ ΜΙΑΣ ΥΠΕΡΑΓΟΡΑΣ (ΣΟΥΠΕΡ ΜΑΡΚΕΤ) ΠΡΟΚΥΨΕΙ ΟΤΙ ΤΗΝ ΕΟΡΤΑΣΤΙΚΗ ΠΕΡΙΟΔΟ ΤΩΝ ΧΡΙΣΤΟΥΓΕΝΝΩΝ ΚΑΙ ΤΟΥ ΝΕΟΥ ΕΤΟΥΣ

ΠΑΡΑΤΗΡΕΙΤΑΙ ΑΥΞΗΣΗ ΤΩΝ ΠΩΛΗΣΕΩΝ ΣΟΚΟΛΑΤΟΕΙΔΩΝ ΤΟΤΕ Η ΔΙΕΥΘΥΝΣΗ ΘΑ ΠΡΟΣΠΑΘΗΣΕΙ ΝΑ ΠΡΟΣΑΡΜΟΣΕΙ ΤΗΝ ΠΡΟΩΘΗΤΙΚΗ ΤΗΣ ΣΤΡΑΤΗΓΙΚΗ ΣΤΗΝ ΠΡΟΣΕΛΚΥΣΗ ΚΑΤΑΝΑΛΩΤΩΝ ΣΟΚΟΛΑΤΟΕΙΔΩΝ. ΧΑΡΑΚΤΗΡΙΣΤΙΚΟ ΤΩΝ ΜΟΝΤΕΛΩΝ ΑΥΤΩΝ ΕΙΝΑΙ ΤΟ ΟΤΙ Η ΑΠΟΤΕΛΕΣΜΑΤΙΚΟΤΗΤΑ ΤΟΥΣ ΕΙΝΑΙ ΕΥΘΕΩΣ ΑΝΑΛΟΓΗ ΤΟΥ ΠΛΗΘΟΥΣ ΤΩΝ ΔΙΑΘΕΣΙΜΩΝ ΣΤΟΙΧΕΙΩΝ. ΑΥΤΟ ΣΗΜΑΙΝΕΙ ΟΤΙ ΠΡΟΚΕΙΜΕΝΟΥ ΝΑ ΠΑΡΑΧΘΟΥΝ ΕΥΣΤΟΧΑ ΚΑΙ ΑΚΡΙΒΗ ΠΡΟΤΥΠΑ ΧΡΕΙΑΖΕΤΑΙ ΝΑ ΥΠΑΡΧΕΙ ΔΙΑΘΕΣΙΜΟΤΗΤΑ ΑΡΚΕΤΩΝ ΩΦΕΛΙΜΩΝ ΔΕΔΟΜΕΝΩΝ.

· ΠΕΡΙΓΡΑΦΙΚΑ ΜΟΝΤΕΛΑ (DESCRIPTIVE MODELS). ΠΡΟΚΕΙΤΑΙ ΓΙΑ ΜΟΝΤΕΛΑ ΠΟΥ ΣΤΟΧΕΥΟΥΝ ΣΤΗΝ ΑΝΑΔΥΣΗ ΤΑΣΕΩΝ ΚΑΙ ΚΑΝΟΝΩΝ ΠΟΥ ΒΑΣΙΖΟΝΤΑΙ ΣΕ ΤΡΕΧΟΥΣΕΣ ΠΑΡΑΤΗΡΗΣΕΙΣ ΚΑΙ ΠΟΥ ΔΕΝ ΕΙΝΑΙ ΡΗΤΑ ΚΑΙ ΠΡΟΦΑΝΩΣ ΔΙΑΘΕΣΙΜΑ. ΟΙ ΠΑΡΑΤΗΡΗΣΕΙΣ ΑΥΤΕΣ ΣΤΟΙΧΕΙΟΘΕΤΟΥΝ ΚΑΝΟΝΕΣ ΟΙ ΟΠΟΙΟΙ ΕΝΔΕΧΕΤΑΙ ΝΑ ΠΕΡΙΛΑΜΒΑΝΟΥΝ ΚΑΙ ΕΝΑ ΠΟΣΟΣΤΟ ΥΠΟΚΕΙΜΕΝΙΚΟΤΗΤΑΣ ΑΝΑΛΟΓΑ ΜΕ ΤΟ ΠΟΙΟΣ ΠΡΑΓΜΑΤΟΠΟΙΕΙ ΤΗΝ ΜΕΛΕΤΗ ΚΑΙ ΜΕ ΠΟΙΟ ΑΝΤΙΚΕΙΜΕΝΙΚΟ ΣΚΟΠΟ. ΈΤΣΙ ΟΙ ΚΑΝΟΝΕΣ ΑΥΤΟΙ ΜΠΟΡΕΙ ΝΑ ΣΥΝΤΙΘΕΝΤΑΙ ΥΠΟ ΤΗΝ ΕΠΙΡΡΟΗ ΕΝΟΣ Η ΠΕΡΙΣΣΟΤΕΡΩΝ ΑΠΟ ΤΟΥΣ ΕΞΗΣ ΠΑΡΑΓΟΝΤΕΣ: ΒΕΒΑΙΟΤΗΤΑ ΚΑΙ ΨΕΥΔΟΒΕΒΑΙΟΤΗΤΑ, ΠΡΟΣΔΟΚΩΜΕΝΟ ΟΦΕΛΟΣ, ΥΠΟΣΧΕΘΕΙΣΑ ΠΡΟΟΠΤΙΚΗ Η ΤΗΝ ΜΕΤΑΝΟΙΑ ΓΙΑ ΑΠΟΦΑΣΗ ΠΟΥ ΛΗΦΘΗΚΕ ΣΤΟ ΠΑΡΕΛΘΟΝ ΚΑΙ ΟΔΗΓΗΣΕ ΣΕ ΑΡΝΗΤΙΚΑ ΑΠΟΤΕΛΕΣΜΑΤΑ. ΓΙΑ ΠΑΡΑΔΕΙΓΜΑ ΕΣΤΩ ΠΑΡΑΤΗΡΕΙΤΑΙ ΟΤΙ Η ΠΟΡΕΙΑ ΤΟΥ ΧΡΗΜΑΤΙΣΤΗΡΙΟΥ ΜΕΤΑ ΑΠΟ ΜΙΑ ΑΡΝΗΤΙΚΗ ΚΑΜΠΗ ΟΔΗΓΕΙΤΑΙ ΕΚ ΝΕΟΥ ΑΜΕΣΑ ΣΕ ΑΝΟΔΙΚΕΣ ΤΑΣΕΙΣ ΑΛΛΑ ΣΤΟ ΠΑΡΕΛΘΟΝ ΕΝΑΣ ΑΡΙΘΜΟΣ ΕΠΙΧΕΙΡΗΣΕΩΝ ΚΑΤΑΣΤΡΑΦΗΚΑΝ ΛΟΓΩ ΜΕΓΑΛΩΝ ΕΠΕΝΔΥΣΕΩΝ ΠΟΥ ΕΚΑΝΑΝ ΑΚΟΛΟΥΘΩΝΤΑΣ ΑΥΤΟΝ ΤΟΝ ΚΑΝΟΝΑ. ΜΙΑ ΤΡΙΤΗ ΕΠΙΧΕΙΡΗΣΗ ΠΟΥ ΓΝΩΡΙΖΕΙ ΤΑ ΔΕΔΟΜΕΝΑ ΑΥΤΑ

ΕΝΔΕΧΟΜΕΝΩΣ ΘΑ ΠΑΡΑΞΕΙ ΕΝΑΝ ΚΑΝΟΝΑ ΠΟΥ ΘΑ ΠΡΟΣΤΑΞΕΙ ΣΕ ΤΕΤΟΙΕΣ ΠΕΡΙΟΔΟΥΣ ΝΑ ΠΡΑΓΜΑΤΟΠΟΙΟΥΝΤΑΙ ΣΥΓΚΡΑΤΗΜΕΝΕΣ ΕΠΕΝΔΥΣΕΙΣ ΚΑΙ ΜΕ ΒΑΣΗ ΤΟΝ ΚΑΝΟΝΑ ΑΥΤΟ ΝΑ ΑΠΟΦΑΣΙΖΕΙ ΣΤΟ ΜΕΛΛΟΝ.

ΜΕ ΒΑΣΗ ΤΗΝ ΜΟΡΦΗ ΠΟΥ ΠΑΡΟΥΣΙΑΖΕΤΑΙ ΤΟ ΑΠΟΤΕΛΕΣΜΑ ΤΟΥΣ ΟΙ ΤΕΧΝΙΚΕΣ ΕΞΟΡΥΞΗΣ ΔΕΔΟΜΕΝΩΝ ΔΙΑΚΡΙΝΟΝΤΑΙ ΣΕ:

- ΤΕΧΝΙΚΕΣ ΥΠΟ ΕΠΙΒΛΕΨΗ: ΠΡΟΚΕΙΤΑΙ ΓΙΑ ΤΕΧΝΙΚΕΣ ΟΠΟΥ Η ΟΛΗ ΠΟΡΕΙΑ ΤΗΣ ΕΞΟΡΥΞΗΣ ΔΕΔΟΜΕΝΩΝ ΕΠΑΝΑΠΡΟΣΔΙΟΡΙΖΕΤΑΙ ΚΑΤΑ ΤΗΝ ΕΞΕΛΙΞΗ ΤΗΣ ΚΑΙ ΑΝΑΛΟΓΑ ΜΕ ΤΗΝ ΤΡΕΧΟΥΣΑ ΠΟΡΕΙΑ ΤΗΣ. ΣΤΙΣ ΤΕΧΝΙΚΕΣ ΑΥΤΕΣ ΠΕΡΙΛΑΜΒΑΝΟΝΤΑΙ Η ΚΑΤΗΓΟΡΙΟΠΟΙΗΣΗ (CLASSIFICATION) ΤΩΝ ΔΕΔΟΜΕΝΩΝ ΣΕ ΚΛΑΣΕΙΣ ΑΝΑΛΟΓΑ ΜΕ ΤΙΣ ΤΙΜΕΣ ΤΩΝ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ ΤΟΥΣ ΚΑΙ Η ΠΡΟΒΛΕΨΗ (PREDICTION) ΜΕ ΒΑΣΗ ΚΑΘΟΡΙΣΜΕΝΑ ΠΡΟΤΥΠΑ.
- ΧΩΡΙΣ ΕΠΙΒΛΕΨΗ: ΠΡΟΚΕΙΤΑΙ ΓΙΑ ΤΕΧΝΙΚΕΣ ΟΙ ΟΠΟΙΕΣ ΣΤΟ ΤΕΛΟΣ ΤΗΣ ΔΙΑΔΙΚΑΣΙΑΣ ΕΛΕΓΧΟΥΝ ΑΝ ΑΠΕΧΟΥΝ ΚΑΙ ΠΟΣΟ ΑΠΟ ΤΑ ΠΡΟΣΔΟΚΟΜΕΝΑ ΑΠΟΤΕΛΕΣΜΑΤΑ. ΣΤΙΣ ΤΕΧΝΙΚΕΣ ΑΥΤΕΣ ΠΕΡΙΛΑΜΒΑΝΟΝΤΑΙ Η ΟΜΑΔΟΠΟΙΗΣΗ (CLUSTERING) ΤΩΝ ΔΕΔΟΜΕΝΩΝ ΜΕ ΣΧΕΤΙΚΑ ΚΟΙΝΑ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΚΑΙ ΟΙ ΣΥΣΧΕΤΙΣΕΙΣ ΤΩΝ ΔΕΔΟΜΕΝΩΝ ΜΕ ΒΑΣΗ ΤΙΣ ΤΙΜΕΣ ΤΩΝ ΕΠΙΜΕΡΟΥΣ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ ΤΟΥΣ [1][2][21].

1.2 ΠΕΡΙΓΡΑΦΗ

Ο ΣΤΡΑΤΗΓΙΚΟΣ ΣΤΟΧΟΣ ΤΩΝ ΜΕΘΟΔΟΛΟΓΙΩΝ ΤΗΣ ΕΞΟΡΥΞΗΣ ΔΕΔΟΜΕΝΩΝ ΕΙΝΑΙ Η ΔΥΝΑΤΟΤΗΤΑ ΠΑΡΑΓΩΓΗΣ ΓΝΩΣΗΣ ΑΠΟ ΣΥΝΟΛΑ ΔΕΔΟΜΕΝΩΝ. Ο ΣΤΡΑΤΗΓΙΚΟΣ ΑΥΤΟΣ ΣΤΟΧΟΣ ΔΥΝΑΤΑΙ ΝΑ ΠΡΑΓΜΑΤΟΠΟΙΗΘΕΙ ΜΕ ΤΗΝ ΕΚΠΛΗΡΩΣΗ ΤΩΝ ΕΞΗΣ ΤΑΚΤΙΚΩΝ ΣΤΟΧΩΝ:

- ΠΡΟΒΛΕΨΗ: ΜΕΛΕΤΑΤΑΙ Η ΣΥΜΠΕΡΙΦΟΡΑ ΤΩΝ ΤΙΜΩΝ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ ΕΓΓΡΑΦΩΝ ΠΟΥ ΠΡΟΕΡΧΟΝΤΑΙ ΑΠΟ ΔΙΑΦΟΡΑ ΣΥΝΟΛΑ ΔΕΔΟΜΕΝΩΝ ΚΑΙ ΠΩΣ ΔΙΑΜΟΡΦΩΝΟΝΤΑΙ ΟΙ ΜΕΤΑΞΥ ΤΟΥΣ ΣΧΕΣΕΙΣ. ΑΠΟ ΤΗΝ ΜΕΛΕΤΗ ΑΥΤΗ ΔΙΕΡΕΥΝΑΤΑΙ ΤΟ ΑΝ ΕΝΥΠΑΡΧΕΙ ΚΑΠΟΙΟ ΠΡΟΤΥΠΟ ΤΟ ΟΠΟΙΟ ΑΠΟ ΤΙΣ ΤΙΜΕΣ ΣΥΓΚΕΚΡΙΜΕΝΩΝ ΠΑΡΑΜΕΤΡΩΝ ΝΑ ΕΙΝΑΙ ΕΦΙΚΤΟ ΝΑ ΕΞΑΧΘΕΙ ΧΩΡΙΣ ΝΑ ΕΙΝΑΙ ΓΝΩΣΤΗ Η ΤΙΜΗ ΜΙΑ ΧΡΗΣΙΜΗΣ ΠΑΡΑΜΕΤΡΟΥ. ΣΤΗΝ ΠΡΑΓΜΑΤΙΚΟΤΗΤΑ ΓΙΝΕΤΑΙ ΠΡΟΣΠΑΘΕΙΑ ΝΑ ΑΝΙΧΝΕΥΘΕΙ Η ΥΠΑΡΞΗ ΚΑΝΟΝΩΝ ΙΚΑΝΩΝ ΝΑ ΠΡΟΣΔΙΟΡΙΖΟΥΝ ΤΙΜΕΣ ΠΑΡΑΜΕΤΡΩΝ ΑΝ ΕΙΝΑΙ ΓΝΩΣΤΕΣ ΑΛΛΕΣ ΣΧΕΤΙΖΟΜΕΝΕΣ ΜΕ ΑΥΤΗ Η ΑΚΟΜΑ ΚΑΙ ΤΗΣ ΙΔΙΑΣ ΣΕ ΔΕΔΟΜΕΝΕΣ ΣΥΝΘΗΚΕΣ. Η ΠΡΟΒΛΕΨΗ ΑΥΤΗ ΑΠΟΤΕΛΕΙ ΖΩΤΙΚΗΣ ΣΗΜΑΣΙΑΣ ΕΡΓΑΛΕΙΟ ΓΙΑ ΤΗΝ ΧΑΡΑΞΗ ΤΗΣ ΜΕΛΛΟΝΤΙΚΗΣ ΠΟΛΙΤΙΚΗΣ ΤΩΝ ΟΡΓΑΝΙΣΜΩΝ ΜΕ ΤΟ ΜΙΚΡΟΤΕΡΟ ΔΥΝΑΤΟ ΡΙΣΚΟ Η ΓΙΑ ΤΟΝ ΠΡΟΣΑΝΑΤΟΛΙΣΜΟ ΤΗΣ ΕΠΙΣΤΗΜΟΝΙΚΗΣ ΕΡΕΥΝΑΣ.
- ΑΝΑΓΝΩΡΙΣΗ: Η ΣΥΜΠΕΡΙΦΟΡΑ ΚΑΙ ΟΙ ΣΧΕΣΕΙΣ ΤΩΝ ΤΙΜΩΝ ΤΩΝ ΠΑΡΑΜΕΤΡΩΝ ΣΤΙΣ ΠΕΡΙΠΤΩΣΕΙΣ ΑΥΤΕΣ ΕΙΝΑΙ ΣΕ ΘΕΣΗ ΝΑ ΠΡΟΣΔΙΟΡΙΣΟΥΝ ΤΗΝ ΥΠΑΡΞΗ ΚΑΠΟΙΟΥ ΦΑΙΝΟΜΕΝΟΥ (ΣΥΜΒΑΝ, ΔΡΑΣΤΗΡΙΟΤΗΤΑ, ΔΙΑΜΟΡΦΩΘΕΙΣΑ ΚΑΤΑΣΤΑΣΗ).
- ΠΕΡΙΓΡΑΦΗ: ΣΤΙΣ ΠΕΡΙΠΤΩΣΕΙΣ ΑΥΤΕΣ Ο ΣΤΟΧΟΣ ΕΙΝΑΙ ΝΑ ΔΙΑΜΟΡΦΩΘΕΙ ΜΙΑ ΓΕΝΙΚΗ ΠΕΡΙΓΡΑΦΗ Η ΟΠΟΙΑ ΝΑ ΤΑΥΤΙΖΕΤΑΙ ΜΕ ΤΗΝ ΜΕΓΙΣΤΗ ΔΥΝΑΤΗ ΑΚΡΙΒΕΙΑ ΜΕ ΕΝΑ ΜΕΓΑΛΟ ΣΥΝΟΛΟ ΔΕΔΟΜΕΝΩΝ. ΜΕ ΤΟΝ ΤΡΟΠΟ ΑΥΤΟ ΜΕΓΑΛΟΙ ΟΓΚΟΙ ΔΕΔΟΜΕΝΩΝ ΜΠΟΡΟΥΝ ΝΑ ΠΕΡΙΓΡΑΦΟΥΝ ΜΕ ΠΕΡΙΕΚΤΙΚΕΣ ΚΑΙ ΣΥΜΠΥΚΝΩΜΕΝΕΣ ΑΠΕΙΚΟΝΙΣΕΙΣ ΚΑΙ ΝΑ ΕΙΝΑΙ ΕΤΣΙ ΕΥΚΟΛΟΤΕΡΑ ΕΚΜΕΤΑΛΛΕΥΣΙΜΟΙ.
- ΤΑΞΙΝΟΜΗΣΗ: Η ΤΑΞΙΝΟΜΗΣΗ ΤΩΝ ΜΕΓΑΛΩΝ ΟΓΚΩΝ ΔΕΔΟΜΕΝΩΝ ΣΥΝΙΣΤΑΤΑΙ ΣΤΟΝ ΟΡΙΣΜΟ ΚΛΑΣΕΩΝ ΔΕΔΟΜΕΝΩΝ ΚΑΘΩΣ ΚΑΙ ΤΟΝ ΠΡΟΣΔΙΟΡΙΣΜΟ ΤΟΥ ΚΑΘΕ

ΣΤΟΙΧΕΙΟΥ ΤΩΝ ΣΥΝΟΛΩΝ ΔΕΔΟΜΕΝΩΝ ΤΗΣ ΚΛΑΣΗΣ ΣΤΗΝ ΟΠΟΙΑ ΑΝΗΚΕΙ. Η ΕΝΤΑΞΗ ΣΤΗΝ ΚΛΑΣΗ ΓΙΝΕΤΑΙ ΜΕ ΒΑΣΗ ΤΙΣ ΤΙΜΕΣ ΠΟΥ ΕΧΕΙ ΣΤΙΣ ΠΑΡΑΜΕΤΡΟΥΣ ΠΟΥ ΤΟ ΤΑΥΤΟΠΟΙΟΥΝ ΚΑΙ ΤΟ ΧΑΡΑΚΤΗΡΙΖΟΥΝ. Η ΤΑΞΙΝΟΜΗΣΗ ΔΙΕΥΚΟΛΥΝΕΙ ΤΗΝ ΔΙΑΧΕΙΡΙΣΗ ΠΟΛΥΠΛΗΘΩΝ ΣΥΝΟΛΩΝ ΑΦΟΥ ΑΥΤΗ – ΑΝ ΕΙΝΑΙ ΑΡΚΟΥΝΤΟΣ ΑΚΡΙΒΗΣ – ΜΕΙΩΝΕΙ ΤΗΝ ΠΟΛΥΠΛΟΚΟΤΗΤΑ ΤΗΣ ΑΝΤΙΜΕΤΩΠΙΣΗΣ ΤΩΝ ΔΙΑΦΟΡΕΤΙΚΩΝ ΠΕΡΙΣΤΑΣΕΩΝ.

ΑΠΟ ΤΟΝ ΟΡΙΣΜΟ ΤΗΣ ΕΞΟΡΥΞΗΣ ΔΕΔΟΜΕΝΩΝ ΔΙΑΦΑΙΝΕΤΑΙ ΟΤΙ ΕΙΝΑΙ ΜΙΑ ΔΙΑΔΙΚΑΣΙΑ Η ΟΠΟΙΑ ΠΕΡΙΛΑΜΒΑΝΕΙ ΤΑ ΕΞΗΣ ΕΠΙΜΕΡΟΥΣ ΣΤΑΔΙΑ:

- ΣΤΑΤΙΣΤΙΚΗ ΕΠΕΞΕΡΓΑΣΙΑ: ΤΑ ΣΥΝΟΛΑ ΤΩΝ ΔΕΔΟΜΕΝΩΝ ΕΛΕΓΧΟΝΤΑΙ ΚΑΙ ΑΞΙΟΛΟΓΟΥΝΤΑΙ ΜΕ ΒΑΣΗ ΣΤΑΤΙΣΤΙΚΕΣ ΜΕΤΡΙΚΕΣ ΠΡΟΚΕΙΜΕΝΟΥ ΝΑ ΕΝΤΟΠΙΣΤΟΥΝ ΠΡΟΤΥΠΑ ΠΟΥ ΝΑ ΕΙΝΑΙ ΙΚΑΝΑ ΝΑ ΠΡΟΒΛΕΠΟΥΝ ΤΗΝ ΣΥΜΠΕΡΙΦΟΡΑ ΤΩΝ ΤΙΜΩΝ ΤΟΥΣ. ΑΝΑΛΟΓΑ ΜΕ ΤΗΝ ΦΥΣΗ ΤΩΝ ΟΝΤΟΤΗΤΩΝ ΠΟΥ ΕΞΕΤΑΖΟΝΤΑΙ ΧΡΗΣΙΜΟΠΟΙΟΥΝΤΑΙ ΚΑΤΑΛΛΗΛΑ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΤΟΥΣ. ΣΤΙΣ ΤΙΜΕΣ ΤΟΥΣ ΕΦΑΡΜΟΖΟΝΤΑΙ ΚΑΤΑΛΛΗΛΕΣ ΓΙΑ ΚΑΘΕ ΠΕΡΙΠΤΩΣΗ ΣΥΝΑΡΤΗΣΕΙΣ ΚΑΙ ΑΝΑΛΥΟΝΤΑΙ ΤΑ ΑΠΟΤΕΛΕΣΜΑΤΑ ΤΟΥΣ ΜΕ ΒΑΣΗ ΚΑΘΟΡΙΣΜΕΝΑ ΦΙΛΤΡΑ ΜΕ ΤΡΟΠΟ ΤΕΤΟΙΟ ΩΣΤΕ ΝΑ ΕΙΝΑΙ ΕΦΙΚΤΟ ΝΑ ΕΝΤΟΠΙΣΤΟΥΝ ΣΥΓΚΕΚΡΙΜΕΝΕΣ ΤΑΣΕΙΣ ΤΩΝ ΔΕΔΟΜΕΝΩΝ.
- ΑΝΑΔΡΟΜΙΚΗ ΑΝΑΛΥΣΗ: ΑΠΟ ΤΟ ΠΑΡΑΠΑΝΩ ΒΗΜΑ ΕΧΟΥΝ ΠΑΡΑΧΘΕΙ ΣΤΑΤΙΣΤΙΚΑ ΠΡΟΤΥΠΑ ΤΗΣ ΣΥΜΠΕΡΙΦΟΡΑΣ ΤΩΝ ΔΕΔΟΜΕΝΩΝ. ΑΥΤΑ ΕΛΕΓΧΟΝΤΑΙ ΚΑΙ ΑΞΙΟΛΟΓΟΥΝΤΑΙ ΣΕ ΣΧΕΣΗ ΜΕ ΣΧΕΤΙΚΑ ΠΡΟΤΥΠΑ ΠΟΥ ΕΧΟΥΝ ΑΝΑΠΤΥΧΘΕΙ ΚΑΙ ΕΚΠΟΝΗΘΕΙ ΜΕ ΒΑΣΗ ΠΑΛΑΙΟΤΕΡΕΣ ΑΝΤΙΣΤΟΙΧΕΣ ΔΙΑΔΙΚΑΣΙΕΣ ΚΑΙ ΤΑ ΟΠΟΙΑ ΕΧΟΥΝ ΕΛΕΓΧΘΕΙ ΕΠΑΡΚΩΣ ΓΙΑ ΤΗΝ ΕΓΚΥΡΟΤΗΤΑ ΤΟΥΣ. ΑΠΟ ΤΟΝ ΕΛΕΓΧΟ ΑΥΤΟ ΠΡΟΚΥΠΤΕΙ Ο ΒΑΘΜΟΣ ΣΥΝΑΦΕΙΑΣ ΤΩΝ ΝΕΩΝ ΔΕΔΟΜΕΝΩΝ ΜΕ ΤΑ ΕΓΚΥΡΑ

ΠΡΟΤΥΠΑ ΩΣΤΕ ΝΑ ΚΑΘΟΡΙΣΤΕΙ ΤΟ ΤΡΟΠΟΣ ΜΕ ΤΟΝ ΟΠΟΙΟΝ ΠΡΕΠΕΙ ΝΑ ΑΝΤΙΜΕΤΩΠΙΣΤΕΙ Η ΠΕΡΑΙΤΕΡΩ ΕΠΕΞΕΡΓΑΣΙΑ ΤΟΥΣ.

- ΜΟΝΤΕΛΑ ΚΑΙ ΓΛΩΣΣΕΣ ΑΝΑΠΑΡΑΣΤΑΣΗΣ: ΑΝΑΠΤΥΣΣΟΝΤΑΙ ΤΡΟΠΟΙ ΑΝΑΠΑΡΑΣΤΑΣΗΣ ΤΩΝ ΔΕΔΟΜΕΝΩΝ ΚΑΙ ΤΩΝ ΠΛΗΡΟΦΟΡΙΩΝ ΠΟΥ ΕΝΥΠΑΡΧΟΥΝ ΣΕ ΑΥΤΑ ΣΕ ΜΟΡΦΗ ΣΑΦΗ ΚΑΙ ΚΑΤΑΝΟΗΤΗ ΑΠΟ ΤΟΥΣ ΔΙΑΦΟΡΩΝ ΤΥΠΩΝ ΧΕΙΡΙΣΤΕΣ ΤΟΥΣ. ΟΙ ΑΠΕΙΚΟΝΙΣΕΙΣ ΑΥΤΕΣ ΕΙΝΑΙ ΠΡΟΦΑΝΩΣ ΠΡΟΚΑΘΟΡΙΣΜΕΝΕΣ ΩΣΤΕ ΝΑ ΕΙΝΑΙ ΣΕ ΘΕΣΗ ΝΑ ΠΕΡΙΓΡΑΦΟΥΝ ΠΛΗΡΩΣ ΚΑΙ ΜΕ ΑΚΡΙΒΕΙΑ ΤΗΝ ΕΓΓΕΝΗ ΣΗΜΑΣΙΑ ΚΑΙ ΟΥΣΙΑ ΤΩΝ ΔΕΔΟΜΕΝΩΝ. ΟΙ ΑΠΕΙΚΟΝΙΣΕΙΣ ΣΥΝΗΘΩΣ ΕΙΝΑΙ ΓΡΑΦΙΚΕΣ ΑΛΛΑ ΣΕ ΟΡΙΣΜΕΝΕΣ ΠΕΡΙΠΤΩΣΕΙΣ ΜΠΟΡΕΙ ΝΑ ΕΙΝΑΙ ΚΑΙ ΛΕΚΤΙΚΕΣ Η ΣΥΝΔΥΑΣΜΟΣ ΤΩΝ ΜΟΡΦΩΝ ΑΥΤΩΝ.
- ΕΠΕΞΕΡΓΑΣΙΑ ΚΑΙ ΔΙΑΧΕΙΡΙΣΗ ΤΩΝ ΙΔΙΩΝ ΤΩΝ ΔΕΔΟΜΕΝΩΝ: ΕΦΟΣΟΝ ΕΧΕΙ ΟΛΟΚΛΗΡΩΘΕΙ Η ΑΝΑΛΥΣΗ ΤΩΝ ΣΥΝΟΛΩΝ ΤΩΝ ΔΕΔΟΜΕΝΩΝ ΚΑΙ ΕΧΕΙ ΑΠΟΦΑΣΙΣΤΕΙ ΟΤΙ ΕΙΝΑΙ ΚΑΤΑΛΛΗΛΑ ΓΙΑ ΤΗΝ ΑΠΟΣΤΟΛΗ ΤΩΝ ΧΕΙΡΙΣΤΩΝ ΤΟΥΣ, ΥΦΙΣΤΑΝΤΑΙ ΕΠΕΞΕΡΓΑΣΙΑ ΠΡΟΚΕΙΜΕΝΟΥ ΝΑ ΠΡΟΚΥΨΟΥΝ ΧΡΗΣΙΜΕΣ ΠΛΗΡΟΦΟΡΙΕΣ ΚΑΙ ΧΡΗΣΙΜΟΠΟΙΟΥΝΤΑΙ ΑΝΑΛΟΓΑ. Η ΕΠΕΞΕΡΓΑΣΙΑ ΠΟΙΚΙΛΕΙ ΑΝΑΛΟΓΑ ΜΕ ΤΗΝ ΦΥΣΗ, ΤΗΝ ΕΚΤΑΣΗ ΚΑΙ ΤΗΝ ΠΟΛΥΠΛΟΚΟΤΗΤΑ ΤΩΝ ΔΕΔΟΜΕΝΩΝ ΑΛΛΑ ΚΑΙ ΤΗΝ ΕΠΙΘΥΜΗΤΗ ΜΟΡΦΗ ΤΗΣ ΕΞΟΔΟΥ ΚΑΙ ΤΟ ΕΙΔΟΣ ΤΩΝ ΠΡΟΣΔΟΚΩΜΕΝΩΝ ΠΛΗΡΟΦΟΡΙΩΝ.

Η ΕΞΟΡΥΞΗ ΔΕΔΟΜΕΝΩΝ ΕΙΝΑΙ ΣΤΗΝ ΠΡΑΓΜΑΤΙΚΟΤΗΤΑ ΕΝΑ ΑΠΟ ΤΑ ΣΤΑΔΙΑ ΤΗΣ ΔΙΑΔΙΚΑΣΙΑΣ ΑΠΟΚΤΗΣΗΣ ΤΗΣ ΓΝΩΣΗΣ. Η ΑΝΑΚΑΛΥΨΗ ΓΝΩΣΗΣ ΑΠΟΤΕΛΕΙ ΜΙΑ ΔΙΑΔΙΚΑΣΙΑ ΠΟΥ ΞΕΚΙΝΑΕΙ ΑΠΟ ΤΗΝ ΠΡΟΣΒΑΣΗ ΣΕ ΜΕΓΑΛΑ ΣΥΝΟΛΑ ΔΕΔΟΜΕΝΩΝ ΚΑΙ ΚΑΤΑΛΗΓΕΙ ΣΕ ΑΠΟΚΤΗΣΗ ΓΝΩΣΗΣ. ΤΑ ΣΤΑΔΙΑ ΤΗΣ ΔΙΑΔΙΚΑΣΙΑΣ ΑΥΤΗΣ ΣΥΝΟΠΤΙΚΑ ΕΙΝΑΙ:

- ΕΠΙΛΟΓΗ ΚΑΙ ΣΥΓΚΕΝΤΡΩΣΗ ΤΩΝ ΔΕΔΟΜΕΝΩΝ: ΣΤΟ ΒΗΜΑ ΑΥΤΟ ΑΡΧΙΚΑ ΕΝΤΟΠΙΖΟΝΤΑΙ ΟΙ ΠΗΓΕΣ ΑΠΟ ΤΙΣ ΟΠΟΙΕΣ ΘΑ ΑΝΤΛΗΘΟΥΝ ΤΑ ΥΠΟ ΜΕΛΕΤΗ ΔΕΔΟΜΕΝΑ. Η ΕΠΙΛΟΓΗ ΜΠΟΡΕΙ ΝΑ ΓΙΝΕΙ ΜΕ ΔΙΑΦΟΡΑ ΚΡΙΤΗΡΙΑ ΠΟΥ ΕΧΟΥΝ ΚΥΡΙΩΣ ΝΑ ΚΑΝΟΥΝ ΜΕ ΤΗΝ ΦΥΣΗ ΤΟΥΣ. ΩΣΤΟΣΟ ΤΟ ΚΡΙΤΗΡΙΟ ΠΟΥ ΕΞΕΤΑΖΕΤΑΙ ΣΕ ΟΛΕΣ ΣΧΕΔΟΝ ΤΙΣ ΠΕΡΙΠΤΩΣΕΙΣ ΕΙΝΑΙ Η ΑΞΙΟΠΙΣΤΙΑ ΤΩΝ ΠΗΓΩΝ. ΕΙΝΑΙ ΑΠΑΡΑΙΤΗΤΗ ΠΡΟΫΠΟΘΕΣΗ ΓΙΑ ΤΗΝ ΠΑΡΑΓΩΓΗ ΩΦΕΛΙΜΗΣ ΓΝΩΣΗΣ ΣΤΟ ΤΕΛΟΣ ΤΗΣ ΟΛΗΣ ΔΙΑΔΙΚΑΣΙΑΣ, ΝΑ ΧΡΗΣΙΜΟΠΟΙΗΘΕΙ ΠΡΩΤΟΓΕΝΕΣ ΥΛΙΚΟ (ΔΕΔΟΜΕΝΑ) ΠΟΥ ΑΝΤΙΚΑΤΟΠΤΡΙΖΟΥΝ ΤΗΝ ΠΡΑΓΜΑΤΙΚΟΤΗΤΑ.
- ΚΑΘΑΡΙΣΜΟΣ ΔΕΔΟΜΕΝΩΝ: Η ΕΙΣΟΔΟΣ ΣΤΟ ΣΤΑΔΙΟ ΑΥΤΟ ΤΗΣ ΔΙΑΔΙΚΑΣΙΑΣ ΕΙΝΑΙ ΜΕΓΑΛΕΣ ΠΟΣΟΤΗΤΕΣ ΔΕΔΟΜΕΝΩΝ ΤΑ ΟΠΟΙΑ ΣΕ ΜΕΓΑΛΟ ΠΟΣΟΣΤΟ ΔΕΝ ΠΑΡΟΥΣΙΑΖΟΥΝ ΣΥΝΟΧΗ Η ΔΕΝ ΕΧΟΥΝ ΚΑΠΟΙΑ ΠΡΟΦΑΝΗ ΣΗΜΑΣΙΑ. ΕΠΙΣΗΣ ΜΠΟΡΕΙ ΝΑ ΜΗΝ ΕΙΝΑΙ ΠΛΗΡΗ Η ΝΑ ΕΙΝΑΙ ΜΕΡΙΚΩΣ ΚΑΤΕΣΤΡΑΜΜΕΝΑ. ΑΥΤΟ ΜΠΟΡΕΙ ΝΑ ΣΥΜΒΑΙΝΕΙ ΓΙΑ ΠΟΙΚΙΛΟΥΣ ΛΟΓΟΥΣ ΠΟΥ ΠΕΡΙΛΑΜΒΑΝΟΥΝ ΕΛΛΙΠΗ ΦΡΟΝΤΙΔΑ ΤΟΥΣ ΕΚ ΜΕΡΟΥΣ ΤΟΥ ΙΔΙΟΚΤΗΤΗ ΤΟΥΣ, ΣΦΑΛΜΑΤΑ ΚΑΤΑ ΤΗΝ ΜΕΤΑΔΟΣΗ – ΜΕΤΑΦΟΡΑ ΤΟΥΣ, ΚΑΚΟΒΟΥΛΗ ΕΠΕΜΒΑΣΗ ΚΑ. ΤΑ ΔΕΔΟΜΕΝΑ ΑΥΤΑ ΕΙΝΑΙ ΠΡΟΦΑΝΩΣ ΜΗ ΩΦΕΛΙΜΑ ΓΙΑ ΠΕΡΑΙΤΕΡΩ ΔΙΕΡΕΥΝΗΣΗ ΚΑΙ ΕΠΕΞΕΡΓΑΣΙΑ ΚΑΙ ΜΠΟΡΕΙ ΝΑ ΟΔΗΓΗΣΟΥΝ ΣΕ ΚΑΤΑΡΡΕΥΣΗ ΤΗΣ ΟΛΗΣ ΔΙΑΔΙΚΑΣΙΑΣ Η ΣΤΗΝ ΠΑΡΑΓΩΓΗ ΕΣΦΑΛΜΕΝΩΝ ΣΥΜΠΕΡΑΣΜΑΤΩΝ. ΈΤΣΙ ΕΠΙΧΕΙΡΕΙΤΑΙ ΕΚΚΑΘΑΡΙΣΗ ΤΩΝ ΣΥΝΟΛΩΝ ΔΕΔΟΜΕΝΩΝ ΑΠΟ ΤΑ ΠΕΡΙΤΤΑ - ΑΧΡΗΣΤΑ ΣΤΟΙΧΕΙΑ ΩΣΤΕ ΣΤΟ ΤΕΛΟΣ ΤΗΣ ΔΙΕΡΓΑΣΙΑΣ ΝΑ ΑΠΟΜΕΙΝΟΥΝ ΜΟΝΟ ΕΚΕΙΝΑ ΤΑ ΔΕΔΟΜΕΝΑ ΤΑ ΟΠΟΙΑ ΣΧΕΤΙΖΟΝΤΑΙ ΜΕ ΤΟΝ ΑΝΤΙΚΕΙΜΕΝΙΚΟ ΣΚΟΠΟ ΚΑΙ ΠΟΥ ΕΙΝΑΙ ΣΕ ΜΟΡΦΗ ΚΑΙ ΚΑΤΑΣΤΑΣΗ ΤΕΤΟΙΑ ΠΟΥ ΝΑ ΕΠΙΤΡΕΠΟΥΝ ΤΗΝ ΔΙΑΧΕΙΡΙΣΗ ΚΑΙ ΤΗΝ ΕΠΕΞΕΡΓΑΣΙΑ ΤΟΥΣ.

- ΕΝΟΠΟΙΗΣΗ ΔΕΔΟΜΕΝΩΝ: ΣΤΟ ΣΤΑΔΙΟ ΑΥΤΟ ΕΙΣΕΡΧΟΝΤΑΙ ΜΟΝΟ ΩΦΕΛΙΜΑ ΔΕΔΟΜΕΝΑ ΚΑΘΩΣ ΤΟ ΣΥΝΟΛΟ ΕΧΕΙ ΚΑΘΑΡΙΣΤΕΙ. ΣΤΙΣ ΠΕΡΙΣΣΟΤΕΡΕΣ ΠΕΡΙΠΤΩΣΕΙΣ ΤΑ ΔΕΔΟΜΕΝΑ ΠΡΟΕΡΧΟΝΤΑΙ ΑΠΟ ΔΙΑΦΟΡΕΤΙΚΕΣ ΠΗΓΕΣ ΚΑΙ ΜΠΟΡΕΙ ΝΑ ΒΡΙΣΚΟΝΤΑΙ ΣΕ ΠΟΙΚΙΛΕΣ ΜΟΡΦΕΣ ΕΙΤΕ ΓΙΑ ΛΟΓΟΥΣ ΠΟΥ ΑΠΤΟΝΤΑΤΩΝ ΥΠΟΔΟΜΩΝ ΠΟΥ ΧΡΗΣΙΜΟΠΟΙΟΥΝΤΑΙ ΓΙΑ ΤΗΝ ΜΕΤΑΦΟΡΑ ΤΟΥΣ, ΕΙΤΕ ΓΙΑ ΤΙΣ ΑΠΑΙΤΗΣΕΙΣ ΑΠΟΘΗΚΕΥΣΗΣ ΤΟΥΣ ΕΙΤΕ ΑΚΟΜΑ ΚΑΙ ΓΙΑ ΤΗΝ ΙΔΙΑ ΤΗΝ ΦΥΣΗ ΤΟΥΣ. ΠΡΟΚΕΙΜΕΝΟΥ ΟΜΩΣ ΝΑ ΥΠΟΣΤΟΥΝ ΕΝΙΑΙΑ ΕΠΕΞΕΡΓΑΣΙΑ ΕΙΝΑΙ ΑΝΑΓΚΗ ΝΑ ΒΡΙΣΚΟΝΤΑΙ ΟΛΑ ΤΑ ΠΡΟΣ ΜΕΛΕΤΗ ΔΕΔΟΜΕΝΑ ΣΤΗΝ ΙΔΙΑ ΜΟΡΦΗ. ΈΤΣΙ ΣΤΟ ΒΗΜΑ ΑΥΤΟ ΤΗΣ ΔΙΑΔΙΚΑΣΙΑΣ ΕΠΙΧΕΙΡΕΙΤΑΙ Η ΣΥΓΚΕΝΤΡΩΣΗ ΚΑΙ Η ΟΜΟΓΕΝΟΠΟΙΗΣΗ ΤΩΝ ΔΕΔΟΜΕΝΩΝ. ΤΗΣ ΟΜΟΓΕΝΟΠΟΙΗΣΗΣ ΑΥΤΗΣ ΠΡΟΗΓΕΙΤΑΙ Ο ΠΡΟΣΔΙΟΡΙΣΜΟΣ ΤΩΝ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ ΒΑΣΗ ΤΩΝ ΟΠΟΙΩΝ ΜΠΟΡΟΥΝ ΝΑ ΧΑΡΑΚΤΗΡΙΣΤΟΥΝ. ΤΟ ΣΥΝΟΛΟ ΑΥΤΩΝ ΤΩΝ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ ΘΑ ΑΠΟΤΕΛΕΣΕΙ ΣΤΗΝ ΟΥΣΙΑ ΤΟΝ ΚΟΙΝΟ ΚΩΔΙΚΑ ΤΗΣ ΟΜΟΓΕΝΟΥΣ ΜΟΡΦΟΠΟΙΗΣΗΣ ΤΟΥΣ ΚΑΙ Η ΤΑΥΤΟΤΗΤΑ ΤΟΥ ΘΑ ΠΡΟΣΔΙΟΡΙΖΕΤΑΙ ΑΠΟ ΤΙΣ ΤΙΜΕΣ ΑΥΤΩΝ ΤΩΝ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ.
- ΕΠΙΛΟΓΗ ΔΕΔΟΜΕΝΩΝ: ΑΠΟ ΤΟ ΑΡΧΙΚΟ ΣΤΑΔΙΟ ΑΦΑΙΡΕΘΗΚΑΝ ΤΑ ΜΗ ΕΚΜΕΤΑΛΛΕΥΣΙΜΑ ΔΕΔΟΜΕΝΑ. ΩΣΤΟΣΟ ΣΤΟ ΣΥΝΟΛΟ ΠΑΡΑΜΕΝΟΥΝ ΔΕΔΟΜΕΝΑ ΤΩΝ ΟΠΟΙΩΝ Η ΣΗΜΑΣΙΑ ΕΙΝΑΙ ΜΑΚΡΙΑ ΑΠΟ ΤΟ ΑΝΤΙΚΕΙΜΕΝΟ ΤΟ ΟΠΟΙΟ ΕΞΕΤΑΖΕΤΑΙ ΚΑΙ ΕΚΤΙΜΑΤΑΙ ΟΤΙ ΘΑ ΕΞΥΠΗΡΕΤΗΣΟΥΝ ΤΗΝ ΣΥΝΟΛΙΚΗ ΔΙΑΔΙΚΑΣΙΑ ΕΞΟΡΥΞΗΣ ΓΝΩΣΗΣ ΛΙΓΟ Η ΚΑΙ ΚΑΘΟΛΟΥ. ΣΤΗΝ ΠΡΟΣΠΑΘΕΙΑ ΝΑ ΔΗΜΙΟΥΡΓΗΘΕΙ ΕΝΑ ΟΣΟ ΤΟ ΔΥΝΑΤΟΝ ΠΕΡΙΣΣΟΤΕΡΟ ΠΕΡΙΕΚΤΙΚΟ ΚΑΙ ΣΥΜΠΑΓΕΣ ΣΥΝΟΛΟ ΔΕΔΟΜΕΝΩΝ, ΕΙΝΑΙ ΑΠΑΙΤΗΣΗ ΤΑ ΔΕΔΟΜΕΝΑ ΑΥΤΑ ΝΑ ΜΗΝ ΠΑΡΑΜΕΝΟΥΝ ΣΤΟ ΣΥΝΟΛΟ. ΈΤΣΙ ΣΕ ΑΥΤΟ ΤΟ ΣΤΑΔΙΟ ΑΦΑΙΡΟΥΝΤΑΙ

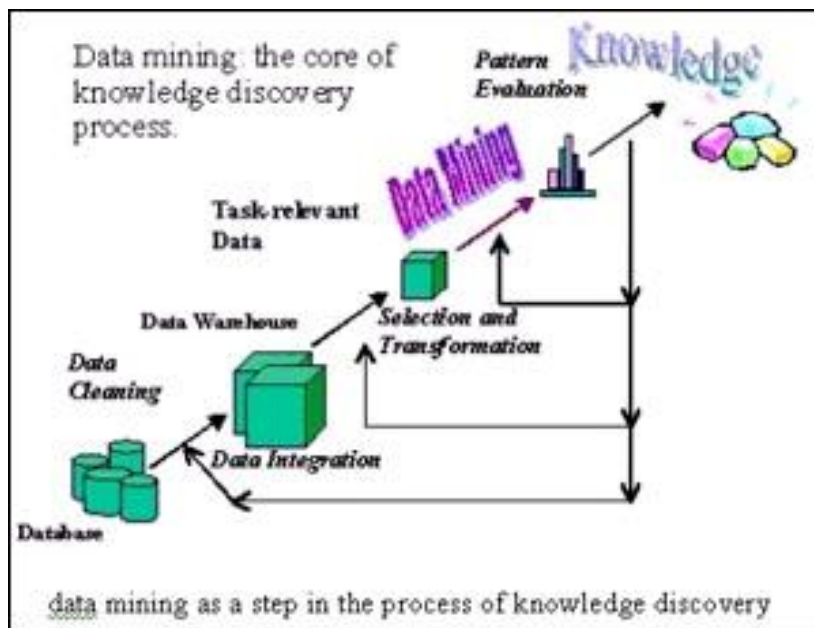
ΑΠΟ ΤΟ ΠΡΟΣ ΜΕΛΕΤΗ ΣΥΝΟΛΟ ΔΕΔΟΜΕΝΩΝ ΕΚΕΙΝΑ ΤΑ ΟΠΟΙΑ ΚΡΙΝΕΤΑΙ ΟΤΙ ΔΕΝ ΣΧΕΤΙΖΟΝΤΑΙ ΣΕ ΕΠΑΡΚΗ ΒΑΘΜΟ ΜΕ ΤΟ ΑΝΤΙΚΕΙΜΕΝΟ ΠΟΥ ΕΞΕΤΑΖΕΤΑΙ.

- ΜΕΤΑΣΧΗΜΑΤΙΣΜΟΣ ΔΕΔΟΜΕΝΩΝ: ΔΕΔΟΜΕΝΟΥ ΟΤΙ ΠΡΟΚΕΙΤΑΙ ΣΤΑ ΔΕΔΟΜΕΝΑ ΝΑ ΕΦΑΡΜΟΣΤΟΥΝ ΣΥΓΚΕΚΡΙΜΕΝΕΣ ΤΕΧΝΙΚΕΣ ΕΞΟΡΥΞΗΣ ΘΑ ΠΡΕΠΕΙ ΑΥΤΑ ΝΑ ΒΡΙΣΚΟΝΤΑΙ ΣΕ ΜΟΡΦΗ ΠΟΥ ΝΑ ΕΠΙΤΡΕΠΟΥΝ ΤΗΝ ΑΠΟΔΟΤΙΚΗ ΕΦΑΡΜΟΓΗ ΤΟΥΣ. ΕΙΝΑΙ ΠΟΛΥ ΠΙΘΑΝΟ ΚΑΠΟΙΑ ΑΠΟ ΤΑ ΔΕΔΟΜΕΝΑ ΝΑ ΜΗΝ ΕΙΝΑΙ ΣΕ ΜΟΡΦΗ ΚΑΤΑΛΛΗΛΗ ΓΙΑ ΤΗΝ ΠΕΡΑΙΤΕΡΩ ΕΞΟΡΥΞΗ ΤΟΥΣ ΟΠΟΤΕ ΣΤΟ ΒΗΜΑ ΑΥΤΟ ΠΡΑΓΜΑΤΟΠΟΙΕΙΤΑΙ Η ΜΕΤΑΤΡΟΠΗ ΤΟΥΣ ΣΕ ΜΟΡΦΗ ΚΑΤΑΛΛΗΛΗ.
- ΕΞΟΡΥΞΗ ΔΕΔΟΜΕΝΩΝ: ΣΤΟ ΒΗΜΑ ΑΥΤΟ ΠΑΡΑΓΟΝΤΑΙ ΧΡΗΣΙΜΑ ΠΡΟΤΥΠΑ ΤΑ ΟΠΟΙΑ ΕΝΔΕΧΟΜΕΝΩΣ ΘΑ ΧΡΗΣΙΜΟΠΟΙΗΘΟΥΝ ΓΙΑ ΤΗΝ ΑΝΑΚΑΛΥΨΗ ΤΗΣ ΓΝΩΣΗΣ. ΤΑ ΠΡΟΤΥΠΑ ΑΥΤΑ ΕΠΙΔΙΩΚΕΤΑΙ ΝΑ ΕΙΝΑΙ ΚΥΡΙΩΣ ΑΝΑΓΝΩΡΙΣΙΜΑ ΑΠΟ ΤΟΝ ΑΝΘΡΩΠΟ, ΝΑ ΕΙΝΑΙ ΣΑΦΗ, ΑΚΡΙΒΗ ΚΑΙ ΠΡΟΦΑΝΗ. ΑΠΟΤΕΛΟΥΝ ΣΥΝΗΘΩΣ ΑΠΟΤΕΛΕΣΜΑ ΑΥΤΟΜΑΤΩΝ Η ΗΜΙΑΥΤΟΜΑΤΩΝ ΕΠΕΞΕΡΓΑΣΤΙΚΩΝ ΔΙΑΔΙΚΑΣΙΩΝ ΤΩΝ ΔΕΔΟΜΕΝΩΝ ΤΟΥ ΠΡΟΗΓΟΥΜΕΝΟΥ ΒΗΜΑΤΟΣ.
- ΑΞΙΟΛΟΓΗΣΗ ΠΡΟΤΥΠΩΝ: ΕΙΣΟΔΟΣ ΣΤΟ ΣΤΑΔΙΟ ΑΥΤΟ ΕΙΝΑΙ ΤΑ ΠΡΟΤΥΠΑ ΠΟΥ ΠΑΡΑΧΘΗΚΑΝ ΑΠΟ ΤΗΝ ΕΞΟΡΥΞΗ ΔΕΔΟΜΕΝΩΝ. ΤΑ ΠΡΟΤΥΠΑ ΑΥΤΑ ΑΞΙΟΛΟΓΟΥΝΤΑΙ ΩΣ ΠΡΟΣ ΤΗΝ ΟΡΘΟΤΗΤΑ ΤΟΥΣ ΒΑΣΗ ΚΑΤΑΛΛΗΛΕΣ ΜΕΤΡΙΚΕΣ ΚΑΙ ΜΕ ΤΗΝ ΧΡΗΣΗ ΕΛΕΓΜΕΝΩΝ ΩΣ ΠΡΟΣ ΤΗΝ ΑΠΟΤΕΛΕΣΜΑΤΙΚΟΤΗΤΑ ΤΟΥΣ ΜΕΘΟΔΩΝ. ΑΠΟ ΤΗΝ ΑΞΙΟΛΟΓΗΣΗ ΑΥΤΗ ΠΑΡΑΓΕΤΑΙ ΠΟΡΙΣΜΑ ΣΧΕΤΙΚΑ ΜΕ ΤΗΝ ΚΑΤΑΛΛΗΛΟΛΗΤΑ ΤΩΝ ΠΡΟΤΥΠΩΝ ΑΥΤΩΝ ΝΑ ΧΡΗΣΙΜΟΠΟΙΟΥΝΤΑΙ ΣΤΟ ΕΞΗΣ ΓΙΑ ΤΗΝ ΠΑΡΑΓΩΓΗ ΓΝΩΣΗΣ ΑΠΟ ΜΕΛΛΟΝΤΙΚΑ ΣΥΝΑΦΗ ΣΥΝΟΛΑ ΔΕΔΟΜΕΝΩΝ. ΤΑ ΠΡΟΤΥΠΑ ΠΟΥ ΘΑ ΥΙΟΘΕΤΗΘΟΥΝ ΘΑ ΠΡΕΠΕΙ ΝΑ ΕΧΟΥΝ ΥΠΟΣΤΕΙ ΠΛΗΡΗ ΕΛΕΓΧΟ ΓΙΑ ΤΗΝ

ΑΞΙΟΛΟΓΗΣΗ ΤΟΥΣ ΑΛΛΑ ΚΑΙ ΝΑ ΕΛΕΓΧΟΝΤΑΙ ΤΑΚΤΙΚΑ ΓΙΑ ΤΗΝ ΕΓΚΥΡΟΤΗΤΑ ΤΟΥΣ.

- ΑΝΑΠΑΡΑΣΤΑΣΗ ΓΝΩΣΗΣ: ΠΡΟΚΕΙΤΑΙ ΓΙΑ ΤΗΝ ΑΝΑΠΑΡΑΣΤΑΣΗ ΤΗΣ ΠΡΟΚΥΠΤΟΥΣΑΣ ΓΝΩΣΗΣ ΑΠΟ ΤΗΝ ΔΙΑΔΙΚΑΣΙΑ ΜΕ ΜΟΡΦΗ ΑΝΑΓΝΩΣΙΜΗ ΑΠΟ ΤΟΥΣ ΑΡΜΟΔΙΟΥΣ ΧΕΙΡΙΣΤΕΣ ΤΗΣ. ΑΥΤΟ ΤΟ ΒΗΜΑ ΧΡΗΣΙΜΟΠΟΙΕΙ ΤΕΧΝΙΚΕΣ ΑΠΕΙΚΟΝΙΣΗΣ ΓΙΑ ΝΑ ΒΟΗΘΗΣΕΙ ΤΟΥΣ ΧΡΗΣΤΕΣ ΝΑ ΚΑΤΑΝΟΗΣΟΥΝ ΚΑΙ ΝΑ ΕΡΜΗΝΕΥΣΟΥΝ ΤΑ ΑΠΟΤΕΛΕΣΜΑΤΑ ΤΗΣ ΕΞΟΡΥΞΗΣ ΔΕΔΟΜΕΝΩΝ. ΟΙ ΓΡΑΦΙΚΕΣ ΜΕΘΟΔΟΙ ΑΠΕΙΚΟΝΙΣΗΣ ΕΙΝΑΙ ΣΥΧΝΑ ΠΡΟΤΙΜΗΤΕΕΣ ΑΦΟΥ ΕΧΟΥΝ ΤΗΝ ΔΥΝΑΤΟΤΗΤΑ ΜΕ ΠΑΡΑΣΙΤΙΚΟ ΤΡΟΠΟ ΝΑ ΠΡΟΒΑΛΛΟΥΝ ΜΕΓΑΛΑ ΠΟΣΑ ΠΛΗΡΟΦΟΡΙΑΣ ΣΕ ΜΙΚΡΗΣ ΣΧΕΤΙΚΑ ΕΚΤΑΣΗΣ ΕΓΓΡΑΦΩΝ (ΜΕ ΤΟΝ ΟΡΟ ΕΓΓΡΑΦΟ ΕΝΝΟΕΙΤΑΙ ΚΑΘΕ ΜΕΣΟ ΣΤΟ ΟΠΟΙΟ ΜΠΟΡΕΙ ΝΑ ΑΠΟΤΥΠΩΘΕΙ ΓΡΑΦΙΚΗ ΑΝΑΠΑΡΑΣΤΑΣΗ ΟΠΩΣ ΧΕΙΡΟΓΡΑΦΟ, ΑΡΧΕΙΟ ΚΤΛ.).

ΣΤΟ ΠΑΡΑΚΑΤΩ ΣΧΗΜΑ ΦΑΙΝΕΤΑΙ ΤΟ ΠΟΥ ΤΟΠΟΘΕΤΕΙΤΑΙ Η ΕΞΟΡΥΞΗ ΔΕΔΟΜΕΝΩΝ ΣΤΗΝ ΔΙΑΔΙΚΑΣΙΑ ΑΠΟΚΤΗΣΗΣ ΤΗΣ ΓΝΩΣΗΣ [3][4][21].



ΕΙΚΟΝΑ2: ΔΙΑΔΙΚΑΣΙΑ ΕΞΟΡΥΞΗΣ ΓΝΩΣΗΣ (ΠΗΓΗ: OSMAR R. ΖΑΙΑΝΕ, INTRODUCTION TO DATA MINING,

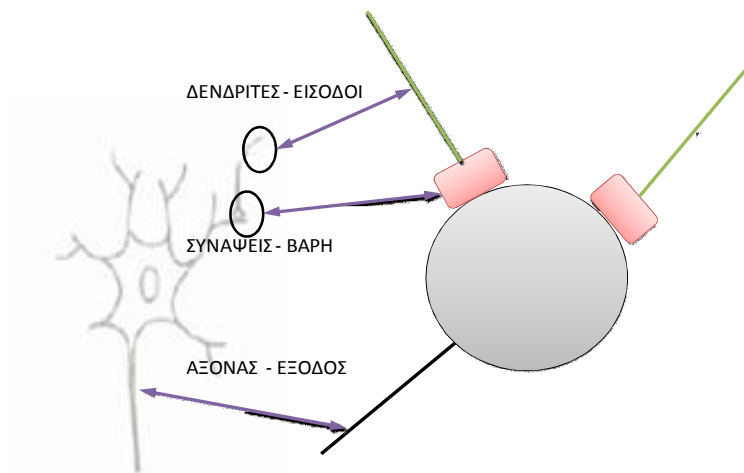
ΟΙ ΜΕΘΟΔΟΛΟΓΙΕΣ ΠΟΥ ΧΡΗΣΙΜΟΠΟΙΟΥΝΤΑΙ ΣΤΗΝ ΔΙΑΔΙΚΑΣΙΑ ΚΑΤΗΓΟΡΙΟΠΟΙΟΥΝΤΑΙ ΩΣ ΑΚΟΛΟΥΘΩΣ:

- ΤΕΧΝΗΤΑ ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ
- ΔΕΝΔΡΑ ΑΠΟΦΑΣΕΩΝ
- ΓΕΝΕΤΙΚΟΙ ΑΛΓΟΡΙΘΜΟΙ
- ΜΕΘΟΔΟΣ ΤΟΥ ΚΟΝΤΙΝΟΤΕΡΟΥ ΓΕΙΤΟΝΑ
- ΓΡΑΦΙΚΗ ΑΝΑΛΥΣΗ

1.3 ΜΕΘΟΔΟΛΟΓΙΕΣ

ΤΕΧΝΗΤΑ ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ (ARTIFICIAL NEURAL NETWORKS)

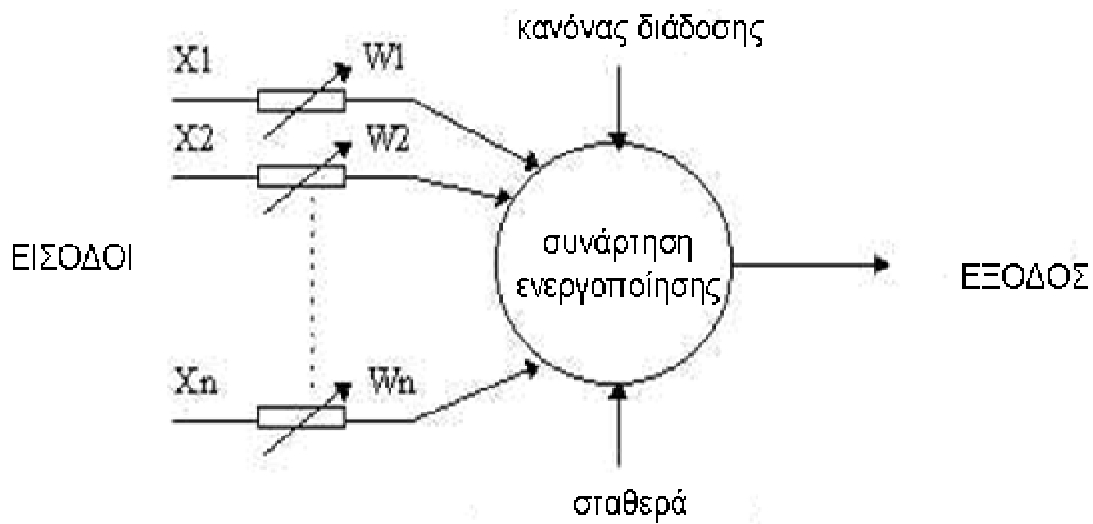
ΤΑ ΤΕΧΝΗΤΑ ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ ΕΙΝΑΙ ΜΑΘΗΜΑΤΙΚΑ ΜΟΝΤΕΛΑ ΕΠΕΞΕΡΓΑΣΙΑΣ ΚΑΙ ΑΞΙΟΛΟΓΗΣΗΣ ΠΛΗΡΟΦΟΡΙΑΣ ΠΟΥ Η ΛΕΙΤΟΥΡΓΙΑ ΒΑΣΙΖΕΤΑΙ ΣΤΗΝ ΜΕΛΕΤΗ ΤΗΣ ΛΕΙΤΟΥΡΓΙΑΣ ΤΩΝ ΒΙΟΛΟΓΙΚΩΝ ΝΕΥΡΩΝΩΝ.



ΕΙΚΟΝΑ 3: ΑΝΤΙΣΤΟΙΧΙΑ ΒΙΟΛΟΓΙΚΩΝ ΚΑΙ ΤΕΧΝΗΤΩΝ ΝΕΥΡΩΝΙΚΩΝ ΔΙΚΤΥΩΝ

Ένας απλός ορισμός για τα τεχνητά νευρωνικά δίκτυα δόθηκε από έναν από τους πρώτους επιστήμονες που μελέτησαν την λειτουργία τους, τον Dr. Robert Hecht-Nielsen, ο οποίος τα χαρακτήρισε ως υπολογιστικά συστήματα που αποτελούνται από μια σειρά απλών, ισχυρά συνδεδεμένων επεξεργαστικών οντοτήτων, που επεξεργάζονται πληροφορίες σχετικές με την αντίδραση της δυναμικής τους κατάστασης σε ερεθίσματα του περιβάλλοντος τους. Η ονομασία τους προέρχεται από την ομοιότητα της λειτουργίας τους με το ανθρώπινο κεντρικό νευρικό σύστημα.

Το βασικότερο συστατικό ενός νευρωνικού δικτύου είναι ο νευρώνας. Στον νευρώνα εφαρμόζονται σταθμισμένοι εισοδοί για μια συνάρτηση ενεργοποίησης. Η συνάρτηση αυτή ακολουθώντας συγκεκριμένους και προκαθορισμένους κανόνες διαδοχής και εφαρμόζοντας μια σταθερή τιμή, προκαλεί ή αναστέλλει την ενεργοποίηση του νευρώνα, προς την δημιουργία εξόδου.



ΕΙΚΟΝΑ 4: ΜΟΡΦΗ ΤΕΧΝΗΤΟΥ ΝΕΥΡΩΝΑ (ΠΗΓΗ:KNOWINO, ARTIFICIALNEURALNETWORK, [HTTP://WWW.TAU.AC.IL/~TSIREL/DUMP/STATIC/KNOWINO.ORG/WIKI/ARTIFICIAL_NEURAL_NETWORK.HTML](http://www.tau.ac.il/~tsirel/dump/static/knowino.org/wiki/artificial_neural_network.html))

ΣΕ ΕΝΑ ΝΕΥΡΩΝΙΚΟ ΔΙΚΤΥΟ ΠΕΡΙΛΑΜΒΑΝΟΝΤΑΙ ΕΝΑΣ ΠΛΗΘΟΣ ΝΕΥΡΩΝΩΝ ΟΙ ΟΠΟΙΟΙ ΣΥΝΔΕΟΝΤΑΙ ΚΑΤΑΛΛΗΛΑ ΩΣΤΕ ΝΑ ΣΥΝΕΡΓΑΖΟΝΤΑΙ ΜΕ ΣΤΟΧΟ ΤΗΝ ΕΚΤΕΛΕΣΗ ΚΑΘΟΡΙΣΜΕΝΟΥ ΕΡΓΟΥ. ΚΑΘΕ ΝΕΥΡΩΝΑΣ ΜΠΟΡΕΙ ΝΑ ΔΕΧΕΤΑΙ ΤΙΜΕΣ ΠΑΡΑΜΕΤΡΩΝ ΕΙΣΟΔΟΥ ΕΙΤΕ ΑΠΟ ΤΟ ΠΕΡΙΒΑΛΛΟΝ ΤΟΥ ΔΙΚΤΥΟΥ ΕΙΤΕ ΑΠΟ ΕΝΑΝ Η ΠΕΡΙΣΣΟΤΕΡΟΥΣ ΑΛΛΟΥΣ ΝΕΥΡΩΝΕΣ ΤΟΥ ΙΔΙΟΥ ΔΙΚΤΥΟΥ ΚΑΙ ΧΡΗΣΙΜΟΠΟΙΕΙ ΤΙΣ ΤΙΜΕΣ ΑΥΤΕΣ, ΓΙΑ ΝΑ ΠΑΡΑΓΕΙ ΔΕΔΟΜΕΝΑ ΤΑ ΟΠΟΙΑ ΕΙΤΕ ΔΙΑΒΙΒΑΖΕΙ ΣΤΟ ΠΕΡΙΒΑΛΛΟΝ ΕΙΤΕ ΣΕ ΑΛΛΟΥΣ ΝΕΥΡΩΝΕΣ. Η ΣΥΜΠΕΡΙΦΟΡΑ ΤΩΝ ΝΕΥΡΩΝΩΝ ΕΙΝΑΙ ΔΥΝΑΜΙΚΗ ΚΑΙ ΔΙΑΜΟΡΦΩΝΕΤΑΙ ΕΞΑΡΤΩΜΕΝΗ ΣΕ ΜΕΓΑΛΟ ΒΑΘΜΟ ΑΠΟ ΤΗΝ ΠΑΡΕΛΘΟΥΣΑ ΕΜΠΕΙΡΙΑ. ΠΕΡΝΑΕΙ ΔΗΛΑΔΗ ΑΠΟ ΜΙΑ ΔΙΑΔΙΚΑΣΙΑ ΜΑΘΗΣΗΣ ΚΑΤΑ ΤΗΝ ΟΠΟΙΑ ΕΠΑΝΑΛΑΜΒΑΝΟΜΕΝΑ ΕΚΠΑΙΔΕΥΕΤΑΙ ΣΤΗΝ ΑΠΟΔΟΤΙΚΗ ΠΑΡΑΓΩΓΗ ΔΕΔΟΜΕΝΩΝ ΕΞΟΔΟΥ.

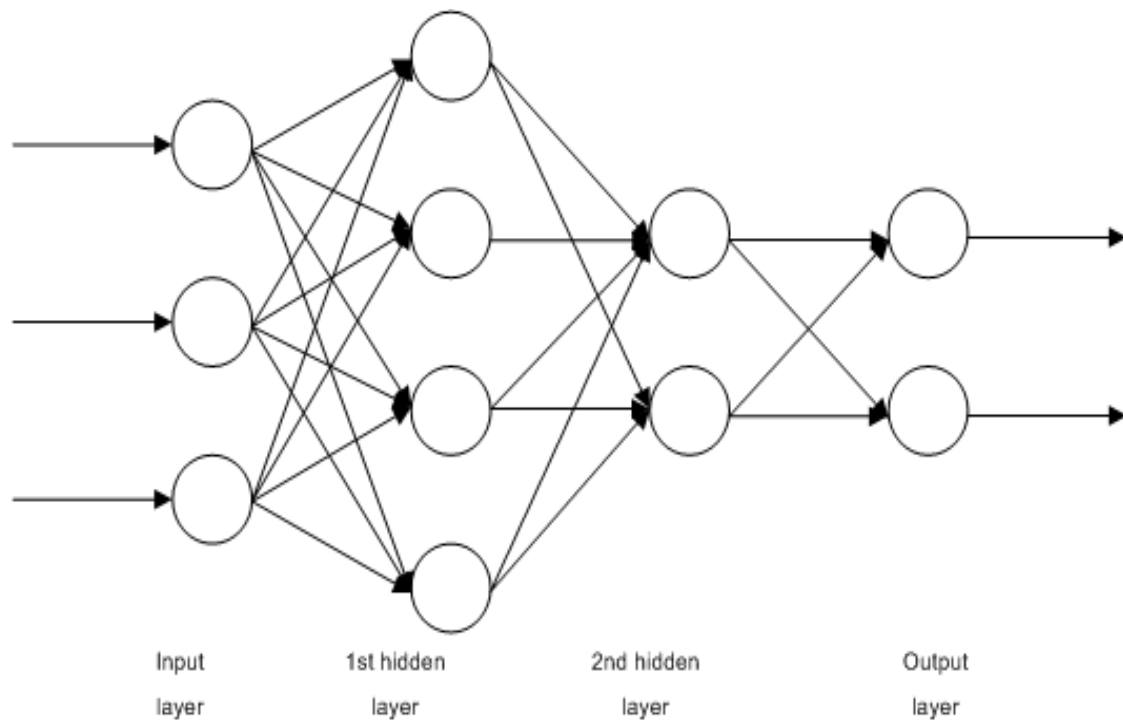
ΣΤΑ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΤΟΥΣ ΠΕΡΙΛΑΜΒΑΝΟΝΤΑΙ:

- Η ΔΥΝΑΤΟΤΗΤΑ ΠΡΟΣΑΡΜΟΓΗΣ ΤΟΥΣ ΣΤΙΣ ΣΥΝΘΗΚΕΣ ΤΟΥ ΠΕΡΙΒΑΛΛΟΝΤΟΣ ΤΟΥΣ
- Η ΑΠΟΚΤΗΣΗΣ ΓΝΩΣΗΣ – ΕΜΠΕΙΡΙΑΣ ΑΠΟ ΠΡΟΤΕΡΕΣ ΕΠΙΚΡΑΤΟΥΣΕΣ ΣΥΝΘΗΚΕΣ.

- Η ΔΥΝΑΤΟΤΗΤΑ ΠΑΡΑΛΛΗΛΗΣ ΕΠΕΞΕΡΓΑΣΙΑΣ.

Η ΤΥΠΙΚΗ ΔΟΜΗ ΕΝΟΣ ΝΕΥΡΩΝΙΚΟΥ ΔΙΚΤΥΟΥ ΠΕΡΙΛΑΜΒΑΝΕΙ ΤΡΙΑ ΕΠΙΠΕΔΑ ΛΕΙΤΟΥΡΓΙΚΩΝ ΜΟΝΑΔΩΝ:

- ΤΟ ΕΠΙΠΕΔΟ ΕΙΣΟΔΟΥ ΤΩΝ ΠΑΡΑΜΕΤΡΩΝ ΛΕΙΤΟΥΡΓΙΑΣ ΤΟΥΣ ΟΠΟΥ ΚΑΘΕ ΣΤΟΙΧΕΙΟ ΤΟΥ ΑΝΤΙΠΡΟΣΩΠΕΥΕΤΑΙ ΜΕ ΜΙΑ ΑΝΕΞΑΡΤΗΤΗ ΜΕΤΑΒΛΗΤΗ. ΣΤΟ ΕΠΙΠΕΔΟ ΑΥΤΟ ΠΕΡΙΛΑΜΒΑΝΟΝΤΑΙ ΟΙ ΜΟΝΑΔΕΣ ΕΙΣΟΔΟΥ ΤΩΝ ΔΕΔΟΜΕΝΩΝ. ΣΤΙΣ ΜΟΝΑΔΕΣ ΑΥΤΕΣ ΕΙΣΕΡΧΟΝΤΑΙ ΤΑ ΣΗΜΑΤΑ ΠΟΥ ΛΑΜΒΑΝΟΝΤΑΙ ΑΠΟ ΤΟ ΠΕΡΙΒΑΛΛΟΝ ΤΟΥ ΔΙΚΤΥΟΥ ΠΡΟΚΕΙΜΕΝΟΥ ΝΑ ΠΡΟΩΘΗΘΟΥΝ ΜΕΣΩ ΚΑΤΑΛΛΗΛΩΝ ΟΔΕΥΣΕΩΝ ΣΤΟ ΕΠΟΜΕΝΟ ΕΠΙΠΕΔΟ.
- ΤΟ ΕΠΙΠΕΔΟ ΤΗΣ ΕΞΟΔΟΥ ΤΗΣ ΠΛΗΡΟΦΟΡΙΑΣ ΠΟΥ ΠΑΡΑΓΕΤΑΙ ΣΑΝ ΑΠΟΤΕΛΕΣΜΑ ΤΗΣ ΕΠΕΞΕΡΓΑΣΙΑΣ ΤΩΝ ΕΙΣΟΔΩΝ. Η ΠΛΗΡΟΦΟΡΙΑ ΔΙΑΧΕΕΤΑΙ ΜΕΣΩ ΤΩΝ ΜΟΝΑΔΩΝ ΕΞΟΔΟΥ ΟΙ ΜΟΝΑΔΕΣ ΑΥΤΕΣ ΑΠΟΔΙΔΟΥΝ ΤΟ ΠΑΡΑΓΟΜΕΝΟ ΕΚ ΤΟΥ ΤΕΧΝΗΤΟΥ ΝΕΥΡΩΝΙΚΟΥ ΔΙΚΤΥΟΥ ΑΠΟΤΕΛΕΣΜΑ ΣΤΟ ΠΕΡΙΒΑΛΛΟΝ.
- ΤΟ ΕΠΙΠΕΔΟ ΕΠΕΞΕΡΓΑΣΙΑΣ ΤΟΥ ΟΠΟΙΟΥ Η ΠΟΛΥΠΛΟΚΟΤΗΤΑ ΜΠΟΡΕΙ ΝΑ ΠΟΙΚΙΛΕΙ ΚΑΘΩΣ ΜΠΟΡΕΙ ΝΑ ΕΙΝΑΙ ΑΠΛΟ Η ΣΥΝΘΕΤΟ. ΣΕ ΑΥΤΟ ΓΕΝΙΚΑ ΠΡΑΓΜΑΤΟΠΟΙΕΙΤΑΙ Η ΕΠΕΞΕΡΓΑΣΙΑ ΤΩΝ ΕΙΣΟΔΩΝ ΚΑΙ ΠΑΡΑΓΕΤΑΙ Η ΕΞΟΔΟΣ.



ΕΙΚΟΝΑ 5: ΜΟΡΦΗ ΝΕΥΡΩΝΙΚΟΥ ΔΙΚΤΥΟΥ (ΠΗΓΗ:CTLPREDUSER INFORMATION, [HTTP://WWW.IMTECH.RES.IN/RAGHAV/CTLPRED/ABOUT.HTML](http://www.imtech.res.in/raghava/ctlpred/about.html))

ΟΙ ΛΕΙΤΟΥΡΓΙΚΕΣ ΜΟΝΑΔΕΣ ΤΟΥ ΝΕΥΡΩΝΙΚΟΥΔΙΚΤΥΟΥ ΕΠΙΚΟΙΝΩΝΟΥΝ ΜΕΤΑΞΥ ΤΟΥΣ ΜΕ ΤΗΝ ΑΠΟΣΤΟΛΗ ΜΗΝΥΜΑΤΩΝ ΜΕΣΩ ΤΩΝ ΣΤΑΘΜΙΣΜΕΝΩΝ ΜΕΤΑΞΥ ΤΟΥΣ ΣΥΝΔΕΣΕΩΝ. ΟΙ ΣΥΝΔΕΣΕΙΣ ΑΥΤΕΣ ΑΚΟΛΟΥΘΟΥΝ ΚΑΘΟΡΙΣΜΕΝΟΥΣ ΚΑΝΟΝΕΣ ΓΙΑ ΤΗΝ ΔΙΑΜΟΡΦΩΣΗ ΤΗΣ ΑΠΟΚΡΙΣΗΣ ΣΕ ΜΙΑ ΕΙΣΟΔΟ. ΟΤΑΝ ΕΝΑ ΔΙΑΝΥΣΜΑ ΕΙΣΟΔΩΝ ΠΕΡΝΑΕΙ ΣΤΟ ΝΕΥΡΩΝΙΚΟ ΔΙΚΤΥΟ ΑΠΟ ΤΟ ΕΠΙΠΕΔΟ ΕΙΣΟΔΟΥ, ΠΟΛΛΑΠΛΑΣΙΑΖΕΤΑΙ ΑΠΟ ΤΟ ΑΝΤΙΣΤΟΙΧΟ ΔΙΑΝΥΣΜΑ ΒΑΡΩΝ ΚΑΙ ΤΑ ΣΤΟΙΧΕΙΑ ΤΟΥ ΠΑΡΑΓΟΜΕΝΟΥ ΔΙΑΝΥΣΜΑΤΟΣ ΠΡΟΣΤΙΘΕΝΤΑΙ ΣΕ ΜΙΑ ΣΤΑΘΕΡΑ ΔΗΜΙΟΥΡΓΩΝΤΑΣ ΤΟΝ ΒΑΘΜΟ ΕΝΕΡΓΟΠΟΙΗΣΗΣ ΤΩΝ ΝΕΥΡΩΝΩΝ. ΣΥΝΗΘΩΣ ΟΙ ΝΕΥΡΩΝΕΣ ΤΟΥ ΔΙΚΤΥΟΥ ΔΙΑΤΑΣΣΟΝΤΑΙ ΣΕ ΠΑΡΑΛΛΗΛΑ ΕΠΙΠΕΔΑ ΕΝΩ ΥΠΑΡΧΟΥΝ ΚΑΙ ΠΕΡΙΠΤΩΣΕΙΣ ΠΟΥ ΠΕΡΙΟΡΙΖΟΝΤΑΙ ΣΕ ΕΝΑ ΜΟΝΟ ΕΠΙΠΕΔΟ. ΥΠΑΡΧΟΥΝ ΜΟΝΤΕΛΑ ΝΕΥΡΩΝΙΚΩΝ ΔΙΚΤΥΩΝ ΤΑ ΟΠΟΙΑ ΔΙΑΧΕΟΥΝ ΤΑ ΣΗΜΑΤΑ ΤΩΝ ΝΕΥΡΩΝΩΝ ΠΡΟΣ ΜΙΑ ΚΑΤΕΥΘΥΝΣΗ ΠΡΟΣ ΤΗΝ ΕΞΟΔΟ ΑΛΛΑ ΚΑΙ ΜΟΝΤΕΛΑ ΤΑ ΟΠΟΙΑ

ΑΝΑΤΡΟΦΟΔΟΤΟΥΝ ΤΙΣ ΕΣΦΑΛΜΕΝΕΣ ΕΞΟΔΟΥΣ ΩΣΤΕ ΝΑ ΒΕΛΤΙΩΝΟΥΝ ΤΟ ΠΑΡΑΓΟΜΕΝΟ ΑΠΟΤΕΛΕΣΜΑ.

Η ΒΑΣΙΚΟΤΕΡΗ ΔΙΑΔΙΚΑΣΙΑ ΤΟΥΣ ΕΙΝΑΙ Η ΜΑΘΗΣΗ ΠΟΥ ΟΔΗΓΕΙ ΤΟ ΤΕΧΝΗΤΟ ΝΕΥΡΩΝΙΚΟ ΔΙΚΤΥΟ ΝΑ ΠΑΡΑΓΕΙ ΤΙΣ ΠΡΟΣΔΟΚΩΜΕΝΕΣ ΕΞΟΔΟΥΣ ΑΝΑΦΟΡΙΚΑ ΣΕ ΣΥΓΚΕΚΡΙΜΕΝΕΣ ΕΙΣΟΔΟΥΣ. ΑΥΤΟ ΓΙΝΕΤΑΙ ΜΕ ΤΗΝ ΑΝΑΖΗΤΗΣΗ ΤΟΥ ΤΡΟΠΟΥ ΠΡΟΣΑΡΜΟΓΗΣ ΤΩΝ ΕΦΑΡΜΟΖΟΜΕΝΩΝ ΣΤΙΣ ΕΙΣΟΔΟΥΣ ΒΑΡΩΝ ΩΣΤΕ ΝΑ ΠΑΡΑΓΕΤΑΙ ΤΟ ΠΡΟΣΔΟΚΩΜΕΝΟ ΑΠΟΤΕΛΕΣΜΑ. Η ΕΚΠΑΙΔΕΥΣΗ ΕΝΟΣ ΝΕΥΡΩΝΙΚΟΥ ΔΙΚΤΥΟΥ ΜΠΟΡΕΙ ΝΑ ΓΙΝΕΙ ΜΕ ΕΠΙΒΛΕΨΗ Η ΧΩΡΙΣ ΕΠΙΒΛΕΨΗ. ΣΤΗΝ ΠΡΩΤΗ ΠΕΡΙΠΤΩΣΗ ΓΙΑ ΚΑΘΕ ΕΙΣΟΔΟ ΠΡΟΣΔΙΟΡΙΖΕΤΑΙ ΜΙΑ ΑΝΑΜΕΝΟΜΕΝΗ ΕΞΟΔΟΣ ΚΑΙ ΔΙΑΔΟΧΙΚΑ ΤΑ ΒΑΡΗ ΤΩΝ ΕΙΣΟΔΟΥ ΑΝΑΠΡΟΣΑΡΜΟΖΟΝΤΑΙ ΩΣΤΕ ΝΑ ΕΛΑΧΙΣΤΟΠΟΙΗΘΕΙ Η ΜΕΣΗ ΤΕΤΡΑΓΩΝΙΚΗ ΑΠΟΚΛΙΣΗ ΤΩΝ ΠΡΑΓΜΑΤΙΚΩΝ ΑΠΟ ΤΙΣ ΑΝΑΜΕΝΟΜΕΝΕΣ ΕΞΟΔΟΥΣ. ΣΤΗΝ ΔΕΥΤΕΡΗ ΠΕΡΙΠΤΩΣΗ ΕΞΕΤΑΖΟΝΤΑΙ ΟΙ ΣΥΣΧΕΤΙΣΕΙΣ ΤΩΝ ΕΙΣΟΔΩΝ ΚΑΙ ΓΙΝΕΤΑΙ ΠΡΟΣΠΑΘΕΙΑ ΝΑ ΟΜΑΔΟΠΟΙΗΘΟΥΝ. ΜΕΤΑ ΤΗΝ ΟΛΟΚΛΗΡΩΣΗ ΤΗΣ ΜΑΘΗΣΗΣ ΕΛΕΓΧΕΤΑΙ ΑΝ ΣΕ ΦΥΣΙΟΛΟΓΙΚΕΣ ΕΠΙΚΡΑΤΟΥΣΕΣ ΣΥΝΘΗΚΕΣ ΚΑΙ ΑΝΑΛΟΓΑ ΜΕ ΤΙΣ ΕΙΣΟΔΟΥΣ ΠΑΡΑΓΟΝΤΑΙ ΤΑ ΑΝΑΜΕΝΟΜΕΝΑ ΑΠΟΤΕΛΕΣΜΑΤΑ. ΑΝ ΑΥΤΟ ΔΕΝ ΣΥΜΒΑΙΝΕΙ ΤΡΟΠΟΠΟΙΕΙΤΑΙ Η ΑΡΧΙΤΕΚΤΟΝΙΚΗ ΤΟΥ ΤΕΧΝΗΤΟΥ ΝΕΥΡΩΝΙΚΟΥ ΔΙΚΤΥΟΥ ΚΑΙ ΕΠΑΝΑΛΑΜΒΑΝΕΤΑΙ Η ΔΙΑΔΙΚΑΣΙΑ ΤΗΣ ΜΑΘΗΣΗΣ.

ΤΑ ΠΛΕΟΝΕΚΤΗΜΑΤΑ ΤΩΝ ΤΕΧΝΗΤΩΝ ΝΕΥΡΩΝΙΚΩΝ ΔΙΚΤΥΩΝ ΕΙΝΑΙ:

- Η ΙΚΑΝΟΤΗΤΑ ΤΟΥΣ ΝΑ ΔΕΧΟΝΤΑΙ ΡΥΘΜΙΖΟΜΕΝΕΣ ΠΑΡΑΜΕΤΡΟΥΣ.
- Η ΔΥΝΑΤΟΤΗΤΑ ΝΑ ΕΚΤΕΛΟΥΝ ΠΑΡΑΛΛΗΛΗ ΕΠΕΞΕΡΓΑΣΙΑ ΤΩΝ ΕΙΣΟΔΩΝ ΤΟΥΣ.
- Η ΔΥΝΑΤΟΤΗΤΑ ΚΑΤΑΝΟΜΗΣ ΤΟΥ ΣΥΝΟΛΟΥ ΤΩΝ ΠΛΗΡΟΦΟΡΙΩΝ ΣΕ ΔΙΑΦΟΡΕΤΙΚΕΣ ΕΞΟΔΟΥΣ.

ΈΤΣΙ Η ΜΕΛΕΤΗ ΚΑΙ ΠΡΟΣΑΡΜΟΓΗ ΤΗΣ ΛΕΙΤΟΥΡΓΙΑΣ ΤΩΝ ΒΙΟΛΟΓΙΚΩΝ ΝΕΥΡΩΝΙΚΩΝ ΔΙΚΤΥΩΝ ΣΕ ΤΕΧΝΙΚΑ ΕΡΓΑ

ΑΠΕΛΕΥΘΕΡΩΝΕΙ ΣΗΜΑΝΤΙΚΕΣ ΔΥΝΑΤΟΤΗΤΕΣ ΓΙΑ ΑΥΞΗΣΗ ΤΗΣ ΑΠΟΔΟΤΙΚΟΤΗΤΑΣ ΑΛΛΑ ΚΑΙ ΤΗΣ ΠΟΙΟΤΗΤΑΣ ΤΟΥ ΠΑΡΑΓΟΜΕΝΟΥ ΑΠΟΤΕΛΕΣΜΑΤΟΣ. ΕΦΑΡΜΟΓΕΣ ΕΝΤΟΠΙΖΟΝΤΑΙ ΣΕ ΠΟΛΛΟΥΣ ΤΟΜΕΙΣ ΤΗΣ ΕΠΙΣΤΗΜΗΣ ΚΑΙ ΓΕΝΙΚΟΤΕΡΑ ΤΗΣ ΑΝΘΡΩΠΙΝΗΣ ΔΡΑΣΤΗΡΙΟΤΗΤΑΣ, ΟΠΩΣ ΣΤΗΝ ΑΕΡΟΔΙΑΣΤΗΜΙΚΗ, ΣΤΑ ΟΙΚΟΝΟΜΙΚΑ, ΣΤΙΣ ΚΑΤΑΣΚΕΥΕΣ, ΣΕ ΣΤΡΑΤΙΩΤΙΚΕΣ ΕΦΑΡΜΟΓΕΣ, ΣΤΗΝ ΙΑΤΡΙΚΗ Κ.Α. [5][6]

ΔΕΝΔΡΑ ΑΠΟΦΑΣΕΩΝ (DECISION TREES)

ΤΑ ΔΕΝΔΡΑ ΑΠΟΦΑΣΕΩΝ ΑΠΟΤΕΛΟΥΝ ΙΣΧΥΡΟ ΓΡΑΦΙΚΟ ΕΡΓΑΛΕΙΟ ΓΙΑ ΤΗΝ ΛΗΨΗ ΑΠΟΦΑΣΕΩΝ. ΕΙΝΑΙ ΚΑΤΑΛΛΗΛΑ ΓΙΑ ΤΗΝ ΑΝΑΠΑΡΑΣΤΑΣΗ ΚΑΙ ΑΝΑΛΥΣΗ ΠΡΟΒΛΗΜΑΤΩΝ ΤΩΝ ΟΠΟΙΩΝ Η ΕΠΙΛΥΣΗ ΠΕΡΝΑΕΙ ΑΠΟ ΤΗΝ ΔΙΑΔΟΧΙΚΗ ΛΗΨΗ ΑΠΟΦΑΣΕΩΝ. ΣΕ ΑΥΤΑ ΑΠΕΙΚΟΝΙΖΕΤΑΙ ΜΕ ΧΡΟΝΟΛΟΓΙΚΗ ΣΕΙΡΑ, ΟΙ ΕΝΑΛΛΑΚΤΙΚΕΣ ΔΙΑΘΕΣΙΜΕΣ ΕΠΙΛΟΓΕΣ ΤΟΥ ΑΡΜΟΔΙΟΥ ΓΙΑ ΛΗΨΗ ΑΠΟΦΑΣΗΣ ΚΑΙ ΤΩΝ ΣΥΝΕΠΕΙΩΝ ΤΟΥΣ. ΤΑ ΔΕΝΔΡΑ ΑΠΟΦΑΣΕΩΝ ΠΑΡΟΥΣΙΑΖΟΝΤΑΙ ΜΕ ΜΕΓΑΛΥΤΕΡΗ ΛΕΠΤΟΜΕΡΕΙΑ ΣΤΟ ΕΠΟΜΕΝΟ ΚΕΦΑΛΑΙΟ [7].

ΓΕΝΕΤΙΚΟΙ ΑΛΓΟΡΙΘΜΟΙ

ΟΙ ΓΕΝΕΤΙΚΟΙ ΑΛΓΟΡΙΘΜΟΙ (GENETICALGORITHMS): ΠΡΟΚΕΙΤΑΙ ΓΙΑ ΕΝΑ ΜΟΝΤΕΛΟ ΕΠΙΛΥΣΗΣ ΠΡΟΒΛΗΜΑΤΩΝ ΠΟΥ ΒΑΣΙΖΕΤΑΙ ΣΤΙΣ ΑΡΧΕΣ ΤΗΣ ΦΥΣΙΚΗΣ ΕΞΕΛΙΞΗΣ, ΣΤΙΣ ΠΡΟΣΑΡΜΟΣΤΙΚΕΣ ΕΚΕΙΝΕΣ ΔΙΑΔΙΚΑΣΙΕΣ ΚΑΤΑ ΤΗΝ ΕΞΕΛΙΞΗ ΤΩΝ ΟΠΟΙΩΝ ΕΝΑ ΣΥΝΟΛΟ ΧΡΩΜΟΣΩΜΑΤΩΝ ΚΑΘΟΡΙΖΕΙ ΤΑ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΕΝΟΣ ΝΕΟΥ ΠΛΗΘΥΣΜΟΥ. ΤΑ ΧΡΩΜΟΣΩΜΑΤΑ ΑΝΑΠΑΡΑΓΟΝΤΑΙ, ΜΕΤΑΛΛΑΣΣΟΝΤΑΙ ΚΑΙ ΔΙΑΣΤΑΥΡΩΝΟΝΤΑΙ ΜΕ ΣΚΟΠΟ ΝΑ ΠΑΡΑΧΘΟΥΝ ΒΕΛΤΙΣΤΟΙ ΑΠΟΓΟΝΟΙ, ΟΙ ΟΠΟΙΟΙ ΣΤΟ ΤΕΛΟΣ ΤΗΣ ΔΙΑΔΙΚΑΣΙΑΣ ΘΑ ΕΠΙΚΡΑΤΗΣΟΥΝ ΕΝΑΝΤΙ ΤΩΝ ΥΠΟΛΟΙΠΩΝ (ΚΑΠΟΙΟΙ ΑΠΟ ΤΟΥΣ ΟΠΟΙΟΥΣ ΔΕΝ ΘΑ ΜΠΟΡΕΣΟΥΝ ΝΑ ΑΝΤΑΠΟΚΡΙΘΟΥΝ ΣΤΙΣ ΑΛΛΑΓΕΣ ΤΟΥ ΠΕΡΙΒΑΛΛΟΝΤΟΣ ΤΟΥΣ). ΟΙ ΕΠΙΚΡΑΤΟΥΝΤΕΣ ΑΠΟΓΟΝΟΙ ΘΕΩΡΕΙΤΑΙ ΟΤΙ ΕΙΝΑΙ ΕΚΕΙΝΟΙ ΜΕ ΤΑ ΚΑΤΑΛΛΗΛΟΤΕΡΑ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΕΝΑΝΤΙ ΤΟΥ ΠΕΡΙΒΑΛΛΟΝΤΟΣ ΤΟΥΣ ΚΑΙ ΤΩΝ ΜΕΤΑΒΟΛΩΝ ΤΟΥ. ΟΙ ΓΕΝΕΤΙΚΟΙ ΑΛΓΟΡΙΘΜΟΙ ΚΑΝΟΥΝ ΧΡΗΣΗ ΤΩΝ ΠΑΡΑΚΑΤΩ ΟΡΩΝ:

- ΓΟΝΙΔΙΟ: ΕΙΝΑΙ Η ΒΑΣΙΚΗ ΓΕΝΕΤΙΚΗ ΔΟΜΙΚΗ ΜΟΝΑΔΑ ΠΟΥ ΑΝΑΠΑΡΙΣΤΑ ΕΝΑ ΧΑΡΑΚΤΗΡΙΣΤΙΚΟ ΜΙΑΣ ΟΝΤΟΤΗΤΑΣ. ΤΑ ΓΟΝΙΔΙΑ ΑΝΤΙΣΤΟΙΧΟΥΝ ΣΕ ΠΑΡΑΜΕΤΡΟΥΣ ΛΥΣΕΩΝ ΠΡΟΒΛΗΜΑΤΩΝ.

- ΧΡΩΜΟΣΩΜΑ: ΑΠΟΤΕΛΟΥΝ ΣΥΛΛΟΓΗ ΑΠΟ ΠΕΠΕΡΑΣΜΕΝΟ ΑΡΙΘΜΟ ΓΟΝΙΔΙΩΝ ΚΑΙ ΑΝΤΙΣΤΟΙΧΟΥΝ ΣΕ ΜΙΑ ΥΠΟΨΗΦΙΑ ΛΥΣΗ ΤΟΥ ΠΡΟΒΛΗΜΑΤΟΣ.
- ΣΥΝΑΡΤΗΣΗ ΚΑΤΑΛΛΗΛΟΛΗΤΑΣ: ΕΙΝΑΙ Η ΑΝΤΙΚΕΙΜΕΝΙΚΗ ΣΥΝΑΡΤΗΣΗ ΠΟΥ ΕΧΕΙ ΣΑΝ ΠΑΡΑΜΕΤΡΟΥΣ ΧΡΩΜΟΣΩΜΑ ΚΑΙ ΣΑΝ ΕΞΟΔΟ ΜΙΑ ΤΙΜΗ ΠΟΥ ΠΡΟΣΔΙΟΡΙΖΕΙ ΤΟ ΒΑΘΜΟ ΚΑΤΑΛΛΗΛΟΛΗΤΑΣ ΤΟΥ ΓΙΑ ΝΑ ΧΡΗΣΙΜΟΠΟΙΗΘΕΙ ΓΙΑ ΑΝΑΠΑΡΑΓΩΓΗ. ΣΤΗΝ ΠΡΑΓΜΑΤΙΚΟΤΗΤΑ Η ΣΥΝΑΡΤΗΣΗ ΑΥΤΗ ΧΡΗΣΙΜΟΠΟΙΕΙΤΑΙ ΓΙΑ ΤΟΝ ΚΑΘΟΡΙΣΜΟ ΤΟΥ ΒΑΘΜΟΥ ΚΑΤΑΛΛΗΛΟΛΗΤΑΣ ΤΗΣ ΛΥΣΗΣ ΠΟΥ ΑΝΤΙΠΡΟΣΩΠΕΥΕΙ ΤΟ ΧΡΩΜΟΣΩΜΑ.
- ΓΟΝΕΙΣ: ΕΙΝΑΙ ΟΙ ΠΡΟΓΟΝΟΙ (ΧΡΩΜΟΣΩΜΑΤΑ) ΤΩΝ ΟΠΟΙΩΝ ΤΑ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ (ΓΟΝΙΔΙΑ) ΘΑ ΣΥΝΔΥΑΣΤΟΥΝ ΠΡΟΚΕΙΜΕΝΟΥ ΝΑ ΔΙΑΜΟΡΦΩΘΟΥΝ ΤΑ ΝΕΑ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΤΩΝ ΜΕΛΩΝ ΝΕΑΣ ΓΕΝΙΑΣ. Η ΕΠΙΛΟΓΗ ΤΟΥΣ ΓΙΝΕΤΑΙ ΒΑΣΗ ΤΗΝ ΤΙΜΗ ΤΗΣ ΣΥΝΑΡΤΗΣΗΣ ΚΑΤΑΛΛΗΛΟΛΗΤΑΣ ΜΕ ΣΚΟΠΟ ΝΑ ΕΠΙΛΕΓΟΥΝ ΠΕΡΙΣΣΟΤΕΡΕΣ ΦΟΡΕΣ ΕΚΕΙΝΟΙ ΠΟΥ ΠΑΡΟΥΣΙΑΖΟΥΝ ΥΨΗΛΟΤΕΡΕΣ ΤΙΜΕΣ.
- ΑΠΟΓΟΝΟΙ: ΟΙ ΑΠΟΓΟΝΟΙ ΔΗΜΙΟΥΡΓΟΥΝΤΑΙ ΑΠΟ ΤΗΝ ΕΦΑΡΜΟΓΗ ΣΤΟΧΑΣΤΙΚΩΝ ΤΕΛΕΣΤΩΝ ΣΤΟ ΕΠΙΛΕΓΜΕΝΟ ΖΕΥΓΟΣ ΓΟΝΕΩΝ. ΠΡΟΚΥΠΤΟΥΝ ΑΠΟ ΤΗΝ ΕΦΑΡΜΟΓΗ ΕΠΙ ΤΩΝ ΧΡΩΜΟΣΩΜΑΤΩΝ ΓΕΝΕΤΙΚΩΝ ΤΕΛΕΣΤΩΝ ΚΑΙ ΜΕΤΑΦΟΡΑ ΤΩΝ ΑΠΟΤΕΛΕΣΜΑΤΩΝ ΣΤΟΝ ΝΕΟ ΠΛΗΘΥΣΜΟ. ΟΙ ΤΕΛΕΣΤΕΣ ΑΥΤΟΙ ΜΠΟΡΕΙ ΝΑ ΕΙΝΑΙ:
 - ΔΙΑΣΤΑΥΡΩΣΗΣ (CROSSOVER): ΣΥΝΙΣΤΑΤΑΙ ΣΤΗΝ ΑΝΤΑΛΛΑΓΗ ΓΕΝΕΤΙΚΟΥ ΥΛΙΚΟΥ ΜΕΤΑΞΥ ΔΥΟ ΠΡΟΓΟΝΩΝ.
 - ΜΕΤΑΛΛΑΞΗΣ (MUTATION): ΕΙΝΑΙ Η ΑΛΛΑΓΗ ΤΩΝ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ ΤΟΥ ΓΕΝΕΤΙΚΟΥ ΥΛΙΚΟΥ ΠΟΥ ΠΡΑΓΜΑΤΟΠΟΙΕΙΤΑΙ ΣΑΝ ΣΥΝΕΠΕΙΑ ΤΗΣ ΠΡΟΣΑΡΜΟΓΗΣ ΣΕ ΝΕΑ ΠΕΡΙΒΑΝΤΟΛΟΓΙΚΑ ΔΕΔΟΜΕΝΑ.

- ΑΝΤΙΣΤΡΟΦΗΣ (INVERSION): ΕΙΝΑΙ Η ΑΛΛΑΓΗ ΣΤΗΝ ΔΙΑΤΑΞΗ ΤΩΝ ΓΟΝΙΔΙΩΝ ΣΕ ΕΝΑ ΧΡΩΜΟΣΩΜΑ.
- ΕΠΙΛΟΓΗΣ (SELECTION): ΕΙΝΑΙ Η ΠΡΑΞΗ ΤΗΣ ΔΙΑΛΟΓΗΣ ΕΚΕΙΝΩΝ ΤΩΝ ΧΡΩΜΟΣΩΜΑΤΩΝ ΠΟΥ ΘΑ ΧΡΗΣΙΜΟΠΟΙΗΘΟΥΝ ΓΙΑ ΤΟΝ ΠΡΟΣΔΙΟΡΙΣΜΟ ΤΩΝ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ ΤΩΝ ΑΠΟΓΟΝΩΝ.
- ΠΛΗΘΥΣΜΟΣ: Ο ΠΛΗΘΥΣΜΟΣ ΑΠΟΤΕΛΕΙΤΑΙ ΑΠΟ ΜΙΑ ΟΜΑΔΑ ΠΙΘΑΝΩΝ ΛΥΣΕΩΝ (ΣΥΛΛΟΓΗ ΧΡΩΜΟΣΩΜΑΤΩΝ) ΚΑΙ ΧΡΗΣΙΜΟΠΟΙΟΥΝΤΑΙ ΓΙΑ ΤΟΝ ΠΡΟΣΔΙΟΡΙΣΜΟ ΤΗΣ ΒΕΛΤΙΣΤΗΣ ΛΥΣΗΣ ΕΝΟΣ ΠΡΟΒΛΗΜΑΤΟΣ. ΚΑΘΕ ΕΚΤΕΛΕΣΗ ΤΟΥ ΓΕΝΕΤΙΚΟΥ ΑΛΓΟΡΙΘΜΟΥ ΟΝΟΜΑΖΕΤΑΙ ΓΕΝΙΑ. Η ΕΚΤΕΛΕΣΗ ΤΟΥ ΑΛΓΟΡΙΘΜΟΥ ΠΕΡΙΛΑΜΒΑΝΕΙ ΚΑΤΑΛΛΗΛΗ ΧΡΗΣΗ ΤΩΝ ΠΡΑΞΕΩΝ ΤΗΣ ΕΠΙΛΟΓΗΣ, ΤΗΣ ΔΙΑΣΤΑΥΡΩΣΗΣ ΚΑΙ ΤΗΣ ΜΕΤΑΛΛΑΞΗΣ ΚΑΙ ΟΛΟΚΛΗΡΩΝΕΤΑΙ ΜΕ ΤΗΝ ΠΛΗΡΗ ΑΝΤΙΚΑΤΑΣΤΑΣΗ ΤΗΣ ΤΡΕΧΟΥΣΑΣ ΓΕΝΙΑΣ ΑΠΟ ΤΗΝ ΔΗΜΙΟΥΡΓΗΘΗΣΑ. Ο ΑΛΓΟΡΙΘΜΟΣ ΟΛΟΚΛΗΡΩΝΕΤΑΙ ΟΤΑΝ ΟΙ ΣΤΟΧΟΙ ΠΟΥ ΕΧΟΥΝ ΤΕΘΕΙ ΩΣ ΠΡΟΣ ΤΗΝ ΠΡΟΣΔΟΚΩΜΕΝΗ ΛΥΣΗ ΕΧΟΥΝ ΠΡΟΣΕΓΓΙΣΤΕΙ.

ΣΕ ΓΕΝΙΚΕΣ ΓΡΑΜΜΕΣ Η ΜΟΡΦΗ ΤΩΝ ΓΕΝΕΤΙΚΩΝ ΑΛΓΟΡΙΘΜΩΝ ΕΧΟΥΝ ΤΗΝ ΕΞΗΣ ΜΟΡΦΗ:

1. ΑΡΧΙΚΟΠΟΙΗΣΗ ΠΛΗΘΥΣΜΟΥ.
2. ΥΠΟΛΟΓΙΣΜΟΣ ΤΗΣ ΤΙΜΗΣ ΤΗΣ ΣΥΝΑΡΤΗΣΗΣ ΚΑΤΑΛΛΗΛΟΛΗΤΑΣ ΓΙΑ ΚΑΘΕ ΣΤΟΙΧΕΙΟ ΤΟΥ ΠΛΗΘΥΣΜΟΥ.
3. ΑΝΤΙΣΤΟΙΧΙΑ ΓΟΝΕΩΝ ΠΡΟΣ ΔΗΜΙΟΥΡΓΙΑ ΝΕΩΝ ΣΤΟΙΧΕΙΩΝ ΠΛΗΘΥΣΜΟΥ.
4. ΑΞΙΟΛΟΓΗΣΗ ΝΕΩΝ ΣΤΟΙΧΕΙΩΝ ΚΑΙ ΚΑΤΑΧΩΡΗΣΗ ΣΤΟΝ ΠΛΗΘΥΣΜΟ ΜΕ ΠΑΡΑΛΛΗΛΗ ΑΠΟΜΑΚΡΥΝΣΗ ΤΡΕΧΟΥΣΩΝ ΑΔΥΝΑΜΩΝ ΣΤΟΙΧΕΙΩΝ.
5. ΑΝ ΕΧΕΙ ΕΠΕΛΘΕΙ ΤΟ ΕΠΙΘΥΜΗΤΟ ΑΠΟΤΕΛΕΣΜΑ, ΟΛΟΚΛΗΡΩΣΗ, ΔΙΑΦΟΡΕΤΙΚΑ ΜΕΤΑΒΑΣΗ ΣΤΟ ΒΗΜΑ 3.

ΕΠΑΓΩΓΗ ΚΑΝΟΝΩΝ (RULEINDUCTION)

ΕΙΝΑΙ ΜΙΑ ΜΕΘΟΔΟΛΟΓΙΑ ΚΑΤΑ ΤΗΝ ΟΠΟΙΑ ΑΝΑΖΗΤΟΥΝΤΑΙ ΚΑΝΟΝΕΣ ΚΑΤΑΤΑΞΗΣ ΤΩΝ ΔΕΔΟΜΕΝΩΝ ΜΕ ΒΑΣΗ ΣΤΑΤΙΣΤΙΚΕΣ ΜΕΤΡΗΣΕΙΣ. ΣΤΗΝ ΠΡΑΓΜΑΤΙΚΟΤΗΤΑ ΑΠΟΤΕΛΕΙ ΜΙΑ ΕΝΑΛΛΑΚΤΙΚΗ ΠΡΟΣΕΓΓΙΣΗ ΤΩΝ ΔΕΝΔΡΩΝ ΑΠΟΦΑΣΗΣ ΟΠΟΥ Η ΚΑΤΑΤΑΞΗ ΤΩΝ ΔΕΔΟΜΕΝΩΝ ΣΕ ΚΛΑΣΕΙΣ ΓΙΝΕΤΑΙ ΜΕ ΒΑΣΗ ΚΑΝΟΝΕΣ ΤΗΣ ΜΟΡΦΗΣ:

IF<ΣΥΝΘΗΚΗ>THEN<ΚΑΤΕΤΑΞΕ ΣΤΗΝ ΚΛΑΣΗ>ENDIF

ΟΙ ΚΑΝΟΝΕΣ ΜΠΟΡΕΙ ΝΑ ΕΙΝΑΙ:

- ΠΡΟΤΑΣΙΑΚΟΙ: ΜΠΟΡΕΙ ΝΑ ΠΡΟΚΥΨΟΥΝ ΑΠΟ ΑΛΛΕΣ ΜΟΡΦΕΣ ΑΝΑΠΑΡΑΣΤΑΣΗΣ (ΔΕΝΔΡΑ ΑΠΟΦΑΣΗΣ, ΓΕΝΕΤΙΚΟΙ ΑΛΓΟΡΙΘΜΟΙ) ΚΑΙ ΔΕΝ ΣΥΜΠΕΡΙΛΑΜΒΑΝΟΥΝ ΜΕΤΑΒΛΗΤΕΣ ΜΕ ΣΥΝΕΠΕΙΑ ΝΑ ΑΔΥΝΑΤΟΥΝ ΝΑ ΑΝΑΠΑΡΑΣΤΗΣΟΥΝ ΓΕΝΙΚΕΣ ΣΧΕΣΕΙΣ ΑΝΑΜΕΣΑ ΣΕ ΤΙΜΕΣ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ.
- ΚΑΤΗΓΟΡΗΜΑΤΙΚΟΙ ΠΡΩΤΗΣ ΤΑΞΕΩΣ: ΑΝΤΙΘΕΤΑ ΑΠΟ ΤΟΥΣ ΠΡΟΤΑΣΙΑΚΟΥΣ ΕΧΟΥΝ ΤΗΝ ΔΥΝΑΤΟΤΗΤΑ ΝΑ ΑΝΑΠΑΡΙΣΤΟΥΝ ΓΕΝΙΚΕΣ ΣΧΕΣΕΙΣ ΜΕΣΩ ΜΕΤΑΒΛΗΤΩΝ.

ΜΙΑ ΥΠΟΘΕΣΗ ΚΑΛΥΠΤΕΤΑΙ ΑΠΟ ΕΝΑΝ ΚΑΝΟΝΑ R ΑΝ ΚΑΙ ΜΟΝΟ ΑΝ ΚΑΘΕ ΣΥΝΘΗΚΗ ΤΟΥ R ΙΚΑΝΟΠΟΙΕΙΤΑΙ ΑΠΟ ΤΗΝ ΑΝΤΙΣΤΟΙΧΗ ΤΙΜΗ ΜΕΤΑΒΛΗΤΗΣ. Η ΚΑΤΑΣΤΑΣΗ ΠΟΥ ΕΞΕΤΑΖΕΤΑΙ ΚΑΛΥΠΤΕΤΑΙ ΠΛΗΡΩΣ ΑΠΟ ΕΝΑ ΣΥΝΟΛΟ ΚΑΝΟΝΩΝ ΑΝ ΓΙΑ ΚΑΘΕ ΥΠΟΘΕΣΗ ΤΗΣ ΚΑΤΑΣΤΑΣΗΣ ΑΥΤΗΣ ΥΠΑΡΧΕΙ ΕΝΑΝ ΚΑΝΟΝΑΣ ΠΟΥ ΤΗΝ ΙΚΑΝΟΠΟΙΕΙ [13][14].

ΜΕΘΟΔΟΣ ΤΟΥ ΚΟΝΤΙΝΟΤΕΡΟΥ ΓΕΙΤΟΝΑ (NEARESTNEIGHBORMETHOD)

ΠΡΟΚΕΙΤΑΙ ΓΙΑ ΜΙΑ ΑΠΟ ΤΙΣ ΠΑΛΑΙΟΤΕΡΕΣ ΜΕΘΟΔΟΥΣ ΓΙΑ ΤΗΝ ΕΞΟΡΥΞΗ ΔΕΔΟΜΕΝΩΝ. ΕΙΝΑΙ ΜΙΑ ΜΕΘΟΔΟΣ ΠΡΟΒΛΕΨΗΣ ΠΟΥ ΠΡΟΣΟΜΟΙΑΖΕΙ ΜΕ ΣΥΣΤΑΔΟΠΟΙΗΣΗ. ΒΑΣΙΖΕΤΑΙ ΣΤΗΝ ΑΝΑΛΥΣΗ ΤΟΥ ΙΣΤΟΡΙΚΟΥ ΤΙΜΩΝ ΠΑΡΑΜΕΤΡΩΝ ΠΟΥ ΒΡΙΣΚΟΝΤΑΙ ΚΟΝΤΑ ΣΤΙΣ ΠΡΟΣ ΕΞΕΤΑΣΗ ΤΙΜΕΣ ΚΑΘΩΣ ΚΑΙ ΣΤΗΝ ΓΕΝΙΚΗ ΠΑΡΑΔΟΧΗ ΟΤΙ ΜΙΑ

ΣΥΓΚΕΚΡΙΜΕΝΗ ΟΝΤΟΤΗΤΑ ΕΙΝΑΙ ΠΛΗΣΙΕΣΤΕΡΑ ΣΕ ΜΙΑ ΑΛΛΗ ΟΝΤΟΤΗΤΑ ΣΕ ΣΧΕΣΗ ΜΕ ΜΙΑ ΤΡΙΤΗ. Η ΠΡΟΒΛΕΨΗ ΑΥΤΗ ΜΑΘΗΜΑΤΙΚΑ ΕΚΦΡΑΖΕΤΑΙ ΑΠΟ ΤΗΝ ΣΧΕΣΗ:

$$\hat{y} = \frac{1}{K} \sum_{x_i \in N_k(x)} y_i$$

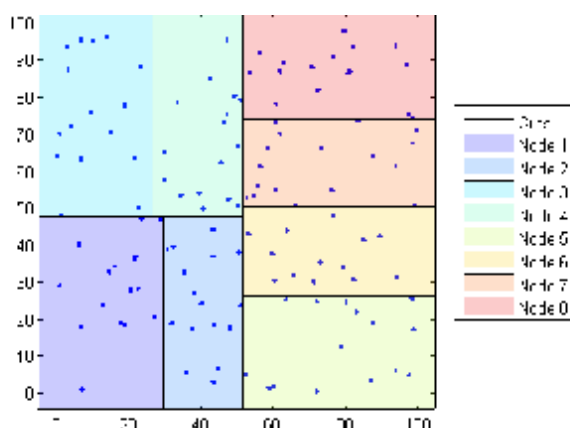
ΟΠΟΥ $N_{k(x)}$ ΕΙΝΑΙ ΤΟ ΣΥΝΟΛΟ ΠΟΥ ΠΕΡΙΕΧΕΙ ΤΙΣ Κ ΠΙΟ ΚΟΝΤΙΝΕΣ ΠΑΡΑΤΗΡΗΣΕΙΣ ΣΤΟ Χ ΓΙΑ ΤΟ ΟΠΟΙΟ ΕΙΝΑΙ ΕΠΙΘΥΜΗΤΟ ΝΑ ΓΙΝΕΙ ΠΡΟΒΛΕΨΗ.

ΔΙΑΙΣΘΗΤΙΚΑ ΣΚΟΠΟΣ ΕΙΝΑΙ ΝΑ ΓΙΝΕΙ ΠΡΟΒΛΕΨΗ ΤΗΣ ΤΙΜΗΣ ΜΙΑΣ ΠΑΡΑΜΕΤΡΟΥ ΣΕ ΕΠΙΠΕΔΑ ΠΟΥ ΒΡΙΣΚΟΝΤΑΙ ΚΟΝΤΑ ΣΕ ΕΝΑΝ ΜΕΣΟ ΟΡΟ ΠΑΡΕΛΘΟΝΤΩΝ ΠΑΡΑΤΗΡΗΣΕΩΝ ΤΗΣ ΤΙΜΗΣ ΜΙΑ ΑΛΛΗΣ ΠΑΡΑΜΕΤΡΟΥ. ΚΟΜΒΙΚΟΙ ΣΥΝΤΕΛΕΣΤΕΣ ΤΗΣ ΜΕΘΟΔΟΥ ΕΙΝΑΙ Ο ΑΡΙΘΜΟΣ ΤΩΝ ΚΟΝΤΙΝΩΝ ΠΑΡΑΤΗΡΗΣΕΩΝ ΚΑΘΩΣ ΚΑΙ Η ΑΠΟΣΤΑΣΗ ΑΝΑΦΟΡΑΣ ΓΙΑ ΤΟΝ ΧΑΡΑΚΤΗΡΙΣΜΟ ΤΩΝ ΚΟΝΤΙΝΩΝ ΠΑΡΑΤΗΡΗΣΕΩΝ. Η ΕΠΙΛΟΓΗ ΤΗΣ ΑΠΟΣΤΑΣΗΣ ΕΙΝΑΙ ΣΥΝΑΡΤΗΣΗ ΤΗΣ ΦΥΣΗΣ ΤΟΥ ΠΡΟΒΛΗΜΑΤΟΣ ΠΟΥ ΑΝΤΙΜΕΤΩΠΙΖΕΤΑΙ ΚΑΘΕ ΦΟΡΑ ΑΛΛΑ ΚΑΙ ΤΗΣ ΕΠΙΘΥΜΗΤΗΣ ΑΚΡΙΒΕΙΑΣ ΤΟΥ ΑΠΟΤΕΛΕΣΜΑΤΟΣ. ΠΡΟΚΕΙΜΕΝΟΥ ΝΑ ΕΠΙΤΥΓΧΑΝΕΤΑΙ ΜΕΓΑΛΥΤΕΡΗ ΑΚΡΙΒΕΙΑ ΣΤΑΘΜΙΖΟΝΤΑΙ ΟΙ ΠΑΡΑΤΗΡΗΣΕΙΣ ΜΕ ΒΑΡΗ ΑΝΤΙΣΤΡΟΦΩΣ ΑΝΑΛΟΓΑ ΜΕ ΤΗΝ ΑΠΟΣΤΑΣΗ ΤΩΝ ΠΑΡΑΤΗΡΗΣΕΩΝ. ΣΕ ΠΕΡΙΠΤΩΣΕΙΣ ΠΟΥ ΕΙΝΑΙ ΑΠΑΙΤΗΤΟ ΝΑ ΓΙΝΕΙ ΚΑΤΑΤΑΞΗ ΤΩΝ ΠΑΡΑΤΗΡΗΣΕΩΝ, ΔΕΝ ΥΠΟΛΟΓΙΖΕΤΑΙ Ο ΜΕΣΟΣ ΟΡΟΣ ΤΩΝ ΠΑΡΑΤΗΡΗΣΕΩΝ ΑΛΛΑ ΕΝΤΟΠΙΖΟΝΤΑΙ ΟΙ ΟΜΑΔΕΣ ΕΚΕΙΝΕΣ ΤΩΝ ΠΑΡΑΤΗΡΗΣΕΩΝ ΜΕ ΤΗ ΜΕΓΑΛΥΤΕΡΗ ΣΥΧΝΟΤΗΤΑ ΣΤΙΣ ΚΟΝΤΙΝΟΤΕΡΕΣ. ΈΤΣΙ ΟΙ ΕΛΕΓΧΟΜΕΝΕΣ ΠΑΡΑΤΗΡΗΣΕΙΣ ΚΑΤΑΤΑΣΣΟΝΤΑΙ ΣΕ ΕΚΕΙΝΕΣ ΤΙΣ ΟΜΑΔΕΣ ΜΕ ΤΗΝ ΜΕΓΑΛΥΤΕΡΗ ΣΥΧΝΟΤΗΤΑ (Η ΣΧΕΤΙΚΗ ΣΥΧΝΟΤΗΤΑ).

Η ΜΕΘΟΔΟΣ ΑΚΟΛΟΥΘΕΙ ΤΗΝ ΥΠΟΘΕΣΗ ΟΤΙ ΑΝΤΙΚΕΙΜΕΝΑ ΤΑ ΟΠΟΙΑ ΕΧΟΥΝ ΤΙΜΕΣ ΠΑΡΑΜΕΤΡΩΝ ΠΟΥ ΒΡΙΣΚΟΝΤΑΙ ΚΟΝΤΑ ΕΙΝΑΙ ΠΟΛΥ ΠΙΘΑΝΟΝ ΚΑΙ ΟΙ ΤΙΜΕΣ ΤΩΝ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ ΤΟΥΣ ΠΟΥ ΠΡΕΠΕΙ ΝΑ ΠΡΟΒΛΕΦΘΟΥΝ ΝΑ ΕΙΝΑΙ ΕΠΙΣΗΣ ΚΟΝΤΑ. ΟΙ ΠΕΡΙΟΡΙΣΜΟΙ ΠΟΥ ΕΜΠΕΡΙΕΧΕΙ Η ΜΕΘΟΔΟΣ ΑΥΤΗ ΕΧΟΥΝ ΝΑ ΚΑΝΟΥΝ ΜΕ ΤΟΝ

ΜΕΓΑΛΟ ΣΧΕΤΙΚΑ ΧΡΟΝΟ ΠΟΥ ΑΠΑΙΤΕΙΤΑΙ ΓΙΑ ΤΗΝ ΕΚΠΑΙΔΕΥΣΗ ΤΟΥ ΑΛΓΟΡΙΘΜΟΥ ΑΛΛΑ ΚΑΙ ΜΕ ΤΟ ΜΕΓΑΛΟ ΜΕΓΕΘΟΣ ΤΩΝ ΔΕΙΓΜΑΤΩΝ ΠΟΥ ΑΠΑΙΤΟΥΝΤΑΙ ΓΙΑ ΤΗΝ ΠΑΡΑΓΩΓΗ ΑΚΡΙΒΩΝ ΑΠΟΤΕΛΕΣΜΑΤΩΝ. ΓΙΑ ΤΗΝ ΑΠΟΔΟΤΙΚΗ ΑΝΤΙΜΕΤΩΠΙΣΗ ΤΩΝ ΠΕΡΙΟΡΙΣΜΩΝ ΑΥΤΩΝ ΕΧΟΥΝ ΠΡΟΤΑΘΕΙ ΔΙΑΦΟΡΕΣ ΠΡΟΣΕΓΓΙΣΕΙΣ, ΟΠΩΣ ΜΕΙΩΣΗ ΤΩΝ ΔΙΑΣΤΑΣΕΩΝ ΤΩΝ ΔΕΔΟΜΕΝΩΝ, ΧΡΗΣΗ ΚΑΤΑΛΛΗΛΩΝ ΔΟΜΩΝ ΔΕΔΟΜΕΝΩΝ ΓΙΑ ΤΑΧΥΤΕΡΗ ΕΠΕΞΕΡΓΑΣΙΑ, ΤΗΝ ΑΠΟΜΑΚΡΥΝΣΗ ΤΟΥ ΘΟΡΥΒΟΥ ΤΩΝ ΔΕΔΟΜΕΝΩΝ (ΔΕΔΟΜΕΝΑ ΠΟΥ ΕΙΝΑΙ ΠΡΟΦΑΝΕΣ ΟΤΙ ΔΕΝ ΜΠΟΡΟΥΝ ΝΑ ΕΝΤΑΧΘΟΥΝ ΣΕ ΚΑΠΟΙΑ ΚΛΑΣΗ).

Ο ΠΙΟ ΧΑΡΑΚΤΗΡΙΣΤΙΚΟΣ ΕΚΠΡΟΣΩΠΟΣ ΤΩΝ ΑΛΓΟΡΙΘΜΩΝ ΑΥΤΩΝ ΕΙΝΑΙ Ο K-NEARESTNEIGHBORS Ο ΟΠΟΙΟΣ ΒΑΣΙΖΕΤΑΙ ΣΤΗΝ ΠΑΡΑΔΟΧΗ ΟΤΙ ΚΑΘΕ ΠΑΡΑΤΗΡΗΣΗ ΜΠΟΡΕΙ ΝΑ ΑΝΑΠΑΡΑΣΤΑΘΕΙ ΣΑΝ ΣΗΜΕΙΟ ΣΕ ΕΥΚΛΕΙΔΕΙΟ ΧΩΡΟ ΤΟΣΩΝ ΔΙΑΣΤΑΣΕΩΝ ΟΣΕΣ ΚΑΙ ΟΙ ΠΑΡΑΜΕΤΡΟΙ ΤΗΣ. ΟΙ K ΚΟΝΤΙΝΟΤΕΡΟΙ ΓΕΙΤΟΝΕΣ ΕΝΤΟΠΙΖΟΝΤΑΙ ΚΑΙ ΥΠΟΛΟΓΙΖΟΝΤΑΙ ΜΕ ΒΑΣΗ ΤΗΝ ΜΕΤΑΞΥ ΤΩΝ ΣΗΜΕΙΩΝ ΕΥΚΛΕΙΔΕΙΑ ΑΠΟΣΤΑΣΗ. ΣΤΗΝ ΕΠΟΜΕΝΗ ΕΙΚΟΝΑ ΠΑΡΟΥΣΙΑΖΕΤΑΙ ΜΙΑ ΓΡΑΦΙΚΗ ΑΝΑΠΑΡΑΣΤΑΣΗ ΚΑΤΗΓΟΡΙΟΠΟΙΗΣΗΣ ΠΟΥ ΒΑΣΙΣΤΗΚΕ ΣΤΗΝ ΜΕΘΟΔΟ ΤΟΥ ΚΟΝΤΙΝΟΤΕΡΟΥ ΓΕΙΤΟΝΑ. ΣΤΟ ΣΥΓΚΕΚΡΙΜΕΝΟ ΣΧΗΜΑ ΦΑΙΝΕΤΑΙ Η ΔΙΑΛΟΓΗ ΤΩΝ ΠΑΡΑΤΗΡΗΣΕΩΝ ΣΤΟ ΣΕ 8 ΠΕΡΙΟΧΕΣ ΣΤΟ ΕΠΙΠΕΔΟ ΤΩΝ ΔΥΟ ΔΙΑΣΤΑΣΕΩΝ [16][17].



ΕΙΚΟΝΑ 6: ΚΑΤΗΓΟΡΙΟΠΟΙΗΣΗ ΜΕ ΤΗΝ ΜΕΘΟΔΟ ΤΟΥ ΚΟΝΤΙΝΟΤΕΡΟΥ ΓΕΙΤΟΝΑ (ΠΗΓΗ: MATHWORKS, CLASSIFICATIONUSINGNEARESTNEIGHBORS, [HTTP://WWW.MATHWORKS.COM/HELP/STATS/CLASSIFICATION-USING-NEAREST-NEIGHBORS.HTML](http://www.mathworks.com/help/stats/classification-using-nearest-neighbors.html)).

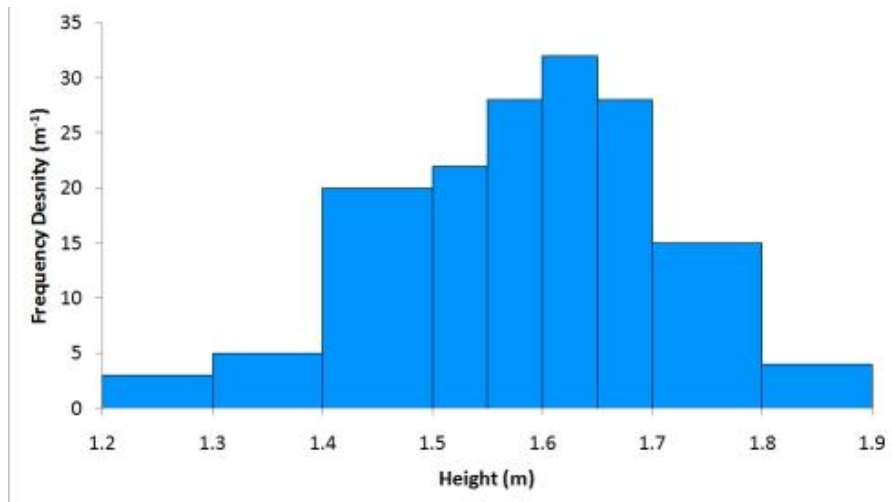
ΓΡΑΦΙΚΗ ΑΝΑΛΥΣΗ (DATAVISUALIZATION)

ΤΑ ΔΕΔΟΜΕΝΑ ΑΠΕΙΚΟΝΙΖΟΝΤΑΙ ΜΕ ΤΗΝ ΧΡΗΣΗ ΓΡΑΦΙΚΩΝ ΜΕΘΟΔΩΝ ΚΑΙ ΕΤΣΙ ΠΑΡΕΧΕΤΑΙ Η ΔΥΝΑΤΟΤΗΤΑ ΣΤΟΥΣ ΑΝΑΛΥΤΕΣ ΝΑ ΕΚΤΙΜΗΣΟΥΝ ΤΗΝ ΣΥΜΠΕΡΙΦΟΡΑ ΤΟΥΣ ΔΙΑΙΣΘΗΤΙΚΑ. Η ΓΡΑΦΙΚΗ ΑΝΑΛΥΣΗ ΣΤΗΡΙΖΕΤΑΙ ΣΤΗΝ ΠΑΡΟΥΣΙΑΣΗ ΤΩΝ ΔΕΔΟΜΕΝΩΝ ΣΕ ΜΙΑ ΕΙΚΟΝΟΓΡΑΦΙΚΗ Η ΓΡΑΦΙΚΗ ΜΟΡΦΗ. ΕΙΝΑΙ ΑΡΚΕΤΑ ΑΠΟΔΟΤΙΚΗ ΜΕΘΟΔΟΛΟΓΙΑ ΛΟΓΩ ΤΟΥ ΤΡΟΠΟΥ ΜΕ ΤΟΝ ΟΠΟΙΟ ΑΝΘΡΩΠΙΝΟΣ ΕΓΚΕΦΑΛΟΣ ΕΠΕΞΕΡΓΑΖΕΤΑΙ ΤΙΣ ΠΛΗΡΟΦΟΡΙΕΣ, ΑΦΟΥ ΕΙΝΑΙ ΕΥΚΟΛΟΤΕΡΟ ΝΑ ΑΝΤΙΛΑΜΒΑΝΕΤΑΙ ΤΗΝ ΕΝΝΟΙΑ ΚΑΙ ΤΗ ΣΗΜΑΣΙΑ ΠΟΛΛΩΝ ΣΗΜΕΙΩΝ ΔΕΔΟΜΕΝΩΝ, ΟΤΑΝ ΕΜΦΑΝΙΖΟΝΤΑΙ ΣΕ ΧΑΡΤΕΣ ΚΑΙ ΓΡΑΦΗΜΑΤΑ ΣΕ ΣΧΕΣΗ ΜΕ ΤΗΝ ΜΕΛΕΤΗ ΤΩΝ ΑΝΤΙΣΤΟΙΧΩΝ ΠΙΝΑΚΩΝ ΔΕΔΟΜΕΝΩΝ Η ΚΕΙΜΕΝΟΥ. ΕΙΔΙΚΟΤΕΡΑ ΣΕ ΠΕΡΙΠΤΩΣΕΙΣ ΜΕΓΑΛΟΥ ΟΓΚΟΥ ΔΕΔΟΜΕΝΩΝ (ΠΟΥ ΣΥΝΗΘΩΣ ΧΡΗΣΙΜΟΠΟΙΟΥΝΤΑΙ ΣΤΙΣ ΔΙΑΔΙΚΑΣΙΕΣ ΕΞΟΡΥΞΗΣ ΔΕΔΟΜΕΝΩΝ) ΕΙΝΑΙ ΠΟΛΥ ΕΥΚΟΛΟΤΕΡΟ ΝΑ ΕΝΤΟΠΙΣΤΟΥΝ ΤΑ ΥΠΟΒΟΣΚΟΝΤΑ ΠΡΟΤΥΠΑ.

ΟΙ ΑΠΕΙΚΟΝΙΣΕΙΣ ΜΕΤΑΦΕΡΟΥΝ ΤΙΣ ΠΛΗΡΟΦΟΡΙΕΣ ΜΕ ΚΑΘΟΛΙΚΟ ΤΡΟΠΟ, ΚΑΘΙΣΤΩΝΤΑΣ ΠΟΛΥ ΑΠΛΗ ΤΗ ΔΙΑΝΟΜΗ ΤΟΥΣ ΣΕ ΤΡΙΤΟΥΣ ΚΑΙ ΤΗΝ ΑΝΤΑΛΛΑΓΗ ΙΔΕΩΝ. Η ΑΠΕΙΚΟΝΙΣΗ ΔΕΔΟΜΕΝΩΝ ΠΑΡΟΥΣΙΑΖΕΙ ΤΑ ΔΕΔΟΜΕΝΑ ΜΕ ΤΡΟΠΟ ΠΟΥ Ο ΑΝΑΛΥΤΗΣ ΜΠΟΡΕΙ ΕΥΚΟΛΑ ΝΑ ΕΡΜΗΝΕΥΣΕΙ, ΕΞΟΙΚΟΝΟΜΩΝΤΑΣ ΧΡΟΝΟ ΚΑΙ ΕΝΕΡΓΕΙΑ.

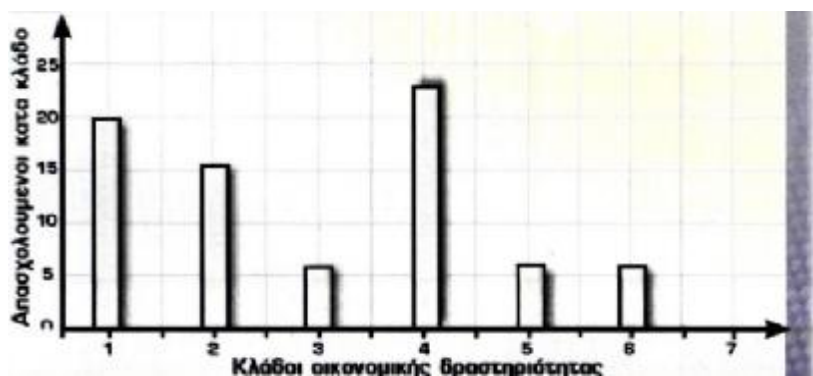
ΟΙ ΓΡΑΦΙΚΕΣ ΑΠΕΙΚΟΝΙΣΕΙΣ ΠΟΥ ΧΡΗΣΙΜΟΠΟΙΟΥΝΤΑΙ ΣΥΧΝΟΤΕΡΑ ΕΙΝΑΙ:

- ΓΙΑ ΤΗΝ ΜΕΛΕΤΗ ΤΟ ΤΡΟΠΟΥ ΚΑΤΑΝΟΜΗΣ ΤΩΝ ΤΙΜΩΝ ΜΙΑΣ ΠΑΡΑΜΕΤΡΟΥ:
 - ΙΣΤΟΓΡΑΜΜΑΤΑ: ΧΡΗΣΙΜΟΠΟΙΟΥΝΤΑΙ ΓΙΑ ΤΗ ΓΡΑΦΙΚΗ ΑΠΕΙΚΟΝΙΣΗ ΠΟΣΟΤΙΚΩΝ ΚΑΤΑΝΟΜΩΝ, ΑΠΟΤΕΛΟΥΝΤΑΙ ΑΠΟ ΔΙΑΔΟΧΙΚΑ ΟΡΘΟΓΩΝΙΑ, ΠΟΥ ΕΧΟΥΝ ΒΑΣΕΙΣ ΙΣΕΣ ΜΕ ΤΑ ΔΙΑΣΤΗΜΑΤΑ ΤΩΝ ΤΑΞΕΩΝ ΤΟΠΟΘΕΤΗΜΕΝΕΣ ΠΑΝΩ ΣΤΟΝ ΟΡΙΖΟΝΤΙΟ ΑΞΟΝΑ. ΤΟ ΕΜΒΑΔΟΝ ΚΑΘΕ ΟΡΘΟΓΩΝΙΟΥ ΙΣΟΥΤΑΙ ΜΕ ΤΗ ΣΥΧΝΟΤΗΤΑ ΤΗΣ ΑΝΤΙΣΤΟΙΧΗΣ ΤΑΞΕΩΣ.



ΕΙΚΟΝΑ7: ΠΑΡΑΔΕΙΓΜΑΙΣΤΟΓΡΑΜΜΑΤΟΣ (ΠΗΓΗ: ΚΕΝ CECIRE, HISTOGRAMS: CONSTRUCTION, ANALYSIS AND UNDERSTANDING, [HTTP://QUARKNET.FNAL.GOV/TOOLKITS/ATI/HISTOGRAMS.HTML](http://quarknet.fnal.gov/toolkits/ati/histograms.html))

- ΔΙΑΓΡΑΜΜΑΤΑ ΚΑΤΑΝΟΜΗΣ ΣΥΧΝΟΤΗΤΩΝ: ΣΤΑ ΔΙΑΓΡΑΜΜΑΤΑ ΑΥΤΑ ΚΑΤΑΓΡΑΦΟΝΤΑΙ ΟΙ ΑΠΟΛΥΤΕΣ ΤΙΜΕΣ ΠΟΥ ΛΑΜΒΑΝΟΥΝ ΟΙ ΚΑΤΑΜΕΤΡΗΘΗΣΕΣ ΠΑΡΑΜΕΤΡΟΙ ΚΑΙ ΑΝΑΠΑΡΙΣΤΑΝΤΑΙ ΜΕ ΡΑΒΔΟΥΣ ΥΨΟΥΣ ΑΝΑΛΟΓΟΥ ΤΗΣ ΤΙΜΗΣ ΤΟΥΣ.

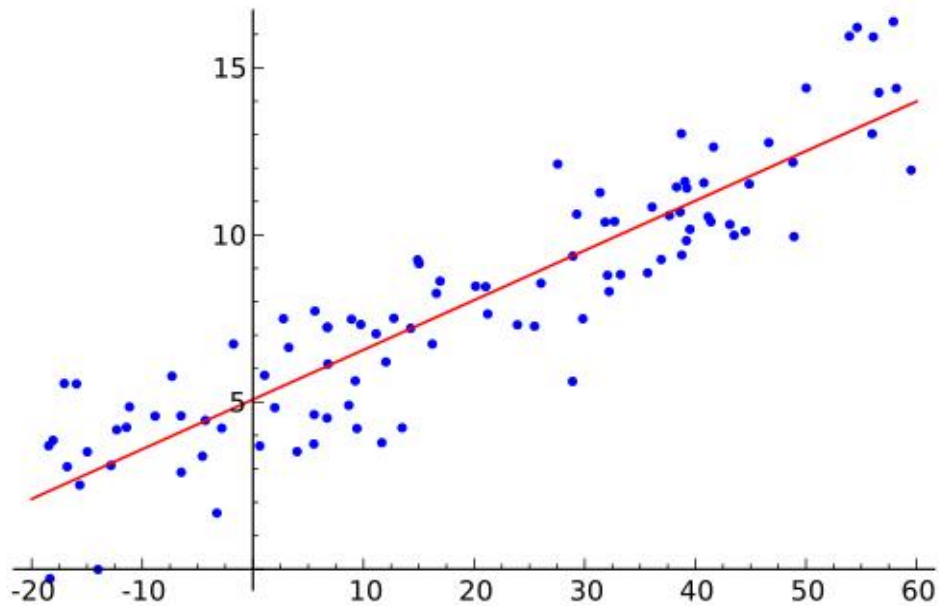


ΕΙΚΟΝΑ 8: ΔΙΑΓΡΑΜΜΑ ΚΑΤΑΝΟΜΗΣ ΣΥΧΝΟΤΗΤΩΝ (ΠΗΓΗ: ΣΧΟΛΙΚΑ ΔΙΑΔΡΑΣΤΙΚΑ ΒΙΒΛΙΑ, ΕΙΣΑΓΩΓΗ ΣΤΗΝ ΣΤΑΤΙΣΤΙΚΗ, [HTTP://EBOOKS.EDU.GR/MODULES/EBOOK/SHOW.PHP/DSGL-C125/494/3206,13027/](http://ebooks.edu.gr/modules/ebook/show.php/dsGL-C125/494/3206,13027/))

- ΑΚΙΔΩΤΑ ΔΙΑΓΡΑΜΜΑΤΑ: ΕΙΝΑΙ ΔΙΑΓΡΑΜΜΑΤΑ ΠΟΥ ΧΡΗΣΙΜΟΠΟΙΟΥΝΤΑΙ ΚΥΡΙΩΣ ΓΙΑ ΤΗ ΓΡΑΦΙΚΗ ΑΠΕΙΚΟΝΙΣΗ ΠΟΙΟΤΙΚΩΝ ΔΕΔΟΜΕΝΩΝ. ΑΠΟΤΕΛΟΥΝΤΑΙ ΑΠΟ ΓΡΑΦΙΚΕΣ

ΣΤΗΛΕΣ ΠΟΥ ΤΟ ΥΨΟΣ ΕΙΝΑΙ ΑΝΑΛΟΓΟ ΤΗΣ ΤΙΜΗΣ ΜΙΑΣ ΠΑΡΑΜΕΤΡΟΥ

- ΠΙΤΕΣ: ΣΥΝΘΕΤΟΥΝ ΕΝΑΝ ΚΥΚΛΟ ΧΩΡΙΣΜΕΝΟ ΣΕ ΤΟΜΕΙΣ ΚΑΘΕ ΕΝΑΣ ΕΚ ΤΩΝ ΟΠΟΙΩΝ ΑΝΤΙΣΤΟΙΧΕΙ ΣΕ ΜΙΑ ΤΙΜΗ ΤΗΣ ΜΕΤΑΒΛΗΤΗΣ (ΤΟ ΣΥΝΟΛΟ ΤΩΝ ΕΠΙΩ ΜΕΡΟΥΣ ΤΙΜΩΝ ΣΥΝΘΕΤΟΥΝ ΤΟ ΟΛΟΝ ΤΟΥ ΣΥΝΟΛΟΥ). Η ΓΩΝΙΑ ΤΟΥ ΚΑΘΕ ΚΥΚΛΙΚΟΥ ΤΟΜΕΑ ΕΙΝΑΙ ΑΝΑΛΟΓΗ ΜΕ ΤΗΝ ΑΝΤΙΣΤΟΙΧΗ ΣΧΕΤΙΚΗ ΣΥΧΝΟΤΗΤΑ ΤΗΣ ΤΙΜΗΣ ΤΗΣ ΚΑΘΕ ΜΕΤΑΒΛΗΤΗΣ.
- ΓΙΑ ΤΗΝ ΜΕΤΑΒΟΛΗ ΤΩΝ ΜΕΓΕΘΩΝ ΣΤΟΝ ΧΡΟΝΟ Ο ΟΡΙΖΟΝΤΙΟΣ ΑΞΟΝΑΣ ΧΡΗΣΙΜΟΠΟΙΕΙΤΑΙ ΩΣ ΑΞΟΝΑΣ ΜΕΤΡΗΣΕΩΣ ΤΟΥ ΧΡΟΝΟΥ ΚΑΙ Ο ΚΑΘΕΤΟΣ ΩΣ ΑΞΟΝΑΣ ΜΕΤΡΗΣΕΩΣ ΤΗΣ ΕΞΕΤΑΖΟΜΕΝΗΣ ΜΕΤΑΒΛΗΤΗΣ
- ΓΙΑ ΤΗΝ ΑΛΛΗΛΕΠΙΔΡΑΣΗ ΔΥΟ ΠΑΡΑΜΕΤΡΩΝ ΜΕ ΣΥΝΕΧΕΣ ΠΕΔΙΟ ΤΙΜΩΝ ΧΡΗΣΙΜΟΠΟΙΟΥΝΤΑΙ ΤΑ ΔΙΑΓΡΑΜΜΑΤΑ ΔΙΑΣΠΟΡΑΣ. ΒΡΙΣΚΟΥΝ ΕΦΑΡΜΟΓΗ ΚΥΡΙΩΣ ΣΕ ΠΕΡΙΠΤΩΣΕΙΣ ΠΟΥ ΣΚΟΠΟΣ ΕΙΝΑΙ ΝΑ ΕΝΤΟΠΙΣΤΕΙ Ο ΤΡΟΠΟΣ ΜΕ ΤΟΝ ΟΠΟΙΟ ΑΛΛΗΛΟΕΠΙΔΡΟΥΝ ΟΙ ΤΙΜΕΣ ΔΥΟ ΜΕΤΑΒΛΗΤΩΝ. ΕΙΝΑΙ Η ΓΡΑΦΙΚΗ ΑΠΕΙΚΟΝΙΣΗ ΤΩΝ ΣΗΜΕΙΩΝ ΣΤΟ ΕΠΙΠΕΔΟ ΠΟΥ ΑΝΤΙΣΤΟΙΧΟΥΝ ΣΤΙΣ ΤΙΜΕΣ ΖΕΥΓΟΥΣ ΜΕΤΑΒΛΗΤΩΝ. Η ΘΕΣΗ ΤΩΝ ΣΗΜΕΙΩΝ ΣΤΟ ΕΠΙΠΕΔΟ ΣΗΜΑΤΟΔΟΤΕΙ ΜΙΑ ΤΑΣΗ ΣΤΗΝ ΣΧΕΣΗ ΜΕΤΑΞΥ ΤΩΝ ΜΕΤΑΒΛΗΤΩΝ.



ΕΙΚΟΝΑ 9: ΠΑΡΑΔΕΙΓΜΑ ΔΙΑΓΡΑΜΜΑΤΟΣ ΔΙΑΣΠΟΡΑΣ (ΠΗΓΗ:WIKIPEDIA, ΑΠΛΗ ΓΡΑΜΜΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ, [HTTPS://EL.WIKIPEDIA.ORG/WIKI/%CE%91%CF%80%CE%BB%CE%AE_%CE%B3%CF%81%CE%B1%CE%BC%CE%BC%CE%B9%CE%BA%CE%AE_%CF%80%CE%B1%CE%BB%CE%B9%CE%BD%CE%B4%CF%81%CF%8C%CE%BC%CE%B7%CF%83%CE%B7](https://el.wikipedia.org/wiki/%CE%91%CF%80%CE%BB%CE%AE_%CE%B3%CF%81%CE%B1%CE%BC%CE%BC%CE%B9%CE%BA%CE%AE_%CF%80%CE%B1%CE%BB%CE%B9%CE%BD%CE%B4%CF%81%CF%8C%CE%BC%CE%B7%CF%83%CE%B7))

- ΓΙΑ ΤΗΝ ΑΝΑΛΥΣΗ ΑΛΛΗΛΟΕΠΙΔΡΑΣΗΣ ΜΙΑΣ ΠΟΙΟΤΙΚΗΣ ΜΕΤΑΒΛΗΤΗΣ ΜΕ ΜΙΑ ΠΟΣΟΤΙΚΗ Η ΔΥΟ ΤΟΥ ΙΔΙΟΥ ΤΥΠΟΥ ΜΕΤΑΒΛΗΤΕΣ ΧΡΗΣΙΜΟΠΟΙΕΙΤΑΙ Ο ΠΙΝΑΚΑΣ ΣΥΝΑΦΕΙΑΣ ΠΟΥ ΕΙΝΑΙ Η ΠΑΡΑΘΕΣΗ ΦΥΣΙΚΩΝ ΑΡΙΘΜΩΝ ΣΕ ΕΝΑ ΔΙΣΔΙΑΣΤΑΤΟ ΠΙΝΑΚΑ ΤΩΝ ΔΕΙΓΜΑΤΩΝ ΤΙΜΩΝ ΠΟΥ ΛΑΜΒΑΝΟΝΤΑΙ ΓΙΑ ΕΝΑ ΣΥΝΟΛΟ ΠΑΡΑΜΕΤΡΩΝ. Η ΣΥΝΑΦΕΙΑ ΤΩΝ ΤΙΜΩΝ ΤΩΝ ΔΙΑΦΟΡΩΝ ΠΑΡΑΜΕΤΡΩΝ ΓΙΑ ΤΟ ΣΥΝΟΛΟ ΤΩΝ ΠΑΡΑΜΕΤΡΩΝ ΕΛΕΓΧΕΤΑΙ ΓΙΑ ΤΗΝ ΚΑΤΑΤΑΞΗ ΤΩΝ ΟΝΤΟΤΗΤΩΝ ΠΟΥ ΑΝΤΙΠΡΟΣΩΠΕΥΟΝΤΑΙ ΑΠΟ ΤΑ ΔΕΙΓΜΑΤΑ ΣΕ ΚΑΤΗΓΟΡΙΕΣ [18][19].

	POMS T	POMS D	POMS V	POMS F	BDI
Λιμφοκύτταρα	-0,1636	-0,2789	-0,3472	0,1718	-0,3482
T4 Λιμφοκύτταρα	-0,3633	-0,3038	-0,3558	-0,0539	-0,3669
T8 Λιμφοκύτταρα	-0,2297	-0,2645	-0,1462	0,0703	-0,4348
Αναλογία T4 : T8	0,1955	0,1404	0,0696	-0,0529	0,3003

ΕΙΚΟΝΑ 10: ΜΟΡΦΗ ΠΙΝΑΚΑ ΣΥΝΑΦΕΙΑΣ (ΠΗΓΗ:ΕΓΚΕΦΑΛΟΣ, [HTTP://WWW.ENKEPHALOS.GR/FULL/39-2-05-1G.HTM](http://www.enkephalos.gr/full/39-2-05-1g.htm))

1.4 ΜΟΝΤΕΛΑ ΠΛΗΡΟΦΟΡΙΩΝ

ΌΠΟΙΑ ΜΕΘΟΔΟΛΟΓΙΑ ΚΑΙ ΝΑ ΧΡΗΣΙΜΟΠΟΙΗΘΕΙ, ΑΠΟ ΤΗΝ ΔΙΑΔΙΚΑΣΙΑ ΤΗΣ ΕΞΟΡΥΞΗΣ ΔΕΔΟΜΕΝΩΝ ΣΤΟΧΟΣ ΕΙΝΑΙ ΝΑ ΠΑΡΑΧΘΟΥΝ ΤΑ ΕΞΗΣ ΜΟΝΤΕΛΑ ΠΛΗΡΟΦΟΡΙΩΝ:

- ΣΥΣΧΕΤΙΣΕΙΣ (ASSOCIATIONS): ΑΦΟΡΟΥΝ ΤΙΣ ΣΥΝΔΕΣΕΙΣ ΜΕΤΑΞΥ ΤΩΝ ΠΑΡΑΤΗΡΟΥΜΕΝΩΝ ΜΕΤΑΒΛΗΤΩΝ ΚΑΙ ΤΟ ΚΑΤΑ ΠΟΣΟ ΑΥΤΕΣ ΕΙΝΑΙ ΙΚΑΝΕΣ ΝΑ ΟΔΗΓΗΣΟΥΝ ΣΕ ΠΛΗΡΟΦΟΡΙΑ ΜΕ ΑΣΦΑΛΕΙΑ. ΚΑΤΑ ΤΟ ΜΟΝΤΕΛΟ ΑΥΤΟ ΟΙ ΤΙΜΕΣ ΠΟΥ ΛΑΜΒΑΝΟΥΝ ΔΥΟ Η ΠΕΡΙΣΣΟΤΕΡΕΣ ΜΕΤΑΒΛΗΤΕΣ ΣΕ ΕΝΑ ΜΕΓΑΛΟ ΣΥΝΟΛΟ ΔΕΔΟΜΕΝΩΝ ΕΝΔΕΧΟΜΕΝΩΣ ΝΑ ΚΡΥΒΟΥΝ ΕΝΑ ΜΕΤΑΞΥ ΤΟΥΣ ΣΥΣΧΕΤΙΣΤΙΚΟ ΜΟΤΙΒΟ. ΣΤΙΣ ΠΕΡΙΠΤΩΣΕΙΣ ΑΥΤΕΣ ΣΚΟΠΟΣ ΤΗΣ ΕΞΟΡΥΞΗΣ ΔΕΔΟΜΕΝΩΝ ΕΙΝΑΙ ΝΑ ΕΝΤΟΠΙΣΕΙ ΤΙΣ ΠΑΡΑΜΕΤΡΟΥΣ ΑΥΤΕΣ ΚΑΙ ΣΤΗΝ ΣΥΝΕΧΕΙΑ ΝΑ ΚΑΘΟΡΙΣΕΙ ΤΗ ΣΥΝΑΡΤΗΣΗ, ΠΟΥ ΔΕΔΟΜΕΝΩΝ ΤΩΝ ΤΙΜΩΝ ΤΩΝ ΑΝΙΧΝΕΥΘΕΝΤΩΝ ΠΑΡΑΜΕΤΡΩΝ, ΕΙΝΑΙ ΣΕ ΘΕΣΗ ΝΑ ΠΑΡΑΞΕΙ ΤΗΝ ΤΙΜΗ ΚΑΠΟΙΑΣ ΑΛΛΗΣ. ΠΑΡΑΔΕΙΓΜΑ ΕΝΟΣ ΤΕΤΟΙΟΥ ΜΟΝΤΕΛΟΥ ΕΙΝΑΙ ΟΙ ΜΗΧΑΝΙΣΜΟΙ ΤΩΝ ΔΙΚΤΥΑΚΩΝ ΤΟΠΩΝ ΗΛΕΚΤΡΟΝΙΚΩΝ ΑΓΟΡΩΝ. ΟΙ ΑΓΟΡΕΣ ΤΩΝ ΠΕΛΑΤΩΝ ΚΑΤΑΓΡΑΦΟΝΤΑΙ ΚΑΙ ΔΗΜΙΟΥΡΓΟΥΝΤΑΙ ΒΑΣΕΙΣ ΔΕΔΟΜΕΝΩΝ ΜΕ ΠΟΛΛΕΣ ΕΓΓΡΑΦΕΣ. ΚΑΤΑΛΛΗΛΕΣ ΛΕΙΤΟΥΡΓΙΕΣ ΕΛΕΓΧΟΥΝ ΤΙΣ ΠΑΡΕΛΘΟΥΣΕΣ ΑΓΟΡΕΣ ΚΑΙ ΔΗΜΙΟΥΡΓΟΥΝ ΚΑΝΟΝΕΣ. ΟΙ ΚΑΝΟΝΕΣ ΑΥΤΟΙ ΑΚΟΛΟΥΘΟΥΝΤΑΙ ΣΕ ΜΕΛΛΟΝΤΙΚΕΣ

ΑΓΟΡΕΣ ΓΙΑ ΝΑ ΠΡΟΤΕΙΝΟΥΝ ΣΤΟΥΣ ΕΠΙΔΟΞΟΥΣ ΑΓΟΡΑΣΤΕΣ ΕΠΙΠΛΕΟΝ ΠΡΟΪΟΝΤΑ ΚΑΙ ΥΠΗΡΕΣΙΕΣ ΠΟΥ ΕΧΟΥΝ ΣΤΟ ΠΑΡΕΛΘΟΝ ΕΠΙΛΕΞΕΙ ΑΓΟΡΑΣΤΕΣ ΠΟΥ ΑΓΟΡΑΣΑΝ ΤΑ ΤΡΕΧΟΝΤΑ ΕΠΙΛΕΓΜΕΝΑ.

- ΑΚΟΛΟΥΘΙΕΣ (SEQUENCES): ΠΡΟΚΕΙΤΑΙ ΓΙΑ ΜΟΡΦΗ ΠΛΗΡΟΦΟΡΙΑΣ ΠΟΥ ΕΧΕΙ ΣΑΝ ΠΑΡΑΜΕΤΡΟ ΤΗΣ ΤΟΝ ΧΡΟΝΟ. ΈΤΣΙ ΑΠΕΙΚΟΝΙΖΟΥΝ ΤΗΝ ΕΞΕΛΙΞΗ ΤΩΝ ΜΕΤΑΒΛΗΤΩΝ ΣΕ ΣΧΕΣΗ ΜΕ ΤΟ ΔΙΑΡΡΕΟΝ ΧΡΟΝΙΚΟ ΔΙΑΣΤΗΜΑ Η ΤΑ ΧΡΟΝΙΚΑ ΣΤΙΓΜΙΟΤΥΠΑ. ΟΙ ΑΚΟΛΟΥΘΙΕΣ ΠΕΡΙΛΑΜΒΑΝΟΥΝ ΔΙΑΔΟΧΙΚΕΣ ΑΣΥΝΕΧΕΙΣ ΤΙΜΕΣ ΠΟΥ ΒΑΣΙΖΟΝΤΑΙ ΣΕ ΧΡΟΝΙΚΑ ΣΥΝΕΧΟΜΕΝΑ ΔΕΔΟΜΕΝΑ ΠΑΡΑΤΗΡΗΣΕΩΝ ΠΟΥ ΕΞΑΡΤΩΝΤΑΙ ΜΕΤΑΞΥ ΤΟΥΣ. Η ΕΝΑΛΛΑΓΗ ΤΩΝ ΚΑΤΑΣΤΑΣΕΩΝ ΠΟΥ ΠΑΡΑΤΗΡΕΙΤΑΙ ΣΕ ΑΥΤΕΣ ΤΙΣ ΔΙΑΔΟΧΕΣ ΑΝΑΛΥΟΝΤΑΙ ΩΣΤΕ ΝΑ ΠΡΟΚΥΨΕΙ ΕΝΑ ΠΡΟΤΥΠΟ ΓΙΑ ΤΗΝ ΠΕΡΙΓΡΑΦΗ ΜΕΛΛΟΝΤΙΚΩΝ ΚΑΤΑΣΤΑΣΕΩΝ Η ΓΙΑ ΤΗΝ ΕΞΕΛΙΞΗ ΝΕΩΝ ΟΜΟΕΙΔΩΝ ΟΝΤΟΤΗΤΩΝ. ΚΥΡΙΕΣ ΕΦΑΡΜΟΓΕΣ ΤΟΥ ΜΟΝΤΕΛΟΥ ΣΥΝΑΝΤΩΝΤΑΙ ΣΤΗΝ ΕΞΕΤΑΣΗ ΤΗΣ ΣΥΜΠΕΡΙΦΟΡΑΣ ΤΩΝ ΧΡΗΣΤΩΝ ΤΟΥ ΔΙΑΔΙΚΤΥΟΥ ΚΑΙ ΤΗΣ ΑΝΑΛΥΣΗΣ ΤΟΥ DNA.
- ΈΝΤΑΞΗ ΣΕ ΚΛΑΣΕΙΣ (CLASSIFICATIONS): ΤΑ ΔΕΙΓΜΑΤΑ ΠΟΥ ΕΞΕΤΑΖΟΝΤΑΙ ΜΕ ΒΑΣΗ ΤΑ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΤΟΥ ΚΑΤΑΤΑΣΣΟΝΤΑΙ ΣΕ ΚΛΑΣΕΙΣ. ΜΕ ΤΟΝ ΤΡΟΠΟ ΑΥΤΟ ΓΙΝΕΤΑΙ ΕΥΧΕΡΕΣΤΕΡΑ Η ΠΡΟΒΛΕΨΗ ΤΗΣ ΣΥΜΠΕΡΙΦΟΡΑΣ ΤΟΥ ΔΕΙΓΜΑΤΟΣ ΣΕ ΔΙΑΦΟΡΕΤΙΚΑ ΕΡΕΘΙΣΜΑΤΑ.
- ΣΥΣΤΑΔΟΠΟΙΗΣΗ (CLUSTERING): ΤΑ ΔΕΔΟΜΕΝΑ ΚΑΤΑΝΕΜΟΝΤΑΙ ΣΕ ΟΜΑΔΕΣ. ΟΙ ΜΕΘΟΔΟΛΟΓΙΕΣ ΕΞΟΡΥΞΗΣ ΔΕΔΟΜΕΝΩΝ ΣΤΙΣ ΠΕΡΙΠΤΩΣΕΙΣ ΑΥΤΕΣ ΣΤΟΧΕΥΟΥΝ ΣΤΗΝ ΑΝΙΧΝΕΥΣΗ ΤΗΣ ΣΧΕΣΗΣ ΜΕΤΑΞΥ ΟΝΤΟΤΗΤΩΝ ΟΙ ΟΠΟΙΕΣ ΛΑΜΒΑΝΟΥΝ ΕΝΑ ΣΥΓΚΕΚΡΙΜΕΝΟ ΕΥΡΟΣ ΤΙΜΩΝ ΣΕ – ΕΠΙΣΗΣ – ΣΥΓΚΕΚΡΙΜΕΝΕΣ ΠΑΡΑΜΕΤΡΟΥΣ. ΧΑΡΑΚΤΗΡΙΣΤΙΚΟ ΤΩΝ ΑΛΓΟΡΙΘΜΩΝ ΠΟΥ ΕΞΕΛΙΣΣΟΝΤΑΙ ΠΡΟΣ ΑΥΤΟ ΤΟ

ΑΠΟΤΕΛΕΣΜΑ ΕΙΝΑΙ ΝΑ ΒΑΣΙΖΟΝΤΑΙ ΣΕ ΕΝΑΝ ΑΡΙΘΜΟ ΕΠΑΝΑΛΗΨΕΩΝ ΜΕΧΡΙ ΤΟ ΜΟΝΤΕΛΟ ΝΑ ΣΥΓΚΛΙΝΕΙ ΚΑΙ ΟΙ ΟΜΑΔΕΣ ΝΑ ΓΙΝΟΥΝ ΔΙΑΚΡΙΤΕΣ. ΈΤΣΙ Η ΕΠΕΞΕΡΓΑΣΙΑ ΠΟΥ ΥΦΙΣΤΑΝΤΑΙ ΓΙΑ ΤΗΝ ΠΑΡΑΓΩΓΗ ΣΥΜΠΕΡΑΣΜΑΤΩΝ ΓΙΝΕΤΑΙ ΤΑΧΥΤΕΡΑ ΚΑΙ ΕΥΚΟΛΟΤΕΡΑ ΜΕ ΠΟΛΥ ΜΙΚΡΕΣ ΑΠΟΚΛΙΣΕΙΣ ΣΤΗΝ ΑΚΡΙΒΕΙΑ ΠΟΥ ΕΠΙΤΥΓΧΑΝΕΤΑΙ. ΕΦΑΡΜΟΖΕΤΑΙ ΣΥΝΗΘΩΣ ΣΕ ΠΟΛΥΠΛΟΚΑ ΠΡΟΒΛΗΜΑΤΑ Η ΣΕ ΣΥΝΟΛΑ ΔΕΔΟΜΕΝΩΝ ΜΕΓΑΛΑ Η/ΚΑΙ ΙΣΧΥΡΑ ΑΝΟΜΟΙΟΓΕΝΗ.

- ΠΡΟΓΝΩΣΗ (FORECASTING): ΚΟΙΝΟΣ ΠΑΡΟΝΟΜΑΣΤΗΣ ΟΛΩΝ ΤΩΝ ΜΕΘΟΔΩΝ ΕΞΟΡΥΞΗΣ ΔΕΔΟΜΕΝΩΝ ΕΙΝΑΙ Η ΠΡΟΒΛΕΨΗ ΤΗΣ ΣΥΜΠΕΡΙΦΟΡΑΣ ΤΟΥ ΑΝΤΙΚΕΙΜΕΝΟΥ ΤΗΣ ΜΕΛΕΤΗΣ. ΣΤΟΧΟΣ ΕΙΝΑΙ Η ΠΑΡΑΓΩΓΗ ΠΡΟΤΥΠΩΝ ΤΑ ΟΠΟΙΑ ΘΑ ΕΙΝΑΙ ΣΕ ΘΕΣΗ ΝΑ ΠΑΡΕΧΟΥΝ ΜΙΑ ΑΣΦΑΛΗ ΕΚΤΙΜΗΣΗ ΓΙΑ ΤΙΣ ΜΕΛΛΟΝΤΙΚΕΣ ΤΙΜΕΣ ΤΩΝ ΠΡΟΣ ΕΞΕΤΑΣΗ ΑΝΤΙΚΕΙΜΕΝΩΝ. ΟΙ ΑΛΓΟΡΙΘΜΟΙ ΤΟΥ ΜΟΝΤΕΛΟΥ ΑΥΤΟΥ ΛΑΜΒΑΝΟΥΝ ΣΑΝ ΕΙΣΟΔΟ ΜΙΑ ΧΡΟΝΟΣΕΙΡΑ ΔΕΔΟΜΕΝΩΝ ΠΟΥ ΠΑΡΟΥΣΙΑΖΟΥΝ ΣΥΓΚΕΚΡΙΜΕΝΕΣ ΤΙΜΕΣ ΚΑΤΑ ΤΟ ΧΡΟΝΙΚΟ ΣΤΙΓΜΙΟΤΥΠΟ ΤΗΣ ΔΕΙΓΜΑΤΟΛΗΨΙΑΣ. ΟΙ ΑΛΓΟΡΙΘΜΟΙ ΑΥΤΟΙ ΠΡΟΣΠΑΘΟΥΝ ΝΑ ΕΝΤΟΠΙΣΟΥΝ ΣΤΙΣ ΔΙΑΔΟΧΙΚΕΣ ΤΙΜΕΣ ΤΩΝ ΠΑΡΑΜΕΤΡΩΝ ΠΕΡΙΟΔΙΚΟΤΗΤΑ Η ΤΑΣΕΙΣ.
- ΠΑΛΙΝΔΡΟΜΗΣΗ (REGRESSION): ΑΠΟΤΕΛΕΙ ΤΕΧΝΙΚΗ ΜΟΝΤΕΛΟΠΟΙΗΣΗΣ ΤΗΣ ΣΧΕΣΗΣ ΜΕΤΑΞΥ ΜΙΑΣ Η ΠΕΡΙΣΣΟΤΕΡΩΝ ΑΝΕΞΑΡΤΗΤΩΝ ΜΕΤΑΒΛΗΤΩΝ ΚΑΙ ΜΙΑΣ ΕΞΑΡΤΗΜΕΝΗΣ. ΣΚΟΠΟΣ ΤΗΣ ΕΙΝΑΙ ΝΑ ΒΡΕΘΟΥΝ ΟΙ ΣΥΝΤΕΛΕΣΤΕΣ ΠΟΥ ΠΡΕΠΕΙ ΝΑ ΕΦΑΡΜΟΣΤΟΥΝ ΣΤΙΣ ΑΝΕΞΑΡΤΗΤΕΣ ΜΕΤΑΒΛΗΤΕΣ ΩΣΤΕ ΝΑ ΑΠΟΔΙΔΕΤΑΙ ΜΕ ΜΕΓΑΛΥΤΕΡΗ ΑΚΡΙΒΕΙΑ Η ΤΙΜΗ ΤΗΣ ΕΞΑΡΤΗΜΕΝΗΣ ΜΕΤΑΒΛΗΤΗΣ. ΕΚΧΩΡΕΙΤΑΙ ΕΝΑ ΣΥΝΟΛΟ ΑΠΟ ΤΙΜΕΣ ΣΕ ΜΙΑ ΟΜΑΔΑ ΜΕΤΑΒΛΗΤΩΝ ΠΡΟΒΛΕΨΗΣ. ΑΝΑΛΟΓΑ ΜΕ ΤΗΝ ΔΙΑΚΥΜΑΝΣΗ ΤΗΣ ΤΙΜΗΣ ΤΗΣ ΕΞΑΡΤΩΜΕΝΗΣ ΜΕΤΑΒΛΗΤΗΣ, ΠΡΟΣΔΙΟΡΙΖΕΤΑΙ ΜΙΑ ΣΥΝΑΡΤΗΣΗ ΠΟΥ

ΣΥΝΔΕΕΙ ΜΕ ΤΗΝ ΕΛΑΧΙΣΤΗ ΔΥΝΑΤΗ ΑΠΟΚΛΙΣΗ ΤΗΝ ΕΞΑΡΤΗΜΕΝΗ ΜΕΤΑΒΛΗΤΗ ΜΕ ΤΙΣ ΑΝΕΞΑΡΤΗΤΕΣ. Η ΣΥΝΑΡΤΗΣΗ ΠΟΥ ΠΡΟΚΥΠΤΕΙ ΑΠΟ ΤΗΝ ΜΕΛΕΤΗ, ΧΡΗΣΙΜΟΠΟΙΕΙΤΑΙ ΓΙΑ ΤΗΝ ΠΡΟΒΛΕΨΗ ΤΩΝ ΤΙΜΩΝ ΤΗΣ ΕΞΑΡΤΗΜΕΝΗΣ ΜΕΤΑΒΛΗΤΗΣ ΓΙΑ ΟΠΟΙΟΝΔΗΠΟΤΕ ΣΥΝΔΥΑΣΜΟ ΤΙΜΩΝ ΤΩΝ ΑΝΕΞΑΡΤΗΤΩΝ ΜΕΤΑΒΛΗΤΩΝ. Η ΓΕΝΙΚΗ ΜΟΡΦΗ ΤΩΝ ΜΟΝΤΕΛΩΝ ΠΑΛΙΝΔΡΟΜΗΣΗΣ ΠΕΡΙΛΑΜΒΑΝΟΥΝ:

- ΜΙΑ ΕΞΑΡΤΗΜΕΝΗ ΜΕΤΑΒΛΗΤΗ Y
 - ΜΙΑ ΣΕΙΡΑ ΑΝΕΞΑΡΤΗΤΩΝ ΜΕΤΑΒΛΗΤΩΝ X .
 - ΜΙΑ ΣΕΙΡΑ ΠΑΡΑΓΟΝΤΩΝ B ΠΟΥ ΡΥΘΜΙΖΟΥΝ ΤΗΝ ΣΧΕΣΗ ΤΗΣ ΕΞΑΡΤΗΜΕΝΗΣ ΜΕΤΑΒΛΗΤΗΣ ΜΕ ΤΙΣ ΑΝΕΞΑΡΤΗΤΕΣ. ΠΡΟΚΕΙΤΑΙ ΣΥΝΗΘΩΣ ΓΙΑ ΤΟ ΖΗΤΟΥΜΕΝΟ.
- ΠΕΡΙΛΗΨΗ (SUMMARIZATION): ΑΠΟΤΕΛΕΙ ΣΥΝΟΠΤΙΚΗ ΚΑΙ ΠΕΡΙΕΚΤΙΚΗ ΠΕΡΙΓΡΑΦΗ ΤΩΝ ΑΠΟΤΕΛΕΣΜΑΤΩΝ ΤΗΣ ΜΕΛΕΤΗΣ ΤΟΥ ΔΕΙΓΜΑΤΟΣ ΤΩΝ ΔΕΔΟΜΕΝΩΝ. ΕΙΝΑΙ ΜΙΑ ΒΑΣΙΚΗ ΕΝΝΟΙΑ ΤΗΣ ΕΞΟΡΥΞΗΣ ΔΕΔΟΜΕΝΩΝ ΠΟΥ ΠΕΡΙΛΑΜΒΑΝΕΙ ΤΕΧΝΙΚΕΣ ΓΙΑ ΤΗΝ ΕΥΡΕΣΗ ΜΙΑ ΣΥΜΠΑΓΟΥΣ ΠΕΡΙΓΡΑΦΗΣ ΕΝΟΣ ΣΥΝΟΛΟΥ ΔΕΔΟΜΕΝΩΝ. ΜΕΡΙΚΕΣ ΑΠΟ ΤΙΣ ΜΕΘΟΔΟΥΣ ΠΟΥ ΧΡΗΣΙΜΟΠΟΙΟΥΝΤΑΙ ΕΙΝΑΙ ΟΙ ΠΑΡΟΥΣΙΑΣΕΙΣ ΜΕΣΩ ΠΙΝΑΚΩΝ, ΟΙ ΕΦΑΡΜΟΓΕΣ ΣΤΑΤΙΣΤΙΚΩΝ ΜΕΤΡΩΝ (ΜΕΣΗ ΤΙΜΗ, ΤΥΠΙΚΕΣ ΑΠΟΚΛΙΣΕΙΣ), Η ΑΝΑΠΤΥΞΗ ΕΚΘΕΣΕΩΝ-ΑΝΑΦΟΡΩΝ ΚΑΙ Η ΟΠΤΙΚΟΠΟΙΗΣΗ ΤΩΝ ΔΕΔΟΜΕΝΩΝ. ΣΥΧΝΑ ΕΝΤΟΠΙΖΕΤΑΙ ΜΕ ΤΗΝ ΜΟΡΦΗ ΠΙΝΑΚΩΝ Η ΣΥΝΤΟΜΩΝ ΑΝΑΦΟΡΩΝ.
- ΜΟΝΤΕΛΟΠΟΙΗΣΗ ΕΞΑΡΤΗΣΕΩΝ (DEPENDENCY MODELING): ΟΙ ΕΞΑΡΤΗΣΕΙΣ ΠΟΥ ΠΡΟΕΚΥΨΑΝ ΑΠΟ ΤΗΝ ΜΕΛΕΤΗ ΜΕΤΑΞΥ ΤΟΥ ΔΕΙΓΜΑΤΟΣ ΚΑΙ ΤΟΥ ΠΡΟΣ ΜΕΛΕΤΗ ΑΝΤΙΚΕΙΜΕΝΟΥ, ΜΟΝΤΕΛΟΠΟΙΕΙΤΑΙ ΩΣΤΕ ΝΑ ΕΙΝΑΙ ΑΝΑΓΝΩΣΙΜΟ ΜΕ ΣΑΦΗΝΕΙΑ ΑΠΟ ΤΟΝ ΑΡΜΟΔΙΟ ΧΕΙΡΙΣΤΗ ΤΩΝ ΔΕΔΟΜΕΝΩΝ. ΤΟ ΜΟΝΤΕΛΟ ΓΕΝΙΚΑ ΧΑΡΑΚΤΗΡΙΖΕΤΑΙ

ΑΠΟ ΤΗΝ ΑΠΛΟΤΗΤΑ ΤΗΣ ΜΟΡΦΗΣ ΤΟΥ ΚΑΙ ΤΗΝ ΠΕΡΙΓΡΑΦΙΚΗ ΤΟΥ ΑΝΤΙΚΕΙΜΕΝΙΚΟΤΗΤΑ.

- ΑΝΙΧΝΕΥΣΗ ΜΕΤΑΒΟΛΩΝ ΚΑΙ ΑΠΟΚΛΙΣΕΩΝ (CHANGE AND DEVIATIONDETECTION): ΤΑ ΜΟΝΤΕΛΑ ΑΥΤΑ ΕΝΤΟΠΙΖΟΥΝ ΤΡΟΠΟΠΟΙΗΣΕΙΣ ΠΟΥ ΕΝΔΕΧΟΜΕΝΩΣ ΕΧΟΥΝ ΣΥΜΒΕΙΣ ΣΤΗΝ ΣΥΜΠΕΡΙΦΟΡΑ ΤΩΝ ΔΕΔΟΜΕΝΩΝ. ΛΑΜΒΑΝΟΥΝ ΥΠ' ΟΨΗ ΚΑΙ ΤΟ ΙΣΤΟΡΙΚΟ ΜΕΤΡΗΣΕΩΝ ΠΑΡΕΛΘΟΝΤΩΝ ΣΧΕΤΙΚΩΝ ΜΕΛΕΤΩΝ [3][20][21].

1.5 ΕΦΑΡΜΟΓΕΣ

ΣΕ ΠΟΛΛΕΣ ΠΕΡΙΠΤΩΣΕΙΣ ΠΛΗΡΟΦΟΡΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ ΥΠΑΡΧΕΙ Η ΑΝΑΓΚΗ ΠΡΟΣΠΕΛΑΣΗΣ ΣΕ ΜΕΓΑΛΟΥ ΟΓΚΟΥ ΣΥΝΟΛΑ ΔΕΔΟΜΕΝΩΝ. ΤΕΤΟΙΑ ΣΥΝΟΛΑ ΜΠΟΡΕΙ ΝΑ ΑΦΟΡΟΥΝ:

- ΚΥΒΕΡΝΗΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ: ΟΙ ΚΥΒΕΡΝΗΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ ΣΤΗΝ ΜΕΓΑΛΗ ΠΛΕΙΟΝΟΤΗΤΑ ΤΟΥΣ ΧΕΙΡΙΖΟΝΤΑΙ ΜΕΓΑΛΟΥΣ ΟΓΚΟΥΣ ΔΕΔΟΜΕΝΩΝ. ΤΑ ΑΠΟΤΕΛΕΣΜΑΤΑ ΤΗΣ ΜΕΛΕΤΗΣ ΤΟΥΣ ΣΕ ΠΟΛΛΕΣ ΠΕΡΙΠΤΩΣΕΙΣ ΑΠΟΤΕΛΕΙ ΕΝΑ ΑΠΟ ΤΑ ΒΑΣΙΚΟΤΕΡΑ ΚΡΙΤΗΡΙΑ ΓΙΑ ΤΗΝ ΧΑΡΑΞΗ ΤΗΣ ΚΥΒΕΡΝΗΤΙΚΗΣ ΠΟΛΙΤΙΚΗΣ. ΈΤΣΙ ΧΡΕΙΑΖΕΤΑΙ ΝΑ ΓΙΝΟΝΤΑΙ ΓΡΗΓΟΡΕΣ ΚΑΙ ΑΣΦΑΛΕΙΣ ΕΚΤΙΜΗΣΕΙΣ ΤΩΝ ΤΑΣΕΩΝ ΠΟΥ ΕΠΙΚΡΑΤΟΥΝ ΣΤΗΝ ΚΟΙΝΩΝΙΑ ΑΛΛΑ ΚΑΙ ΣΤΟ ΔΙΕΘΝΕΣ ΠΕΡΙΒΑΛΛΟΝ ΩΣΤΕ Η ΠΟΛΙΤΙΚΗ ΑΥΤΗ ΝΑ ΜΠΟΡΕΙ ΝΑ ΠΡΟΣΑΡΜΟΖΕΤΑΙ ΑΡΜΟΝΙΚΑ ΣΤΙΣ ΜΕΛΛΟΝΤΙΚΕΣ ΕΠΙΚΡΑΤΟΥΣΕΣ ΚΑΤΑΣΤΑΣΕΙΣ.
- ΙΑΤΡΙΚΕΣ ΕΦΑΡΜΟΓΕΣ: Η ΙΑΤΡΙΚΗ ΕΙΝΑΙ ΕΝΑΣ ΚΛΑΔΟΣ ΤΗΣ ΕΠΙΣΤΗΜΗΣ ΟΠΟΥ ΟΙ ΤΕΧΝΙΚΕΣ ΤΗΣ ΕΞΟΡΥΞΗΣ ΔΕΔΟΜΕΝΩΝ ΠΡΟΣΦΕΡΟΥΝ ΣΠΟΥΔΑΙΕΣ ΛΥΣΕΙΣ ΣΤΗΝ ΔΙΑΓΝΩΣΗ ΑΛΛΑ ΚΑΙ ΤΟΝ ΤΡΟΠΟ ΑΝΤΙΜΕΤΩΠΙΣΗΣ ΤΩΝ ΔΙΑΦΟΡΩΝ ΥΓΕΙΟΝΟΜΙΚΩΝ ΠΕΡΙΣΤΑΤΙΚΩΝ. ΤΑ ΙΑΤΡΙΚΑ ΙΣΤΟΡΙΚΑ ΕΙΝΑΙ ΜΙΑ ΤΕΡΑΣΤΙΑ ΔΕΞΑΜΕΝΗ ΔΕΔΟΜΕΝΩΝ ΤΑ ΟΠΟΙΑ ΜΠΟΡΟΥΝ ΝΑ ΜΕΛΕΤΗΘΟΥΝ ΜΕ ΠΟΙΚΙΛΟΥΣ ΤΡΟΠΟΥΣ ΚΑΙ ΝΑ ΟΔΗΓΗΣΟΥΝ ΣΕ ΣΥΜΠΕΡΑΣΜΑΤΑ ΣΧΕΤΙΚΑ ΜΕ ΤΗΝ ΕΞΕΛΙΞΗ ΤΩΝ ΘΕΡΑΠΕΥΤΙΚΩΝ ΜΕΘΟΔΩΝ, ΤΩΝ

ΧΡΗΣΙΜΟΠΟΙΟΥΜΕΝΩΝ ΑΝΑ ΠΕΡΙΠΤΩΣΗ ΦΑΡΜΑΚΕΥΤΙΚΩΝ ΣΚΕΥΑΣΜΑΤΩΝ ΚΑΙ ΑΛΛΩΝ ΖΗΤΗΜΑΤΩΝ ΙΑΤΡΙΚΟΥ ΕΝΔΙΑΦΕΡΟΝΤΟΣ.

- ΟΙΚΟΝΟΜΙΚΕΣ ΣΥΝΑΛΛΑΓΕΣ: Η ΟΙΚΟΝΟΜΙΑ ΕΙΝΑΙ ΕΝΑΣ ΚΛΑΔΟΣ ΤΗΣ ΑΝΘΡΩΠΙΝΗΣ ΔΡΑΣΤΗΡΙΟΤΗΤΑΣ ΠΟΥ ΧΡΗΣΙΜΟΠΟΙΕΙ ΑΡΚΕΤΑ ΤΗΝ ΕΞΟΡΥΞΗ ΔΕΔΟΜΕΝΩΝ. ΒΑΣΙΚΕΣ ΠΑΡΑΜΕΤΡΟΙ ΓΙΑ ΤΗΝ ΠΡΟΒΛΕΨΗ ΤΩΝ ΟΙΚΟΝΟΜΙΚΩΝ ΤΑΣΕΩΝ ΚΑΙ ΜΕΓΕΘΩΝ ΑΠΟΤΕΛΟΥΝ ΤΡΕΧΟΝΤΑ ΚΑΙ ΠΑΡΕΛΘΟΝΤΑ ΔΕΔΟΜΕΝΑ. ΑΥΤΑ ΥΦΙΣΤΑΝΤΑΙ ΚΑΤΑΛΛΗΛΗ ΕΠΕΞΕΡΓΑΣΙΑ ΠΡΟΚΕΙΜΕΝΟΥ ΝΑ ΠΑΡΑΞΟΥΝ ΓΝΩΣΗ ΙΚΑΝΗ ΝΑ ΟΔΗΓΗΣΕΙ ΣΕ ΟΡΘΕΣ ΕΠΕΝΔΥΤΙΚΕΣ ΕΠΙΛΟΓΕΣ.
- ΑΣΤΥΝΟΜΙΚΗ ΈΡΕΥΝΑ: Η ΑΣΤΥΝΟΜΙΚΗ ΕΡΕΥΝΑ ΟΠΟΙΑΣΔΗΠΟΤΕ ΚΛΙΜΑΚΑΣ ΕΦΑΡΜΟΖΕΙ ΤΙΣ ΜΕΘΟΔΟΛΟΓΙΕΣ ΤΗΣ ΕΞΟΡΥΞΗΣ ΔΕΔΟΜΕΝΩΝ ΠΡΟΚΕΙΜΕΝΟΥ ΝΑ ΔΗΜΙΟΥΡΓΗΣΕΙ ΓΝΩΣΗ ΙΚΑΝΗ ΝΑ ΤΗΝ ΠΡΟΣΑΝΑΤΟΛΙΖΕΙ ΚΑΤΑΛΛΗΛΑ. ΜΕ ΒΑΣΗ ΤΗΝ ΓΝΩΣΗ ΑΥΤΗ ΜΟΡΦΟΠΟΙΟΥΝΤΑΙ ΚΑΙ ΑΝΑΠΤΥΣΣΟΝΤΑΙ ΤΑ ΠΡΟΦΙΛ ΤΩΝ ΣΥΜΜΕΤΕΧΟΝΤΩΝ ΠΑΡΑΓΟΝΤΩΝ ΥΠΟΘΕΣΕΩΝ ΠΟΥ ΕΡΕΥΝΟΥΝ ΟΙ ΑΣΤΥΝΟΜΙΚΕΣ ΑΡΧΕΣ. ΈΤΣΙ ΔΙΕΥΚΟΛΥΝΟΝΤΑΙ ΣΤΟ ΝΑ ΠΡΟΣΔΙΟΡΙΖΟΥΝ ΤΙΣ ΠΙΘΑΝΕΣ ΔΙΑΔΡΟΜΕΣ ΕΠΙΛΥΣΗΣ ΚΑΙ ΝΑ ΤΙΣ ΑΞΙΟΛΟΓΗΣΟΥΝ.
- ΕΠΙΣΤΗΜΟΝΙΚΗ ΈΡΕΥΝΑ: ΕΙΝΑΙ ΤΟΜΕΑΣ ΟΠΟΥ ΚΑΤ' ΕΞΟΧΗΣ ΧΡΗΣΙΜΟΠΟΙΕΙΤΑΙ Η ΕΞΟΡΥΞΗ ΔΕΔΟΜΕΝΩΝ. ΑΠΟΤΕΛΕΙ ΣΗΜΑΝΤΙΚΟ ΕΡΓΑΛΕΙΟ ΤΗΣ ΒΑΣΙΚΗΣ ΕΡΕΥΝΑΣ ΠΟΥ ΑΠΟΣΚΟΠΕΙ ΣΤΗΝ ΑΝΑΒΑΘΜΙΣΗ ΚΑΙ ΕΠΕΚΤΑΣΗ ΤΗΣ ΓΝΩΣΗΣ. ΩΣ ΕΡΕΥΝΑ ΝΟΕΙΤΑΙ Η ΣΥΣΤΗΜΑΤΙΚΗ, ΑΝΤΙΚΕΙΜΕΝΙΚΗ ΚΑΙ ΕΞΑΚΡΙΒΩΜΕΝΗ ΑΝΑΖΗΤΗΣΗ ΠΛΗΡΟΦΟΡΙΩΝ ΠΡΟΣ ΕΠΙΛΥΣΗ ΚΑΠΟΙΟΥ ΠΡΟΒΛΗΜΑΤΟΣ. ΟΙ ΜΕΘΟΔΟΛΟΓΙΕΣ ΤΗΣ ΕΞΟΡΥΞΗΣ ΔΕΔΟΜΕΝΩΝ ΧΡΗΣΙΜΟΠΟΙΟΥΝΤΑΙ ΚΥΡΙΩΣ ΣΤΙΣ ΠΕΙΡΑΜΑΤΙΚΕΣ ΕΡΕΥΝΕΣ ΟΠΟΥ ΕΞΕΤΑΖΟΝΤΑΙ ΚΑΙ ΑΝΑΖΗΤΟΥΝΤΑΙ ΟΙ ΤΥΠΟΙ ΣΥΣΧΕΤΙΣΕΩΝ ΜΕΤΑΞΥ

ΠΑΡΑΜΕΤΡΩΝ ΠΟΥ ΣΥΜΜΕΤΕΧΟΥΝ ΣΤΗΝ ΕΞΕΛΙΞΗ ΦΑΙΝΟΜΕΝΩΝ.

- ΚΟΙΝΩΝΙΚΗ ΔΙΚΤΥΩΣΗ: Η ΚΟΙΝΩΝΙΚΗ ΔΙΚΤΥΩΣΗ ΕΧΕΙ ΓΝΩΡΙΣΕΙ ΣΗΜΑΝΤΙΚΗ ΑΝΑΠΤΥΞΗ ΤΗΝ ΤΕΛΕΥΤΑΙΑ ΔΕΚΑΕΤΙΑ. ΚΑΘΩΣ ΣΥΝΕΠΕΣΕ ΜΕ ΤΗΝ ΡΑΓΔΑΙΑ ΑΝΑΠΤΥΞΗ ΤΩΝ ΤΕΧΝΟΛΟΓΙΩΝ ΠΟΥ ΣΧΕΤΙΖΟΝΤΑΙ ΜΕ ΤΟ ΔΙΑΔΙΚΤΥΟ ΠΟΥ ΕΚΑΝΕ ΤΙΣ ΕΦΑΡΜΟΓΕΣ ΤΟΥ ΠΡΟΣΙΤΕΣ ΣΕ ΜΕΓΑΛΟ ΜΕΡΟΣ ΤΟΥ ΠΛΗΘΥΣΜΟΥ, ΟΙ ΑΝΤΙΣΤΟΙΧΟΙ ΤΟΠΟΙ ΕΓΙΝΑΝ ΙΔΙΑΙΤΕΡΑ ΔΗΜΟΦΙΛΕΙΣ ΜΕ ΑΠΟΤΕΛΕΣΜΑ ΜΕΓΑΛΟ ΜΕΡΟΣ ΤΟΥ ΠΛΗΘΥΣΜΟΥ ΤΟΥ ΔΙΑΔΙΚΤΥΟΥ ΝΑ ΧΡΗΣΙΜΟΠΟΙΕΙ ΤΟ ΔΙΑΔΙΚΤΥΟ ΑΠΟΚΛΕΙΣΤΙΚΑ ΓΙΑ ΤΗΝ ΧΡΗΣΗ ΤΟΥΣ. ΣΗΜΕΡΑ ΠΟΛΥ ΜΙΚΡΟ ΠΟΣΟΣΤΟ ΑΝΘΡΩΠΩΝ ΔΕΝ ΕΙΝΑΙ ΕΓΓΕΓΡΑΜΜΕΝΟ ΣΕ ΚΑΠΟΙΑ ΕΦΑΡΜΟΓΗ ΚΟΙΝΩΝΙΚΗΣ ΔΙΚΤΥΩΣΗΣ. ΤΑ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΚΑΙ Η ΣΥΜΠΕΡΙΦΟΡΑ ΤΩΝ ΧΡΗΣΤΩΝ ΑΠΟΤΕΛΟΥΝ ΑΝΤΙΚΕΙΜΕΝΑ ΜΕΛΕΤΗΣ ΠΡΟΚΕΙΜΕΝΟΥ ΝΑ ΑΝΙΧΝΕΥΟΝΤΑΙ ΟΙ ΤΑΣΕΙΣ ΣΥΓΚΕΚΡΙΜΕΝΩΝ ΟΜΑΔΩΝ ΤΟΥΣ. Η ΓΝΩΣΗ ΠΟΥ ΑΠΟΚΤΑΤΑΙ ΜΕ ΤΟΝ ΤΡΟΠΟ ΑΥΤΟ ΜΠΟΡΕΙ ΝΑ ΧΡΗΣΙΜΟΠΟΙΗΘΕΙ ΓΙΑ ΤΗΝ ΧΑΡΑΞΗ ΣΤΡΑΤΗΓΙΚΩΝ ΜΑΡΚΕΤΙΝΓΚ, ΤΗΝ ΠΡΟΣΑΡΜΟΓΗ ΤΗΣ ΠΟΛΙΤΙΚΗ ΤΩΝ ΚΥΒΕΡΝΗΣΕΩΝ ΣΤΙΣ ΤΑΣΕΙΣ ΤΗΣ ΚΟΙΝΗΣ ΓΝΩΜΗΣ ΚΑΙ ΣΕ ΠΛΗΘΟΣ ΑΛΛΩΝ ΠΕΡΙΠΤΩΣΕΩΝ.
- ΔΙΑΔΙΚΤΥΑΚΕΣ ΑΝΑΖΗΤΗΣΕΙΣ: ΟΙ ΑΝΑΖΗΤΗΣΕΙΣ ΣΤΟ ΔΙΑΔΙΚΤΥΟ ΕΙΝΑΙ ΕΝΑΣ ΑΚΟΜΑ ΤΟΜΕΑΣ ΟΠΟΥ ΒΡΙΣΚΟΥΝ ΕΦΑΡΜΟΓΕΣ ΟΙ ΤΕΧΝΙΚΕΣ ΕΞΟΡΥΞΗΣ ΔΕΔΟΜΕΝΩΝ. ΣΥΝΗΘΕΣΤΕΡΗ ΕΦΑΡΜΟΓΗ ΑΠΟΤΕΛΕΙ Η ΜΕΛΕΤΗ ΤΩΝ ΑΠΟΤΕΛΕΣΜΑΤΩΝ ΠΟΥ ΙΚΑΝΟΠΟΙΟΥΝ ΧΡΗΣΤΕΣ ΠΟΥ ΥΠΟΒΑΛΛΟΥΝ ΣΥΓΚΕΚΡΙΜΕΝΑ ΚΡΙΤΗΡΙΑ ΑΝΑΖΗΤΗΣΗΣ ΩΣΤΕ ΣΤΟ ΜΕΛΛΟΝ ΤΑ ΑΠΟΤΕΛΕΣΜΑΤΑ ΑΥΤΑ ΝΑ ΠΑΡΟΥΣΙΑΖΟΝΤΑΙ ΣΕ ΥΨΗΛΟΤΕΡΕΣ ΘΕΣΕΙΣ.
- ΕΠΕΞΕΡΓΑΣΙΑ ΕΙΚΟΝΑΣ: ΟΙ ΨΗΦΙΑΚΕΣ ΕΙΚΟΝΕΣ ΑΥΞΑΝΟΜΕΝΗΣ ΤΗΣ ΠΟΙΟΤΗΤΑΣ ΤΟΥΣ, ΑΠΑΙΤΟΥΝ ΜΕΓΑΛΟΥΣ ΑΠΟΘΗΚΕΥΤΙΚΟΥΣ ΧΩΡΟΥΣ ΓΙΑ ΤΗΝ ΔΙΑΤΗΡΗΣΗ ΤΟΥΣ. ΠΑΡΑΛΛΗΛΑ Η ΜΕΤΑΔΟΣΗ ΤΟΥΣ ΕΠΙΣΗΣ ΑΠΑΙΤΕΙ

ΣΥΝΔΕΣΕΙΣ ΥΨΗΛΩΝ ΔΥΝΑΤΟΤΗΤΩΝ ΩΣΤΕ ΝΑ ΕΙΝΑΙ ΠΟΙΟΤΙΚΑ ΑΠΟΔΕΚΤΗ. ΣΥΝΕΠΕΙΑ ΑΥΤΩΝ ΕΙΝΑΙ Η ΑΝΑΖΗΤΗΣΗ ΑΠΟΔΟΤΙΚΩΝ ΜΕΘΟΔΩΝ ΣΥΜΠΙΕΣΗΣ ΤΩΝ ΔΕΔΟΜΕΝΩΝ ΠΟΥ ΑΠΟΘΗΚΕΥΟΝΤΑΙ ΣΤΑ ΑΡΧΕΙΑ ΕΙΚΟΝΩΝ ΩΣΤΕ ΝΑ ΕΙΝΑΙ ΕΦΙΚΤΗ Η ΜΕΙΩΣΗ ΤΟΥ ΜΕΓΕΘΟΥΣ ΤΟΥΣ ΧΩΡΙΣ ΩΣΤΟΣΟ ΝΑ ΕΠΗΡΕΑΖΕΤΑΙ ΣΕ ΜΕΓΑΛΟ ΒΑΘΜΟ Η ΠΟΙΟΤΗΤΑ ΤΟΥΣ. ΑΡΚΕΤΕΣ ΤΕΧΝΙΚΕΣ ΒΑΣΙΖΟΝΤΑΙ ΣΕ ΜΕΘΟΔΟΛΟΓΙΕΣ ΕΞΟΡΥΞΗΣ ΔΕΔΟΜΕΝΩΝ ΟΠΟΥ ΑΝΑΖΗΤΟΥΝΤΑΙ ΣΧΕΣΕΙΣ ΜΕΤΑΞΥ ΣΧΕΔΙΑΣΤΙΚΩΝ ΠΡΟΤΥΠΩΝ ΩΣΤΕ ΝΑ ΑΠΟΦΕΥΓΕΤΑΙ Η ΡΗΤΗ ΑΠΟΘΗΚΕΥΣΗ ΟΛΩΝ ΤΩΝ ΔΕΔΟΜΕΝΩΝ ΠΟΥ ΠΕΡΙΓΡΑΦΟΥΝ ΤΗΝ ΕΙΚΟΝΑ ΜΕ ΤΗΝ ΕΠΙΤΥΧΗ ΠΡΟΒΛΕΨΗ ΚΑΠΟΙΩΝ ΑΠΟ ΑΥΤΑ.

- ΕΜΠΟΡΙΚΕΣ ΠΡΟΩΘΗΤΙΚΕΣ ΕΝΕΡΓΕΙΕΣ: Η ΕΞΑΡΣΗ ΤΟΥ ΚΑΤΑΝΑΛΩΤΙΣΜΟΥ ΠΟΥ ΠΑΡΑΤΗΡΕΙΤΑΙ ΑΜΕΙΩΤΗ ΤΙΣ ΤΕΛΕΥΤΑΙΕΣ ΤΕΣΣΕΡΙΣ ΔΕΚΑΕΤΙΕΣ ΔΗΜΙΟΥΡΓΗΣΕ ΣΥΝΘΗΚΕΣ ΕΝΤΟΝΟΥ ΑΝΤΑΓΩΝΙΣΜΟΥ ΣΤΗΝ ΔΙΑΘΕΣΗ ΑΓΑΘΩΝ ΚΑΙ ΥΠΗΡΕΣΙΩΝ. ΈΤΣΙ ΚΑΙ ΣΕ ΣΥΝΔΥΑΣΜΟ ΜΕ ΤΗΝ ΠΑΡΑΤΕΤΑΜΕΝΗ ΠΑΓΚΟΣΜΙΑ ΟΙΚΟΝΟΜΙΚΗ ΚΡΙΣΗ ΔΗΜΙΟΥΡΓΗΘΗΚΕ Η ΑΝΑΓΚΗ, ΟΙ ΠΡΟΩΘΗΤΙΚΕΣ ΕΝΕΡΓΕΙΕΣ ΝΑ ΕΙΝΑΙ ΑΚΡΩΣ ΑΠΟΔΟΤΙΚΕΣ ΩΣΤΕ ΝΑ ΕΠΙΤΥΓΧΑΝΕΤΑΙ ΤΟ ΜΕΓΙΣΤΟ ΔΥΝΑΤΟ ΑΠΟΤΕΛΕΣΜΑ ΜΕ ΤΗΝ ΜΙΚΡΟΤΕΡΗ ΔΙΑΦΗΜΙΣΤΙΚΗ ΔΑΠΑΝΗ. ΓΙΑ ΝΑ ΕΠΙΤΕΥΧΘΕΙ ΑΥΤΟ ΧΡΗΣΙΜΟΠΟΙΟΥΝΤΑΙ ΚΑΙ ΜΕΘΟΔΟΛΟΓΙΕΣ ΕΞΟΡΥΞΗΣ ΔΕΔΟΜΕΝΩΝ ΠΡΟΚΕΙΜΕΝΟΥ ΝΑ ΠΡΟΣΔΙΟΡΙΖΟΝΤΑΙ ΤΑ ΣΤΟΙΧΕΙΑ ΑΥΤΑ ΠΟΥ ΜΠΟΡΕΙ ΝΑ ΔΙΑΜΟΡΦΩΝΟΥΝ ΤΙΣ ΤΑΣΕΙΣ ΤΗΣ ΑΓΟΡΑΣ Η ΝΑ ΑΝΙΧΝΕΥΟΝΤΑΙ ΑΥΤΕΣ [21].

1.6 ΕΞΟΡΥΞΗ ΔΕΔΟΜΕΝΩΝ ΚΑΙ ΙΔΙΩΤΙΚΟΤΗΤΑ

Η ΑΠΟΔΟΣΗ ΤΩΝ ΜΕΘΟΔΩΝ ΕΞΟΡΥΞΗΣ ΔΕΔΟΜΕΝΩΝ ΕΞΑΡΤΑΤΑΙ ΣΕ ΜΕΓΑΛΟ ΒΑΘΜΟ ΑΠΟ ΤΟ ΕΙΔΟΣ ΚΑΙ ΤΗΝ ΠΟΣΟΤΗΤΑ ΤΩΝ ΧΡΗΣΙΜΟΠΟΙΟΥΜΕΝΩΝ ΔΕΔΟΜΕΝΩΝ. ΣΕ ΑΡΚΕΤΕΣ ΠΕΡΙΠΤΩΣΕΙΣ Η ΛΗΨΗ ΙΚΑΝΟΠΟΙΗΤΙΚΟΥ ΑΠΟΤΕΛΕΣΜΑΤΟΣ ΑΠΑΙΤΕΙ ΤΗΝ ΛΗΨΗ ΚΑΙ

ΕΠΕΞΕΡΓΑΣΙΑ ΕΥΑΙΣΘΗΤΩΝ ΠΡΟΣΩΠΙΚΩΝ ΔΕΔΟΜΕΝΩΝ ΜΕΛΩΝ ΟΜΑΔΩΝ ΠΟΥ ΜΕΛΕΤΩΝΤΑΙ. ΣΤΗΝ ΣΥΓΧΡΟΝΗ ΕΠΟΧΗ ΕΙΝΑΙ ΑΚΟΜΑ ΣΤΟ ΣΤΑΔΙΟ ΤΗΣ ΜΕΛΕΤΗΣ Η ΑΝΑΖΗΤΗΣΗ ΤΡΟΠΩΝ ΓΙΑ ΤΗΝ ΧΡΗΣΗ ΤΩΝ ΤΕΧΝΟΛΟΓΙΚΩΝ ΕΠΙΤΕΥΓΜΑΤΩΝ ΣΤΗΝ ΔΙΑΧΕΙΡΙΣΗ ΤΕΤΟΙΟΥ ΕΙΔΟΥΣ ΔΕΔΟΜΕΝΩΝ ΧΩΡΙΣ ΝΑ ΘΙΓΕΤΑΙ Η ΙΔΙΩΤΙΚΟΤΗΤΑ ΤΩΝ ΥΠΟΚΕΙΜΕΝΩΝ. Η ΙΔΙΩΤΙΚΟΤΗΤΑ ΘΙΓΕΤΑΙ ΟΤΑΝ Η ΕΞΟΥΣΙΑ ΔΕΔΟΜΕΝΩΝ ΜΠΟΡΕΙ – ΕΚΟΥΣΙΑ Η ΑΚΟΥΣΙΑ – ΝΑ ΘΕΣΕΙ ΣΕ ΚΙΝΔΥΝΟ ΤΗΝ ΙΔΙΩΤΙΚΗ (ΠΡΟΣΩΠΙΚΗ) ΖΩΗ ΤΟΥ ΥΠΟΚΕΙΜΕΝΟΥ.

ΟΙ ΑΥΤΟΜΑΤΟΠΟΙΗΜΕΝΟΙ ΤΡΟΠΟΙ ΣΥΛΛΟΓΗΣ ΔΕΔΟΜΕΝΩΝ ΕΧΟΥΝ ΤΗΝ ΔΥΝΑΤΟΤΗΤΑ ΝΑ ΕΛΕΓΧΟΥΝ ΜΕΓΑΛΟΥΣ ΟΓΚΟΥΣ ΗΛΕΚΤΡΟΝΙΚΩΝ ΠΗΓΩΝ, ΝΑ ΕΠΕΞΕΡΓΑΖΟΝΤΑΙ ΤΑ ΠΕΡΙΕΧΟΜΕΝΑ ΤΟΥΣ ΚΑΙ ΝΑ ΕΞΑΓΟΥΝ ΠΛΗΡΟΦΟΡΙΕΣ ΣΧΕΤΙΚΕΣ ΜΕ «ΑΘΩΕΣ» ΤΑΣΕΙΣ ΤΩΝ ΥΠΟΚΕΙΜΕΝΩΝ ΤΩΝ ΔΕΔΟΜΕΝΩΝ ΑΥΤΩΝ. ΠΑΡΑΛΛΗΛΑ ΜΕ ΒΑΣΗ ΤΑ ΙΔΙΑ ΣΤΟΙΧΕΙΑ ΚΑΙ ΤΙΣ ΙΔΙΕΣ ΕΠΕΞΕΡΓΑΣΙΕΣ ΕΙΝΑΙ ΔΥΝΑΤΟΝ ΝΑ ΠΡΟΣΔΙΟΡΙΣΤΟΥΝ ΠΡΟΣΩΠΙΚΟΥ ΧΑΡΑΚΤΗΡΑ ΠΛΗΡΟΦΟΡΙΕΣ ΓΙΑ ΑΝΘΡΩΠΟΥΣ ΟΠΩΣ Η ΟΙΚΟΝΟΜΙΚΗ ΚΑΤΑΣΤΑΣΗ, ΔΡΑΣΤΗΡΙΟΤΗΤΕΣ, ΠΟΛΙΤΙΚΕΣ ΠΕΠΟΙΘΗΣΕΙΣ ΚΑ. ΟΤΑΝ ΣΤΟΙΧΕΙΑ ΠΟΥ ΣΧΕΤΙΖΟΝΤΑΙ ΜΕ ΑΝΘΡΩΠΟΥΣ ΧΡΗΣΙΜΟΠΟΙΟΥΝΤΑΙ ΧΩΡΙΣ ΤΗΝ ΣΥΓΚΑΤΑΘΕΣΗ ΤΟΥ ΓΙΑ ΣΚΟΠΟΥΣ ΔΙΑΦΟΡΕΤΙΚΟΥΣ ΣΕ ΣΧΕΣΗ ΜΕ ΑΥΤΟΥΣ ΓΙΑ ΤΟΥΣ ΟΠΟΙΟΥΣ ΤΑ ΓΝΩΣΤΟΠΟΙΗΣΕ Η/ΚΑΙ ΠΕΡΝΑΝΕ ΣΤΗΝ ΚΑΤΟΧΗ ΤΡΙΤΩΝ, ΤΟΤΕ ΘΕΩΡΕΙΤΑΙ ΟΤΙ ΠΑΡΑΒΙΑΖΕΤΑΙ Η ΙΔΙΩΤΙΚΟΤΗΤΑ ΤΟΥΣ. ΣΗΜΕΡΑ ΕΙΝΑΙ ΔΥΣΚΟΛΟ - ΑΝ ΟΧΙ ΑΚΑΤΟΡΘΩΤΟ – ΕΝΑ ΑΤΟΜΟ ΝΑ ΜΠΟΡΕΙ ΝΑ ΕΙΝΑΙ ΣΙΓΟΥΡΟ ΟΤΙ ΤΑ ΠΡΟΣΩΠΙΚΑ ΣΤΟΙΧΕΙΑ ΠΟΥ ΠΑΡΑΧΩΡΗΣΕ ΓΙΑ ΕΝΑ ΣΚΟΠΟ ΟΤΙ ΔΕΝ ΘΑ ΧΡΗΣΙΜΟΠΟΙΗΘΟΥΝ ΚΑΙ ΓΙΑ ΚΑΠΟΙΟ ΑΛΛΟ ΣΚΟΠΟ Η ΑΠΟ ΚΑΠΟΙΑ ΑΛΛΗ ΟΝΤΟΤΗΤΑ (ΔΙΑΦΟΡΕΤΙΚΗ ΑΠΟ ΑΥΤΗΝ ΣΤΗΝ ΟΠΟΙΑ ΤΑ ΔΙΕΘΕΣΕ). ΈΧΟΥΝ ΑΝΑΦΕΡΘΕΙ ΠΟΛΛΕΣ ΠΕΡΙΠΤΩΣΕΙΣ ΣΤΟΙΧΕΙΩΝ ΙΣΤΟΡΙΚΟΥ ΑΓΟΡΩΝ, ΕΠΙΣΚΕΨΗΣ ΔΙΚΤΥΑΚΩΝ ΤΟΠΩΝ, ΙΑΤΡΙΚΩΝ ΔΕΔΟΜΕΝΩΝ, ΟΙΚΟΝΟΜΙΚΩΝ ΣΤΟΙΧΕΙΩΝ ΠΟΥ ΣΥΣΧΕΤΙΖΟΝΤΑΝ ΜΕ ΕΝΑ ΣΥΝΟΛΟ ΑΝΘΡΩΠΩΝ ΤΑ ΟΠΟΙΑ ΔΙΑΤΙΘΟΝΤΑΝ (ΣΥΧΝΑ ΕΝΑΝΤΙ ΑΜΟΙΒΗΣ) ΠΡΟΚΕΙΜΕΝΟΥ ΝΑ ΧΡΗΣΙΜΟΠΟΙΗΘΟΥΝ ΓΙΑ ΔΙΑΔΙΚΑΣΙΕΣ ΞΕΝΕΣ ΣΕ ΣΧΕΣΗ ΜΕ ΑΥΤΕΣ ΓΙΑ ΤΙΣ ΟΠΟΙΕΣ ΣΥΛΛΕΧΘΗΚΑΝ ΑΡΧΙΚΑ. ΤΟ ΘΕΣΜΙΚΟ ΠΛΑΙΣΙΟ ΠΟΥ ΔΙΕΠΕΙ ΤΑ ΘΕΜΑΤΑ ΕΞΟΥΣΗΣ ΔΕΔΟΜΕΝΩΝ ΣΕ ΕΛΛΑΔΑ ΚΑΙ ΕΥΡΩΠΗ

ΘΕΩΡΕΙΤΑΙ ΕΛΛΙΠΕΣ ΚΑΘΩΣ ΔΕΝ ΕΧΕΙ ΠΡΟΛΑΒΕΙ ΤΙΣ ΕΞΕΛΙΞΕΙΣ ΣΤΙΣ ΤΕΧΝΟΛΟΓΙΕΣ ΤΗΣ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΩΝ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ. ΣΤΗΝ ΠΡΟΣΠΑΘΕΙΑ ΕΦΑΡΜΟΓΗΣ ΜΙΑ ΑΠΟΔΟΤΙΚΗΣ ΠΟΛΙΤΙΚΗΣ ΓΙΑ ΤΗΝ ΠΡΟΣΤΑΣΙΑ ΤΩΝ ΠΡΟΣΩΠΙΚΩΝ ΔΕΔΟΜΕΝΩΝ ΑΠΟ ΤΙΣ ΔΙΑΔΙΚΑΣΙΕΣ ΕΞΟΡΥΞΗΣ ΔΕΔΟΜΕΝΩΝ ΕΙΝΑΙ ΠΙΘΑΝΟ ΤΑ ΣΥΓΚΕΝΤΡΩΜΕΝΑ ΔΕΔΟΜΕΝΑ ΝΑ ΜΗΝ ΕΙΝΑΙ ΔΥΝΑΤΟΝ ΑΝ ΤΥΧΟΥΝ ΕΚΜΕΤΑΛΛΕΥΣΗΣ. ΑΥΤΟ ΜΠΟΡΕΙ ΝΑ ΣΥΜΒΕΙ ΣΤΗΝ ΠΕΡΙΠΤΩΣΗ ΚΑΤΑ ΤΗΝ ΟΠΟΙΑ ΟΙ ΜΕΛΕΤΗΤΕΣ ΤΟΥΣ ΠΕΡΙΟΡΙΣΤΟΥΝ ΑΠΟ ΙΣΧΥΡΑ ΔΕΣΜΕΥΤΙΚΟΥΣ ΚΑΝΟΝΕΣ ΩΣ ΠΡΟΣ ΤΗΝ ΧΡΗΣΗ ΤΟΥΣ ΠΟΥ ΘΑ ΤΟΥΣ ΑΠΑΓΟΡΕΥΕΙ ΝΑ ΠΡΑΓΜΑΤΟΠΟΙΗΣΟΥΝ ΤΙΣ ΑΠΑΙΤΟΥΜΕΝΕΣ ΕΝΕΡΓΕΙΕΣ ΕΠΙ ΑΥΤΩΝ ΚΑΙ ΠΟΥ ΘΑ ΕΙΝΑΙ ΙΚΑΝΕΣ ΝΑ ΟΔΗΓΗΣΟΥΝ ΣΕ ΧΡΗΣΙΜΑ ΣΥΜΠΕΡΑΣΜΑΤΑ.

ΣΕ ΠΡΩΤΗ ΦΑΣΗ Ο ΟΡΓΑΝΙΣΜΟΣ ΓΙΑ ΟΙΚΟΝΟΜΙΚΗ ΣΥΝΕΡΓΑΣΙΑ ΚΑΙ ΑΝΑΠΤΥΞΗ (ΟΟΣΑ) ¹ΚΑΘΟΡΙΣΕ ΚΑΝΟΝΕΣ ΠΟΥ ΔΙΝΕΙ ΤΗΝ ΔΥΝΑΤΟΤΗΤΑ ΣΤΟΝ ΚΑΘΕ ΑΝΘΡΩΠΟ ΝΑ ΟΡΙΖΕΙ ΤΟ ΑΝ ΘΑ ΕΠΙΤΡΕΠΕΙ Η ΟΧΙ ΤΗΝ ΔΙΑΘΕΣΗ ΤΩΝ ΠΡΟΣΩΠΙΚΩΝ ΤΟΥ ΣΤΟΙΧΕΙΩΝ ΚΑΙ ΝΑ ΑΠΑΙΤΕΙ ΤΗΝ ΕΠΙΓΝΩΣΗ ΚΑΙ ΔΕΣΜΕΥΣΗ ΓΙΑ ΤΟΝ ΤΡΟΠΟ ΜΕ ΤΟΝ ΟΠΟΙΟ ΜΠΟΡΕΙ ΝΑ ΧΡΗΣΙΜΟΠΟΙΗΘΟΥΝ. ΠΑΡΑΛΛΗΛΑ ΟΡΙΖΕΤΑΙ ΟΤΙ ΑΝ ΤΑ ΠΡΟΣΩΠΙΚΑ ΔΕΔΟΜΕΝΑ ΧΡΗΣΙΜΟΠΟΙΗΘΟΥΝ ΘΑ ΠΡΕΠΕΙ ΝΑ ΕΙΝΑΙ ΠΡΟΣΦΑΤΑ ΚΑΙ ΕΓΚΥΡΑ ΩΣΤΕ ΝΑ ΑΠΟΦΕΥΓΕΤΑΙ Η ΠΕΡΙΠΤΩΣΗ Η ΕΠΕΞΕΡΓΑΣΙΑ ΤΟΥΣ ΝΑ ΔΗΜΙΟΥΡΓΕΙ ΕΣΦΑΛΜΕΝΑ ΣΥΜΠΕΡΑΣΜΑΤΑ ΠΟΥ ΕΝΔΕΧΕΤΑΙ ΝΑ ΘΙΓΟΥΝ ΤΑ ΥΠΟΚΕΙΜΕΝΑ ΤΟΥΣ [21].

ΣΕ ΣΥΝΔΥΑΣΜΟ ΜΕ ΤΙΣ ΔΙΑΤΑΞΕΙΣ ΑΥΤΕΣ ΕΙΝΑΙ ΣΗΜΑΝΤΙΚΟ ΟΙ ΦΟΡΕΙΣ ΠΟΥ ΕΜΠΛΕΚΟΝΤΑΙ ΣΤΗΝ ΔΙΑΤΗΡΗΣΗ ΚΑΙ ΕΠΕΞΕΡΓΑΣΙΑ ΕΥΑΙΣΘΗΤΩΝ ΠΡΟΣΩΠΙΚΩΝ ΔΕΔΟΜΕΝΩΝ ΝΑ ΕΞΑΣΦΑΛΙΖΟΥΝ ΤΗΝ ΑΣΦΑΛΕΙΑ ΤΟΥΣ. ΘΑ ΠΡΕΠΕΙ ΔΗΛΑΔΗ ΝΑ ΕΞΑΣΦΑΛΙΖΟΥΝ ΟΤΙ ΔΕΝ ΠΡΟΚΕΙΤΑΙ ΝΑ ΥΠΟΣΤΟΥΝ ΑΛΛΟΙΩΣΗ Η ΟΤΙ ΔΕΝ ΠΡΟΚΕΙΤΑΙ ΝΑ ΠΕΡΑΣΟΥΝ ΣΤΗΝ ΚΑΤΟΧΗ ΑΝΑΡΜΟΔΙΩΝ ΟΝΤΟΤΗΤΩΝ. Η ΠΡΟΣΤΑΣΙΑ ΑΠΟ ΚΑΚΟΒΟΥΛΗ ΕΞΟΡΥΞΗ ΔΕΔΟΜΕΝΩΝ ΑΠΟΤΕΛΕΙ ΜΙΑ ΕΞΑΙΡΕΤΙΚΑ

¹ΔΙΕΘΝΗΣ ΟΡΓΑΝΙΣΜΟΣ ΕΚΕΙΝΩΝ ΤΩΝ ΑΝΑΠΤΥΓΜΕΝΩΝ ΧΩΡΩΝ ΠΟΥ ΥΠΟΣΤΗΡΙΖΟΥΝ ΤΙΣ ΑΡΧΕΣ ΤΗΣ ΑΝΤΙΠΡΟΣΩΠΕΥΤΙΚΗΣ ΔΗΜΟΚΡΑΤΙΑΣ ΚΑΙ ΤΗΣ ΟΙΚΟΝΟΜΙΑΣ ΤΗΣ ΕΛΕΥΘΕΡΗΣ ΑΓΟΡΑΣ. Ο ΟΡΓΑΝΙΣΜΟΣ ΠΑΡΕΧΕΙ ΕΝΑ ΠΕΡΙΒΑΛΛΟΝ ΟΠΟΥ ΟΙ ΚΥΒΕΡΝΗΣΕΙΣ ΜΠΟΡΟΥΝ ΝΑ ΣΥΓΚΡΙΝΟΥΝ ΕΦΑΡΜΟΓΕΣ ΠΟΛΙΤΙΚΗΣ, ΝΑ ΒΡΟΥΝ ΑΠΑΝΤΗΣΕΙΣ ΣΤΑ ΚΟΙΝΑ ΠΡΟΒΛΗΜΑΤΑ, ΝΑ ΠΡΟΣΔΙΟΡΙΣΟΥΝ ΤΙΣ ΚΑΛΕΣ ΠΡΑΚΤΙΚΕΣ ΚΑΙ ΝΑ ΣΥΝΤΟΝΙΣΟΥΝ ΤΙΣ ΕΣΩΤΕΡΙΚΕΣ ΚΑΙ ΔΙΕΘΝΕΙΣ ΠΟΛΙΤΙΚΕΣ.

ΔΥΣΚΟΛΗ ΔΙΑΔΙΚΑΣΙΑ ΚΑΤΑ ΤΗΝ ΟΠΟΙΑ ΘΑ ΠΡΕΠΕΙ ΝΑ ΛΑΜΒΑΝΟΝΤΑΙ ΥΠ' ΟΨΗ ΤΑ ΕΞΗΣ:

- ΜΕ ΤΗΝ ΕΞΟΡΥΞΗ ΔΕΔΟΜΕΝΩΝ ΕΙΝΑΙ ΔΥΝΑΤΟΝ ΝΑ ΑΝΙΧΝΕΥΘΟΥΝ ΠΡΟΤΥΠΑ ΚΑΙ ΤΑΣΕΙΣ ΓΙΑ ΤΗΝ ΣΥΜΠΕΡΙΦΟΡΑ ΚΑΙ ΤΙΣ ΤΑΣΕΙΣ ΤΩΝ ΟΝΤΟΤΗΤΩΝ ΠΟΥ ΜΕΛΕΤΩΝΤΑΙ ΧΩΡΙΣ ΝΑ ΧΡΕΙΑΖΕΤΑΙ ΝΑ ΑΝΑΚΤΗΘΕΙ ΠΛΗΡΟΦΟΡΙΑ ΠΟΥ ΝΑ ΤΙΣ ΠΡΟΣΔΙΟΡΙΖΕΙ ΡΗΤΑ.
- ΟΙ ΜΕΘΟΔΟΛΟΓΙΕΣ ΕΞΟΡΥΞΗΣ ΔΕΔΟΜΕΝΩΝ ΜΠΟΡΕΙ ΝΑ ΕΙΝΑΙ ΕΠΙΤΥΧΕΙΣ ΒΑΣΙΖΟΜΕΝΕΣ ΣΕ ΣΥΝΔΥΑΣΜΟ ΣΥΝΟΛΩΝ ΔΕΔΟΜΕΝΩΝ ΤΑ ΟΠΟΙΑ ΜΕΜΟΝΩΜΕΝΑ ΕΝΔΕΧΕΤΑΙ ΝΑ ΜΗΝ ΕΧΟΥΝ ΑΞΙΑ.
- ΔΕΝ ΕΙΝΑΙ ΓΕΝΙΚΑ ΠΡΟΒΛΕΨΙΜΟ ΤΟ ΑΝ ΚΑΙ ΠΟΙΑ ΣΥΝΟΛΑ ΔΕΔΟΜΕΝΩΝ ΜΠΟΡΟΥΝ ΝΑ ΟΔΗΓΗΣΟΥΝ ΣΕ ΕΠΙΤΥΧΗΜΕΝΕΣ ΔΙΑΔΙΚΑΣΙΕΣ ΕΞΟΡΥΞΗΣ ΔΕΔΟΜΕΝΩΝ.

ΆΛΛΕΣ ΤΕΧΝΙΚΕΣ ΠΟΥ ΕΦΑΡΜΟΖΟΝΤΑΙ ΠΡΟΚΕΙΜΕΝΟΥ ΝΑ ΓΙΝΕΤΑΙ ΣΕΒΑΣΤΗ Η ΙΔΩΤΙΚΟΤΗΤΑ ΤΩΝ ΥΠΟΚΕΙΜΕΝΩΝ ΤΩΝ ΔΙΑΔΙΚΑΣΙΩΝ ΕΞΟΡΥΞΗΣ ΔΕΔΟΜΕΝΩΝ ΕΙΝΑΙ:

- Η ΠΡΟΣΠΑΘΕΙΑ ΔΙΑΤΗΡΗΣΗΣ ΤΗΣ ΑΝΩΝΥΜΙΑΣ ΤΩΝ ΣΥΜΜΕΤΕΧΟΝΤΩΝ ΔΕΙΓΜΑΤΩΝ. ΑΠΟΦΕΥΓΕΤΑΙ ΔΗΛΑΔΗ ΝΑ ΛΑΜΒΑΝΕΤΑΙ ΚΑΙ ΝΑ ΚΑΤΑΓΡΑΦΕΤΑΙ ΟΠΟΙΟΔΗΠΟΤΕ ΣΤΟΙΧΕΙΟ ΜΠΟΡΕΙ ΝΑ ΕΙΝΑΙ ΙΚΑΝΟ ΝΑ ΠΡΟΣΔΙΟΡΙΣΕΙ ΤΗΝ ΤΑΥΤΟΤΗΤΑ ΚΑΠΟΙΟΥ ΣΤΟΙΧΕΙΟΥ ΤΟΥ ΔΕΙΓΜΑΤΟΣ ΚΑΙ ΤΟ ΟΠΟΙΟ ΔΕΝ ΑΠΟΤΕΛΕΙ ΟΥΣΙΑΣΤΙΚΗ ΠΑΡΑΜΕΤΡΟ ΤΗΣ ΟΛΗΣ ΔΙΑΔΙΚΑΣΙΑΣ.
- ΠΕΡΙΟΡΙΣΜΟΙ ΤΩΝ ΕΡΩΤΗΜΑΤΩΝ. ΜΕ ΤΗΝ ΤΕΧΝΙΚΗ ΑΥΤΗ ΠΕΡΙΟΡΙΖΕΤΑΙ Ο ΑΡΙΘΜΟΣ ΤΩΝ ΕΡΩΤΗΜΑΤΩΝ ΑΝΑΚΤΗΣΗΣ ΣΤΟΙΧΕΙΩΝ ΑΠΟ ΤΑ ΣΥΓΚΕΝΤΡΩΜΕΝΑ ΔΕΔΟΜΕΝΑ ΣΕ ΕΚΕΙΝΑ ΤΑ ΟΠΟΙΑ ΕΙΝΑΙ ΖΩΤΙΚΑ Η ΟΥΣΙΩΔΗ ΓΙΑ ΤΗΝ ΔΙΑΔΙΚΑΣΙΑ.
- ΔΥΝΑΜΙΚΗ ΔΕΙΓΜΑΤΟΛΗΨΙΑ. Η ΤΕΧΝΙΚΗ ΑΥΤΗ ΕΦΑΡΜΟΖΕΤΑΙ ΠΡΟΚΕΙΜΕΝΟΥ ΝΑ ΜΕΙΩΝΕΤΑΙ ΚΑΤΑ ΤΟ

ΔΥΝΑΤΟΝ ΠΕΡΙΣΣΟΤΕΡΟ ΤΟ ΜΕΓΕΘΟΣ ΤΟΥ ΔΕΙΓΜΑΤΟΣ ΠΟΥ ΧΡΗΣΙΜΟΠΟΙΕΙΤΑΙ. Η ΜΕΙΩΣΗΣ ΑΥΤΗ ΠΡΑΓΜΑΤΟΠΟΙΕΙΤΑΙ ΜΕΧΡΙ ΤΟ ΣΗΜΕΙΟ ΕΚΕΙΝΟ ΚΑΤΑ ΤΟ ΟΠΟΙΟ ΕΙΝΑΙ ΕΦΙΚΤΗ Η ΠΑΡΑΓΩΓΗ ΑΣΦΑΛΩΝ ΣΥΜΠΕΡΑΣΜΑΤΩΝ ΧΩΡΙΣ ΣΗΜΑΝΤΙΚΑ ΣΦΑΛΜΑΤΑ.

- ΠΡΟΣΘΗΚΗ ΤΕΧΝΗΤΟΥ ΘΟΡΥΒΟΥ ΣΤΟ ΔΕΙΓΜΑ. ΠΟΛΛΕΣ ΦΟΡΕΣ Ο ΘΟΡΥΒΟΣ ΣΤΑ ΔΕΔΟΜΕΝΑ ΕΙΝΑΙ ΕΠΙΘΥΜΗΤΟΣ ΩΣΤΕ ΑΠΟΦΕΥΓΕΤΑΙ Η ΑΠΟΚΑΛΥΨΗ ΚΑΠΟΙΩΝ ΑΠΟ ΤΑ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΤΩΝ ΜΟΝΑΔΩΝ ΤΟΥΣ. ΈΤΣΙ ΕΠΙΔΙΩΚΕΤΑΙ Η ΔΙΑΤΑΡΑΞΗ ΜΕΜΟΝΩΜΕΝΩΝ ΔΕΔΟΜΕΝΩΝ ΜΕ ΤΡΟΠΟ ΤΕΤΟΙΟ ΠΟΥ ΣΥΝΟΛΙΚΑ ΝΑ ΜΗΝ ΜΕΤΑΒΑΛΛΟΝΤΑΙ ΤΑ ΣΤΑΤΙΣΤΙΚΑ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΤΟΥ ΣΥΝΟΛΟΥ. ΜΙΑ ΑΛΛΗ ΜΟΡΦΗ ΤΗΣ ΤΕΧΝΙΚΗΣ ΑΥΤΗΣ ΘΕΛΕΙ ΤΙΣ ΤΙΜΕΣ ΤΩΝ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ ΤΩΝ ΜΟΝΑΔΩΝ ΤΟΥ ΔΕΙΓΜΑΤΟΣ ΝΑ ΑΝΤΑΛΛΑΣΣΟΝΤΑΙ, ΧΩΡΙΣ ΕΤΣΙ ΝΑ ΕΠΗΡΕΑΖΟΥΝ ΤΙΣ ΤΙΜΕΣ ΠΟΥ ΠΡΟΚΥΠΤΟΥΝ ΑΠΟ ΤΗΝ ΣΥΝΟΛΙΚΗ ΤΟΥ ΑΝΑΛΥΣΗ.

2. ΔΕΝΔΡΑ ΑΠΟΦΑΣΕΩΝ

2.1 ΠΕΡΙΓΡΑΦΗ

Η ΛΕΙΤΟΥΡΓΙΑ ΤΩΝ ΟΡΓΑΝΙΣΜΩΝ ΠΡΟΫΠΟΘΕΤΕΙ ΤΗΝ ΛΗΨΗ ΑΠΟΦΑΣΕΩΝ ΣΕ ΟΛΑ ΤΑ ΕΠΙΠΕΔΑ ΤΟΥ ΑΛΛΑ ΚΥΡΙΩΣ ΣΕ ΣΤΡΑΤΗΓΙΚΟ ΕΠΙΠΕΔΟ. Η ΑΝΑΓΚΗ ΑΥΤΗ ΠΡΟΚΥΠΤΕΙ ΩΣ ΣΥΝΕΠΕΙΑ ΜΕΤΑΒΟΛΩΝ ΣΤΗΝ ΠΡΟΓΡΑΜΜΑΤΙΣΜΕΝΗ ΡΟΗ ΤΩΝ ΔΙΑΔΙΚΑΣΙΩΝ Η ΕΚΤΑΚΤΩΝ ΚΑΤΑΣΤΑΣΕΩΝ ΠΟΥ ΕΠΙΚΡΑΤΟΥΝ (ΘΕΤΙΚΩΝ Η ΑΡΝΗΤΙΚΩΝ). ΣΤΙΣ ΠΕΡΙΠΤΩΣΕΙΣ ΑΥΤΕΣ ΧΡΕΙΑΖΕΤΑΙ ΝΑ ΕΠΑΝΑΠΡΟΣΔΙΟΡΙΣΤΕΙ Η ΣΥΜΠΕΡΙΦΟΡΑ ΤΟΥ ΟΡΓΑΝΙΣΜΟΥ ΜΕ ΛΗΨΗ ΑΠΟΦΑΣΕΩΝ. Η ΠΟΙΟΤΗΤΑ ΤΩΝ ΑΠΟΦΑΣΕΩΝ ΑΥΤΩΝ ΕΙΝΑΙ Η ΠΑΡΑΜΕΤΡΟΣ ΤΗΣ ΣΥΝΑΡΤΗΣΗΣ ΠΟΥ ΕΞΟΔΟ ΕΧΕΙ ΤΗΝ ΑΠΟΔΟΤΙΚΟΤΗΤΑ ΤΟΥ. ΟΙ ΠΑΡΑΜΕΤΡΟΙ ΠΟΥ ΕΠΗΡΕΑΖΟΥΝ ΤΗΝ ΠΟΙΟΤΗΤΑ ΤΩΝ ΑΠΟΦΑΣΕΩΝ ΕΙΝΑΙ Η ΕΥΣΤΟΧΙΑ ΤΟΥΣ ΚΑΙ Η ΤΑΧΥΤΗΤΑ ΛΗΨΗΣ ΤΟΥΣ. ΤΑ ΣΤΑΔΙΑ ΛΗΨΗΣ ΑΠΟΦΑΣΕΩΝ ΕΙΝΑΙ:

- ΑΝΙΧΝΕΥΣΗ ΚΑΙ ΠΡΟΣΔΙΟΡΙΣΜΟΣ ΤΟΥ ΠΡΟΒΛΗΜΑΤΟΣ, ΤΗΣ ΑΠΕΙΛΗΣ Η ΤΗΣ ΕΥΚΑΙΡΙΑΣ ΠΟΥ ΠΑΡΟΥΣΙΑΖΕΤΑΙ.
- ΚΑΘΟΡΙΣΜΟΣ ΤΩΝ ΕΝΑΛΛΑΚΤΙΚΩΝ ΛΥΣΕΩΝ ΚΑΙ ΕΠΙΛΟΓΩΝ ΠΟΥ ΜΠΟΡΟΥΝ ΝΑ ΕΦΑΡΜΟΣΤΟΥΝ.
- ΚΑΘΟΡΙΣΜΟΣ ΤΩΝ ΚΡΙΤΗΡΙΩΝ ΑΞΙΟΛΟΓΗΣΗΣ ΤΩΝ ΕΝΑΛΛΑΚΤΙΚΩΝ ΛΥΣΕΩΝ ΚΑΙ ΕΠΙΛΟΓΩΝ ΚΑΙ ΑΞΙΟΛΟΓΗΣΗ ΤΟΥΣ ΜΕ ΤΑ ΚΡΙΤΗΡΙΑ ΑΥΤΑ.
- ΕΠΙΛΟΓΗ ΤΗΣ ΒΕΛΤΙΣΤΗΣ ΕΦΙΚΤΗΣ ΛΥΣΗΣ.

ΤΙΣ ΠΕΡΙΣΣΟΤΕΡΕΣ ΦΟΡΕΣ ΤΑ ΣΤΕΛΕΧΗ ΤΩΝ ΟΡΓΑΝΙΣΜΩΝ ΒΡΙΣΚΟΝΤΑΙ ΜΠΡΟΣΤΑ ΣΕ ΠΡΟΒΛΗΜΑΤΑ ΑΠΟΦΑΣΕΩΝ ΔΙΑΚΡΙΤΩΝ ΕΠΙΛΟΓΩΝ. ΑΥΤΟ ΣΗΜΑΙΝΕΙ ΟΤΙ ΚΑΘΕ ΑΠΟΦΑΣΗ ΠΟΥ ΠΡΕΠΕΙ ΝΑ ΛΗΦΘΕΙ ΠΕΡΙΛΑΜΒΑΝΕΙ ΠΕΠΕΡΑΣΜΕΝΟ ΠΛΗΘΟΣ ΔΙΑΚΡΙΤΩΝ ΕΝΑΛΛΑΚΤΙΚΩΝ ΕΠΙΛΟΓΩΝ. ΤΑ ΠΡΟΒΛΗΜΑΤΑ ΑΥΤΑ ΕΠΙΛΥΟΝΤΑΙ ΓΕΝΙΚΑ ΑΚΟΛΟΥΘΟΝΤΑΣ ΤΗΝ ΠΑΡΑΚΑΤΩ ΔΙΑΔΙΚΑΣΙΑ:

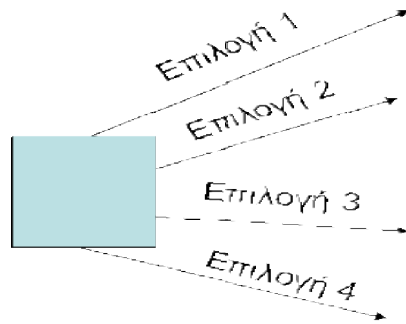
- ΜΟΝΤΕΛΟΠΟΙΗΣΗ ΤΟΥ ΠΡΟΒΛΗΜΑΤΟΣ ΑΠΟΦΑΣΕΩΝ ΣΕ ΑΠΛΟΥΣΤΕΥΜΕΝΗ ΜΟΡΦΗ ΠΟΥ ΝΑ ΜΠΟΡΕΙ ΝΑ ΤΥΧΕΙ ΕΠΕΞΕΡΓΑΣΙΑΣ ΑΠΟ ΗΛΕΚΤΡΟΝΙΚΟ ΥΠΟΛΟΓΙΣΤΗ.
- ΕΠΙΛΥΣΗ ΤΟΥ ΜΟΝΤΕΛΟΠΟΙΗΜΕΝΟΥ ΠΡΟΒΛΗΜΑΤΟΣ ΜΕ ΤΟΝ ΠΡΟΣΔΙΟΡΙΣΜΟ ΤΗΣ ΒΕΛΤΙΣΤΗΣ ΑΠΟΔΟΣΗΣ ΔΙΑΔΟΧΗΣ ΕΠΙΛΟΓΩΝ.
- ΑΝΑΛΥΣΗ ΕΠΙΚΙΝΔΥΝΟΤΗΤΑΣ ΤΗΣ ΠΑΡΑΠΑΝΩ ΔΙΑΔΟΧΗΣ ΕΠΙΛΟΓΩΝ.
- ΑΝΑΛΥΣΗ ΕΥΑΙΣΘΗΣΙΑΣ ΟΠΟΥ ΕΛΕΓΧΕΤΑΙ ΚΑΤΑ ΠΟΣΟ ΜΙΚΡΕΣ ΑΛΛΑΓΕΣ ΣΤΙΣ ΠΑΡΑΜΕΤΡΟΥΣ ΤΟΥ ΠΡΟΒΛΗΜΑΤΟΣ ΜΠΟΡΕΙ ΝΑ ΟΔΗΓΗΣΕΙ ΣΕ ΜΕΤΑΒΟΛΕΣ ΤΗΣ ΒΕΛΤΙΣΤΗΣ ΔΙΑΔΟΧΗΣ ΕΠΙΛΟΓΩΝ.

ΓΙΑ ΤΗΝ ΜΟΝΤΕΛΟΠΟΙΗΣΗ ΤΩΝ ΠΡΟΒΛΗΜΑΤΩΝ ΑΠΟΦΑΣΗΣ ΔΙΑΚΡΙΤΩΝ ΕΠΙΛΟΓΩΝ ΣΥΧΝΑ ΧΡΗΣΙΜΟΠΟΙΟΥΝΤΑΙ ΤΑ ΔΕΝΔΡΑ ΑΠΟΦΑΣΕΩΝ. ΤΑ ΔΕΝΔΡΑ ΑΠΟΦΑΣΕΩΝ ΕΙΝΑΙ ΓΝΩΣΤΑ ΚΑΙ ΜΕ ΤΟΝ ΟΡΟ ΔΙΑΓΡΑΜΜΑΤΑ ΑΠΟΦΑΣΕΩΝ. ΑΠΟΤΕΛΟΥΝ ΜΙΑ ΓΡΑΦΙΚΗ ΑΝΑΠΑΡΑΣΤΑΣΗ ΤΗΣ ΑΝΑΛΥΣΗΣ ΤΩΝ ΠΡΟΒΛΗΜΑΤΩΝ ΚΑΙ ΤΗΝ ΠΟΡΕΙΑ ΕΠΙΛΥΣΗΣ ΤΟΥΣ. Η ΚΕΝΤΡΙΚΗ ΙΔΕΑ ΣΤΗΝ ΟΠΟΙΑ ΒΑΣΙΖΟΝΤΑΙ ΕΙΝΑΙ ΟΤΙ Η ΕΠΙΛΥΣΗ ΤΩΝ ΠΡΟΒΛΗΜΑΤΩΝ ΠΕΡΝΑΕΙ ΑΠΟ ΤΗΝ ΔΙΑΔΟΧΙΚΗ ΛΗΨΗ ΑΠΟΦΑΣΕΩΝ ΜΕΧΡΙ ΝΑ ΔΗΜΙΟΥΡΓΗΘΟΥΝ ΟΙ ΚΑΤΑΛΛΗΛΕΣ ΠΡΟΫΠΟΘΕΣΕΙΣ ΓΙΑ ΤΗΝ ΕΠΙΛΥΣΗ. ΈΤΣΙ ΜΕ ΔΕΝΔΡΟΕΙΔΗ ΜΟΡΦΗ ΠΑΡΟΥΣΙΑΖΟΝΤΑΙ ΟΙ ΔΙΑΔΟΧΙΚΕΣ ΑΠΟΦΑΣΕΙΣ ΚΑΙ ΤΑ ΜΕΤΑΒΛΗΤΑ ΑΠΟΤΕΛΕΣΜΑ ΣΕ ΔΕΔΟΜΕΝΑ ΧΡΟΝΙΚΑ ΔΙΑΣΤΗΜΑΤΑ. ΌΠΩΣ ΚΑΘΕ ΔΕΝΔΡΟΕΙΔΗΣ ΔΟΜΗ, ΜΕΙΩΝΕΙ ΤΗΝ ΠΟΛΥΠΛΟΚΟΤΗΤΑ ΤΩΝ ΠΡΟΒΛΗΜΑΤΩΝ ΚΑΘΩΣ ΤΟ ΑΝΑΛΥΕΙ ΣΕ ΕΠΙΜΕΡΟΥΣ ΜΙΚΡΟΤΕΡΑ.

ΚΑΘΕ ΔΕΝΔΡΟ ΑΠΟΦΑΣΗΣ ΠΕΡΙΛΑΜΒΑΝΕΙ:

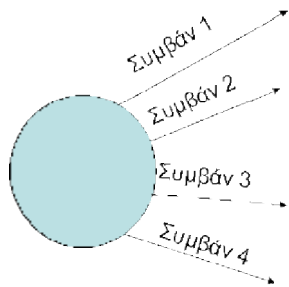
- ΚΟΜΒΟΥΣ ΑΠΟΦΑΣΗΣ: ΣΥΝΗΘΩΣ ΣΥΜΒΟΛΙΖΟΝΤΑΙ ΜΕ ΤΕΤΡΑΓΩΝΟ. ΠΑΡΙΣΤΑΝΟΥΝ ΚΑΤΑΣΤΑΣΕΙΣ – ΟΡΟΣΗΜΑ ΣΤΑ ΟΠΟΙΑ ΣΤΕΛΕΧΟΣ Η ΦΟΡΕΑΣ ΤΟΥ ΟΡΓΑΝΙΣΜΟΥ ΚΑΛΕΙΤΑΙ ΝΑ ΛΑΒΕΙ ΜΙΑ ΑΠΟΦΑΣΗ ΑΠΟ ΕΝΑ ΠΕΠΕΡΑΣΜΕΝΟ ΣΥΝΟΛΟ ΔΙΑΘΕΣΙΜΩΝ. ΟΙ ΔΙΑΘΕΣΙΜΕΣ ΕΠΙΛΟΓΕΣ ΑΝΑΠΑΡΙΣΤΑΝΤΑΙ ΜΕ ΒΕΛΗ ΠΟΥ ΕΧΟΥΝ ΣΑΝ

ΑΦΕΤΗΡΙΑ ΤΟ ΤΕΤΡΑΓΩΝΩΝ ΑΥΤΟ. ΣΤΑ ΒΕΛΗ ΑΝΑΓΡΑΦΟΝΤΑΙ ΧΑΡΑΚΤΗΡΙΣΤΙΚΟΙ ΠΑΡΑΜΕΤΡΟΙ ΤΗΣ ΕΠΙΛΟΓΗΣ (ΑΝ ΥΠΑΡΧΟΥΝ). ΣΤΗΝ ΠΑΡΑΚΑΤΩ ΕΙΚΟΝΑ ΦΑΙΝΕΤΑΙ ΕΝΑ ΤΕΤΟΙΟ ΟΡΟΣΗΜΟ ΜΕ ΤΕΣΣΕΡΙΣ ΕΝΑΛΛΑΚΤΙΚΕΣ ΑΠΟΦΑΣΕΙΣ.



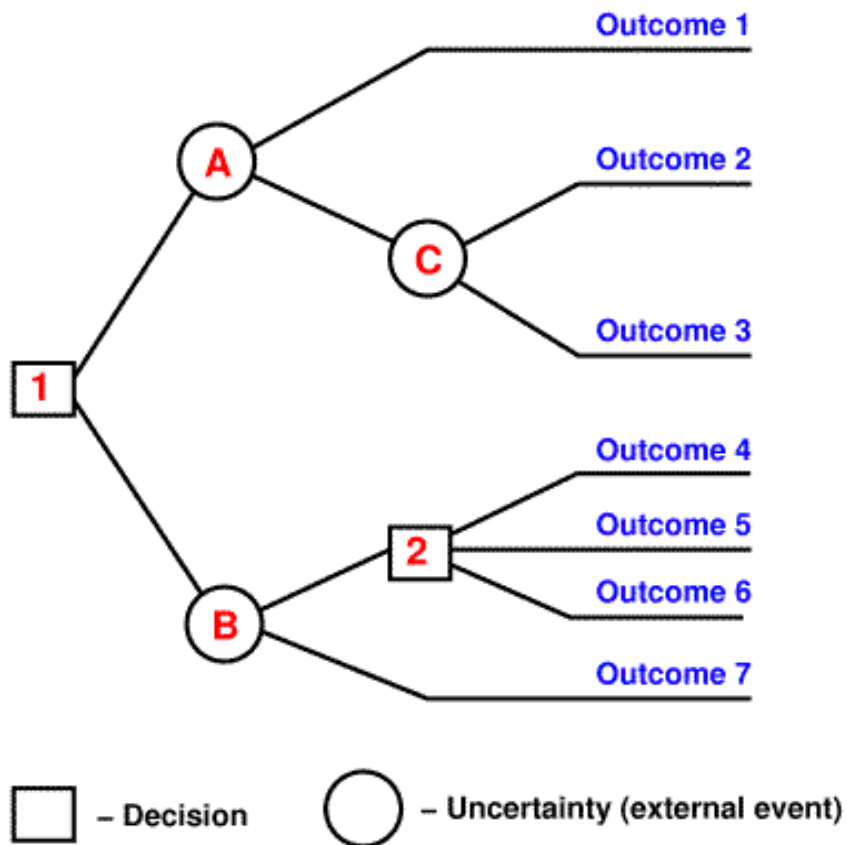
ΕΙΚΟΝΑ 11: ΚΟΜΒΟΣ ΑΠΟΦΑΣΗΣ ΣΕ ΔΕΝΔΡΟ ΑΠΟΦΑΣΗΣ (ΠΗΓΗ:BRHAM YOUNG UNIVERSITY, DECISIONTREES, [HTTP://WIKI.CS.BYU.EDU/CS-677SP2010/DECISION-THEORETIC-GRAPHICAL-MODELS](http://wiki.cs.byu.edu/cs-677sp2010/decision-theoretic-graphical-models))

- ΚΟΜΒΟΥΣ ΤΥΧΗΣ: ΣΥΜΒΟΛΙΖΟΝΤΑΙ ΜΕ ΚΥΚΛΟ ΚΑΙ ΑΝΑΠΑΡΙΣΤΟΥΝ ΟΡΟΣΗΜΑ ΣΤΑ ΟΠΟΙΑ ΑΝΑΜΕΝΕΤΑΙ Η ΕΝΔΕΧΕΤΑΙ ΝΑ ΣΥΜΒΕΙ ΕΝΑ ΑΠΡΟΓΡΑΜΜΑΤΙΣΤΟ ΓΕΓΟΝΟΣ ΠΟΥ ΠΕΡΙΛΑΜΒΑΝΕΤΑΙ ΣΕ ΕΝΑ ΠΕΠΕΡΑΣΜΕΝΟ ΣΥΝΟΛΟ. ΤΑ ΓΕΓΟΝΟΤΑ ΑΥΤΑ ΑΝΑΠΑΡΙΣΤΑΝΤΑΙ ΜΕ ΒΕΛΗ ΠΟΥ ΕΧΟΥΝ ΣΑΝ ΑΦΕΤΗΡΙΑ ΤΟΝ ΚΟΜΒΟ ΤΥΧΗΣ ΜΕ ΤΟΝ ΟΠΟΙΟ ΣΧΕΤΙΖΟΝΤΑΙ. ΣΤΗΝ ΠΑΡΑΚΑΤΩ ΕΙΚΟΝΑ ΠΑΡΟΥΣΙΑΖΕΤΑΙ ΕΝΑΣ ΚΟΜΒΟΣ ΤΥΧΗΣ ΜΕ ΤΕΣΣΕΡΙΣ ΔΥΝΑΤΕΣ ΕΠΙΛΟΓΕΣ.



ΕΙΚΟΝΑ 12: ΚΟΜΒΟΣ ΤΥΧΗΣ ΣΕ ΔΕΝΔΡΟ ΑΠΟΦΑΣΗΣ (ΠΗΓΗ:BRIHAM YOUNG UNIVERSITY, DECISIONTREES, [HTTP://WIKI.CS.BYU.EDU/CS-677SP2010/DECISION-THEORETIC-GRAPHICAL-MODELS](http://wiki.cs.byu.edu/cs-677sp2010/decision-theoretic-graphical-models))

Η ΑΝΑΠΤΥΞΗ ΕΝΟΣ ΔΕΝΔΡΟΥ ΑΠΟΦΑΣΗΣ ΞΕΚΙΝΑ ΑΠΟ ΑΡΙΣΤΕΡΑ ΠΡΟΣ ΤΑ ΔΕΞΙΑ ΜΕ ΤΗΝ ΣΧΕΔΙΑΣΗ ΕΝΟΣ Η ΠΕΡΙΣΣΟΤΕΡΩΝ ΚΟΜΒΩΝ ΑΠΟΦΑΣΗΣ ΜΕ ΤΙΣ ΣΧΕΤΙΚΕΣ ΜΕ ΑΥΤΟΥΣ ΕΝΑΛΛΑΚΤΙΚΕΣ ΕΠΙΛΟΓΕΣ. ΚΑΘΕ ΜΙΑ ΑΠΟ ΤΙΣ ΕΠΙΛΟΓΕΣ ΑΥΤΕΣ ΚΑΤΑΛΗΓΕΙ ΣΕ ΚΑΠΟΙΟΝ ΑΛΛΟΝ ΚΟΜΒΟ ΑΠΟΦΑΣΗΣ Η ΣΕ ΚΟΜΒΩΝ ΤΥΧΗΣ ΜΕ ΤΙΣ ΕΝΑΛΛΑΚΤΙΚΕΣ ΤΟΥΣ ΕΠΙΛΟΓΕΣ Η ΣΥΜΒΑΝΤΑ ΑΝΤΙΣΤΟΙΧΑ. Η ΔΙΑΔΙΚΑΣΙΑ ΑΥΤΗ ΕΠΑΝΑΛΑΜΒΑΝΕΤΑΙ ΕΞΕΛΙΣΣΟΜΕΝΗ ΑΠΟ ΑΡΙΣΤΕΡΑ ΠΡΟΣ ΤΑ ΔΕΞΙΑ ΜΕΧΡΙ ΤΗΝ ΤΕΛΙΚΗ ΕΠΙΛΥΣΗ ΤΟΥ ΠΡΟΒΛΗΜΑΤΟΣ. ΣΤΗΝ ΕΠΟΜΕΝΗ ΕΙΚΟΝΑ ΦΑΙΝΕΤΑΙ ΜΙΑ ΤΥΠΙΚΗ ΜΟΡΦΗ ΕΝΟΣ ΔΕΝΔΡΟΥ ΑΠΟΦΑΣΗΣ.



ΕΙΚΟΝΑ13: ΜΟΡΦΗ ΔΕΝΔΡΟΥ ΑΠΟΦΑΣΗΣ (ΠΗΓΗ: TIME-MANAGEMENT-GUIDE.COM, USE A DECISION TREE ANALYSIS TO SYSTEMATICALLY ARRIVE AT YOUR SMARTEST CHOICE, [HTTP://WWW.TIME-MANAGEMENT-GUIDE.COM/DECISION-TREE.HTML](http://www.time-management-guide.com/decision-tree.html))

ΑΝΑΛΥΤΙΚΟΤΕΡΑ Η ΔΙΑΔΙΚΑΣΙΑ ΚΑΤΑΣΚΕΥΗΣ ΕΝΟΣ ΔΕΝΔΡΟΥ ΑΠΟΦΑΣΗΣ ΠΕΡΙΛΑΜΒΑΝΕΙ:

- ΠΡΟΣΔΙΟΡΙΣΜΟΣ ΤΩΝ ΟΡΟΣΗΜΩΝ ΟΠΟΥ ΑΠΑΙΤΕΙΤΑΙ ΛΗΨΗ ΑΠΟΦΑΣΗΣ. ΓΙΑ ΚΑΘΕ ΕΝΑ ΑΠΟ ΤΑ ΟΡΟΣΗΜΑ ΑΥΤΑ ΘΑ ΠΡΕΠΕΙ ΝΑ ΠΡΟΣΔΙΟΡΙΣΤΟΥΝ ΚΑΙ ΟΙ ΠΕΠΕΡΑΣΜΕΝΕΣ ΕΝΑΛΛΑΚΤΙΚΕΣ ΕΠΙΛΟΓΕΣ.
- ΠΡΟΣΔΙΟΡΙΣΜΟΣ ΤΩΝ ΟΡΟΣΗΜΩΝ ΟΠΟΥ ΑΝΑΜΕΝΕΤΑΙ Η ΕΝΔΕΧΕΤΑΙ ΝΑ ΠΑΡΟΥΣΙΑΣΤΟΥΝ ΤΥΧΑΙΑ ΣΥΜΒΑΝΤΑ. ΓΙΑ ΚΑΘΕ ΕΝΑ ΑΠΟ ΤΑ ΟΡΟΣΗΜΑ ΑΥΤΑ ΘΑ ΠΡΕΠΕΙ ΝΑ ΠΡΟΣΔΙΟΡΙΣΤΟΥΝ ΚΑΙ ΟΙ ΠΕΠΕΡΑΣΜΕΝΕΣ ΕΝΑΛΛΑΚΤΙΚΕΣ ΚΑΤΑΣΤΑΣΕΙΣ ΠΟΥ ΑΝΑΜΕΝΕΤΑΙ Η ΕΝΔΕΧΕΤΑΙ ΝΑ ΠΑΡΟΥΣΙΑΣΤΟΥΝ.

- ΕΚΤΙΜΗΣΗ ΤΩΝ ΤΙΜΩΝ ΤΩΝ ΠΑΡΑΜΕΤΡΩΝ ΤΟΥ ΠΡΟΒΛΗΜΑΤΟΣ ΠΟΥ ΕΙΝΑΙ ΙΚΑΝΕΣ ΝΑ ΔΙΑΦΟΡΟΠΟΙΟΥΝ ΤΗΝ ΡΟΗ ΚΙΝΗΣΗΣ ΣΤΟ ΔΕΝΔΡΟ ΑΠΟ ΚΟΜΒΟ ΣΕ ΚΟΜΒΟ.
- ΑΝΑΛΥΣΗ ΤΩΝ ΕΝΑΛΛΑΚΤΙΚΩΝ ΔΙΑΔΡΟΜΩΝ ΑΠΟ ΤΗΝ ΑΡΧΗ ΩΣ ΤΟ ΤΕΛΟΣ ΤΟΥ ΔΕΝΔΡΟΥ ΜΕ ΤΗΝ ΧΡΗΣΗ ΚΑΤΑΛΛΗΛΩΝ ΑΛΓΟΡΙΘΜΩΝ.

ΈΝΑ ΔΕΝΔΡΟ ΑΠΟΦΑΣΗΣ ΓΙΑ ΝΑ ΕΙΝΑΙ ΑΠΟΔΟΤΙΚΟ ΚΑΙ ΙΚΑΝΟ ΝΑ ΔΩΣΕΙ ΤΗΝ ΒΕΛΤΙΣΤΗ ΛΥΣΗ ΘΑ ΠΡΕΠΕΙ ΝΑ ΚΑΤΑΣΚΕΥΑΣΤΕΙ ΜΕ ΤΡΟΠΟ ΤΕΤΟΙΟ ΩΣΤΕ ΝΑ ΙΚΑΝΟΠΟΙΟΥΝΤΑΙ ΚΑΤ' ΕΛΑΧΙΣΤΟ ΟΙ ΕΞΗΣ ΒΑΣΙΚΕΣ ΑΡΧΕΣ:

- ΠΡΕΠΕΙ ΣΕ ΑΥΤΟ ΝΑ ΠΡΟΒΛΕΦΘΟΥΝ ΚΑΙ ΝΑ ΠΕΡΙΛΑΜΒΑΝΟΝΤΑΙ ΟΛΕΣ ΟΙ ΔΙΑΦΟΡΕΤΙΚΕΣ ΕΝΑΛΛΑΚΤΙΚΕΣ ΑΠΟΦΑΣΕΙΣ ΠΟΥ ΜΠΟΡΕΙ ΝΑ ΛΗΦΘΟΥΝ ΣΕ ΚΑΘΕ ΟΡΟΣΗΜΟ.
- ΠΡΕΠΕΙ ΝΑ ΠΡΟΒΛΕΦΘΟΥΝ ΚΑΙ ΝΑ ΠΕΡΙΛΑΜΒΑΝΟΝΤΑΙ ΣΕ ΑΥΤΟ ΟΛΑ ΤΑ ΠΙΘΑΝΑ ΟΡΟΣΗΜΑ ΣΤΑ ΟΠΟΙΑ ΜΠΟΡΕΙ ΝΑ ΠΑΡΟΥΣΙΑΣΤΟΥΝ ΚΑΤΑΣΤΑΣΕΙΣ ΕΚΤΟΣ ΤΟΥ ΑΡΧΙΚΟΥ ΠΡΟΓΡΑΜΜΑΤΙΣΜΟΥ.
- ΠΡΕΠΕΙ ΣΕ ΑΥΤΟ ΝΑ ΠΡΟΒΛΕΦΘΟΥΝ ΚΑΙ ΝΑ ΠΕΡΙΛΑΜΒΑΝΟΝΤΑΙ ΟΛΕΣ ΤΑ ΔΙΑΦΟΡΕΤΙΚΑ ΣΥΜΒΑΝΤΑ ΠΟΥ ΜΠΟΡΕΙ ΝΑ ΛΗΦΘΟΥΝ ΣΕ ΚΑΘΕ ΟΡΟΣΗΜΟ.
- ΠΡΕΠΕΙ ΝΑ ΕΞΕΤΑΖΕΤΑΙ ΤΟ ΠΩΣ ΜΠΟΡΕΙ ΝΑ ΕΠΗΡΕΑΖΕΙ ΚΑΘΕ ΠΙΘΑΝΟ ΣΥΜΒΑΝ Η ΕΝΑΛΛΑΚΤΙΚΗ ΑΠΟΦΑΣΗ ΤΗΝ ΜΟΡΦΗ ΤΩΝ ΕΠΟΜΕΝΩΝ ΟΡΟΣΗΜΩΝ.
- ΚΑΘΕ ΕΝΑΛΛΑΚΤΙΚΗ ΕΠΙΛΟΓΗ ΣΕ ΕΝΑΝ ΚΟΜΒΟ ΑΠΟΦΑΣΗΣ Η ΚΑΘΕ ΕΝΑΛΛΑΚΤΙΚΟ ΣΥΜΒΑΝ ΣΕ ΕΝΑΝ ΚΟΜΒΟ ΤΥΧΗΣ ΘΑ ΠΡΕΠΕΙ ΝΑ ΕΙΝΑΙ ΑΜΟΙΒΑΙΑ ΑΠΟΚΛΕΙΣΜΕΝΟ ΜΕ ΚΑΘΕ ΑΝΤΙΣΤΟΙΧΟ ΠΟΥ ΕΧΕΙ ΑΦΕΤΗΡΙΑ ΤΟΝ ΙΔΙΟ ΚΟΜΒΟ.
- Η ΕΠΙΛΟΓΗ ΤΩΝ ΠΑΡΑΜΕΤΡΩΝ ΠΟΥ ΘΑ ΧΡΗΣΙΜΟΠΟΙΗΘΟΥΝ ΓΙΑ ΤΗΝ ΔΗΜΙΟΥΡΓΙΑ ΤΩΝ

ΔΙΑΚΛΑΔΩΣΕΩΝ ΤΟΥ ΔΕΝΔΡΟΥ ΧΡΗΣΟΥΝ ΠΡΟΣΟΧΗΣ ΚΑΤΑ ΤΗΝ ΕΠΙΛΟΓΗ ΤΟΥΣ, ΑΦΟΥ ΠΑΙΖΟΥΝ ΣΗΜΑΝΤΙΚΟ ΡΟΛΟ ΣΤΗΝ ΑΠΟΔΟΣΗ ΤΟΥ. Η ΕΠΙΛΟΓΗ ΤΟΥΣ ΕΧΕΙ ΝΑ ΚΑΝΕΙ ΜΕ ΤΗΝ ΦΥΣΗ ΤΟΥ ΠΡΟΒΛΗΜΑΤΟΣ ΠΟΥ ΕΞΕΤΑΖΕΤΑΙ ΑΛΛΑ ΚΑΙ ΜΕ ΤΗΝ ΕΜΠΕΙΡΙΑ ΤΟΥ ΜΕΛΕΤΗΤΗ.

- Η ΣΕΙΡΑ ΜΕ ΤΗΝ ΟΠΟΙΑ ΘΑ ΕΞΕΤΑΣΤΟΥΝ ΟΙ ΠΑΡΑΜΕΤΡΟΙ ΤΩΝ ΟΡΟΣΗΜΩΝ ΕΠΙΣΗΣ ΠΑΙΖΕΙ ΣΗΜΑΝΤΙΚΟ ΡΟΛΟ ΣΤΟ ΜΕΓΕΘΟΣ ΚΑΙ ΣΤΗΝ ΑΠΟΔΟΣΗ ΤΟΥ ΔΕΝΔΡΟΥ.
- ΟΣΟ ΠΙΟ ΙΣΟΖΥΓΙΣΜΕΝΟ ΕΙΝΑΙ ΤΟ ΔΕΝΔΡΟ ΑΠΟΦΑΣΗΣ ΤΟΣΟ ΛΙΓΟΤΕΡΑ ΕΠΙΠΕΔΑ ΤΟΥ ΕΞΕΤΑΖΟΝΤΑΙ ΚΑΙ ΕΠΟΜΕΝΩΣ ΠΑΡΕΧΕΤΑΙ ΛΥΣΗ ΜΕ ΜΙΚΡΟΤΕΡΗ ΠΟΛΥΠΛΟΚΟΤΗΤΑ.
- ΤΑ ΔΕΔΟΜΕΝΑ ΤΑ ΟΠΟΙΑ ΘΑ ΧΡΗΣΙΜΟΠΟΙΗΘΟΥΝ ΓΙΑ ΤΗΝ ΕΚΠΑΙΔΕΥΣΗ ΤΟΥ ΔΕΝΔΡΟΥ ΑΠΟΦΑΣΗΣ ΩΣΤΕ ΝΑ ΠΙΣΤΟΠΟΙΗΘΕΙ Η ΚΑΤΑΛΛΗΛΟΛΗΤΑ ΤΟΥ ΓΙΑ ΤΗΝ ΑΝΤΙΜΕΤΩΠΙΣΗ ΓΕΝΙΚΕΥΜΕΝΩΝ ΠΡΟΒΛΗΜΑΤΩΝ ΑΠΟΦΑΣΗΣ, ΠΡΕΠΕΙ ΝΑ ΕΠΙΛΕΓΟΥΝ ΠΡΟΣΕΚΤΙΚΑ, ΝΑ ΕΙΝΑΙ ΕΠΑΡΚΗ ΣΕ ΠΟΣΟΤΗΤΑ ΚΑΙ ΝΑ ΑΞΙΟΛΟΓΗΘΟΥΝ ΜΕ ΑΚΡΙΒΕΙΑ.
- ΠΡΙΝ ΤΗΝ ΕΦΑΡΜΟΓΗ ΟΠΟΙΟΔΗΠΟΤΕ ΑΛΓΟΡΙΘΜΟΥ ΕΞΕΤΑΣΗΣ ΤΟΥ ΔΕΝΔΡΟΥ ΑΠΟΦΑΣΗΣ ΘΑ ΠΡΕΠΕΙ ΕΠ' ΑΥΤΟΥ ΝΑ ΕΦΑΡΜΟΣΤΕΙ ΚΛΑΔΕΜΑ ΩΣΤΕ ΝΑ ΑΠΟΜΑΚΡΥΝΘΟΥΝ ΚΛΑΔΟΙ ΠΟΥ ΕΙΤΕ ΕΙΝΑΙ ΠΕΡΙΤΤΟΙ ΕΙΤΕ ΕΠΑΝΑΛΑΜΒΑΝΟΝΤΑΙ. Η ΜΕΙΩΣΗ ΤΟΥ ΜΕΓΕΘΟΥΣ ΤΟΥ ΔΕΝΔΡΟΥ ΜΕΙΩΝΕΙ ΤΗΝ ΠΟΛΥΠΛΟΚΟΤΗΤΑ ΤΗΣ ΕΞΕΤΑΣΗΣ ΤΟΥ.

2.2 ΑΛΓΟΡΙΘΜΟΙ

ΓΙΑ ΤΗΝ ΚΑΤΗΓΟΡΙΟΠΟΙΗΣΗ ΤΩΝ ΔΕΔΟΜΕΝΩΝ ΜΕ ΔΕΝΔΡΑ ΑΠΟΦΑΣΗΣ ΕΧΟΥΝ ΕΠΙΚΡΑΤΗΣΕΙ ΚΑΤΑ ΚΥΡΙΟ ΛΟΓΟ ΤΡΕΙΣ ΑΛΓΟΡΙΘΜΟΙ ΠΟΥ ΜΑΛΙΣΤΑ ΣΥΜΒΑΙΝΕΙ Ο ΕΝΑΣ ΝΑ ΑΠΟΤΕΛΕΙ ΕΞΕΛΙΞΗ

ΤΟΥ ΑΛΛΟΥ. ΟΙ ΑΛΓΟΡΙΘΜΟΙ ΤΩΝ ΔΕΝΔΡΩΝ ΑΠΟΦΑΣΗΣ ΣΕ ΠΡΩΤΟ ΕΠΙΠΕΔΟ ΠΕΡΙΛΑΜΒΑΝΟΥΝ ΔΥΟ ΦΑΣΕΙΣ:

- ΤΗΝ ΦΑΣΗ ΤΗΣ ΑΝΑΠΤΥΞΗΣ ΤΟΥ ΔΕΝΔΡΟΥ ΚΑΤΑ ΤΗΝ ΟΠΟΙΑ ΣΧΗΜΑΤΙΖΕΤΑΙ ΤΟ ΔΕΝΔΡΟ ΑΝΑΛΟΓΑ ΜΕ ΤΙΣ ΕΠΙΤΑΓΕΣ ΤΟΥ ΚΑΘΕ ΑΛΓΟΡΙΘΜΟΥ.
- ΤΗ ΦΑΣΗ ΤΟΥ ΚΛΑΔΕΜΑΤΟΣ ΚΑΤΑ ΤΗΝ ΟΠΟΙΑ ΕΠΙΧΕΙΡΕΙΤΑΙ Η ΜΕΙΩΣΗ ΤΗΣ ΠΟΛΥΠΛΟΚΟΤΗΤΑΣ ΤΟΥ ΔΕΝΔΡΟΥ ΜΕ ΤΗΝ ΑΠΟΚΟΠΗ ΚΛΑΔΩΝ ΟΙ ΟΠΟΙΟΙ ΕΛΑΧΙΣΤΑ ΕΠΗΡΕΑΖΟΥΝ ΤΗΝ ΑΚΡΙΒΕΙΑ ΤΟΥ ΤΕΛΙΚΟΥ ΑΠΟΤΕΛΕΣΜΑΤΟΣ.

ΚΟΜΒΙΚΑ ΣΗΜΕΙΑ ΠΟΥ ΓΕΝΙΚΑ ΚΑΛΟΥΝΤΑΙ ΝΑ ΑΝΤΙΜΕΤΩΠΙΣΟΥΝ ΟΙ ΑΛΓΟΡΙΘΜΟΙ ΤΩΝ ΔΕΝΔΡΩΝ ΑΠΟΦΑΣΗΣ ΕΙΝΑΙ:

- Η ΕΠΙΛΟΓΗ ΤΩΝ ΓΝΩΡΙΣΜΑΤΩΝ ΟΠΟΥ ΘΑ ΓΙΝΕΤΑΙ Η ΔΙΑΣΠΑΣΗ ΤΩΝ ΔΕΔΟΜΕΝΩΝ ΚΑΘΩΣ ΕΠΙΔΙΩΚΕΤΑΙ ΝΑ ΕΠΙΛΕΓΟΝΤΑΙ ΣΕ ΚΑΘΕ ΠΕΡΙΠΤΩΣΗ ΤΑ ΠΕΡΙΣΣΟΤΕΡΟ ΣΗΜΑΝΤΙΚΑ ΓΝΩΡΙΣΜΑΤΑ.
- Η ΣΕΙΡΑ ΜΕ ΤΗΝ ΟΠΟΙΑ ΘΑ ΓΙΝΟΝΤΑΙ ΟΙ ΔΙΑΣΠΑΣΕΙΣ: Η ΑΛΛΗΛΟΥΧΙΑ ΤΩΝ ΔΙΑΣΠΑΣΕΩΝ ΕΠΙΔΙΩΚΕΤΑΙ ΝΑ ΓΙΝΕΤΑΙ ΜΕ ΤΡΟΠΟ ΤΕΤΟΙΟ ΠΟΥ ΝΑ ΑΠΟΦΕΥΓΕΤΑΙ Ο ΕΠΑΝΕΛΕΓΧΟΣ ΓΝΩΡΙΣΜΑΤΩΝ.
- Ο ΤΡΟΠΟΣ ΑΝΤΙΜΕΤΩΠΙΣΗΣ ΤΩΝ ΓΝΩΡΙΣΜΑΤΩΝ ΤΩΝ ΟΠΟΙΩΝ ΟΙ ΤΙΜΕΣ ΕΙΝΑΙ ΣΥΝΕΧΗΣ ΚΑΙ ΟΧΙ ΔΙΑΚΡΙΤΕΣ ΚΑΙ ΠΟΥ ΘΑ ΠΡΕΠΕΙ ΝΑ ΧΡΗΣΙΜΟΠΟΙΗΘΟΥΝ ΣΤΙΣ ΔΙΑΣΠΑΣΕΙΣ.
- Η ΔΟΜΗ ΠΟΥ ΘΑ ΕΧΕΙ ΤΟ ΔΕΝΔΡΟ ΚΑΘΩΣ ΕΠΙΔΙΩΚΕΤΑΙ ΝΑ ΠΡΟΚΥΠΤΟΥΝ ΙΣΟΖΥΓΙΣΜΕΝΑ ΔΕΝΔΡΑ ΜΕ ΟΣΟ ΤΟ ΔΥΝΑΤΟΝ ΛΙΓΟΤΕΡΟ ΥΨΟΣ.
- ΠΟΙΕΣ ΘΑ ΕΙΝΑΙ ΟΙ ΣΥΝΘΗΚΕΣ ΠΟΥ ΘΑ ΕΠΙΒΑΛΛΟΥΝ ΤΟΝ ΤΕΡΜΑΤΙΣΜΟ ΤΟΥ ΑΛΓΟΡΙΘΜΟΥ.
- Η ΒΕΛΤΙΣΤΗ ΠΟΣΟΤΗΤΑ ΔΕΔΟΜΕΝΩΝ ΠΟΥ ΑΠΑΙΤΟΥΝΤΑΙ ΓΙΑ ΤΗΝ ΕΚΠΑΙΔΕΥΣΗ ΤΟΥ ΔΕΝΔΡΟΥ. ΠΟΛΛΑ ΔΕΔΟΜΕΝΑ ΑΥΞΑΝΟΥ ΤΗΝ ΠΟΛΥΠΛΟΚΟΤΗΤΑ ΤΗΣ ΕΚΠΑΙΔΕΥΣΗΣ

ΑΛΛΑ ΠΑΡΑΛΛΗΛΑ ΔΗΜΙΟΥΡΓΟΥΝ ΑΚΡΙΒΕΣΤΕΡΑ ΜΟΝΤΕΛΑ.

- ΤΟ ΑΝ ΠΡΑΓΜΑΤΟΠΟΙΕΙΤΑΙ ΚΛΑΔΕΜΑ ΤΟΥ ΔΕΝΔΡΟΥ ΚΑΙ ΜΕ ΠΟΙΟΝ ΤΡΟΠΟ. ΤΟ ΚΛΑΔΕΜΑ ΕΠΙΔΙΩΚΕΙ ΝΑ ΜΕΙΩΣΕΙ ΤΗΝ ΠΟΛΥΠΛΟΚΟΤΗΤΑ ΤΟΥ ΔΕΝΔΡΟΥ ΜΕ ΤΙΣ ΜΙΚΡΟΤΕΡΕΣ ΔΥΝΑΤΕΣ ΑΠΩΛΕΙΕΣ ΣΕ ΑΚΡΙΒΕΙΑ

ΣΤΙΣ ΕΠΟΜΕΝΕΣ ΠΑΡΑΓΡΑΦΟΥΣ ΠΑΡΟΥΣΙΑΖΟΝΤΑΙ ΟΙ ΑΛΓΟΡΙΘΜΟΙ ΑΥΤΟΙ.

ID3

Ο ΠΙΟ ΔΙΑΔΕΔΟΜΕΝΟΣ ΑΛΓΟΡΙΘΜΟΣ ΓΙΑ ΤΗΝ ΔΗΜΙΟΥΡΓΙΑ ΔΕΝΔΡΩΝ ΑΠΟΦΑΣΗΣ ΕΙΝΑΙ Ο ID3. ΕΙΝΑΙ ΕΠΙΣΗΣ ΓΝΩΣΤΟΣ ΚΑΙ ΩΣ ΑΛΓΟΡΙΘΜΟΣ ΚΑΤΑΣΚΕΥΗΣ ΔΕΝΔΡΩΝ ΑΠΟΦΑΣΗΣ ΜΕ ΕΠΑΓΩΓΗ (DECISION TREE INDUCTIONALGORITHM) ΑΠΟ ΔΕΔΟΜΕΝΑ ΕΚΠΑΙΔΕΥΣΗΣ. ΠΑΡΟΥΣΙΑΣΤΗΚΕ ΤΗΝ ΔΕΚΑΕΤΙΑ ΤΟΥ 1970 ΑΠΟ ΤΟΝ ROSSQUINLAN. Η ΚΕΝΤΡΙΚΗ ΦΙΛΟΣΟΦΙΑ ΤΟΥ ΑΛΓΟΡΙΘΜΟΥ ΕΙΝΑΙ ΝΑ ΕΠΙΛΕΓΕΙ ΠΑΡΑΜΕΤΡΟΥΣ ΤΕΤΟΙΟΥΣ ΟΙ ΟΠΟΙΟΙ ΝΑ ΠΡΟΚΑΛΟΥΝ ΔΙΑΚΛΑΔΩΣΕΙΣ ΣΤΟ ΔΕΝΔΡΟ ΟΙ ΟΠΟΙΕΣ ΝΑ ΕΚΛΥΟΥΝ ΜΕΓΑΛΗ ΠΟΣΟΤΗΤΑ ΠΛΗΡΟΦΟΡΙΑΣ ΜΕ ΣΚΟΠΟ ΑΥΤΕΣ ΝΑ ΕΛΑΧΙΣΤΟΠΟΙΟΥΝΤΑΙ. ΠΡΟΚΕΙΜΕΝΟΥ ΝΑ ΕΚΤΙΜΗΘΕΙ ΤΟ ΠΟΣΟ ΠΛΗΡΟΦΟΡΙΑΣ ΠΟΥ ΕΚΛΥΕΤΑΙ ΣΕ ΚΑΘΕ ΔΙΑΚΛΑΔΩΣΗ ΧΡΗΣΙΜΟΠΟΙΕΙ ΤΙΣ ΕΝΝΟΙΕΣ ΤΗΣ ΕΝΤΡΟΠΙΑΣ ΚΑΙ ΤΟΥ ΚΕΡΔΟΥΣ ΠΛΗΡΟΦΟΡΙΑΣ. ΜΠΟΡΕΙ ΝΑ ΧΑΡΑΚΤΗΡΙΣΤΕΙ ΣΑΝ ΕΝΑΣ ΑΛΓΟΡΙΘΜΟΣ ΑΝΑΡΡΙΧΗΣΗΣ ΛΟΦΩΝ² ΚΑΘΩΣ ΣΕ ΚΑΘΕ ΚΥΚΛΟ ΤΗΣ ΕΞΕΛΙΞΗΣ ΤΟΥ ΕΠΕΚΤΕΙΝΕΙ ΤΟ ΤΡΕΧΟΝ ΔΕΝΔΡΟ ΜΕ ΤΟΝ ΤΟΠΙΚΑ ΚΑΛΥΤΕΡΟ ΤΡΟΠΟ ΚΑΙ ΣΥΝΕΧΙΖΕΙ ΧΩΡΙΣ ΔΥΝΑΤΟΤΗΤΑ ΟΠΙΣΘΟΔΡΟΜΗΣΗΣ.

²ΠΡΟΚΕΙΤΑΙ ΓΙΑ ΚΑΤΗΓΟΡΙΑ ΑΛΓΟΡΙΘΜΩΝ ΑΝΑΖΗΤΗΣΗΣ ΠΟΥ ΣΕ ΚΑΘΕ ΤΟΥΣ ΣΤΑΔΙΟ ΜΕΤΑΒΑΙΝΟΥΝ ΣΤΗΝ ΤΟΠΙΚΑ ΚΑΛΥΤΕΡΗ ΚΑΤΑΣΤΑΣΗ ΣΥΜΦΩΝΑ ΜΕ ΜΙΑ ΔΕΔΟΜΕΝΗ ΕΥΡΗΣΤΙΚΗ ΣΥΝΑΡΤΗΣΗ.

ΕΝΤΡΟΠΙΑ

Η ΕΝΤΡΟΠΙΑ ΕΙΝΑΙ ΜΙΑ ΣΥΝΑΡΤΗΣΗ ΤΟΥ ΠΛΗΘΟΥΣ ΤΩΝ ΠΙΘΑΝΩΝ ΚΑΤΑΣΤΑΣΕΩΝ ΕΝΟΣ ΣΥΣΤΗΜΑΤΟΣ ΚΑΙ ΕΝΑ ΦΥΣΙΚΟ ΜΕΓΕΘΟΣ ΓΙΑ ΤΗΝ ΑΠΟΤΙΜΗΣΗ ΤΗΣ ΒΕΒΑΙΟΤΗΤΑ ΕΝΟΣ ΣΥΣΤΗΜΑΤΟΣ. ΣΤΗΡΙΖΕΤΑΙ ΣΤΗΝ ΘΕΩΡΙΑ ΟΤΙ ΣΕ ΕΝΑ ΚΛΕΙΣΤΟ ΣΥΣΤΗΜΑ Η ΔΙΑΘΕΣΙΜΗ ΕΝΕΡΓΕΙΑ ΔΕΝ ΜΠΟΡΕΙ ΝΑ ΑΥΞΑΝΕΙ Η ΔΕΝ ΜΠΟΡΕΙ ΝΑ ΜΕΙΩΝΕΤΑΙ. Ο SHANNON ΜΟΝΤΕΛΟΠΟΙΗΣΕ ΤΗΝ ΠΛΗΡΟΦΟΡΙΑ ΣΑΝ ΜΙΑ ΣΕΙΡΑ ΑΠΟ ΓΕΓΟΝΟΤΑ ΠΟΥ ΣΥΜΒΑΙΝΟΥΝ ΜΕ ΣΥΓΚΕΚΡΙΜΕΝΕΣ ΠΙΘΑΝΟΤΗΤΕΣ. ΚΑΘΟΡΙΣΕ ΟΤΙ Η ΠΛΗΡΟΦΟΡΙΑ ΠΟΥ ΠΕΡΙΛΑΜΒΑΝΕΙ ΕΝΑ ΓΕΓΟΝΟΣ Χ ΕΞΑΡΤΑΤΑΙ ΜΟΝΟ ΑΠΟ ΤΗΝ ΠΙΘΑΝΟΤΗΤΑ ΤΟΥ ΝΑ ΣΥΜΒΕΙ, ΕΣΤΩ Ρ. Όσο μικρότερο είναι το Ρ τόσο περισσότερη πληροφορία περιέχεται στο ενδεχόμενο Χ. Η μεταβολή του Ρ προκαλεί μεταβολή στην πληροφορία I(P) του ενδεχομένου Χ και για δύο ανεξάρτητα ενδεχόμενα Χ και Υ με πιθανότητες P_X, P_Y αντίστοιχα, ισχύει:

$$I(P_X + P_Y) = I(P_X) + I(P_Y)$$

ΜΕ ΒΑΣΗ ΤΙΣ ΠΑΡΑΠΑΝΩ ΥΠΟΘΕΣΕΙΣ ΑΠΕΔΕΙΞΕ ΟΤΙ ΑΝ Χ ΕΙΝΑΙ ΜΙΑ ΤΥΧΑΙΑ

ΜΕΤΑΒΛΗΤΗ ΣΕ ΕΝΑ ΣΥΝΟΛΟ ΜΕΤΑΒΛΗΤΩΝ Χ ΠΟΥ ΤΟ ΚΑΘΕΝΑ ΣΥΜΒΑΙΝΕΙ ΜΕ ΠΙΘΑΝΟΤΗΤΑ P_X, ΤΟΤΕ Η ΕΝΤΡΟΠΙΑ SHANNON ΤΟΥ ΓΕΓΟΝΟΤΟΣ Χ

ΕΙΝΑΙ $-\log_2(P_X)$.

ΑΝ ΘΕΩΡΗΣΟΥΜΕ ΜΙΑ ΤΥΧΑΙΑ ΜΕΤΑΒΛΗΤΗ Χ ΠΟΥ ΛΑΜΒΑΝΕΙ ΤΙΜΕΣ ΑΠΟ ΕΝΑ ΠΕΠΕΡΑΣΜΕΝΟ ΠΛΗΘΟΣ X₁, X₂, X₃, ..., X_N ΜΕ ΠΙΘΑΝΟΤΗΤΑ P(X=X_i)=P_i, 0 ≤ P_i ≤ 1 ΓΙΑ ΚΑΘΕ I, 1 ≤ I ≤ N ΚΑΙ P₁+P₂+...+P_N=1, ΤΟΤΕ Η ΕΝΤΡΟΠΙΑ ΤΗΣ ΟΡΙΖΕΤΑΙ ΩΣ:

$$H(X) = -\sum_{i=1}^n p_i \lg p_i = \sum_{i=1}^n p_i \lg \frac{1}{p_i}$$

ΚΕΡΔΟΣ ΠΛΗΡΟΦΟΡΙΑΣ

ΤΟ ΚΕΡΔΟΣ ΠΛΗΡΟΦΟΡΙΑΣ ΕΙΝΑΙ ΤΟ ΜΕΓΕΘΟΣ ΤΟ ΟΠΟΙΟ ΔΕΙΧΝΕΙ ΚΑΤΑ ΠΟΣΟ ΘΑ ΜΕΙΩΘΕΙ Η ΕΝΤΡΟΠΙΑ ΤΟΥ ΣΥΝΟΛΟΥ ΤΩΝ ΔΕΔΟΜΕΝΩΝ ΕΚΠΑΙΔΕΥΣΗΣ S , ΜΕ ΕΝΤΡΟΠΙΑ H , ΑΝ ΓΙΑ ΤΗΝ ΔΙΑΣΠΑΣΗ ΤΟΥ ΧΡΗΣΙΜΟΠΟΙΗΘΕΙ ΤΟ ΧΑΡΑΚΤΗΡΙΣΤΙΚΟ A_i ΑΠΟ ΤΟ ΣΥΝΟΛΟ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ A ΜΕ

$$A = \{A_1, A_2, \dots, A_3\}$$

ΚΑΙ ΔΙΝΕΤΑΙ ΑΠΟ ΤΗΝ ΣΧΕΣΗ

$$G(S, A_i) = H - \sum_{u \in A_i} \frac{|S_u|}{|S|} H(S_u)$$

ΟΠΟΥ

S_u , ΤΟ ΣΥΝΟΛΟ ΤΩΝ ΔΕΔΟΜΕΝΩΝ ΜΕ ΤΙΜΗ u ΣΤΟ ΧΑΡΑΚΤΗΡΙΣΤΙΚΟ A_i

$H(S_u)$, Η ΕΝΤΡΟΠΙΑ ΩΣ ΠΡΟΣ ΤΗΝ ΤΙΜΗ u ΤΟΥ A_i

ΜΕΙΩΣΗ ΤΗΣ ΕΝΤΡΟΠΙΑΣ ΣΗΜΑΙΝΕΙ ΕΚΛΥΣΗ ΜΕΓΑΛΥΤΕΡΟΥ ΠΟΣΟΥ ΠΛΗΡΟΦΟΡΙΑΣ. Ο ΠΡΩΤΟΣ ΟΡΟΣ ΤΟΥ ΔΕΥΤΕΡΟΥ ΜΕΛΟΥΣ ΤΗΣ ΣΧΕΣΗΣ ΠΕΡΙΓΡΑΦΕΙ ΤΗΝ ΕΝΤΡΟΠΙΑ ΤΟΥ ΣΥΝΟΛΟΥ ΔΕΔΟΜΕΝΩΝ ΕΝΩ Ο ΔΕΥΤΕΡΟΣ ΠΕΡΙΓΡΑΦΕΙ ΤΗΝ ΕΝΤΡΟΠΙΑ ΤΩΝ ΔΕΔΟΜΕΝΩΝ ΜΕΤΑ ΤΗ ΔΙΑΣΠΑΣΗ ΤΟΥΣ ΣΥΜΦΩΝΑ ΜΕ ΤΗΝ ΤΙΜΗ ΤΟΥ ΧΑΡΑΚΤΗΡΙΣΤΙΚΟΥ S . Ο ΔΕΥΤΕΡΟΣ ΟΡΟΣ ΑΠΟΤΕΛΕΙΤΑΙ ΑΠΟ ΤΟ ΑΘΡΟΙΣΜΑ ΤΗΣ ΕΝΤΡΟΠΙΑΣ ΓΙΑ ΤΟ ΚΑΘΕ ΣΥΝΟΛΟ ΠΟΥ ΠΡΟΚΥΠΤΕΙ ΜΕΤΑ ΤΗ ΔΙΑΣΠΑΣΗ.

Η ΚΕΝΤΡΙΚΗ ΙΔΕΑ ΠΟΥ ΔΙΕΠΕΙ ΤΗΝ ΕΞΕΛΙΞΗ ΤΟΥ ΑΛΓΟΡΙΘΜΟΥ ID3 ΕΙΝΑΙ:

- ΥΠΟΛΟΓΙΣΜΟΣ ΤΗΣ ΕΝΤΡΟΠΙΑΣ ΟΛΩΝ ΤΩΝ ΑΧΡΗΣΙΜΟΠΟΙΗΤΩΝ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ ΣΕ ΣΧΕΣΗ ΜΕ ΤΑ ΔΕΙΓΜΑΤΑ.
- ΕΠΙΛΟΓΗ ΤΟΥ ΧΑΡΑΚΤΗΡΙΣΤΙΚΟΥ ΑΥΤΟΥ ΠΟΥ ΠΑΡΟΥΣΙΑΖΕΙ ΤΗΝ ΜΙΚΡΟΤΕΡΗ ΕΝΤΡΟΠΙΑ Η ΤΟ ΜΕΓΑΛΥΤΕΡΟ ΚΕΡΔΟΣ ΠΛΗΡΟΦΟΡΙΑΣ ΚΑΙ ΔΗΜΙΟΥΡΓΙΑ ΚΟΜΒΟΥ ΓΙΑ ΤΟ ΧΑΡΑΚΤΗΡΙΣΤΙΚΟ ΑΥΤΟ.

ΤΑ ΒΗΜΑΤΑ ΤΟΥ ΑΛΓΟΡΙΘΜΟΥ ΕΙΝΑΙ ΤΑ ΕΞΗΣ:

1. ΕΠΙΛΕΓΕΤΑΙ ΤΟ ΧΑΡΑΚΤΗΡΙΣΤΙΚΟ ΤΟ ΟΠΟΙΟ ΠΑΡΟΥΣΙΑΖΕΙ ΤΗΝ ΜΕΓΑΛΥΤΕΡΗ ΑΠΟΔΟΣΗ ΣΕ ΚΕΡΔΟΣ ΠΛΗΡΟΦΟΡΙΑΣ ΣΤΟ ΣΥΝΟΛΟ ΤΩΝ ΔΕΙΓΜΑΤΩΝ. ΤΟ ΧΑΡΑΚΤΗΡΙΣΤΙΚΟ ΑΥΤΟ ΤΟΠΟΘΕΤΕΙΤΑΙ ΣΤΟΝ ΚΟΜΒΟ ΡΙΖΑ ΤΟΥ ΔΕΝΔΡΟΥ.
2. ΓΙΑ ΚΑΘΕ ΤΙΜΗ ΤΟΥ ΕΠΙΛΕΓΕΝΤΟΣ ΧΑΡΑΚΤΗΡΙΣΤΙΚΟΥ ΔΗΜΙΟΥΡΓΟΥΝΤΑΙ ΠΑΙΔΙΑ ΤΙΣ ΡΙΖΑΣ.
3. ΤΑ ΔΕΔΟΜΕΝΑ ΚΑΤΑΝΕΜΟΝΤΑΙ ΣΤΟΥΣ ΚΟΜΒΟΥΣ ΠΑΙΔΙΑ ΤΗΣ ΡΙΖΑΣ ΑΝΑΛΟΓΑ ΜΕ ΤΗΝ ΤΙΜΗ ΤΟΥ ΧΑΡΑΚΤΗΡΙΣΤΙΚΟΥ ΠΟΥ ΧΡΗΣΙΜΟΠΟΙΗΘΗΚΕ ΣΤΗΝ ΡΙΖΑ.
4. ΓΙΑ ΚΑΘΕ ΚΟΜΒΟ ΕΠΑΝΑΛΑΜΒΑΝΟΝΤΑΙ ΤΑ ΒΗΜΑΤΑ 2 ΚΑΙ 3 ΜΕ ΤΗΝ ΕΠΙΛΟΓΗ ΤΟΥ ΧΑΡΑΚΤΗΡΙΣΤΙΚΟΥ ΝΑ ΕΠΙΛΕΓΕΤΑΙ ΑΠΟ ΤΑ ΔΕΔΟΜΕΝΑ ΠΟΥ ΥΠΑΡΧΟΥΝ ΣΤΟΝ ΚΑΘΕ ΚΟΜΒΟ.
5. ΈΝΑΣ ΚΟΜΒΟΣ ΓΙΝΕΤΑΙ ΦΥΛΛΟ ΟΤΑΝ ΟΛΑ ΤΑ ΔΕΔΟΜΕΝΑ ΠΟΥ ΑΝΗΚΟΥΝ ΣΕ ΑΥΤΟΝ ΑΝΗΚΟΥΝ ΣΤΗΝ ΙΔΙΑ ΚΑΤΗΓΟΡΙΑ. Η ΚΑΤΗΓΟΡΙΑ ΑΥΤΗ ΓΙΝΕΤΑΙ ΚΑΙ Η ΤΙΜΗ ΤΟΥ ΦΥΛΛΟΥ.
6. ΑΝ ΣΕ ΚΑΠΟΙΟ ΒΑΘΟΣ ΤΕΛΕΙΩΣΟΥΝ ΤΑ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΠΡΟΣ ΕΛΕΓΧΟ, ΤΟΤΕ Ο ΚΟΜΒΟΣ ΓΙΝΕΤΑΙ ΤΕΡΜΑΤΙΚΟΣ ΚΑΙ ΣΑΝ ΤΙΜΗ ΠΑΙΡΝΕΙ ΕΚΕΙΝΗ ΠΟΥ ΕΧΕΙ ΤΗΝ ΠΛΕΙΟΨΗΦΙΑ ΜΕ ΒΑΣΗ ΤΑ ΔΕΔΟΜΕΝΑ ΤΟΥ ΚΟΜΒΟΥ ΑΥΤΟΥ.

Ο ΑΛΓΟΡΙΘΜΟΣ ΣΕ ΨΕΥΔΟΓΛΩΣΣΑ ΕΧΕΙ ΟΠΩΣ ΑΚΟΛΟΥΘΩΣ:

```

MAKEDECISIONTREEID3(NODE N, SAMPLES S,
ATTRIBUTES A){
    C = Η ΚΑΤΗΓΟΡΙΑ ΠΟΥ ΕΜΦΑΝΙΖΕΤΑΙ ΣΥΧΝΟΤΕΡΑ
ΣΤΟ S
    N = C

```



```

IF (ΔΕΝ ΥΠΑΡΧΕΙ C' ΣΤΟ S ΜΕ C'!=C){
    EXIT;
}
ELSE {
    ΑΤΟΥΑ ΜΕΓΑΙΝΙΝFORMATION(A)=MAX
    N = A
    FOREACHVITΟΥΑ {
        SI = { ΣΤΟΥ TS: S.A = VI}
        IF (SI != EMPTY){
            N' = N(VI)
            A' = A - {A}
            ΜΑΚΕΔΕCΙSSIONTREE(N' ,SI ,A' )
        }
    }
    ENDIF
    FOR-END
    ENDIF
END

```

ΠΡΟΚΕΙΜΕΝΟΥ ΝΑ ΓΙΝΕΙ ΠΕΡΙΣΣΟΤΕΡΟ ΚΑΤΑΝΟΗΤΟΣ Ο ΑΛΓΟΡΙΘΜΟΣ ΠΑΡΑΤΙΘΕΤΑΙ ΕΝΑ ΠΑΡΑΔΕΙΓΜΑ. ΈΣΤΩ ΕΝΑ ΣΥΝΟΛΟ 15 ΜΑΘΗΤΩΝ. ΓΙΑ ΤΟ ΣΥΝΟΛΟ ΑΥΤΟ ΓΙΝΕΤΑΙ ΠΡΟΣΠΑΘΕΙΑ ΝΑ ΔΙΑΠΙΣΤΩΘΕΙ ΚΑΤΑ ΠΟΣΟ ΤΟ ΦΥΛΛΟ ΚΑΙ Η ΕΝΑΣΧΟΛΗΣΗ ΤΟΥΣ ΜΕ ΤΟΝ ΑΘΛΗΤΙΣΜΟ ΕΠΙΔΡΑ ΣΤΙΣ ΕΠΙΔΟΣΕΙΣ ΤΟΥΣ. ΣΤΟ ΠΑΡΑΚΑΤΩ ΠΙΝΑΚΑ ΠΑΡΑΤΙΘΕΝΤΑΙ ΤΟ ΠΟΙΟΙ ΑΣΧΟΛΟΥΝΤΑΙ ΜΕ ΟΜΑΔΙΚΟ ΑΘΛΗΜΑ, ΠΟΙΟΙ ΜΕ ΑΤΟΜΙΚΟ ΚΑΙ ΠΟΙΟΙ ΔΕΝ ΑΣΧΟΛΟΥΝΤΑΙ ΜΕ ΚΑΠΟΙΟ ΑΘΛΗΜΑ. ΦΑΙΝΕΤΑΙ ΕΠΙΣΗΣ ΤΟ ΦΥΛΛΟ ΤΟΥΣ ΚΑΙ Ο ΜΕΣΟΣ ΟΡΟΣ ΤΩΝ ΕΠΙΔΟΣΕΩΝ ΣΤΑ ΜΑΘΗΜΑΤΑ.

ΜΑΘΗΤΗΣ	ΦΥΛΛΟ	ΑΘΛΗΜΑ	ΕΠΙΔΟΣΗ
1	ΑΓΟΡΙ	ΑΤΟΜΙΚΟ	ΑΡΙΣΤΑ
2	ΚΟΡΙΤΣΙ	ΑΤΟΜΙΚΟ	ΑΡΙΣΤΑ
3	ΑΓΟΡΙ	ΑΤΟΜΙΚΟ	ΚΑΛΑ
4	ΑΓΟΡΙ	ΑΤΟΜΙΚΟ	ΜΕΤΡΙΑ
5	ΑΓΟΡΙ	ΟΜΑΔΙΚΟ	ΑΡΙΣΤΑ
6	ΚΟΡΙΤΣΙ	ΟΜΑΔΙΚΟ	ΑΡΙΣΤΑ
7	ΚΟΡΙΤΣΙ	ΟΜΑΔΙΚΟ	ΚΑΛΑ

8	ΚΟΡΙΤΣΙ	ΚΑΝΕΝΑ	ΑΡΙΣΤΑ
9	ΑΓΟΡΙ	ΚΑΝΕΝΑ	ΑΡΙΣΤΑ
10	ΑΓΟΡΙ	ΚΑΝΕΝΑ	ΚΑΛΑ
11	ΑΓΟΡΙ	ΚΑΝΕΝΑ	ΑΡΙΣΤΑ
12	ΚΟΡΙΤΣΙ	ΚΑΝΕΝΑ	ΑΡΙΣΤΑ
13	ΚΟΡΙΤΣΙ	ΚΑΝΕΝΑ	ΚΑΛΑ
14	ΑΓΟΡΙ	ΚΑΝΕΝΑ	ΜΕΤΡΙΑ
15	ΑΓΟΡΙ	ΚΑΝΕΝΑ	ΜΕΤΡΙΑ

ΤΑ ΔΕΔΟΜΕΝΑ ΠΟΥ ΕΙΝΑΙ ΚΑΤΑΓΕΓΡΑΜΜΕΝΑ ΣΤΟ ΠΑΡΑΠΑΝΩ ΠΙΝΑΚΑ ΦΑΙΝΕΤΑΙ ΟΤΙ ΑΠΟ ΤΟΥΣ 15 ΜΑΘΗΤΕΣ ΟΙ 8 ΕΧΟΥΝ ΑΡΙΣΤΗ ΕΠΙΔΟΣΗ, ΟΙ 4 ΚΑΛΗ ΕΠΙΔΟΣΗ ΚΑΙ ΟΙ ΥΠΟΛΟΙΠΟΙ 3 ΕΧΟΥΝ ΜΕΤΡΙΑ ΕΠΙΔΟΣΗ. ΚΑΤΑ ΣΥΝΕΠΕΙΑ ΕΦΑΡΜΟΖΟΝΤΑΣ ΤΟΝ ΟΡΙΣΜΟ ΤΗΣ ΕΝΤΡΟΠΙΑΣ ΤΟΥ ΑΡΧΙΚΟΥ ΣΥΝΟΛΟΥ ΠΡΟΚΥΠΤΕΙ ΟΤΙ ΑΥΤΗ ΕΙΝΑΙ:

$$0,533 \cdot \text{LOG}(1,875) + 0,267 \cdot \text{LOG}(3,75) + 0,2 \cdot 5 = 0.4384$$

ΠΑΡΑΤΗΡΟΥΜΕ ΟΤΙ Η ΤΙΜΗ ΑΓΟΡΙ ΕΜΦΑΝΙΖΕΤΑΙ ΠΕΡΙΣΣΟΤΕΡΕΣ ΦΟΡΕΣ ΑΠΟ ΚΑΘΕ ΑΛΛΗ. ΈΤΣΙ ΕΠΙΛΕΓΕΤΑΙ ΤΟ ΧΑΡΑΚΤΗΡΙΣΤΙΚΟ ΦΥΛΛΟ ΩΣ ΤΟ ΧΑΡΑΚΤΗΡΙΣΤΙΚΟ ΤΗΣ ΠΡΩΤΗΣ ΔΙΑΣΠΑΣΗΣ.

ΜΑΘΗΤΗΣ	ΦΥΛΛΟ	ΆΘΛΗΜΑ	ΕΠΙΔΟΣΗ
2	ΚΟΡΙΤΣΙ	ΑΤΟΜΙΚΟ	ΑΡΙΣΤΑ
6	ΚΟΡΙΤΣΙ	ΟΜΑΔΙΚΟ	ΑΡΙΣΤΑ
7	ΚΟΡΙΤΣΙ	ΟΜΑΔΙΚΟ	ΚΑΛΑ
8	ΚΟΡΙΤΣΙ	ΚΑΝΕΝΑ	ΑΡΙΣΤΑ
12	ΚΟΡΙΤΣΙ	ΚΑΝΕΝΑ	ΑΡΙΣΤΑ
13	ΚΟΡΙΤΣΙ	ΚΑΝΕΝΑ	ΚΑΛΑ

ΜΑΘΗΤΗΣ	ΦΥΛΛΟ	ΆΘΛΗΜΑ	ΕΠΙΔΟΣΗ
1	ΑΓΟΡΙ	ΑΤΟΜΙΚΟ	ΑΡΙΣΤΑ
3	ΑΓΟΡΙ	ΑΤΟΜΙΚΟ	ΚΑΛΑ
4	ΑΓΟΡΙ	ΑΤΟΜΙΚΟ	ΜΕΤΡΙΑ
5	ΑΓΟΡΙ	ΟΜΑΔΙΚΟ	ΑΡΙΣΤΑ
9	ΑΓΟΡΙ	ΚΑΝΕΝΑ	ΑΡΙΣΤΑ
10	ΑΓΟΡΙ	ΚΑΝΕΝΑ	ΚΑΛΑ
11	ΑΓΟΡΙ	ΚΑΝΕΝΑ	ΑΡΙΣΤΑ

14	ΑΓΟΡΙ	ΚΑΝΕΝΑ	ΜΕΤΡΙΑ
15	ΑΓΟΡΙ	ΚΑΝΕΝΑ	ΜΕΤΡΙΑ

Η ΕΝΤΡΟΠΙΑ ΤΟΥ ΣΥΝΟΛΟΥ ΠΟΥ ΠΕΡΙΛΑΜΒΑΝΕΙ ΤΙΣ ΕΓΓΡΑΦΕΣ ΤΩΝ ΑΓΟΡΙΩΝ ΕΙΝΑΙ:

$$0,444 \text{ LOG}(2,25) + 0,222 \text{ LOG}(4,5) + 0,333 \text{ LOG}(3) = 0,52$$

Η ΕΝΤΡΟΠΙΑ ΤΟΥ ΣΥΝΟΛΟΥ ΠΟΥ ΠΕΡΙΛΑΜΒΑΝΕΙ ΤΙΣ ΕΓΓΡΑΦΕΣ ΤΩΝ ΚΟΡΙΤΣΙΩΝ ΕΙΝΑΙ:

$$0,667 \text{ LOG}(1,5) + 0,333 \text{ LOG}(3) = 0,27$$

Ο ID3 ΠΡΕΠΕΙ ΝΑ ΚΑΘΟΡΙΣΕΙ ΠΟΙΟ ΕΙΝΑΙ ΤΟ ΚΕΡΔΟΣ ΠΛΗΡΟΦΟΡΙΑΣ ΧΡΗΣΙΜΟΠΟΙΩΝΤΑΣ ΑΥΤΗ ΤΗΝ ΔΙΑΣΠΑΣΗ. ΓΙΑ ΝΑ ΓΙΝΕΙ ΑΥΤΟ ΥΠΟΛΟΓΙΖΕΙ ΤΟ ΣΤΑΘΜΙΣΜΕΝΟ ΑΘΡΟΙΣΜΑ ΤΩΝ 2 ΠΑΡΑΠΑΝΩ ΕΝΤΡΟΠΙΩΝ.

$$(0,6 \cdot 0,52) + (0,4 \cdot 0,27) = 1,418$$

ΑΡΑ ΤΟ ΚΕΡΔΟΣ ΠΛΗΡΟΦΟΡΙΑΣ ΧΡΗΣΙΜΟΠΟΙΩΝΤΑΣ ΤΟ ΧΑΡΑΚΤΗΡΙΣΤΙΚΟ «ΦΥΛΟ» ΕΙΝΑΙ

$$0.4384 - 0.42 = 0.0184$$

ΕΚΤΕΛΕΙΤΑΙ ΤΗΝ ΙΔΙΑ ΔΙΑΔΙΚΑΣΙΑ ΓΙΑ ΤΟ ΧΑΡΑΚΤΗΡΙΣΤΙΚΟ ΑΘΛΗΜΑ. ΠΑΡΑΤΗΡΕΙΤΑΙ ΟΤΙ ΥΠΑΡΧΟΥΝ 4 ΜΑΘΗΤΕΣ ΠΟΥ ΑΚΟΛΟΥΘΟΥΝ ΟΜΑΔΙΚΟ ΑΘΛΗΜΑ, 3 ΠΟΥ ΑΚΟΛΟΥΘΟΥΝ ΑΤΟΜΙΚΟ ΚΑΙ 8 ΠΟΥ ΔΕΝ ΑΘΛΟΥΝΤΑΙ. ΑΚΟΛΟΥΘΩΝΤΑΣ ΠΑΡΟΜΟΙΑ ΜΕ ΤΗΝ ΠΑΡΑΠΑΝΩ ΔΙΑΔΙΚΑΣΙΑ ΠΡΟΚΥΠΤΕΙ ΟΤΙ Η ΕΝΤΡΟΠΙΑ ΤΟΥ ΑΡΧΙΚΟΥ ΣΥΝΟΛΟΥ ΕΙΝΑΙ:

$$E=0,4384$$

ΤΟ ΑΡΧΙΚΟ ΣΥΝΟΛΟ ΔΙΑΜΕΡΙΖΕΤΑΙ ΟΠΩΣ ΠΑΡΑΚΑΤΩ

1	ΑΓΟΡΙ	ΑΤΟΜΙΚΟ	ΑΡΙΣΤΑ
2	ΚΟΡΙΤΣΙ	ΑΤΟΜΙΚΟ	ΑΡΙΣΤΑ
3	ΑΓΟΡΙ	ΑΤΟΜΙΚΟ	ΚΑΛΑ
4	ΑΓΟΡΙ	ΑΤΟΜΙΚΟ	ΜΕΤΡΙΑ

5	ΑΓΟΡΙ	ΟΜΑΔΙΚΟ	ΑΡΙΣΤΑ
6	ΚΟΡΙΤΣΙ	ΟΜΑΔΙΚΟ	ΑΡΙΣΤΑ
7	ΚΟΡΙΤΣΙ	ΟΜΑΔΙΚΟ	ΚΑΛΑ

8	ΚΟΡΙΤΣΙ	ΚΑΝΕΝΑ	ΑΡΙΣΤΑ
9	ΑΓΟΡΙ	ΚΑΝΕΝΑ	ΑΡΙΣΤΑ
10	ΑΓΟΡΙ	ΚΑΝΕΝΑ	ΚΑΛΑ
11	ΑΓΟΡΙ	ΚΑΝΕΝΑ	ΑΡΙΣΤΑ
12	ΚΟΡΙΤΣΙ	ΚΑΝΕΝΑ	ΑΡΙΣΤΑ
13	ΚΟΡΙΤΣΙ	ΚΑΝΕΝΑ	ΚΑΛΑ
14	ΑΓΟΡΙ	ΚΑΝΕΝΑ	ΜΕΤΡΙΑ
15	ΑΓΟΡΙ	ΚΑΝΕΝΑ	ΜΕΤΡΙΑ

Η ΕΝΤΡΟΠΙΑ ΓΙΑ ΚΑΘΕ ΟΜΑΔΑ ΑΠΟ ΤΟ ΔΕΙΓΜΑ ΥΠΟΛΟΓΙΖΕΤΑΙ ΜΕ ΠΑΡΟΜΟΙΟ ΤΡΟΠΟ ΜΕ ΤΗΝ ΠΡΩΤΗ ΔΙΑΣΠΑΣΗ ΚΑΙ ΥΠΟΛΟΓΙΖΕΤΑΙ ΟΤΙ ΕΙΝΑΙ:

$$E_1=0,45$$

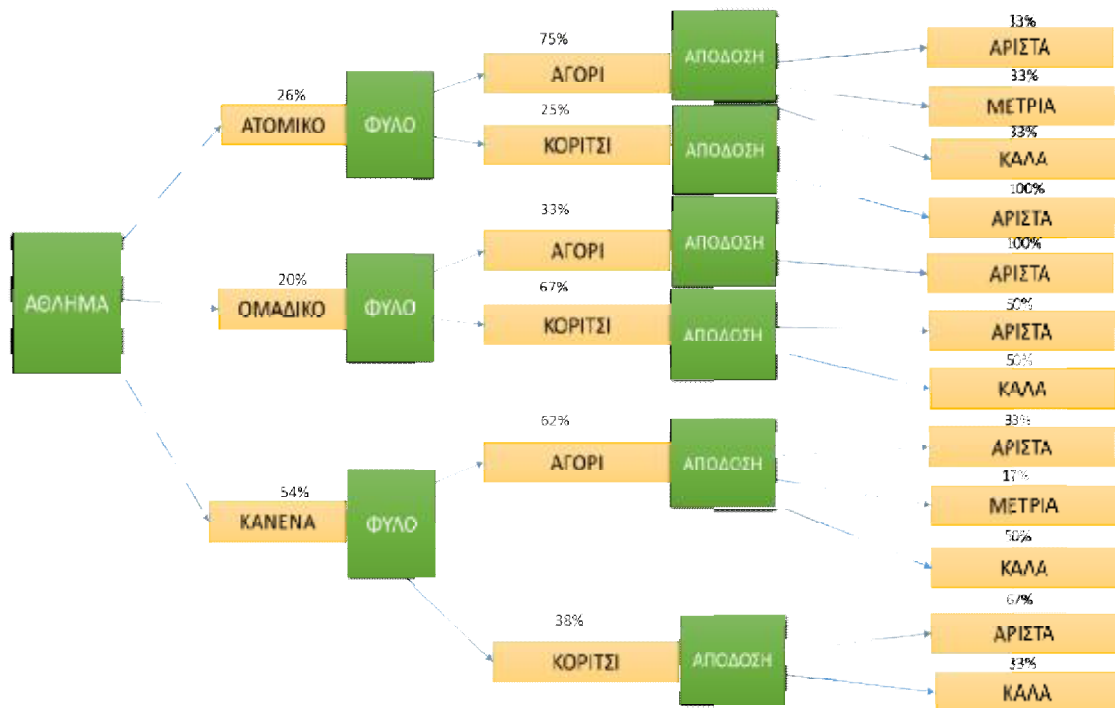
$$E_2=0,27$$

$$E_3=0,45$$

ΥΠΟΛΟΓΙΖΟΝΤΑΣ ΣΤΑΘΜΙΣΜΕΝΟ ΑΘΡΟΙΣΜΑ ΤΩΝ 3 ΠΑΡΑΠΑΝΩ ΕΝΤΡΟΠΙΩΝ ΠΡΟΚΥΠΤΕΙ

$$E' = 0,4165$$

ΈΤΣΙ ΤΟ ΚΕΡΔΟΣ ΠΛΗΡΟΦΟΡΙΑΣ ΑΠΟ ΤΗΝ ΔΙΑΣΠΑΣΗ ΑΥΤΗ ΘΑ ΕΙΝΑΙ: 0,021 ΠΟΥ ΕΙΝΑΙ ΜΕΓΑΛΥΤΕΡΟ ΑΠΟ ΤΗΝ ΠΡΟΗΓΟΥΜΕΝΗ ΔΙΑΣΠΑΣΗ. Ο ΠΡΩΤΟΣ ΚΟΜΒΟΣ ΤΟΥ ΔΕΝΔΡΟΥ ΘΑ ΑΠΑΝΤΑ ΣΤΗΝ ΕΡΩΤΗΣΗ «ΤΙ ΑΘΛΗΜΑ ΚΑΝΕΙ Ο ΜΑΘΗΤΗΣ». ΣΤΗΝ ΣΥΝΕΧΕΙΣ Η ΕΡΩΤΗΣΗ ΘΑ ΑΦΟΡΑ ΤΟ ΦΥΛΛΟ ΤΟΥ ΜΑΘΗΤΗ. ΤΟ ΔΕΝΔΡΟ ΑΥΤΟ ΘΑ ΕΧΕΙ ΤΗΝ ΑΚΟΛΟΥΘΗ ΜΟΡΦΗ



ΒΑΣΙΚΟ ΜΕΙΟΝΕΚΤΗΜΑ ΤΟΥ ΑΛΓΟΡΙΘΜΟΥ ΕΙΝΑΙ ΤΟ ΓΕΓΟΝΟΣ ΟΤΙ ΔΕΝ ΜΠΟΡΕΙ ΝΑ ΧΕΙΡΙΣΤΕΙ ΣΥΝΕΧΗΣ ΤΙΜΕΣ. ΣΤΙΣ ΠΕΡΙΠΤΩΣΕΙΣ ΣΥΝΕΧΩΝ ΤΙΜΩΝ ΧΡΕΙΑΖΕΤΑΙ ΝΑ ΟΡΙΣΤΕΙ ΜΗΧΑΝΙΣΜΟΣ ΚΑΤΗΓΟΡΙΟΠΟΙΗΣΗΣ ΤΩΝ ΤΙΜΩΝ ΤΩΝ ΠΑΡΑΜΕΤΡΩΝ, ΚΑΤΙ ΠΟΥ ΩΣΤΟΣΟ ΠΕΡΙΛΑΜΒΑΝΕΙ ΜΕΓΑΛΟ ΠΟΣΟΣΤΟ ΥΠΟΚΕΙΜΕΝΙΚΟΤΗΤΑΣ.

C4.5

Ο ΑΛΓΟΡΙΘΜΟΣ C4.5 ΑΠΟΤΕΛΕΙ ΜΙΑ ΕΞΕΛΙΞΗ ΤΟΥ ID3. ΑΚΟΛΟΥΘΕΙ ΤΗΝ ΓΕΝΙΚΗ ΤΟΥ ΦΙΛΟΣΟΦΙΑ ΑΛΛΑ ΕΝΣΩΜΑΤΩΝΕΙ ΑΠΟΔΟΤΙΚΟΤΕΡΟΥΣ ΜΗΧΑΝΙΣΜΟΥΣ ΣΤΗΝ ΔΙΑΧΕΙΡΙΣΗ ΤΟΥ ΔΕΙΓΜΑΤΟΣ. ΑΠΟΤΕΛΕΣΕ ΑΠΟΤΕΛΕΣΜΑ ΤΗΣ ΜΕΛΕΤΗΣ ΤΟΥ ROSSQUINLAN ΓΙΑ ΤΗΝ ΒΕΛΤΙΩΣΗΣ ΤΟΥ ID3 ΚΑΙ ΠΑΡΟΥΣΙΑΣΤΗΚΕ ΤΟ 1993. ΔΙΑΦΕΡΕΙ ΑΠΟ ΤΟΝ ID3 ΩΣ ΠΡΟΣ ΤΑ ΠΑΡΑΚΑΤΩ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ:

- ΜΠΟΡΕΙ ΝΑ ΧΕΙΡΙΖΕΤΑΙ ΠΑΡΑΜΕΤΡΟΥΣ ΠΟΥ ΛΑΜΒΑΝΟΥΝ ΣΥΝΕΧΕΙΣ ΤΙΜΕΣ. ΓΙΑ ΝΑ ΕΠΙΤΕΥΧΘΕΙ ΑΥΤΟ ΔΙΑΙΡΕΙ ΤΟ ΕΥΡΟΣ ΤΩΝ ΣΥΝΕΧΩΝ ΤΙΜΩΝ ΜΕ ΒΑΣΗ ΕΝΑ Η ΠΕΡΙΣΣΟΤΕΡΑ ΚΑΤΩΦΛΙΑ ΚΑΙ ΔΙΑΜΕΡΙΖΕΙ ΤΑ ΔΕΙΓΜΑΤΑ ΣΤΑ ΔΙΑΣΤΗΜΑΤΑ ΤΙΜΩΝ ΠΟΥ ΔΗΜΙΟΥΡΓΟΥΝΤΑΙ.

- ΜΠΟΡΕΙ ΝΑ ΧΕΙΡΙΖΕΤΑΙ ΑΓΝΩΣΤΕΣ ΤΙΜΕΣ. ΟΙ ΤΙΜΕΣ ΑΥΤΕΣ ΔΕΝ ΠΕΡΙΛΑΜΒΑΝΟΝΤΑΙ ΣΤΟΥΣ ΥΠΟΛΟΓΙΣΜΟΥΣ ΤΗΣ ΕΝΤΡΟΠΙΑΣ ΚΑΙ ΤΗΣ ΠΟΣΟΤΗΤΑΣ ΠΛΗΡΟΦΟΡΙΑΣ. ΓΙΑ ΝΑ ΚΑΤΗΓΟΡΙΟΠΟΙΗΘΕΙ ΜΙΑ ΑΓΝΩΣΤΗ ΤΙΜΗ ΘΑ ΠΡΕΠΕΙ ΑΥΤΗ ΝΑ ΠΡΟΒΛΕΦΘΕΙ ΣΕ ΣΧΕΣΗ ΜΕ ΤΙΣ ΥΠΟΛΟΙΠΕΣ ΤΙΜΕΣ ΤΗ ΙΔΙΑΣ ΠΑΡΑΜΕΤΡΟΥ.
- ΜΠΟΡΕΙ ΝΑ ΧΕΙΡΙΖΕΤΑΙ ΠΑΡΑΜΕΤΡΟΥΣ ΠΟΥ ΣΥΜΜΕΤΕΧΟΥΝ ΣΤΟΥΣ ΥΠΟΛΟΓΙΣΜΟΥΣ ΜΕ ΔΙΑΦΟΡΕΤΙΚΑ ΒΑΡΗ (ΚΟΣΤΗ).
- ΠΡΑΓΜΑΤΟΠΟΙΕΙ ΔΙΑΔΙΚΑΣΙΑ ΚΛΑΔΕΜΑΤΟΣ ΜΕ ΜΙΑ ΑΝΤΙΣΤΡΟΦΗ ΔΙΑΔΙΚΑΣΙΑ ΑΦΑΙΡΕΣΗΣ ΤΩΝ ΚΛΑΔΩΝ ΠΟΥ ΔΕΝ ΠΕΡΙΕΧΟΥΝ ΠΛΗΡΟΦΟΡΙΑ. ΤΟ ΚΛΑΔΕΜΑ ΤΟΥ ΔΕΝΔΡΟΥ ΜΠΟΡΕΙ ΝΑ ΓΙΝΕΙ ΜΕ:
 - ΑΝΤΙΚΑΤΑΣΤΑΣΗ ΥΠΟΔΕΝΔΡΟΥ: ΈΝΑ ΥΠΟΔΕΝΔΡΟ ΜΠΟΡΕΙ ΝΑ ΑΝΤΙΚΑΤΑΣΤΑΘΕΙ ΑΠΟ ΕΝΑ ΦΥΛΛΟ ΑΝ Η ΑΝΤΙΚΑΤΑΣΤΑΣΗ ΑΥΤΗ ΠΡΟΚΑΛΕΙ ΑΝΕΚΤΟ ΣΦΑΛΜΑ.
 - ΑΝΥΨΩΣΗ ΥΠΟΔΕΝΔΡΟΥ: ΑΝΤΙΚΑΘΙΣΤΑΤΑΙ ΕΝΑ ΥΠΟΔΕΝΔΡΟ ΜΕ ΤΟ ΠΙΟ ΠΙΘΑΝΑ ΧΡΗΣΙΜΟΠΟΙΟΥΜΕΝΟ ΥΠΟΔΕΝΔΡΟ ΤΟΥ.

ΚΑΤΑ ΒΑΣΗ Ο ΑΛΓΟΡΙΘΜΟΣ C4.5 ΕΞΕΛΙΣΣΕΤΑΙ ΟΠΩΣ Ο ID3. ΣΕ ΠΡΩΤΗ ΦΑΣΗ ΕΠΙΛΕΓΕΤΑΙ Η ΚΑΤΑΛΛΗΛΟΤΕΡΗ ΠΑΡΑΜΕΤΡΟΣ ΓΙΑ ΝΑ ΧΡΗΣΙΜΟΠΟΙΗΘΕΙ ΩΣ ΡΙΖΑ ΤΟΥ ΔΕΝΔΡΟΥ. ΣΤΗΝ ΣΥΝΕΧΕΙΑ ΚΑΘΕ ΔΙΑΦΟΡΕΤΙΚΗ ΤΙΜΗ ΤΗΣ ΠΑΡΑΜΕΤΡΟΥ ΑΥΤΗΣ ΑΝΤΙΣΤΟΙΧΙΖΕΤΑΙ ΣΕ ΕΝΑΝ ΔΙΑΦΟΡΕΤΙΚΟ ΚΟΜΒΟ ΠΟΥ ΕΙΝΑΙ ΠΑΙΔΙ ΤΗΣ ΡΙΖΑΣ. ΣΤΟΥΣ ΚΟΜΒΟΥΣ ΑΥΤΟΥΣ ΠΕΡΙΛΑΜΒΑΝΟΝΤΑΙ ΤΑ ΔΕΙΓΜΑΤΑ ΤΑ ΟΠΟΙΑ ΕΧΟΥΝ ΤΗΝ ΑΥΤΗ ΤΙΜΗ ΠΑΡΑΜΕΤΡΟΥ. Η ΔΙΑΔΙΚΑΣΙΑ ΑΥΤΗ ΕΠΑΝΑΛΑΜΒΑΝΕΤΑΙ ΑΝΑΔΡΟΜΙΚΑ ΜΕΧΡΙ ΟΛΑ ΤΑ ΔΕΙΓΜΑΤΑ ΠΟΥ ΠΕΡΙΛΑΜΒΑΝΟΝΤΑΙ ΣΤΟΝ ΚΟΜΒΟ ΝΑ ΑΝΗΚΟΥΝ ΣΤΗΝ ΙΔΙΑ ΔΙΑΜΕΡΙΣΗ ΕΙΤΕ ΝΑ ΕΧΟΥΝ ΧΡΗΣΙΜΟΠΟΙΗΘΕΙ ΟΛΟΙ ΟΙ ΠΡΟΣ ΕΞΕΤΑΣΗ ΠΑΡΑΜΕΤΡΟΙ. Η ΕΠΙΛΟΓΗ ΤΟΥ ΚΑΤΑΛΛΗΛΟΤΕΡΟΥ ΧΑΡΑΚΤΗΡΙΣΤΙΚΟΥ ΕΙΝΑΙ ΤΟ ΣΗΜΕΙΟ ΚΛΕΙΔΙ ΤΗΣ ΔΙΑΔΙΚΑΣΙΑΣ ΑΦΟΥ ΑΥΤΟ ΘΑ ΠΡΕΠΕΙ ΝΑ

ΕΞΑΣΦΑΛΙΖΕΙ ΤΗΝ ΔΙΑΙΡΕΣΗ ΤΟΥ ΔΕΙΓΜΑΤΟΣ ΑΠΟΤΕΛΕΣΜΑΤΙΚΑ. ΓΙΑ ΤΗΝ ΜΕΛΕΤΗ ΤΟΥ ΑΛΓΟΡΙΘΜΟΥ ΕΠΙΣΗΣ ΥΠΕΙΣΕΡΧΟΝΤΑΙ ΟΙ ΕΝΝΟΙΕΣ ΤΗΣ ΕΝΤΡΟΠΙΑΣ ΚΑΙ ΤΟΥ ΚΕΡΔΟΥΣ ΠΛΗΡΟΦΟΡΙΑΣ ΟΠΩΣ ΕΠΙΣΗΣ ΚΑΙ ΤΟΥ ΛΟΓΟΥ ΚΕΡΔΟΥΣ ΠΛΗΡΟΦΟΡΙΑΣ. Η ΧΡΗΣΗ ΤΗΣ ΠΟΣΟΤΗΤΑΣ ΠΛΗΡΟΦΟΡΙΑΣ ΑΝ ΚΑΙ ΦΑΙΝΟΜΕΝΙΚΑ ΕΙΝΑΙ ΤΟ ΚΑΤΑΛΛΗΛΟΤΕΡΟ ΜΕΤΡΟ ΓΙΑ ΤΗΝ ΕΠΙΛΟΓΗ ΤΗΣ ΚΑΤΑΛΛΗΛΗΣ ΠΑΡΑΜΕΤΡΟΥ ΠΟΥ ΔΙΑΧΩΡΙΖΕΙ ΤΟ ΔΕΙΓΜΑ ΚΑΤΑ ΤΟΝ ΒΕΛΤΙΣΤΟ ΤΡΟΠΟ ΓΙΑ ΚΑΘΕ ΚΟΜΒΟ ΤΟΥ ΔΕΝΔΡΟΥ ΑΠΟΦΑΣΗΣ. ΣΤΗ ΠΡΑΓΜΑΤΙΚΟΤΗΤΑ ΤΟ ΜΕΓΕΘΟΣ ΑΥΤΟ ΤΕΙΝΕΙ ΝΑ ΠΡΟΚΡΙΝΕΙ ΠΑΡΑΜΕΤΡΟΥΣ ΜΕ ΠΟΛΛΕΣ ΔΙΑΦΟΡΕΤΙΚΕΣ ΤΙΜΕΣ. ΓΙΑ ΤΟΝ ΛΟΓΟ ΑΥΤΟ ΧΡΗΣΙΜΟΠΟΙΕΙΤΑΙ Ο ΛΟΓΟΣ ΤΟΥ ΚΕΡΔΟΥΣ ΠΛΗΡΟΦΟΡΙΑΣ ΠΟΥ ΓΕΝΙΚΑ ΟΔΗΓΕΙ ΣΕ ΜΕΓΑΛΥΤΕΡΟ ΚΕΡΔΟΣ ΠΛΗΡΟΦΟΡΙΑΣ ΚΑΘΩΣ ΟΙ ΔΙΑΣΠΑΣΕΙΣ ΠΟΥ ΠΡΟΚΑΛΕΙ ΔΗΜΙΟΥΡΓΕΙ ΣΥΝΟΛΑ ΕΝΑ ΕΚ ΤΩΝ ΟΠΟΙΩΝ ΕΙΝΑΙ ΑΡΚΕΤΑ ΚΟΝΤΑ ΣΤΟ ΜΕΓΕΘΟΣ ΤΟΥ ΑΡΧΙΚΟΥ. ΑΝ ΘΕΩΡΗΘΕΙ ΕΝΑ ΧΑΡΑΚΤΗΡΙΣΤΙΚΟ D ΠΟΥ ΧΡΗΣΙΜΟΠΟΙΕΙΤΑΙ ΓΙΑ ΤΗΝ ΠΕΡΙΓΡΑΦΗ ΤΩΝ ΜΟΝΑΔΩΝ ΤΟΥ ΣΥΝΟΛΟΥ ΔΕΙΓΜΑΤΩΝ S ΚΑΙ D_i ΕΙΝΑΙ Η ΚΑΘΕ ΜΙΑ ΑΠΟ ΤΙΣ ΤΙΜΕΣ ΠΟΥ ΛΑΜΒΑΝΟΥΝ ΓΙΑ ΤΗΝ ΠΑΡΑΜΕΤΡΟ D ΤΑ ΔΕΙΓΜΑΤΑ, ΤΟΤΕ Ο ΛΟΓΟΣ ΚΕΡΔΟΥΣ ΠΛΗΡΟΦΟΡΙΑΣ ΘΑ ΕΙΝΑΙ ΤΟ ΚΕΡΔΟΣ ΠΛΗΡΟΦΟΡΙΑΣ ΠΟΥ ΕΠΙΤΥΓΧΑΝΕΤΑΙ ΑΠΟ ΤΗΝ ΠΑΡΑΜΕΤΡΟ D ΠΡΟΣ ΤΗΝ ΕΝΤΡΟΠΙΑ ΤΟΥ ΣΥΝΟΛΟΥ ΤΟΥ ΔΕΙΓΜΑΤΟΣ ΣΕ ΣΧΕΣΗ ΜΕ ΤΗΝ ΠΑΡΑΜΕΤΡΟ D . ΜΑΘΗΜΑΤΙΚΑ ΕΚΦΡΑΖΕΤΑΙ ΩΣ ΕΞΗΣ:

$$GainRatio(D, S) = \frac{Gain(D, S)}{H\left(\frac{D_1}{D}, \frac{D_2}{D}, \frac{D_3}{D}, \dots, \frac{D_S}{D}\right)}$$

ΤΑ ΒΗΜΑΤΑ ΤΟΥ ΑΛΓΟΡΙΘΜΟΥ ΕΙΝΑΙ ΤΑ ΕΞΗΣ:

1. ΕΠΙΛΕΓΕΤΑΙ ΤΟ ΧΑΡΑΚΤΗΡΙΣΤΙΚΟ ΤΟ ΟΠΟΙΟ ΠΑΡΟΥΣΙΑΖΕΙ ΤΗΝ ΜΕΓΑΛΥΤΕΡΗ ΑΠΟΔΟΣΗ ΣΕ ΛΟΓΟ ΚΕΡΔΟΥΣ ΠΛΗΡΟΦΟΡΙΑΣ ΣΤΟ ΣΥΝΟΛΟ ΤΩΝ ΔΕΙΓΜΑΤΩΝ. ΤΟ ΧΑΡΑΚΤΗΡΙΣΤΙΚΟ ΑΥΤΟ ΤΟΠΟΘΕΤΕΙΤΑΙ ΣΤΟΝ ΚΟΜΒΟ ΡΙΖΑ ΤΟΥ ΔΕΝΔΡΟΥ.
2. ΤΟ ΧΑΡΑΚΤΗΡΙΣΤΙΚΟ ΤΟ ΟΠΟΙΟ ΠΑΡΟΥΣΙΑΖΕΙ ΤΙΜΗ ΠΟΥ ΕΜΦΑΝΙΖΕΤΑΙ ΣΥΧΝΟΤΕΡΑ ΣΤΟ ΣΥΝΟΛΟ ΤΩΝ ΔΕΙΓΜΑΤΩΝ. ΤΟ ΧΑΡΑΚΤΗΡΙΣΤΙΚΟ ΑΥΤΟ

ΤΟΠΟΘΕΤΕΙΤΑΙ ΣΤΟΝ ΚΟΜΒΟ ΡΙΖΑ ΤΟΥ ΔΕΝΔΡΟΥ.

3. ΓΙΑ ΚΑΘΕ ΤΙΜΗ ΤΟΥ ΕΠΙΛΕΓΕΝΤΟΣ ΧΑΡΑΚΤΗΡΙΣΤΙΚΟΥ ΔΗΜΙΟΥΡΓΟΥΝΤΑΙ ΠΑΙΔΙΑ ΤΙΣ ΡΙΖΑΣ.
4. ΤΑ ΔΕΔΟΜΕΝΑ ΚΑΤΑΝΕΜΟΝΤΑΙ ΣΤΟΥΣ ΚΟΜΒΟΥΣ ΠΑΙΔΙΑ ΤΗΣ ΡΙΖΑΣ ΑΝΑΛΟΓΑ ΜΕ ΤΗΝ ΤΙΜΗ ΤΟΥ ΧΑΡΑΚΤΗΡΙΣΤΙΚΟΥ ΠΟΥ ΧΡΗΣΙΜΟΠΟΙΗΘΗΚΕ ΣΤΗΝ ΡΙΖΑ.
5. ΓΙΑ ΚΑΘΕ ΚΟΜΒΟ ΕΠΑΝΑΛΑΜΒΑΝΟΝΤΑΙ ΤΑ ΒΗΜΑΤΑ 2 ΚΑΙ 3 ΜΕ ΤΗΝ ΕΠΙΛΟΓΗ ΤΟΥ ΧΑΡΑΚΤΗΡΙΣΤΙΚΟΥ ΝΑ ΕΠΙΛΕΓΕΤΑΙ ΑΠΟ ΤΑ ΔΕΔΟΜΕΝΑ ΠΟΥ ΥΠΑΡΧΟΥΝ ΣΤΟΝ ΚΑΘΕ ΚΟΜΒΟ.
6. ΈΝΑΣ ΚΟΜΒΟΣ ΓΙΝΕΤΑΙ ΦΥΛΛΟ ΟΤΑΝ ΟΛΑ ΤΑ ΔΕΔΟΜΕΝΑ ΠΟΥ ΑΝΗΚΟΥΝ ΣΕ ΑΥΤΟΝ ΑΝΗΚΟΥΝ ΣΤΗΝ ΙΔΙΑ ΚΑΤΗΓΟΡΙΑ. Η ΚΑΤΗΓΟΡΙΑ ΑΥΤΗ ΓΙΝΕΤΑΙ ΚΑΙ Η ΤΙΜΗ ΤΟΥ ΦΥΛΛΟΥ.
7. ΑΝ ΣΕ ΚΑΠΟΙΟ ΒΑΘΟΣ ΤΕΛΕΙΩΣΟΥΝ ΤΑ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΠΡΟΣ ΕΛΕΓΧΟ, ΤΟΤΕ Ο ΚΟΜΒΟΣ ΓΙΝΕΤΑΙ ΤΕΡΜΑΤΙΚΟΣ ΚΑΙ ΣΑΝ ΤΙΜΗ ΠΑΙΡΝΕΙ ΕΚΕΙΝΗ ΠΟΥ ΕΧΕΙ ΤΗΝ ΠΛΕΙΟΨΗΦΙΑ ΜΕ ΒΑΣΗ ΤΑ ΔΕΔΟΜΕΝΑ ΤΟΥ ΚΟΜΒΟΥ ΑΥΤΟΥ.

ΣΕ ΨΕΥΔΟΓΛΩΣΣΑ Ο ΑΛΓΟΡΙΘΜΟΣ ΕΧΕΙ ΩΣ ΕΞΗΣ:

```
MAKEDECISIONTREEC4DOT0(NODE N, SAMPLES S,  
ATTRIBUTES A){  
    C = Η ΚΑΤΗΓΟΡΙΑ ΠΟΥ ΕΜΦΑΝΙΖΕΤΑΙ ΣΥΧΝΟΤΕΡΑ ΣΤΟ S  
    N = C  
    IF (ΔΕΝ ΥΠΑΡΧΕΙ C' ΣΤΟ S ΜΕ C'!=C){  
        EXIT;  
    }  
    ELSE {  
        ΑΤΟΥΑ ΜΕΓΑΙΝINFORMATIONRATIO(A)=MAX  
        N = A
```



```

FOREACHVITTOYA {
    SI = { STOY TS: S.A = VI}
    IF (SI != EMPTY){
        N' = N(VI)
                A' = A - {A}
                MAKEDECISSIONTREE(N',SI,A')
    }
}
FOR-END
ENDIF
END

```

C5.0

ΑΠΟΤΕΛΕΙ ΣΥΝΕΧΕΙΑ ΤΟΥ C4.5 ΚΑΙ ΕΠΙΣΗΣ ΠΑΡΟΥΣΙΑΣΤΗΚΕ ΑΠΟ ΤΟΝ ROSSQUINLAN. Ο C5.0 ΕΙΝΑΙ ΜΙΑ ΕΜΠΟΡΙΚΗ ΕΚΔΟΣΗ ΤΟΥ C4.5. ΒΡΙΣΚΕΙ ΕΦΑΡΜΟΓΕΣ ΣΕ ΕΜΠΟΡΙΚΑ ΠΑΚΕΤΑ ΕΦΑΡΜΟΓΩΝ ΕΞΟΡΥΞΗΣ ΓΝΩΣΗΣ. Ο ΠΡΟΣΑΝΑΤΟΛΙΣΜΟΣ ΤΟΥ ΕΙΝΑΙ ΠΡΟΣ ΤΗ ΧΡΗΣΗ ΜΕΓΑΛΩΝ ΣΥΝΟΛΩΝ ΔΕΔΟΜΕΝΩΝ. Η ΦΑΣΗ ΤΗΣ ΕΠΑΓΩΓΗΣ ΕΙΝΑΙ ΟΜΟΙΑ ΜΕ ΑΥΤΗ ΤΟΥ C4.5 ΑΛΛΑ Η ΔΗΜΙΟΥΡΓΙΑ ΚΑΝΟΝΩΝ ΕΙΝΑΙ ΔΙΑΦΟΡΕΤΙΚΗ. ΣΕ ΑΝΤΙΘΕΣΗ ΜΕ ΤΟΝ C4.5, ΟΙ ΑΛΓΟΡΙΘΜΟΙ ΠΟΥ ΧΡΗΣΙΜΟΠΟΙΟΥΝΤΑΙ ΣΤΟΝ C5.0 ΔΕΝ ΕΧΟΥΝ ΓΙΝΕΙ ΓΝΩΣΤΟΙ. ΤΑ ΑΠΟΤΕΛΕΣΜΑΤΑ ΠΟΥ ΕΧΟΥΝ ΔΗΜΟΣΙΕΥΤΕΙ, ΔΕΙΧΝΟΥΝ ΟΤΙ Ο C5.0 ΒΕΛΤΙΩΝΕΙ ΤΗΝ ΧΡΗΣΗ ΤΗΣ ΜΝΗΜΗΣ ΚΑΤΑ 90%, ΤΡΕΧΕΙ ΠΟΛΥ ΠΙΟ ΓΡΗΓΟΡΑ ΑΠΟ ΤΟΝ C4.5 ΚΑΙ ΠΑΡΑΓΕΙ ΠΙΟ ΑΚΡΙΒΕΙΣ ΚΑΝΟΝΕΣ. ΜΙΑ ΠΟΛΥ ΣΗΜΑΝΤΙΚΗ ΒΕΛΤΙΩΣΗ ΣΤΗΝ ΑΚΡΙΒΕΙΑ ΤΟΥ C5.0 ΒΑΣΙΖΕΤΑΙ ΣΤΗΝ ΕΝΙΣΧΥΣΗ (BOOSTING) Η ΟΠΟΙΑ ΕΙΝΑΙ ΜΙΑ ΤΕΧΝΙΚΗ ΠΟΥ ΣΥΝΔΥΑΖΕΙ ΔΙΑΦΟΡΟΥΣ ΚΑΤΗΓΟΡΙΟΠΟΙΗΤΕΣ [7][8][9][10][11].

3. ΧΡΗΣΗ ΤΟΥ WEKA

3.1 ΠΕΡΙΓΡΑΦΗ ΤΟΥ ΕΡΓΑΛΕΙΟΥ

ΤΟ WEKA (WAIKATOENVIRONMENT FOR KNOWLEDGEANALYSIS) ΕΙΝΑΙ ΛΟΓΙΣΜΙΚΟ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ, ΠΟΥ ΑΝΑΠΤΥΧΘΗΚΕ ΑΠΟ ΤΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΤΟΥ WAIKATO ΚΑΙ ΤΟ ΟΠΟΙΟ ΠΡΟΣΦΕΡΕΙ ΕΝΑ ΣΥΝΟΛΟ ΛΕΙΤΟΥΡΓΙΩΝ ΓΙΑ ΤΗΝ ΓΡΑΦΙΚΗ ΠΑΡΟΥΣΙΑΣΗ ΤΗΣ ΕΞΕΛΙΞΗΣ ΑΛΓΟΡΙΘΜΩΝ ΠΟΥ ΣΤΟΧΕΥΟΥΝ ΣΤΗΝ ΑΝΑΛΥΣΗ ΔΕΔΟΜΕΝΩΝ ΚΑΙ ΤΗΝ ΠΑΡΑΓΩΓΗ ΠΡΟΒΛΕΠΤΙΚΩΝ ΜΟΝΤΕΛΩΝ. ΧΑΡΑΚΤΗΡΙΣΤΙΚΟ ΤΟΥ ΕΙΝΑΙ Η ΔΙΑΘΕΣΗ ΦΙΛΙΚΩΝ ΠΡΟΣ ΤΟΝ ΤΕΛΙΚΟ ΧΡΗΣΤΗ ΔΙΕΠΑΦΩΝ, ΚΑΤΙ ΠΟΥ ΤΟ ΚΑΘΙΣΤΑ ΙΔΙΑΙΤΕΡΑ ΔΗΜΟΦΙΛΕΣ ΓΙΑ ΧΡΗΣΗ ΣΕ ΔΙΑΦΟΡΩΝ ΤΥΠΩΝ ΕΡΕΥΝΕΣ. ΤΑ ΕΡΓΑΛΕΙΑ ΤΟΥ (ΦΙΛΤΡΑ, ΚΑΤΗΓΟΡΙΟΠΟΙΗΤΕΣ, ΤΑΞΙΝΟΜΗΤΕΣ, ΣΥΣΧΕΤΙΣΤΕΣ ΚΑΙ ΕΠΙΛΟΓΕΙΣ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ) ΧΡΗΣΙΜΟΠΟΙΟΥΝΤΑΙ ΜΕΣΩ ΓΡΑΦΙΚΩΝ ΔΙΕΠΑΦΩΝ. ΕΙΝΑΙ ΓΡΑΜΜΕΝΟ ΣΕ ΓΛΩΣΣΑ ΠΡΟΓΡΑΜΜΑΤΙΣΜΟΥ JAVA ΚΑΙ ΠΕΡΙΕΧΕΙ ΜΕΘΟΔΟΥΣ ΠΟΥ ΜΠΟΡΟΥΝ ΝΑ ΑΝΤΑΠΟΚΡΙΘΟΥΝ ΣΕ ΑΠΑΙΤΗΣΕΙΣ ΓΙΑ

- ΠΡΟΕΠΕΞΕΡΓΑΣΙΑ ΔΕΔΟΜΕΝΩΝ
- ΤΑΞΙΝΟΜΗΣΗ
- ΣΥΣΤΑΔΟΠΟΙΗΣΗ
- ΑΝΑΚΑΛΥΨΗ ΣΥΣΧΕΤΙΣΤΙΚΩΝ ΚΑΝΟΝΩΝ

ΕΙΝΑΙ ΔΩΡΕΑΝ ΔΙΑΘΕΣΙΜΟ ΣΤΗΝ ΙΣΤΟΣΕΛΙΔΑ [HTTP://WWW.CS.WAIKATO.AC.NZ/ML/WEKA/](http://www.cs.waikato.ac.nz/ml/weka/). ΓΙΑ ΝΑ ΤΡΕΞΕΙ Η ΤΕΛΕΥΤΑΙΑ ΕΚΔΟΣΗ ΘΑ ΠΡΕΠΕΙ ΣΤΟ ΦΙΛΟΞΕΝΟΥΝ ΜΗΧΑΝΗΜΑ ΝΑ ΕΙΝΑΙ ΕΓΚΑΤΕΣΤΗΜΕΝΗ ΕΙΚΟΝΙΚΗ ΜΗΧΑΝΗ JAVA ΕΚΔΟΣΗΣ ΤΟΥΛΑΧΙΣΤΟΝ 1.6.

Ο ΧΡΗΣΤΗΣ ΤΟΥ WEKA ΜΠΟΡΕΙ ΝΑ ΔΙΑΜΟΡΦΩΣΕΙ ΚΑΙ ΝΑ ΤΡΕΞΕΙ ΔΙΑΦΟΡΑ ΜΟΝΤΕΛΑ ΚΑΤΗΓΟΡΙΟΠΟΙΗΣΗΣ ΔΕΔΟΜΕΝΩΝ, ΝΑ ΥΛΟΠΟΙΗΣΕΙ ΣΥΓΚΡΙΤΙΚΕΣ ΜΕΛΕΤΕΣ ΚΑΙ ΕΛΕΓΧΟΥΣ ΣΕ ΣΥΝΟΛΑ

ΔΕΔΟΜΕΝΩΝ. ΤΟ ΛΟΓΙΣΜΙΚΟ ΕΧΕΙ ΕΠΙΣΗΣ ΤΗΝ ΔΥΝΑΤΟΤΗΤΑ ΝΑ ΠΑΡΟΥΣΙΑΖΕΙ ΤΑ ΣΦΑΛΜΑΤΑ ΠΟΥ ΠΑΡΑΓΟΝΤΑΙ ΑΠΟ ΤΙΣ ΚΑΤΗΓΟΡΙΟΠΟΙΗΣΕΙΣ ΠΟΥ ΠΡΑΓΜΑΤΟΠΟΙΕΙ. ΣΤΙΣ ΠΕΡΙΠΤΩΣΕΙΣ ΤΗΣ ΔΗΜΙΟΥΡΓΙΑΣ ΔΕΝΔΡΩΝ ΑΠΟΦΑΣΗΣ, ΥΛΟΠΟΙΕΙ ΚΑΙ ΠΑΡΟΥΣΙΑΖΕΙ ΓΡΑΦΙΚΑ ΤΟ ΠΑΡΑΓΟΜΕΝΟ ΑΠΟΤΕΛΕΣΜΑ.

3.2 ΧΡΗΣΗ

Η ΓΡΑΦΙΚΗ ΔΙΕΠΑΦΗ ΤΟΥ ΕΡΓΑΛΕΙΟΥ ΚΑΤΑΝΕΜΕΙ ΤΙΣ ΛΕΙΤΟΥΡΓΙΕΣ ΤΟΥ ΚΑΙ ΤΙΣ ΠΑΡΟΥΣΙΑΣΕΙΣ ΤΟΥ ΣΕ ΚΑΡΤΕΛΕΣ ΩΣ ΕΞΗΣ:

- **PREPROCESS:** ΠΕΡΙΛΑΜΒΑΝΕΙ ΤΗΝ ΔΙΑΔΙΚΑΣΙΑ ΤΗΣ ΠΡΟΕΠΕΞΕΡΓΑΣΙΑΣ ΤΟΥ ΣΥΝΟΛΟΥ ΔΕΔΟΜΕΝΩΝ ΠΟΥ Θ ΧΡΗΣΙΜΟΠΟΙΗΘΕΙ. Ο ΧΡΗΣΤΗΣ ΕΠΙΛΕΓΕΙ ΤΟ ΑΡΧΕΙΟ ΠΟΥ ΠΕΡΙΛΑΜΒΑΝΕΙ ΤΑ ΔΕΔΟΜΕΝΑ (ΣΕ ΠΟΙΚΙΛΕΣ ΜΟΡΦΕΣ) ΚΑΙ ΜΠΟΡΕΙ ΝΑ ΕΠΙΛΕΞΕΙ ΦΙΛΤΡΑ ΓΙΑ ΤΟΝ ΜΕΤΑΣΧΗΜΑΤΙΣΜΟΣ ΤΩΝ ΔΕΔΟΜΕΝΩΝ ΣΕ ΜΟΡΦΕΣ ΤΕΤΟΙΕΣ ΠΟΥ ΝΑ ΜΠΟΡΕΙ ΝΑ ΤΙΣ ΕΠΕΞΕΡΓΑΣΤΕΙ Ο ΕΠΙΘΥΜΗΤΟΣ ΑΛΓΟΡΙΘΜΟΣ, Η ΑΚΟΜΑ ΚΑΙ ΝΑ ΑΠΟΜΑΚΡΥΝΘΟΥΝ ΔΕΔΟΜΕΝΑ ΤΑ ΟΠΟΙΑ ΕΙΝΑΙ – ΓΙΑ ΚΑΠΟΙΟ ΛΟΓΟ – ΑΚΑΤΑΛΛΗΛΑ.
- **CLASSIFY:** ΠΕΡΙΛΑΜΒΑΝΟΝΤΑΙ ΟΙ ΛΕΙΤΟΥΡΓΙΕΣ ΤΗΣ ΤΑΞΙΝΟΜΗΣΗΣ ΚΑΙ ΠΑΛΙΝΔΡΟΜΗΣΗΣ ΠΟΥ ΕΦΑΡΜΟΖΟΝΤΑΙ ΣΤΑ ΣΥΝΟΛΑ ΔΕΔΟΜΕΝΩΝ. ΜΕ ΤΟΝ ΤΡΟΠΟ ΑΥΤΟΝ ΥΠΟΛΟΓΙΖΕΤΑΙ Η ΑΚΡΙΒΕΙΑ ΤΟΥ ΜΟΝΤΕΛΟΥ ΠΡΟΒΛΕΨΗΣ ΠΟΥ ΠΑΡΑΛΑΜΒΑΝΕΤΑΙ ΣΤΗΝ ΕΞΟΔΟ ΤΗΣ ΔΙΑΔΙΚΑΣΙΑΣ. ΠΑΡΑΛΛΗΛΑ ΠΑΡΕΧΕΙ ΤΑ ΑΠΟΤΕΛΕΣΜΑΤΑ ΤΩΝ ΛΕΙΤΟΥΡΓΙΩΝ ΑΥΤΩΝ ΜΕ ΓΡΑΦΙΚΟ ΤΡΟΠΟ.
- **ASSOCIATE:** ΠΕΡΙΛΑΜΒΑΝΕΙ ΤΗΝ ΛΕΙΤΟΥΡΓΙΑ ΤΗΣ ΑΝΑΚΑΛΥΨΗΣ ΚΑΝΟΝΩΝ ΣΥΣΧΕΤΙΣΕΩΝ ΠΟΥ ΜΠΟΡΕΙ ΝΑ ΥΠΑΡΧΟΥΝ ΜΕΤΑΞΥ ΤΩΝ ΠΑΡΑΜΕΤΡΩΝ – ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ ΤΩΝ ΔΕΔΟΜΕΝΩΝ.

- CLUSTER: ΠΕΡΙΛΑΜΒΑΝΕΙ ΤΙΣ ΛΕΙΤΟΥΡΓΙΕΣ ΤΗΣΣΥΣΤΑΔΟΠΟΙΗΣΗΣ ΠΟΥ ΕΧΕΙ ΤΗΝ ΔΥΝΑΤΟΤΗΤΑ ΝΑ ΕΚΤΕΛΕΣΕΙ ΤΟ ΕΡΓΑΛΕΙΟ.
- SELECTATTRIBUTES: ΠΕΡΙΛΑΜΒΑΝΕΙ ΛΕΙΤΟΥΡΓΙΕΣ ΓΙΑ ΤΟΝ ΚΑΘΟΡΙΣΜΟ ΤΩΝ ΜΕΤΑΒΛΗΤΩΝ ΠΟΥ ΠΑΡΟΥΣΙΑΖΟΥΝ ΤΟ ΚΑΛΥΤΕΡΟ ΕΠΙΠΕΔΟ ΠΡΟΒΛΕΨΗΣ ΣΤΑ ΣΥΝΟΛΑ ΔΕΔΟΜΕΝΩΝ.
- VISUALIZE: ΠΕΡΙΛΑΜΒΑΝΟΝΤΑΙ ΓΡΑΦΙΚΕΣ ΑΠΕΙΚΟΝΙΣΕΙΣ ΤΩΝ ΑΠΟΤΕΛΕΣΜΑΤΩΝ ΤΩΝ ΛΕΙΤΟΥΡΓΙΩΝ ΠΟΥ ΠΡΑΓΜΑΤΟΠΟΙΟΥΝΤΑΙ ΜΕ ΤΟ ΛΟΓΙΣΜΙΚΟ.

ΣΤΟ ΛΟΓΙΣΜΙΚΟ ΠΕΡΙΛΑΜΒΑΝΕΤΑΙ ΕΝΑΣ ΑΡΙΘΜΟΣ ΑΠΟ ΑΛΓΟΡΙΘΜΟΥΣ ΚΑΤΗΓΟΡΙΟΠΟΙΗΣΗΣ ΚΑΙ ΠΡΟΒΛΕΨΗΣ. Ο ΑΛΓΟΡΙΘΜΟΣ C4.5 ΠΕΡΙΛΑΜΒΑΝΕΤΑΙ ΣΤΟ ΣΥΝΟΛΟ ΑΥΤΟ ΜΕ ΤΟ ΟΝΟΜΑ J48 ΥΠΟ ΤΗΝ ΚΛΑΣΗ ΑΛΓΟΡΙΘΜΩΝ WEKA.CLASSIFIERS.TREES. ΑΝ ΘΕΩΡΗΘΕΙ ΟΤΙ ΜΕ ΒΑΣΗ ΕΝΑ ΣΥΝΟΛΟ ΔΕΔΟΜΕΝΩΝ ΧΡΕΙΑΖΕΤΑΙ ΝΑ ΔΗΜΙΟΥΡΓΗΘΕΙ ΕΝΑ ΔΕΝΔΡΟ ΑΠΟ ΦΑΣΗΣ ΑΚΟΛΟΥΘΕΙΤΑΙ Η ΔΙΑΔΙΚΑΣΙΑ ΠΟΥ ΠΕΡΙΓΡΑΦΕΤΑΙ ΑΚΟΛΟΥΘΩΣ. ΜΕ ΤΗΝ ΕΚΚΙΝΗΣΗ ΤΗΣ ΕΦΑΡΜΟΓΗΣ ΠΑΡΟΥΣΙΑΖΕΤΑΙ Η ΑΡΧΙΚΗ ΤΗΣ ΟΘΟΝΗ. ΣΤΗΝ ΣΥΝΕΧΕΙΑ Ο ΧΡΗΣΤΗΣ ΕΠΙΛΕΓΕΙ ΤΗΝ ΚΑΡΤΕΛΑ CLASSIFY ΚΑΙ ΣΤΗΝ ΣΥΝΕΧΕΙΑ ΚΑΝΕΙ ΚΛΙΚ ΣΤΟ ΠΛΗΚΤΡΟ CHOOSE ΓΙΑ ΝΑ ΕΠΙΛΕΓΕΙ Ο ΕΠΙΘΥΜΗΤΟΣ ΑΛΓΟΡΙΘΜΟΣ. ΈΠΕΙΤΑ Ο ΧΡΗΣΤΗΣ ΜΠΟΡΕΙ ΝΑ ΠΡΟΣΔΙΟΡΙΖΕΙ ΤΙΣ ΠΑΡΑΜΕΤΡΟΥΣ ΕΚΤΕΛΕΣΗΣ ΤΟΥ ΑΛΓΟΡΙΘΜΟΥ. ΆΛΛΗ ΕΠΙΛΟΓΗ ΠΟΥ ΜΠΟΡΕΙ ΝΑ ΚΑΝΕΙ Ο ΧΡΗΣΤΗΣ ΕΙΝΑΙ ΤΗΣ ΜΕΘΟΔΟΥ ΕΛΕΓΧΟΥ - ΑΞΙΟΛΟΓΗΣΗΣ ΤΟΥ, ΜΕ ΕΠΙΛΟΓΗ ΜΙΑΣ ΑΠΟ ΑΥΤΕΣ ΠΟΥ ΒΡΙΣΚΟΝΤΑΙ ΎΚΑΤΩ ΑΠΟ ΤΗΝ ΕΤΙΚΕΤΑ TESTOPTIONS. ΑΦΟΥ ΓΙΝΟΥΝ ΟΙ ΕΠΙΛΟΓΕΣ ΑΥΤΕΣ ΠΡΕΠΕΙ ΝΑ ΓΙΝΕΙ ΚΛΙΚ ΣΤΟ ΠΛΗΚΤΡΟ START ΓΙΑ ΝΑ ΔΗΜΙΟΥΡΓΗΘΕΙ ΤΟ ΜΟΝΤΕΛΟ ΚΑΙ ΕΜΦΑΝΙΖΕΤΑΙ Η ASCII ΕΚΔΟΣΗ ΤΟΥ ΔΕΝΔΡΟΥ ΟΠΩΣ ΚΑΙ ΤΑ ΣΧΕΤΙΚΑ ΜΕ ΑΥΤΟ ΣΤΑΤΙΣΤΙΚΑ. ΥΠΑΡΧΕΙ ΕΠΙΣΗΣ Η ΔΥΝΑΤΟΤΗΤΑ ΝΑ ΠΑΡΟΥΣΙΑΣΤΟΥΝ ΤΑ ΑΠΟΤΕΛΕΣΜΑΤΑ ΣΕ ΞΕΧΩΡΙΣΤΟ ΠΑΡΑΘΥΡΟ ΑΝ ΓΙΝΕΙ ΔΕΞΙ ΚΛΙΚ ΣΤΟ ΑΠΟΤΕΛΕΣΜΑ ΚΑΙ ΕΠΙΛΕΓΕΙ Η ΔΥΝΑΤΟΤΗΤΑ "VIEW IN SEPARATEWINDOW" ΑΠΟ ΤΟ ΜΕΝΟΥ ΠΟΥ ΕΜΦΑΝΙΖΕΤΑΙ. ΕΠΙΣΗΣ ΜΕ ΔΕΞΙ ΚΛΙΚ ΣΤΟ ΑΠΟΤΕΛΕΣΜΑ ΚΑΙ ΕΠΙΛΕΓΟΝΤΑΣ "VISUALIZETREE" ΕΜΦΑΝΙΖΕΤΑΙ ΜΙΑ ΓΡΑΦΙΚΗ

ΑΝΑΠΑΡΑΣΤΑΣΗΣ ΤΟΥ ΔΕΝΔΡΟΥ ΑΠΟΦΑΣΗΣ.

ΤΟ ΜΟΝΤΕΛΟ ΠΟΥ ΕΧΕΙ ΔΗΜΙΟΥΡΓΗΘΕΙ ΑΠΟ ΤΗΝ ΔΙΑΔΙΚΑΣΙΑ ΑΥΤΗ ΜΠΟΡΕΙ ΝΑ ΧΡΗΣΙΜΟΠΟΙΗΘΕΙ ΓΙΑ ΤΗΝ ΔΙΑΛΟΓΗ ΝΕΩΝ ΔΕΔΟΜΕΝΩΝ. ΈΤΣΙ ΜΠΟΡΕΙ ΝΑ ΕΠΙΛΕΓΕΙ ΕΝΑ ΝΕΟ ΣΥΝΟΛΟ ΔΕΔΟΜΕΝΩΝ. ΚΑΘΕ ΣΤΟΙΧΕΙΟ ΤΟΥ ΣΥΝΟΛΟΥ ΘΑ ΠΡΕΠΕΙ ΝΑ ΠΕΡΙΓΡΑΦΕΤΑΙ ΑΠΟ ΤΑ ΙΔΙΑ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΠΟΥ ΠΕΡΙΓΡΑΦΟΝΤΑΙ ΚΑΙ ΤΑ ΣΤΟΙΧΕΙΑ ΤΟΥ ΑΡΧΙΚΟΥ ΣΥΝΟΛΟΥ ΕΝΩ ΤΟ ΧΑΡΑΚΤΗΡΙΣΤΙΚΟ ΣΤΟ ΟΠΟΙΟ ΘΑ ΓΙΝΕΙ Η ΚΑΤΗΓΟΡΙΟΠΟΙΗΣΗ ΠΑΡΑΜΕΝΕΙ ΑΓΝΩΣΤΟ. Η ΕΠΙΛΟΓΗ ΤΟΥ ΜΟΝΤΕΛΟΥ ΠΟΥ ΘΑ ΒΑΣΙΣΤΕΙ Η ΚΑΤΗΓΟΡΙΟΠΟΙΗΣΗ ΓΙΝΕΤΑΙ ΑΠΟ ΤΗΝ ΒΑΣΙΚΗ ΟΘΟΝΗ ΤΗΣ ΕΦΑΡΜΟΓΗΣ ΕΠΙΛΕΓΟΝΤΑΣ TESTOPTIONS ΚΑΙ SUPPLIEDTESTSET ΚΑΙ ΣΤΗΝ ΣΥΝΕΧΕΙΑ ΤΟ ΠΛΗΚΤΡΟ SET ΓΙΑ ΝΑ ΓΙΝΕΙ Η ΕΠΙΛΟΓΗ ΤΟΥ ΑΡΧΕΙΟΥ ΤΥΠΟΥ ARFF ΠΟΥ ΑΝΤΙΣΤΟΙΧΕΙ ΣΤΟ ΜΟΝΤΕΛΟ. ΜΕ ΚΛΙΚ ΣΤΟ ΠΛΗΚΤΡΟ START ΕΚΚΙΝΕΙ Η ΚΑΤΗΓΟΡΙΟΠΟΙΗΣΗ ΤΩΝ ΔΕΔΟΜΕΝΩΝ. Η ΚΙΝΗΣΗ ΑΥΤΗ ΔΗΜΙΟΥΡΓΕΙ ΓΙΑ ΑΛΛΗ ΜΙΑ ΦΟΡΑ ΤΑ ΜΟΝΤΕΛΑ ΑΠΟ ΤΑ ΔΕΔΟΜΕΝΑ ΕΚΠΑΙΔΕΥΣΗΣ, ΑΛΛΑ ΑΥΤΗ ΤΗ ΦΟΡΑ ΕΦΑΡΜΟΖΕΤΑΙ ΤΟ ΜΟΝΤΕΛΟ ΣΤΑ ΝΕΑ ΜΗ ΤΑΞΙΝΟΜΗΜΕΝΑ ΔΕΙΓΜΑΤΑ ΤΟΥ ΝΕΟΥ ΑΡΧΕΙΟΥ.

3.3 ΔΟΜΗ ΑΡΧΕΙΩΝ ARFF

ΤΑ ARFF ΑΡΧΕΙΑ, ΤΑ ΟΠΟΙΑ ΜΠΟΡΕΙ ΝΑ ΔΙΑΧΕΙΡΙΣΤΕΙ ΤΟ ΛΟΓΙΣΜΙΚΟ ΑΠΟΤΕΛΟΥΝΤΑΙ ΑΠΟ ΔΥΟ ΜΕΡΗ:

- ΤΗΝ ΚΕΦΑΛΙΔΑ (HEADER): ΣΤΟ ΤΜΗΜΑ ΑΥΤΟ ΠΕΡΙΛΑΜΒΑΝΟΝΤΑΙ ΤΑ ΜΕΤΑΔΕΔΟΜΕΝΑ ΤΟΥ ΣΥΝΟΛΟΥ ΔΕΔΟΜΕΝΩΝ. ΤΑ ΜΕΤΑΔΕΔΟΜΕΝΑ ΑΥΤΑ ΠΕΡΙΛΑΜΒΑΝΟΥΝ:
 - ΤΗΝ ΟΝΟΜΑΣΙΑ ΤΟΥ ΣΥΝΟΛΟΥ ΔΕΔΟΜΕΝΩΝ ΥΠΟ ΤΗΝ ΕΤΙΚΕΤΑ @RELATION (ΠΧ @RELATIONMYRELATION)
 - ΤΙΣ ΔΗΛΩΣΕΙΣ ΤΩΝ ΜΕΤΑΒΛΗΤΩΝ ΥΠΟ ΤΗΝ ΕΤΙΚΕΤΑ @ATTRIBUTE. Η ΔΗΛΩΣΗ ΑΥΤΗ ΠΡΟΣΔΙΟΡΙΖΕΙ ΤΟ

ΟΝΟΜΑ ΤΗΣ ΜΕΤΑΒΛΗΤΗΣ ΚΑΙ ΤΟΝ ΤΥΠΟ ΔΕΔΟΜΕΝΩΝ ΤΩΝ ΤΙΜΩΝ ΤΗΣ (ΠΧ @ATTRIBUTEHEIGHTNUMERIC). ΟΙ ΤΥΠΟΙ ΔΕΔΟΜΕΝΩΝ ΠΟΥ ΜΠΟΡΕΙ ΝΑ ΕΧΕΙ ΜΙΑ ΜΕΤΑΒΛΗΤΗ ΕΙΝΑΙ:

§ ΑΡΙΘΜΗΤΙΚΗ ΚΑΙ ΧΑΡΑΚΤΗΡΙΖΕΤΑΙ ΣΑΝ NUMERIC

§ ΚΑΤΗΓΟΡΗΜΑ ΚΑΙ ΧΑΡΑΚΤΗΡΙΖΕΤΑΙ ΣΑΝ NOMINAL ΚΑΙ ΑΚΟΛΟΥΘΕΙΤΑΙ ΑΠΟ ΤΙΣ ΔΙΑΚΡΙΤΕΣ ΤΙΜΕΣ ΠΟΥ ΜΠΟΡΕΙ ΝΑ ΛΑΒΕΙ

§ ΗΜΕΡΟΜΗΝΙΕΣ ΚΑΙ ΧΑΡΑΚΤΗΡΙΖΟΝΤΑΙ ΣΑΝ DATA

§ ΣΥΜΒΟΛΟΣΕΙΡΕΣ ΠΟΥ ΧΑΡΑΚΤΗΡΙΖΟΝΤΑΙ ΣΑΝ STRING

- ΤΑ ΔΕΔΟΜΕΝΑ (DATA): ΣΤΟ ΤΜΗΜΑ ΑΥΤΟ ΠΕΡΙΛΑΜΒΑΝΕΤΑΙ ΤΟ ΩΦΕΛΙΜΟ ΠΕΡΙΕΧΟΜΕΝΟ ΤΟΥ ΑΡΧΕΙΟΥ ΚΑΙ ΞΕΚΙΝΑΕΙ ΜΕ ΤΗΝ ΔΗΛΩΣΗ DATA. ΑΠΟ ΤΟ ΣΗΜΕΙΟ ΑΥΤΟ ΚΑΙ ΜΕΧΡΙ ΤΟ ΤΕΛΟΣ ΤΟΥ ΑΡΧΕΙΟΥ, ΣΕ ΚΑΘΕ ΓΡΑΜΜΗ ΠΕΡΙΛΑΜΒΑΝΟΝΤΑΙ ΟΙ ΤΙΜΕΣ ΤΩΝ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ. ΟΙ ΤΙΜΕΣ ΔΙΑΧΩΡΙΖΟΝΤΑΙ ΜΕ (,). ΟΙ ΜΗ ΓΝΩΣΤΕΣ ΤΙΜΕΣ ΑΝΑΠΑΡΙΣΤΑΝΤΑΙ ΜΕ (?). ΣΤΟ ΤΕΛΟΣ ΚΑΘΕ ΓΡΑΜΜΗΣ ΥΠΑΡΧΕΙ Η ΔΥΝΑΤΟΤΗΤΑ ΠΡΟΣΘΗΚΗΣ ΣΧΟΛΙΟΥ ΑΦΟΥ ΠΡΟΗΓΗΘΕΙ Ο ΧΑΡΑΚΤΗΡΑ %'. (ΠΧ 4,6,7,BLABLA %COMMENT).

ΣΤΗΝ ΠΑΡΑΚΑΤΩ ΕΙΚΟΝΑ ΠΑΡΑΤΙΘΕΤΑΙ ΕΝΑ ΠΑΡΑΔΕΙΓΜΑ ΕΝΟΣ ARFF ΑΡΧΕΙΟΥ.

```

← → ↻ 🏠 storm.cis.fordham.edu/~gweiss/data-mining/weka-data/cpu.arff
%
% As used by Kilpatrick, D. & Cameron-Jones, M. (1998). Numeric prediction
% using instance-based learning with encoding length selection. In Progress
% in Connectionist-Based Information Systems. Singapore: Springer-Verlag.
%
% Deleted "vendor" attribute to make data consistent with with what we
% used in the data mining book.
%
@relation 'cpu'
@attribute MVEF real
@attribute MMIN real
@attribute MMAX real
@attribute CACH real
@attribute CHMIN real
@attribute CHMAX real
@attribute class real
@data
125,256,6000,256,16,128,198
29,8000,32000,32,8,32,269
29,8000,32000,32,8,32,220
29,8000,32000,32,8,32,172
29,8000,16000,32,8,16,132
26,8000,32000,64,8,32,318
23,16000,32000,64,16,32,367
23,16000,32000,64,16,32,489
23,16000,64000,64,16,32,636
23,32000,64000,128,32,64,1144
400,1000,3000,0,1,2,38
400,512,3500,4,1,6,40
60,2000,8000,65,1,8,92
50,4000,16000,65,1,8,138
350,64,64,0,1,4,10
200,512,18000,0,4,32,55
167,524,2000,8,4,15,19
143,512,5000,0,7,32,28
143,1000,2000,0,5,16,51
110,5000,5000,142,8,64,120
143,1500,6300,0,5,32,30
143,3100,6200,0,5,20,33
143,2300,6200,0,6,64,61
110,3100,6200,0,6,64,76
520,128,6000,0,1,12,25
520,512,2000,4,1,3,89
320,256,6000,0,1,6,33
320,256,3000,4,1,3,27
520,512,5000,4,1,5,77
320,256,5000,4,1,6,27
25,1310,2620,131,12,24,274
25,1310,2620,131,12,24,368
50,2620,10480,30,12,24,32
50,2620,10480,30,12,24,63
56,5240,20970,30,12,24,106
64,5240,20970,30,12,24,208
50,500,2000,0,1,4,20
50,1000,4000,0,1,5,29
50,2000,8000,0,1,5,71
50,1000,4000,0,3,5,26

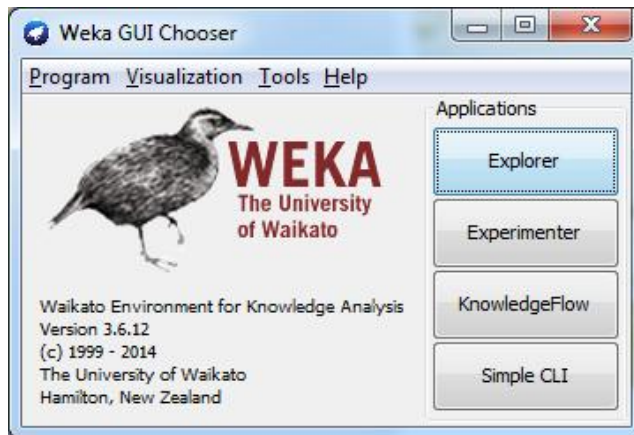
```

ΕΙΚΟΝΑ 14: ΠΑΡΑΔΕΙΓΜΑ ARFF ΑΡΧΕΙΟΥ

3.4 ΜΕΛΕΤΗ ΠΕΡΙΠΤΩΣΗΣ

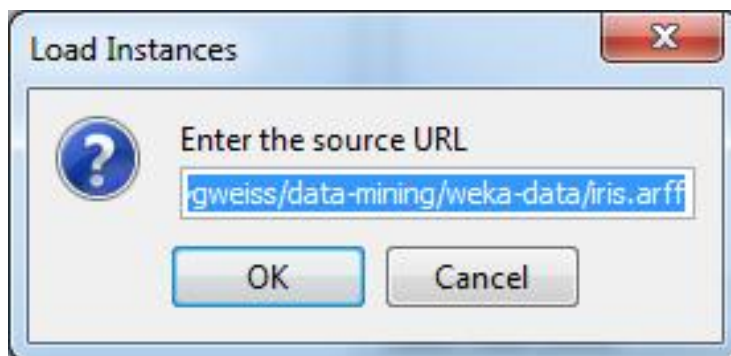
ΠΡΟΚΕΙΜΕΝΟΥ ΝΑ ΕΙΝΑΙ ΠΛΗΡΗΣ Η ΠΑΡΟΥΣΙΑΣΗ ΤΗΣ ΧΡΗΣΗΣ ΤΟΥ ΕΡΓΑΛΕΙΟΥ ΠΑΡΟΥΣΙΑΖΕΤΑΙ ΣΤΗΝ ΣΥΝΕΧΕΙΑ ΜΙΑ ΠΕΡΙΠΤΩΣΗ ΧΡΗΣΗΣ. ΕΠΙΛΕΧΘΗΚΕ ΝΑ ΕΞΕΤΑΣΤΕΙ ΕΝΑ ΣΥΝΟΛΟ ΔΕΔΟΜΕΝΩΝ ΑΠΟ ΤΗΝ ΤΟΠΟΘΕΣΙΑ [HTTP://STORM.CIS.FORDHAM.EDU/~GWEISS/DATA-MINING/DATASETS.HTML](http://storm.cis.fordham.edu/~gweiss/data-mining/datasets.html), ΤΟ ΟΠΟΙΟ ΗΤΑΝ ΗΔΗ ΣΕ ARFF ΜΟΡΦΗ (IRIS.ARFF).

ΑΡΧΙΚΑ ΕΚΚΙΝΕΙ Η ΕΦΑΡΜΟΓΗ ΚΑΙ ΠΑΡΟΥΣΙΑΖΕΤΑΙ Η ΑΡΧΙΚΗ ΤΗΣ ΟΘΟΝΗ. ΣΕ ΑΥΤΗΝ Ο ΧΡΗΣΤΗΣ ΕΠΙΛΕΓΕΙ «EXPLORE».

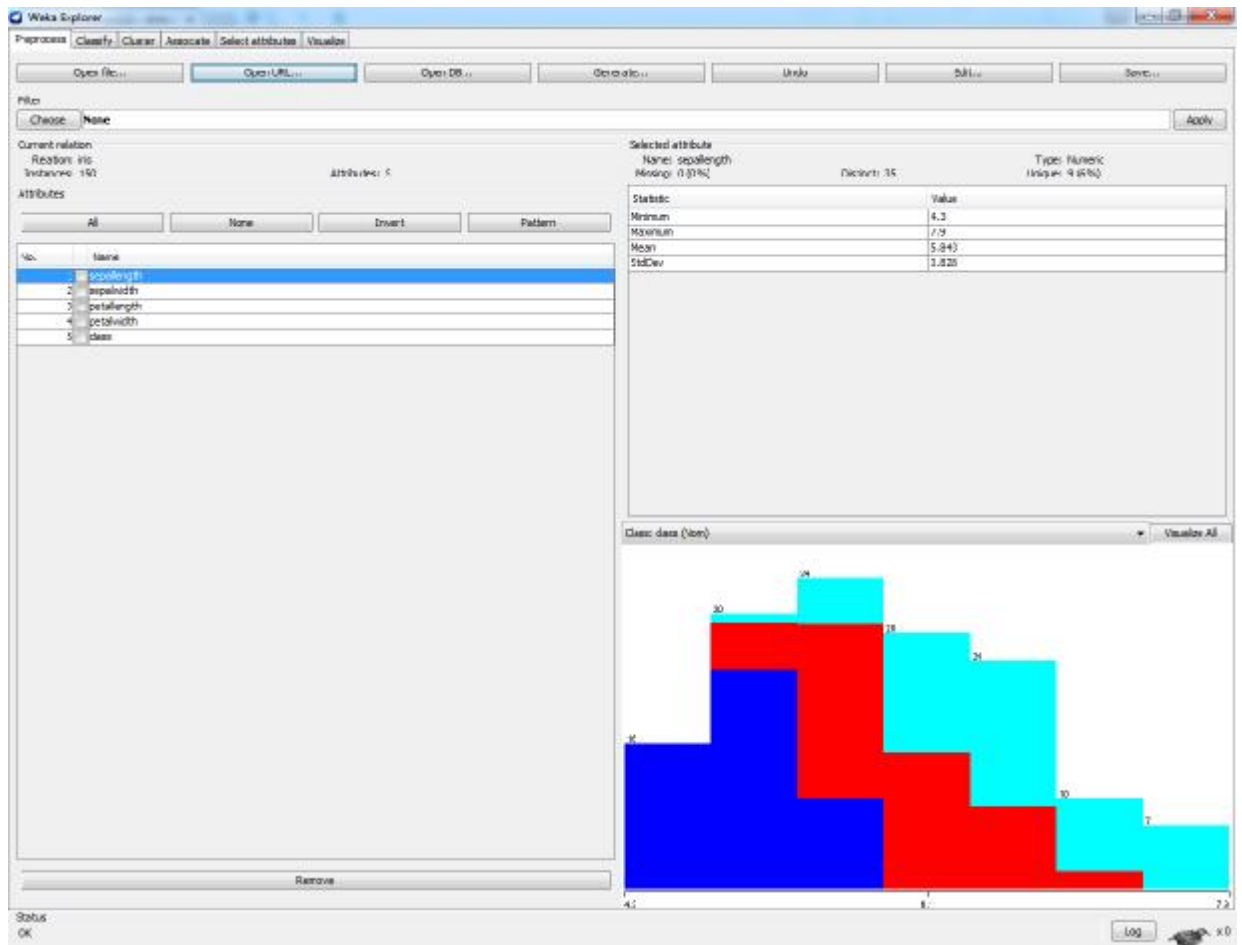


ΕΙΚΟΝΑ 15: ΑΡΧΙΚΗ ΟΘΟΝΗ WEKA

ΣΤΗΝ ΕΠΟΜΕΝΗ ΟΘΟΝΗ Ο ΧΡΗΣΤΗΣ ΚΑΛΕΙΤΑΙ ΝΑ ΕΠΙΛΕΞΕΙ ΤΟ ΑΡΧΕΙΟ ΠΟΥ ΘΑ ΕΞΕΤΑΣΕΙ. ΚΑΝΕΙ ΚΛΙΚ ΣΤΟ ΠΛΗΚΤΡΟ «OPENFILE» ΚΑΙ ΤΟ ΕΠΙΛΕΓΕΙ ΑΝ ΑΥΤΟ ΒΡΙΣΚΕΤΑΙ ΣΤΟΝ ΤΟΠΙΚΟ ΤΟΥ ΔΙΣΚΟ. ΣΕ ΠΕΡΙΠΤΩΣΗ ΟΜΩΣ ΠΟΥ ΒΡΙΣΚΕΤΑΙ ΣΕ ΜΙΑ ΑΠΟΜΑΚΡΥΣΜΕΝΗ ΤΟΠΟΘΕΣΙΑ ΜΠΟΡΕΙ ΝΑ ΤΟ ΕΠΙΛΕΞΕΙ ΚΑΝΟΝΤΑΣ ΚΛΙΚ ΣΤΟ ΠΛΗΚΤΡΟ «OPENURL» ΚΑΙ ΚΑΤΑΧΩΡΩΝΤΑΣ ΤΟ ΕΠΙΘΥΜΗΤΟ URL (ΣΤΗΝ ΕΞΕΤΑΖΟΜΕΝΗ ΠΕΡΙΠΤΩΣΗ ΤΟ [HTTP://STORM.CIS.FORDHAM.EDU/~GWEISS/DATA-MINING/WEKA-DATA/IRIS.ARFF](http://storm.cis.fordham.edu/~gweiss/data-mining/weka-data/iris.arff)).



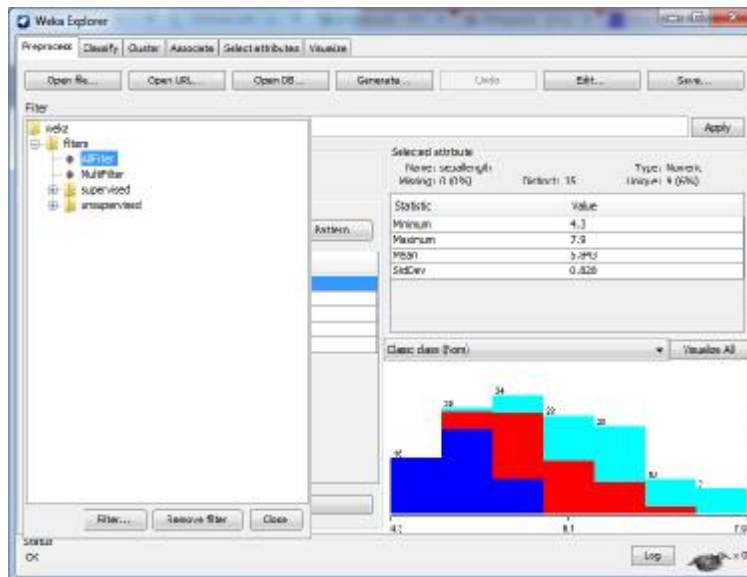
ΕΙΚΟΝΑ 16: ΕΠΙΛΟΓΗ ΣΥΝΟΛΟΥ ΔΕΔΟΜΕΝΩΝ ΑΠΟ ΤΟ ΔΙΑΔΙΚΤΥΟ



ΕΙΚΟΝΑ 17: ΑΝΑΚΤΗΣΗ ΣΤΟΙΧΕΙΩΝ ΤΟΥ ΣΥΝΟΛΟΥ ΔΕΔΟΜΕΝΩΝ

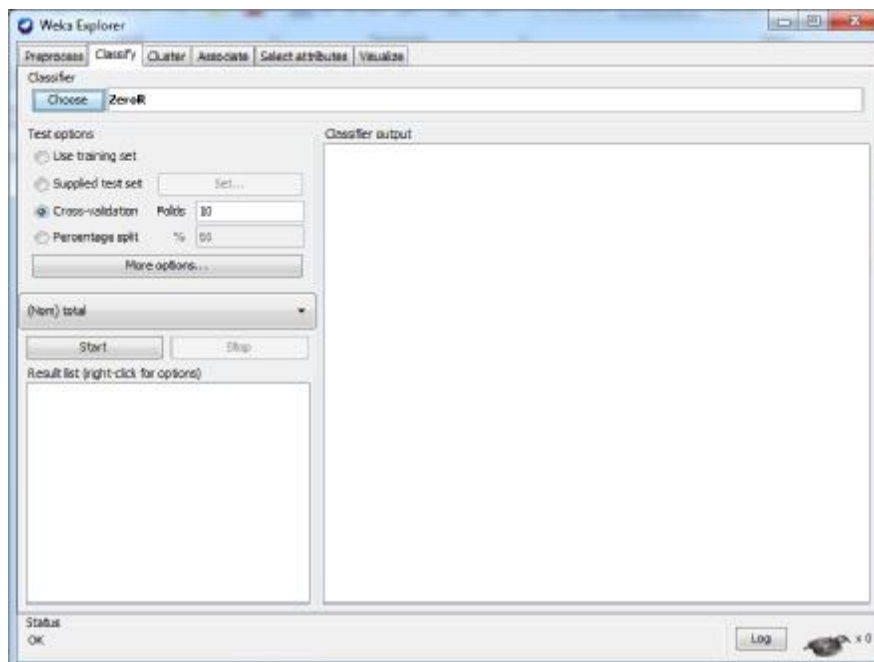
ΚΑΤΑ ΤΗΝ ΠΡΟΕΠΕΞΕΡΓΑΣΙΑ ΠΟΥ ΣΥΝΤΕΛΕΙΤΑΙ ΣΤΗΝ ΠΡΩΤΗ ΚΑΡΤΕΛΑ ΤΗΣ ΕΦΑΡΜΟΓΗΣ ΜΠΟΡΕΙ ΝΑ ΕΠΙΛΕΧΘΟΥΝ ΚΑΙ ΝΑ ΕΦΑΡΜΟΣΤΟΥΝ ΦΙΛΤΡΑ ΣΤΑ ΔΕΔΟΜΕΝΑ ΜΕΣΑ ΑΠΟ ΜΙΑ ΑΡΚΕΤΑ ΜΕΓΑΛΗ ΠΟΙΚΙΛΙΑ. ΓΙΑ ΝΑ ΓΙΝΕΙ ΑΥΤΟ ΘΑ ΠΡΕΠΕΙ ΝΑ ΓΙΝΕΙ ΚΛΙΚ ΣΤΟ ΠΛΗΚΤΡΟ CHOOSE ΚΑΤΩ ΑΠΟ ΤΗΝ ΕΤΙΚΕΤΑ FILTER. ΓΙΑ ΠΑΡΑΔΕΙΓΜΑ ΜΠΟΡΕΙ ΝΑ ΕΠΙΛΕΓΕΙ ΝΑ ΑΦΑΙΡΕΘΕΙ ΚΑΠΟΙΑ ΑΠΟ ΤΙΣ ΔΙΑΘΕΣΙΜΕΣ ΜΕΤΑΒΛΗΤΕΣ Η ΝΑ ΓΙΝΕΙ ΜΕΤΑΒΟΛΗ ΤΩΝ ΣΥΝΕΧΩΝ ΤΙΜΩΝ ΣΕ ΔΙΑΚΡΙΤΕΣ. ΣΤΗΝ ΠΑΡΟΥΣΑ ΕΚΤΕΛΕΣΗ ΕΦΑΡΜΟΣΤΗΚΕ ΤΟ ΔΕΥΤΕΡΟ ΜΕ ΤΗΝ ΕΠΙΛΟΓΗ ΔΙΑΔΟΧΙΚΑ

WEKA.FILTERSà UNSUPERVISEDà ATTRIBUTEà DISCRETIZE

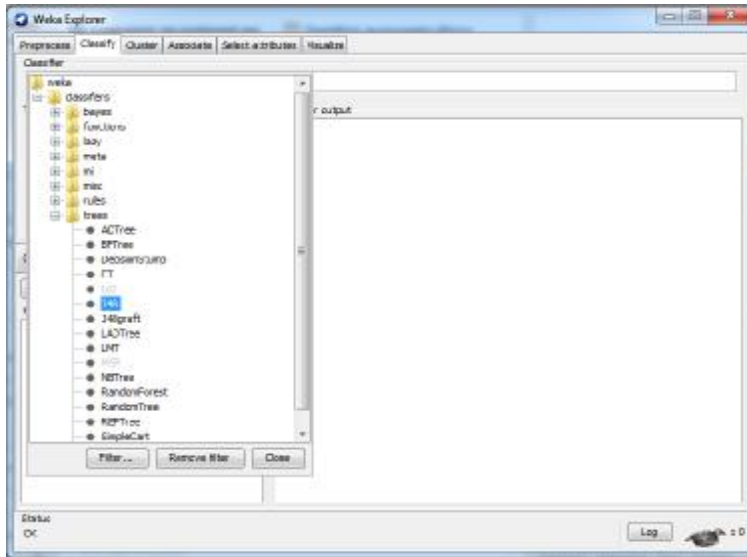


ΕΙΚΟΝΑ 18: ΕΠΙΛΟΓΗ ΦΙΛΤΡΩΝ ΣΤΟ WEKA

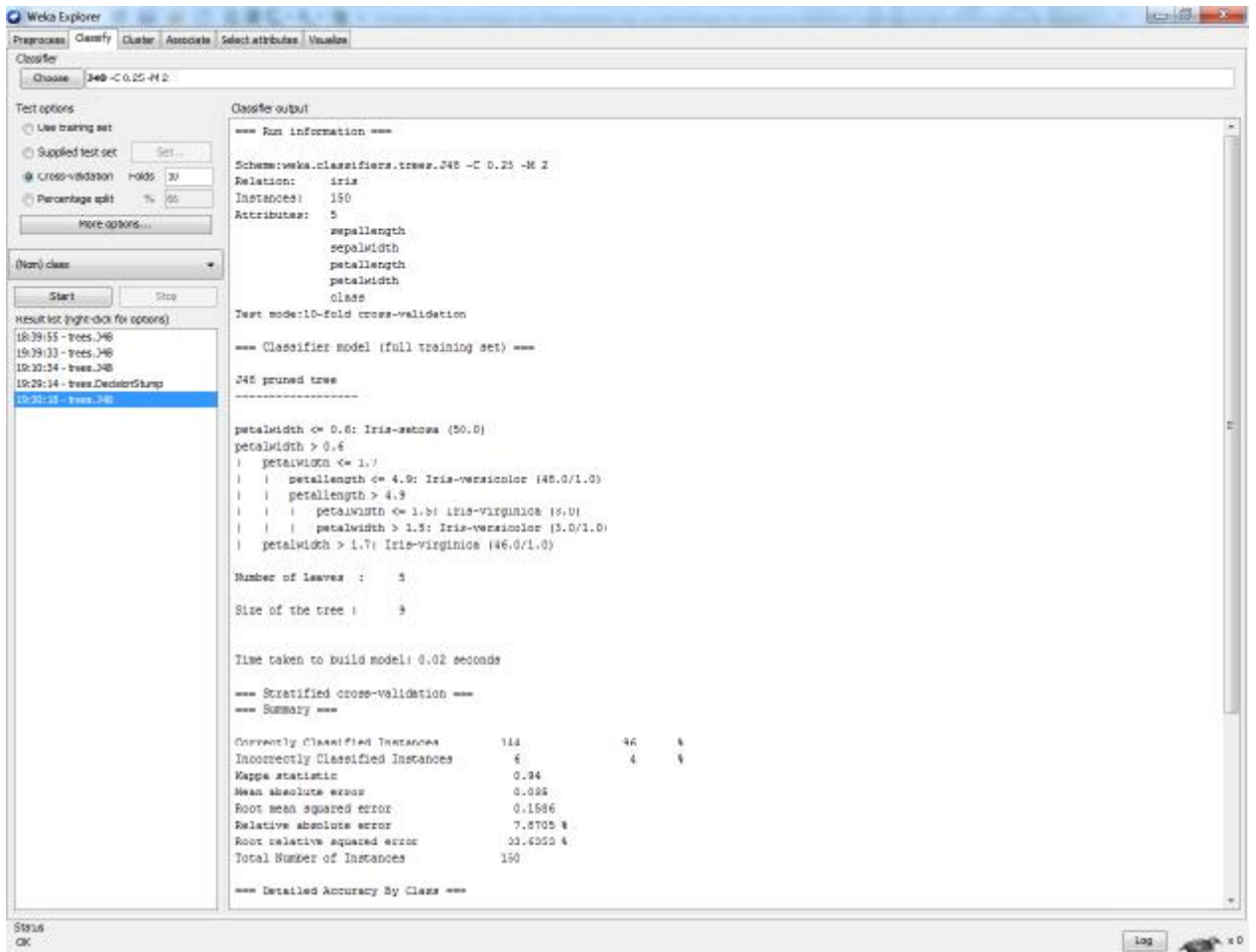
Η ΕΠΟΜΕΝΗ ΚΙΝΗΣΗ ΤΟΥ ΧΡΗΣΤΗ ΕΙΝΑΙ ΝΑ ΚΑΝΕΙ ΚΛΙΚ ΣΤΗΝ ΚΑΡΤΕΛΑ «CLASSIFY».



ΕΠΕΙΤΑ ΧΡΕΙΑΖΕΤΑΙ ΝΑ ΕΠΙΛΕΞΕΙ ΤΟΝ ΕΠΙΘΥΜΗΤΟ ΑΛΓΟΡΙΘΜΟ. ΑΥΤΟ ΓΙΝΕΤΑΙ ΜΕ ΚΛΙΚ ΣΤΟ ΠΛΗΚΤΡΟ «CHOOSE» ΚΑΙ ΕΠΙΛΟΓΗ «J48».



ΠΡΟΚΕΙΜΕΝΟΥ ΝΑ ΓΙΝΕΙ Η ΕΚΤΕΛΕΣΗ ΤΟΥ ΑΛΓΟΡΙΘΜΟΥ ΜΕ ΤΙΣ ΠΡΟΚΑΘΟΡΙΣΜΕΝΕΣ ΡΥΘΜΙΣΕΙΣ ΠΡΕΠΕΙ ΝΑ ΚΑΝΕΙ ΚΛΙΚ ΣΤΟ ΠΛΗΚΤΡΟ «START». ΜΕ ΤΗΝ ΟΛΟΚΛΗΡΩΣΗ ΤΗΣ ΕΚΤΕΛΕΣΗ ΤΟΥ ΑΛΓΟΡΙΘΜΟΥ ΠΑΡΟΥΣΙΑΖΟΝΤΑΙ ΤΑ ΑΠΟΤΕΛΕΣΜΑΤΑ ΣΕ ΜΟΡΦΗ ΑΠΛΟΥ ΚΕΙΜΕΝΟΥ.



ΕΙΚΟΝΑ 19: ΑΠΟΤΕΛΕΣΜΑΤΑ WEKA

ΤΑ ΑΠΟΤΕΛΕΣΜΑΤΑ ΠΟΥ ΕΛΗΦΘΗΣΑΝ ΗΤΑΝ

```

=== RUNINFORMATION ===

SCHEME:WEKA.CLASSIFIERS.TREES.J48 -C 0.25 -M 2
RELATION: IRIS
INSTANCES: 150
ATTRIBUTES: 5
SEPALLENGTH
SEPALWIDTH
PETALLENGTH
PETALWIDTH
CLASS
TEST MODE:10-FOLD CROSS-VALIDATION

=== CLASSIFIER MODEL (FULL TRAINING SET) ===

J48 PRUNED TREE
-----

PETALWIDTH<= 0.6: IRIS-SETOSA (50.0)
PETALWIDTH> 0.6
| PETALWIDTH<= 1.7
| | PETALLENGTH<= 4.9: IRIS-VERSICOLOR (48.0/1.0)
| | PETALLENGTH> 4.9
| | | PETALWIDTH<= 1.5: IRIS-VIRGINICA (3.0)
| | | PETALWIDTH> 1.5: IRIS-VERSICOLOR (3.0/1.0)
| PETALWIDTH> 1.7: IRIS-VIRGINICA (46.0/1.0)

NUMBER OF LEAVES :      5

SIZE OF THE TREE :      9

TIME TAKEN TO BUILD MODEL: 0.02 SECONDS

=== STRATIFIED CROSS-VALIDATION ===
=== SUMMARY ===

CORRECTLY CLASSIFIED INSTANCES      144      96 %
INCORRECTLY CLASSIFIED INSTANCES     6        4 %
KAPPA STATISTIC                      0.94
MEAN ABSOLUTE ERROR                   0.035
ROOT MEAN SQUARED ERROR                0.1586
RELATIVE ABSOLUTE ERROR                7.8705 %
ROOT RELATIVE SQUARED ERROR           33.6353 %
TOTAL NUMBER OF INSTANCES             150

=== DETAILED ACCURACY BY CLASS ===

      TP RATE  FP RATE  PRECISION  RECALL  F-MEASURE  ROC AREA  CLASS
      0.98    0        1        0.98   0.99    0.99    IRIS-SETOSA
      0.94    0.03    0.94    0.94   0.94    0.952   IRIS-VERSICOLOR
      0.96    0.03    0.941   0.96   0.95    0.961   IRIS-VIRGINICA
WEIGHTED AVG.  0.96    0.02    0.96    0.96   0.96    0.968

=== CONFUSION MATRIX ===

A B C <-- CLASSIFIED AS
49 1 0 | A = IRIS-SETOSA
 0 47 3 | B = IRIS-VERSICOLOR
2 48 | C = IRIS-VIRGINICA

```

ΣΕ ΜΙΑ ΑΛΛΗ ΕΚΤΕΛΕΣΗ ΤΟΥ ΙΔΙΟΥ ΑΛΓΟΡΙΘΜΟΥ ΣΤΟ ΙΔΙΟ ΣΥΝΟΛΟ ΔΕΔΟΜΕΝΩΝ ΕΠΙΛΕΧΘΗΚΕ CROSSVALIDATION 5 ΚΑΙ ΖΗΤΗΘΗΚΕ ΝΑ ΛΗΦΘΟΥΝ ΠΕΡΙΣΣΟΤΕΡΕΣ ΠΛΗΡΟΦΟΡΙΕΣ ΣΤΗΝ ΕΞΟΔΟ. ΤΟ ΑΠΟΤΕΛΕΣΜΑ ΗΤΑΝ ΤΟ ΠΑΡΑΚΑΤΩ.

```

=== RUN INFORMATION ===

SCHEME:WEKA.CLASSIFIERS.TREES.J48 -C 0.25 -M 2
RELATION: IRIS
INSTANCES: 150
ATTRIBUTES: 5
SEPALLENGTH
SEPALWIDTH
PETALLENGTH
PETALWIDTH
CLASS
TEST MODE:5-FOLD CROSS-VALIDATION

=== CLASSIFIER MODEL (FULL TRAINING SET) ===

J48 PRUNED TREE
-----

PETALWIDTH<= 0.6: IRIS-SETOSA (50.0)
PETALWIDTH> 0.6
| PETALWIDTH<= 1.7
| | PETALLENGTH<= 4.9: IRIS-VERSICOLOR (48.0/1.0)
| | | PETALLENGTH> 4.9
| | | | PETALWIDTH<= 1.5: IRIS-VIRGINICA (3.0)
| | | | | PETALWIDTH> 1.5: IRIS-VERSICOLOR (3.0/1.0)
| | | | | PETALWIDTH> 1.7: IRIS-VIRGINICA (46.0/1.0)

NUMBER OF LEAVES :      5

SIZE OF THE TREE :      9

TIME TAKEN TO BUILD MODEL: 0 SECONDS

=== PREDICTIONS ON TEST DATA ===

INST#,  ACTUAL, PREDICTED, ERROR, PROBABILITY DISTRIBUTION
1 3:IRIS-VIR 3:IRIS-VIR      0  0.027 *0.973
2 3:IRIS-VIR 3:IRIS-VIR      0  0.027 *0.973
3 3:IRIS-VIR 3:IRIS-VIR      0  0.027 *0.973
4 3:IRIS-VIR 3:IRIS-VIR      0  0.027 *0.973
5 3:IRIS-VIR 3:IRIS-VIR      0  0.027 *0.973
6 3:IRIS-VIR 3:IRIS-VIR      0  0.027 *0.973
7 3:IRIS-VIR 3:IRIS-VIR      0  0.027 *0.973
8 3:IRIS-VIR 3:IRIS-VIR      0  0.027 *0.973
9 3:IRIS-VIR 3:IRIS-VIR      0  0.027 *0.973
10 3:IRIS-VIR 3:IRIS-VIR      0  0.25 *0.75
11 1:IRIS-SET 1:IRIS-SET      *1  0  0
12 1:IRIS-SET 1:IRIS-SET      *1  0  0
13 1:IRIS-SET 1:IRIS-SET      *1  0  0
14 1:IRIS-SET 1:IRIS-SET      *1  0  0
15 1:IRIS-SET 1:IRIS-SET      *1  0  0
16 1:IRIS-SET 1:IRIS-SET      *1  0  0
17 1:IRIS-SET 1:IRIS-SET      *1  0  0
18 1:IRIS-SET 1:IRIS-SET      *1  0  0
19 1:IRIS-SET 1:IRIS-SET      *1  0  0

```

20	1:IRIS-SET 1:IRIS-SET	*1	0	0
21	2:IRIS-VER 2:IRIS-VER	0	*0.974	0.026
22	2:IRIS-VER 2:IRIS-VER	0	*0.974	0.026
23	2:IRIS-VER 2:IRIS-VER	0	*0.974	0.026
24	2:IRIS-VER 2:IRIS-VER	0	*0.974	0.026
25	2:IRIS-VER 2:IRIS-VER	0	*0.974	0.026
26	2:IRIS-VER 2:IRIS-VER	0	*0.974	0.026
27	2:IRIS-VER 2:IRIS-VER	0	*0.974	0.026
28	2:IRIS-VER 2:IRIS-VER	0	*0.974	0.026
29	2:IRIS-VER 3:IRIS-VIR	+ 0	0.25	*0.75
30	2:IRIS-VER 2:IRIS-VER	0	*0.974	0.026
1	3:IRIS-VIR 3:IRIS-VIR	0	0.027	*0.973
2	3:IRIS-VIR 3:IRIS-VIR	0	0.027	*0.973
3	3:IRIS-VIR 3:IRIS-VIR	0	0.027	*0.973
4	3:IRIS-VIR 2:IRIS-VER	+ 0	*1	0
5	3:IRIS-VIR 3:IRIS-VIR	0	0.027	*0.973
6	3:IRIS-VIR 3:IRIS-VIR	0	0.027	*0.973
7	3:IRIS-VIR 3:IRIS-VIR	0	0.027	*0.973
8	3:IRIS-VIR 3:IRIS-VIR	0	0.027	*0.973
9	3:IRIS-VIR 3:IRIS-VIR	0	0.027	*0.973
10	3:IRIS-VIR 3:IRIS-VIR	0	0.027	*0.973
11	1:IRIS-SET 1:IRIS-SET	*1	0	0
12	1:IRIS-SET 1:IRIS-SET	*1	0	0
13	1:IRIS-SET 1:IRIS-SET	*1	0	0
14	1:IRIS-SET 1:IRIS-SET	*1	0	0
15	1:IRIS-SET 1:IRIS-SET	*1	0	0
16	1:IRIS-SET 2:IRIS-VER	+ 0	*1	0
17	1:IRIS-SET 1:IRIS-SET	*1	0	0
18	1:IRIS-SET 1:IRIS-SET	*1	0	0
19	1:IRIS-SET 1:IRIS-SET	*1	0	0
20	1:IRIS-SET 1:IRIS-SET	*1	0	0
21	2:IRIS-VER 2:IRIS-VER	0	*1	0
22	2:IRIS-VER 2:IRIS-VER	0	*1	0
23	2:IRIS-VER 2:IRIS-VER	0	*1	0
24	2:IRIS-VER 2:IRIS-VER	0	*1	0
25	2:IRIS-VER 2:IRIS-VER	0	*1	0
26	2:IRIS-VER 2:IRIS-VER	0	*1	0
27	2:IRIS-VER 2:IRIS-VER	0	*1	0
28	2:IRIS-VER 2:IRIS-VER	0	*1	0
29	2:IRIS-VER 2:IRIS-VER	0	*1	0
30	2:IRIS-VER 2:IRIS-VER	0	*1	0
1	3:IRIS-VIR 3:IRIS-VIR	0	0	*1
2	3:IRIS-VIR 3:IRIS-VIR	0	0	*1
3	3:IRIS-VIR 3:IRIS-VIR	0	0	*1
4	3:IRIS-VIR 3:IRIS-VIR	0	0	*1
5	3:IRIS-VIR 3:IRIS-VIR	0	0	*1
6	3:IRIS-VIR 3:IRIS-VIR	0	0	*1
7	3:IRIS-VIR 3:IRIS-VIR	0	0	*1
8	3:IRIS-VIR 2:IRIS-VER	+ 0	*0.975	0.025
9	3:IRIS-VIR 3:IRIS-VIR	0	0	*1
10	3:IRIS-VIR 3:IRIS-VIR	0	0	*1
11	1:IRIS-SET 1:IRIS-SET	*1	0	0
12	1:IRIS-SET 1:IRIS-SET	*1	0	0
13	1:IRIS-SET 1:IRIS-SET	*1	0	0
14	1:IRIS-SET 1:IRIS-SET	*1	0	0
15	1:IRIS-SET 1:IRIS-SET	*1	0	0
16	1:IRIS-SET 1:IRIS-SET	*1	0	0
17	1:IRIS-SET 1:IRIS-SET	*1	0	0
18	1:IRIS-SET 1:IRIS-SET	*1	0	0
19	1:IRIS-SET 1:IRIS-SET	*1	0	0
20	1:IRIS-SET 1:IRIS-SET	*1	0	0
21	2:IRIS-VER 2:IRIS-VER	0	*0.975	0.025
22	2:IRIS-VER 2:IRIS-VER	0	*0.975	0.025
23	2:IRIS-VER 2:IRIS-VER	0	*0.975	0.025
24	2:IRIS-VER 2:IRIS-VER	0	*0.975	0.025
25	2:IRIS-VER 2:IRIS-VER	0	*0.975	0.025
26	2:IRIS-VER 2:IRIS-VER	0	*0.975	0.025
27	2:IRIS-VER 2:IRIS-VER	0	*0.975	0.025
28	2:IRIS-VER 3:IRIS-VIR	+ 0	0	*1

29 2:IRIS-VER 2:IRIS-VER	0	*0.975	0.025
30 2:IRIS-VER 2:IRIS-VER	0	*0.975	0.025
1 3:IRIS-VIR 3:IRIS-VIR	0	0.027	*0.973
2 3:IRIS-VIR 3:IRIS-VIR	0	0.027	*0.973
3 3:IRIS-VIR 3:IRIS-VIR	0	0.027	*0.973
4 3:IRIS-VIR 3:IRIS-VIR	0	0.027	*0.973
5 3:IRIS-VIR 3:IRIS-VIR	0	0.027	*0.973
6 3:IRIS-VIR 3:IRIS-VIR	0	0	*1
7 3:IRIS-VIR 3:IRIS-VIR	0	0.027	*0.973
8 3:IRIS-VIR 3:IRIS-VIR	0	0.027	*0.973
9 3:IRIS-VIR 3:IRIS-VIR	0	0.027	*0.973
10 3:IRIS-VIR 3:IRIS-VIR	0	0.027	*0.973
11 1:IRIS-SET 1:IRIS-SET	*1	0	0
12 1:IRIS-SET 1:IRIS-SET	*1	0	0
13 1:IRIS-SET 1:IRIS-SET	*1	0	0
14 1:IRIS-SET 1:IRIS-SET	*1	0	0
15 1:IRIS-SET 1:IRIS-SET	*1	0	0
16 1:IRIS-SET 1:IRIS-SET	*1	0	0
17 1:IRIS-SET 1:IRIS-SET	*1	0	0
18 1:IRIS-SET 1:IRIS-SET	*1	0	0
19 1:IRIS-SET 1:IRIS-SET	*1	0	0
20 1:IRIS-SET 1:IRIS-SET	*1	0	0
21 2:IRIS-VER 2:IRIS-VER	0	*0.974	0.026
22 2:IRIS-VER 2:IRIS-VER	0	*0.974	0.026
23 2:IRIS-VER 2:IRIS-VER	0	*0.974	0.026
24 2:IRIS-VER 2:IRIS-VER	0	*0.974	0.026
25 2:IRIS-VER 2:IRIS-VER	0	*0.974	0.026
26 2:IRIS-VER 2:IRIS-VER	0	*0.974	0.026
27 2:IRIS-VER 2:IRIS-VER	0	*0.974	0.026
28 2:IRIS-VER 2:IRIS-VER	0	*0.974	0.026
29 2:IRIS-VER 2:IRIS-VER	0	*0.974	0.026
30 2:IRIS-VER 2:IRIS-VER	0	*0.974	0.026
1 3:IRIS-VIR 3:IRIS-VIR	0	0.026	*0.974
2 3:IRIS-VIR 3:IRIS-VIR	0	0.026	*0.974
3 3:IRIS-VIR 3:IRIS-VIR	0	0.026	*0.974
4 3:IRIS-VIR 3:IRIS-VIR	0	0.026	*0.974
5 3:IRIS-VIR 3:IRIS-VIR	0	0.25	*0.75
6 3:IRIS-VIR 3:IRIS-VIR	0	0.026	*0.974
7 3:IRIS-VIR 3:IRIS-VIR	0	0.026	*0.974
8 3:IRIS-VIR 3:IRIS-VIR	0	0.026	*0.974
9 3:IRIS-VIR 3:IRIS-VIR	0	0.026	*0.974
10 3:IRIS-VIR 3:IRIS-VIR	0	0.026	*0.974
11 1:IRIS-SET 1:IRIS-SET	*1	0	0
12 1:IRIS-SET 1:IRIS-SET	*1	0	0
13 1:IRIS-SET 1:IRIS-SET	*1	0	0
14 1:IRIS-SET 1:IRIS-SET	*1	0	0
15 1:IRIS-SET 1:IRIS-SET	*1	0	0
16 1:IRIS-SET 1:IRIS-SET	*1	0	0
17 1:IRIS-SET 1:IRIS-SET	*1	0	0
18 1:IRIS-SET 1:IRIS-SET	*1	0	0
19 1:IRIS-SET 1:IRIS-SET	*1	0	0
20 1:IRIS-SET 1:IRIS-SET	*1	0	0
21 2:IRIS-VER 2:IRIS-VER	0	*1	0
22 2:IRIS-VER 2:IRIS-VER	0	*1	0
23 2:IRIS-VER 2:IRIS-VER	0	*1	0
24 2:IRIS-VER 2:IRIS-VER	0	*1	0
25 2:IRIS-VER 3:IRIS-VIR	+ 0	0.026	*0.974
26 2:IRIS-VER 2:IRIS-VER	0	*1	0
27 2:IRIS-VER 2:IRIS-VER	0	*1	0
28 2:IRIS-VER 2:IRIS-VER	0	*1	0
29 2:IRIS-VER 2:IRIS-VER	0	*1	0
30 2:IRIS-VER 2:IRIS-VER	0	*1	0
=== STRATIFIED CROSS-VALIDATION ===			
=== SUMMARY ===			
CORRECTLY CLASSIFIED INSTANCES	144	96	%
INCORRECTLY CLASSIFIED INSTANCES	6	4	%
KAPPA STATISTIC	0.94		

```

K&B RELATIVE INFO SCORE      14001.6251 %
K&B INFORMATION SCORE       221.9205 BITS    1.4795 BITS/INSTANCE
CLASS COMPLEXITY | ORDER 0   237.7444 BITS    1.585 BITS/INSTANCE
CLASS COMPLEXITY | SCHEME   3237.8642 BITS    21.5858 BITS/INSTANCE
COMPLEXITY IMPROVEMENT (SF) -3000.1198 BITS  -20.0008 BITS/INSTANCE
MEAN ABSOLUTE ERROR         0.035
ROOT MEAN SQUARED ERROR     0.1582
RELATIVE ABSOLUTE ERROR     7.8842 %
ROOT RELATIVE SQUARED ERROR 33.5577 %
TOTAL NUMBER OF INSTANCES   150

=== DETAILED ACCURACY BY CLASS ===

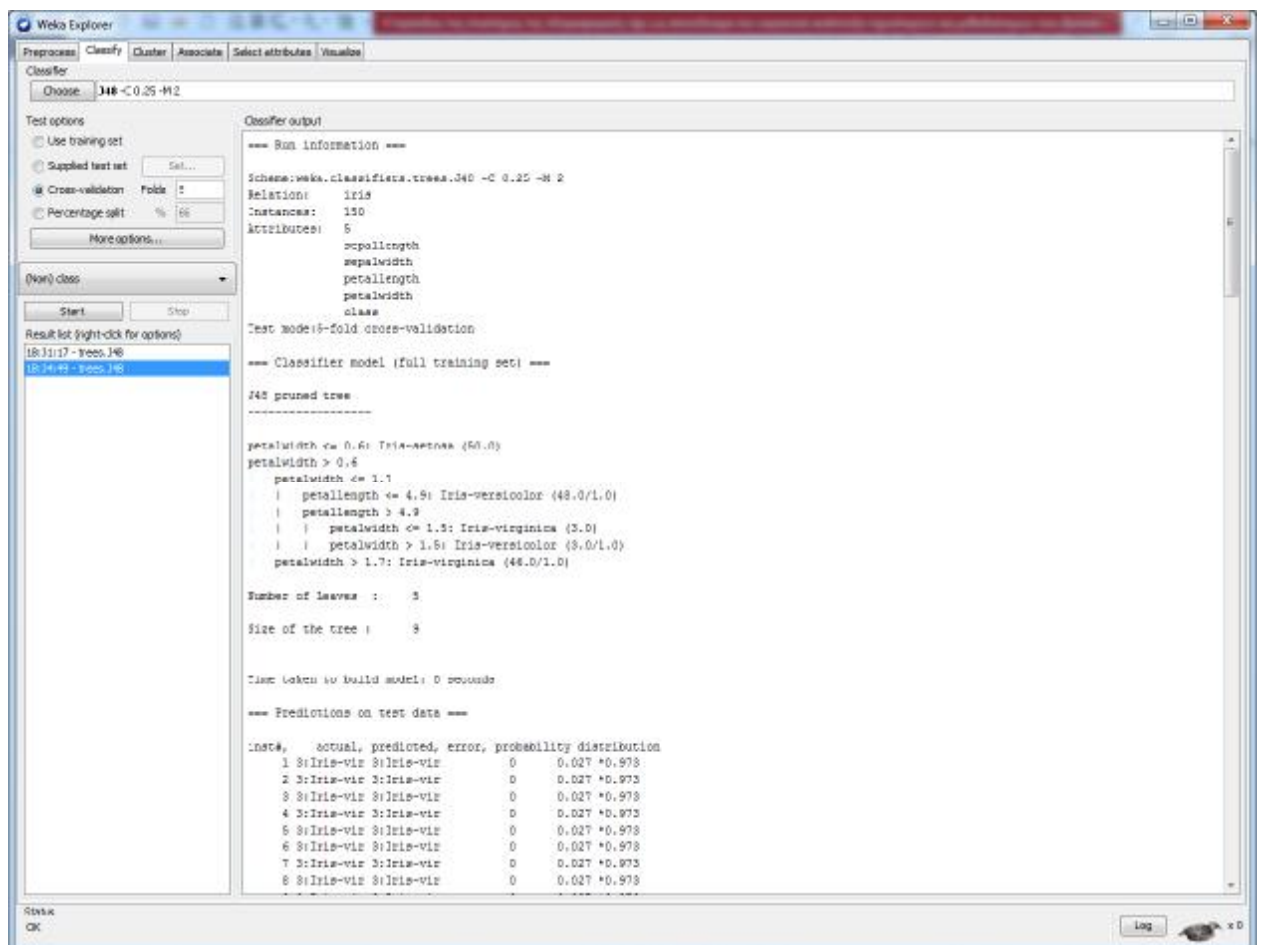
      TP RATE  FP RATE  PRECISION  RECALL  F-MEASURE  ROC AREA  CLASS
      0.98    0        1          0.98   0.99       0.99      IRIS-SETOSA
      0.94    0.03    0.94       0.94   0.94       0.958     IRIS-VERSICOLOR
      0.96    0.03    0.941      0.96   0.95       0.966     IRIS-VIRGINICA
WEIGHTED AVG.  0.96    0.02    0.96       0.96   0.96       0.971

=== CONFUSION MATRIX ===

A B C <-- CLASSIFIED AS
49 1 0 | A = IRIS-SETOSA
 0 47 3 | B = IRIS-VERSICOLOR
 0 2 48 | C = IRIS-VIRGINICA

```

Η ΘΘΟΝΗ ΤΟΥ ΑΠΟΤΕΛΕΣΜΑΤΟΣ ΕΧΕΙ ΤΗΝ ΠΑΡΑΚΑΤΩ ΜΟΡΦΗ.



ΕΙΚΟΝΑ 20: ΘΘΟΝΗ ΑΠΟΤΕΛΕΣΜΑΤΟΣ ΣΤΟ WEKA

ΑΠΟ ΤΑ ΑΠΟΤΕΛΕΣΜΑΤΑ ΠΟΥ ΕΛΗΦΘΗΣΑΝ ΣΤΗΝ ΔΕΥΤΕΡΗ ΕΚΤΕΛΕΣΗ ΔΙΑΚΡΙΝΟΝΤΑΙ ΤΟ ΟΝΟΜΑ ΤΟΥ ΣΥΝΟΛΟΥ ΔΕΔΟΜΕΝΩΝ, ΤΟ ΠΛΗΘΟΣ ΤΩΝ ΠΑΡΑΜΕΤΡΩΝ ΠΟΥ ΕΛΕΓΧΘΗΚΑΝ ΚΑΙ ΤΟ ΟΤΙ ΕΠΙΛΕΧΘΗΚΕ CROSSVALIDATION 5 ΓΙΑ ΤΗΝ ΕΠΙΚΥΡΩΣΗ. ΚΑΘΟΡΙΖΕΙ ΟΤΙ ΤΟ ΔΕΝΔΡΟ ΕΙΝΑΙ ΤΟ

```

PETALWIDTH<= 0.6: IRIS-SETOSA (50.0)
PETALWIDTH> 0.6
| PETALWIDTH<= 1.7
| | PETALLENGTH<= 4.9: IRIS-VERSICOLOR (48.0/1.0)
| | PETALLENGTH> 4.9
| | | PETALWIDTH<= 1.5: IRIS-VIRGINICA (3.0)
| | | PETALWIDTH> 1.5: IRIS-VERSICOLOR (3.0/1.0)
| PETALWIDTH> 1.7: IRIS-VIRGINICA (46.0/1.0)

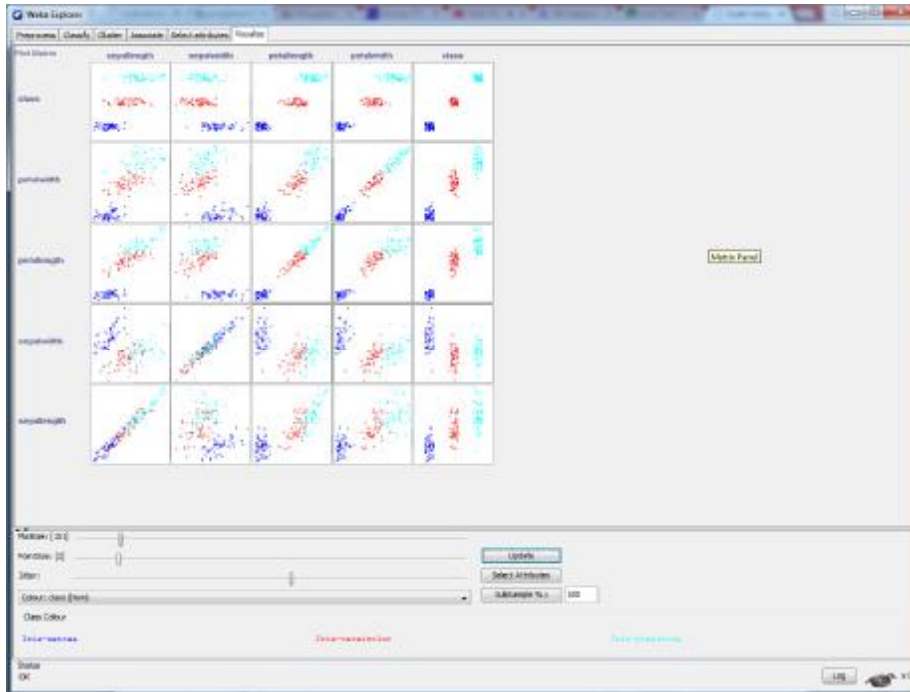
```

ΜΕ 5 ΦΥΛΛΑΚΑΙΜΕΓΕΘΟΣ 9. ΣΤΑ ΦΥΛΛΑ ΤΟΥ ΔΕΝΔΡΟΥ ΦΑΙΝΟΝΤΑΙ ΤΑ ΣΤΙΓΜΙΟΤΥΠΑ ΠΟΥ ΚΑΤΗΓΟΡΙΟΠΟΙΟΥΝΤΑΙ ΣΤΗΝ ΚΑΤΗΓΟΡΙΑ ΑΥΤΗ ΚΑΘΩΣ ΚΑΙ ΠΟΣΑ ΣΤΙΓΜΙΟΤΥΠΑ ΚΑΤΑΤΑΧΘΗΚΑΝ ΕΣΦΑΛΜΕΝΑ ΣΕ ΑΥΤΗΝ. ΠΕΡΙΛΑΜΒΑΝΟΝΤΑΙ ΑΝΑΛΥΤΙΚΕΣ ΠΛΗΡΟΦΟΡΙΕΣ ΓΙΑ ΤΑ ΣΤΙΓΜΙΟΤΥΠΑ ΤΩΝ ΔΕΔΟΜΕΝΩΝ. ΤΕΛΟΣ ΚΑΤΑΓΡΑΦΟΝΤΑΙ ΠΛΗΡΟΦΟΡΙΕΣ ΠΟΥ ΑΦΟΡΟΥΝ ΤΗΝ ΑΠΟΔΟΣΗΣ ΤΟΥ ΔΕΝΔΡΟΥ ΟΠΩΣ:

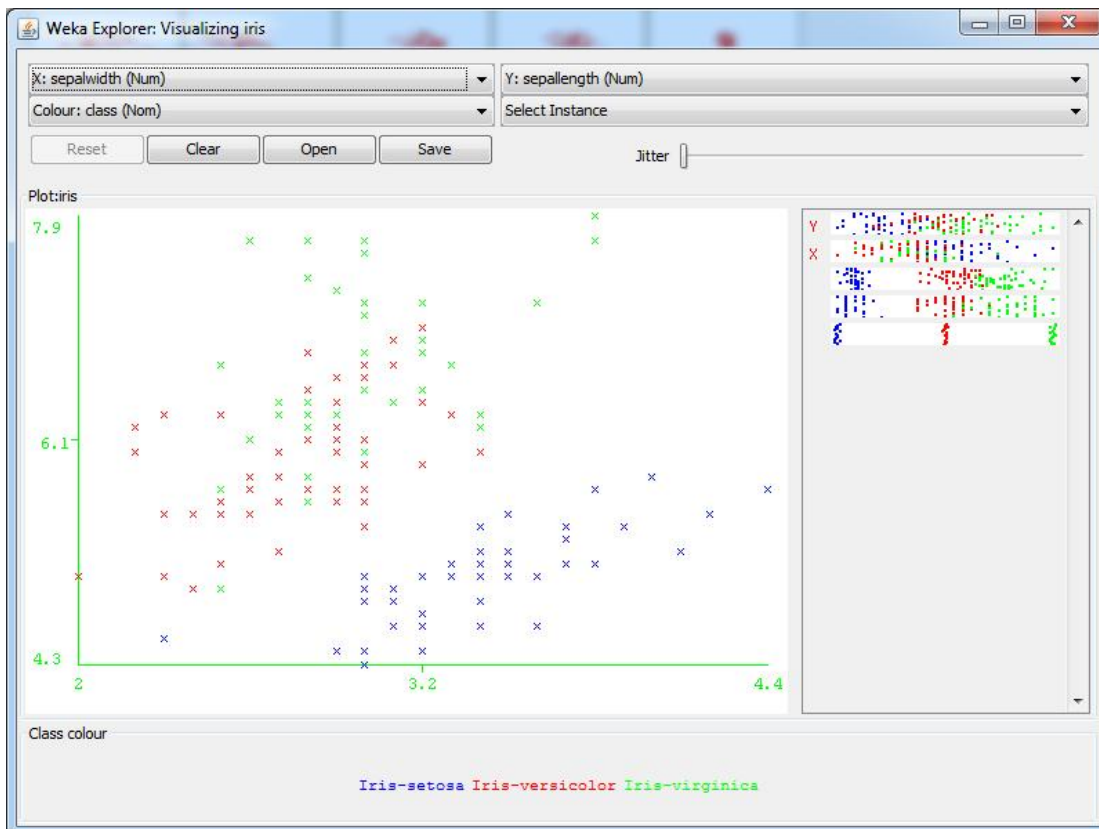
- ΣΩΣΤΑ ΤΑΞΙΝΟΜΗΜΕΝΑ ΣΤΙΓΜΙΟΤΥΠΑ (96%)
- ΕΣΦΑΛΜΕΝΑ ΤΑΞΙΝΟΜΗΜΕΝΑ (4%)
- ΚΑΡΡΑ ΣΤΑΤΙΣΤΙΚΗ (0.94)
- ΜΕΣΟ ΑΠΟΛΥΤΟ ΣΦΑΛΜΑ (0.035)
- ΡΙΖΑ ΤΟΥ ΜΕΣΟΥ ΤΕΤΡΑΓΩΝΙΚΟΥ ΣΦΑΛΜΑΤΟΣ (0,1582)
- ΣΧΕΤΙΚΗ ΑΠΟΛΥΤΟ ΣΦΑΛΜΑ (7,8842%)
- ROOT ΣΧΕΤΙΚΗ ΤΕΤΡΑΓΩΝΙΚΟ ΣΦΑΛΜΑ (33,5577%)
- ΣΥΝΟΛΙΚΟΣ ΑΡΙΘΜΟΣ ΤΩΝ ΣΤΙΓΜΙΟΤΥΠΩΝ (150)

ΤΟ ΕΡΓΑΛΕΙΟ ΠΑΡΕΧΕΙ ΚΑΙ ΓΡΑΦΙΚΗ ΑΠΕΙΚΟΝΙΣΗ ΤΩΝ ΑΠΟΤΕΛΕΣΜΑΤΩΝ ΣΤΗΝ ΚΑΡΤΕΛΑ «VISUALIZE». ΣΕ ΑΥΤΕΣ

ΦΑΙΝΟΝΤΑΙ ΔΙΑΓΡΑΜΜΑΤΙΚΑ Η ΣΥΜΠΕΡΙΦΟΡΑ ΤΩΝ ΤΙΜΩΝ ΤΩΝ ΠΑΡΑΜΕΤΡΩΝ ΣΕ ΑΝΑ ΔΥΟ ΣΧΕΣΕΙΣ.



ΕΙΚΟΝΑ 21: ΓΡΑΦΙΚΗ ΑΠΕΙΚΟΝΙΣΗ ΑΠΟΤΕΛΕΣΜΑΤΩΝ



ΕΙΚΟΝΑ 22: ΛΕΠΤΟΜΕΡΕΙΑ ΓΡΑΦΙΚΗΣ ΑΠΕΙΚΟΝΙΣΗΣ

4. ΣΥΜΠΕΡΑΣΜΑΤΑ

Η ΕΞΕΛΙΞΗ ΤΗΣ ΕΠΙΣΤΗΜΗΣ ΕΙΝΑΙ ΜΙΑ ΑΕΝΑΗ ΔΙΑΔΙΚΑΣΙΑ. ΚΑΘΕ ΑΝΑΒΑΘΜΙΣΗ – ΒΕΛΤΙΩΣΗ – ΚΑΙΝΟΤΟΜΙΑ ΕΧΕΙ ΣΑΝ ΕΠΑΚΟΛΟΥΘΑ ΠΟΛΛΑ ΠΛΕΟΝΕΚΤΗΜΑΤΑ ΣΕ ΤΟΜΕΙΣ ΟΠΟΥ ΕΙΝΑΙ ΔΥΝΑΤΟΝ ΝΑ ΕΦΑΡΜΟΣΤΕΙ. ΠΑΡΑΛΛΗΛΑ ΤΙΣ ΠΕΡΙΣΣΟΤΕΡΕΣ ΦΟΡΕΣ ΣΥΝΟΔΕΥΕΤΑΙ ΑΠΟ ΕΝΑ ΣΥΝΟΛΟ ΖΗΤΗΜΑΤΩΝ – ΑΛΛΟΤΕ ΜΙΚΡΟ ΑΛΛΟΤΕ ΜΕΓΑΛΩΝ – ΠΟΥ ΕΓΕΙΡΟΝΤΑΙ ΚΑΙ ΧΡΗΣΟΥΝ ΑΠΟΤΕΛΕΣΜΑΤΙΚΗΣ ΑΝΤΙΜΕΤΩΠΙΣΗΣ ΩΣΤΕ ΤΑ ΟΦΕΛΗ ΤΟΥΣ ΝΑ ΣΥΝΕΧΙΖΟΥΝ ΝΑ ΕΙΝΑΙ ΕΠΙΘΥΜΗΤΑ. ΈΤΣΙ ΠΑΡΟΥΣΙΑΖΟΝΤΑΙ ΕΚ ΝΕΟΥ ΑΝΑΒΑΘΜΙΣΕΙΣ – ΒΕΛΤΙΩΣΕΙΣ – ΚΑΙΝΟΤΟΜΙΕΣ ΠΡΟΚΕΙΜΕΝΟΥ ΝΑ ΑΝΤΙΜΕΤΩΠΙΣΤΟΥΝ ΖΗΤΗΜΑΤΑ ΠΟΥ ΠΡΟΚΥΠΤΟΥΝ. Η ΑΝΑΔΡΟΜΙΚΗΣ ΑΥΤΗ ΔΙΑΔΙΚΑΣΙΑ ΚΑΘΙΣΤΑ ΤΗΝ ΕΞΕΛΙΞΗ ΣΥΝΕΧΗ. Η ΑΝΑΠΤΥΞΗ ΤΩΝ ΤΕΧΝΟΛΟΓΙΩΝ ΤΗΣ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΩΝ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ ΔΗΜΙΟΥΡΓΗΣΕ ΚΑΤΑΛΛΗΛΕΣ ΣΥΝΘΗΚΕΣ ΓΙΑ ΤΗΝ ΑΠΟΘΗΚΕΥΣΗ ΚΑΙ ΤΗΝ ΕΠΕΞΕΡΓΑΣΙΑ ΜΕΓΑΛΩΝ ΟΓΚΩΝ ΔΕΔΟΜΕΝΩΝ. ΠΑΡΑΛΛΗΛΑ ΔΗΜΙΟΥΡΓΗΘΗΚΕ Η ΑΝΑΓΚΗ ΝΑ ΕΝΤΟΠΙΣΤΟΥΝ ΤΕΧΝΙΚΕΣ ΚΑΙ ΜΕΘΟΔΟΛΟΓΙΕΣ ΓΙΑ ΤΗΝ ΑΠΟΤΕΛΕΣΜΑΤΙΚΗ ΔΙΑΧΕΙΡΙΣΗ ΑΥΤΟΥ ΤΟΥ ΟΓΚΟΥ ΠΛΗΡΟΦΟΡΙΩΝ. Η ΛΥΣΗ ΣΤΑ ΖΗΤΗΜΑΤΑ ΑΥΤΑ ΕΙΝΑΙ Η ΕΦΑΡΜΟΓΗ ΤΩΝ ΜΕΘΟΔΟΛΟΓΙΩΝ ΤΗΣ ΕΞΟΡΥΞΗΣ ΔΕΔΟΜΕΝΩΝ. Ο ΣΤΟΧΟΣ ΠΛΕΟΝ ΤΩΝ ΟΡΓΑΝΙΣΜΩΝ ΔΕΝ ΕΙΝΑΙ Η ΑΠΟΚΤΗΣΗ Η Η ΠΡΟΣΒΑΣΗ ΣΕ ΜΕΓΑΛΑ ΣΥΝΟΛΑ ΔΕΔΟΜΕΝΩΝ ΑΛΛΑ Η ΠΑΡΑΓΩΓΗ ΓΝΩΣΗΣ ΑΠΟ ΑΥΤΑ. Η ΕΡΕΥΝΑ ΣΤΟΝ ΤΟΜΕΣ ΑΥΤΟ ΑΝΕΔΕΙΞΕ ΤΕΧΝΙΚΕΣ ΚΑΙ ΜΕΘΟΔΟΛΟΓΙΕΣ ΟΙ ΟΠΟΙΕΣ ΕΧΟΥΝ ΤΗΝ ΔΥΝΑΤΟΤΗΤΑ ΝΑ ΦΙΛΤΡΑΡΟΥΝ ΤΑ ΩΦΕΛΙΜΑ ΔΕΔΟΜΕΝΑ ΚΑΙ ΜΕΣΩ ΑΥΤΩΝ ΝΑ ΠΑΡΑΓΟΥΝ ΠΡΟΤΥΠΑ ΤΑ ΟΠΟΙΑ ΠΕΡΙΓΡΑΦΟΥΝ ΜΕ ΑΡΚΕΤΑ ΜΕΓΑΛΗ ΑΚΡΙΒΕΙΑ ΤΡΕΧΟΥΣΕΣ ΑΛΛΑ ΚΑΙ ΠΙΘΑΝΕΣ ΜΕΛΛΟΝΤΙΚΕΣ ΚΑΤΑΣΤΑΣΕΙΣ.

ΟΙ ΔΙΑΔΙΚΑΣΙΕΣ ΕΞΟΡΥΞΗΣ ΓΝΩΣΗΣ, ΜΕΡΟΣ ΤΗΣ ΟΠΟΙΑΣ ΕΙΝΑΙ Η ΕΞΟΡΥΞΗ ΔΕΔΟΜΕΝΩΝ ΑΝΑΒΑΘΜΙΖΕΙ ΤΗΝ ΛΕΙΤΟΥΡΓΙΑ ΤΩΝ ΟΡΓΑΝΙΣΜΩΝ. ΤΑ ΠΛΗΡΟΦΟΡΙΑΚΑ ΤΟΥΣ ΣΥΣΤΗΜΑΤΑ ΠΡΟΧΩΡΑΝ ΕΝΑΝ ΒΗΜΑ ΠΑΡΑΠΕΡΑ ΑΠΟ ΤΗΝ ΑΠΟΘΗΚΕΥΣΗ ΚΑΙ ΕΠΕΞΕΡΓΑΣΙΑ

ΔΕΔΟΜΕΝΩΝ ΓΙΑ ΤΗΝ ΠΑΡΑΓΩΓΗΣ ΠΛΗΡΟΦΟΡΙΩΝ, ΑΓΑΘΩΝ ΚΑΙ ΥΠΗΡΕΣΙΩΝ. ΈΧΟΥΝ ΠΛΕΟΝ ΤΗΝ ΔΥΝΑΤΟΤΗΤΑ ΝΑ ΓΝΩΡΙΖΟΥΝ ΚΑΙ ΝΑ ΠΡΟΒΛΕΠΟΥΝ. ΤΑ ΠΛΗΡΟΦΟΡΙΑΚΑ ΣΥΣΤΗΜΑΤΑ ΠΛΕΟΝ ΕΞΟΠΛΙΖΟΝΤΑΙ ΜΕ ΕΥΦΥΪΑ. ΤΑ ΑΠΟΘΗΚΕΥΜΕΝΑ ΔΕΔΟΜΕΝΑ ΑΠΟ ΣΤΑΤΙΚΗ ΠΕΡΙΟΥΣΙΑ ΤΩΝ ΟΡΓΑΝΙΣΜΩΝ ΜΕΤΑΤΡΕΠΟΝΤΑΙ ΣΕ ΔΥΝΑΜΙΚΕΣ ΠΗΓΕΣ ΓΝΩΣΗΣ ΚΑΘΩΣ ΣΥΝΔΥΑΖΟΜΕΝΑ ΕΙΝΑΙ ΙΚΑΝΑ ΝΑ ΑΠΟΚΑΛΥΠΤΟΥΝ ΤΑΣΕΙΣ ΚΑΙ ΚΑΝΟΝΕΣ ΠΟΥ ΔΕΝ ΕΙΝΑΙ ΕΚ ΤΩΝ ΠΡΟΤΕΡΩΝ ΓΝΩΣΤΟΙ Η ΠΡΟΦΑΝΕΙΣ. Η ΕΞΟΡΥΞΗ ΔΕΔΟΜΕΝΩΝ ΘΑ ΑΠΟΤΕΛΕΣΕΙ ΣΤΟ ΜΕΛΛΟΝ ΑΝΤΙΚΕΙΜΕΝΟ ΠΕΡΙΣΣΟΤΕΡΗΣ ΜΕΛΕΤΗΣ ΚΑΙ ΑΝΑΛΥΣΗΣ, ΟΣΟ ΑΝΑΒΑΘΜΙΖΟΝΤΑΙ ΤΑ ΕΡΓΑΛΕΙΑ ΠΟΥ ΠΡΟΣΦΕΡΟΥΝ ΟΙ ΕΠΙΣΤΗΜΕΣ ΤΗΣ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΩΝ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ. ΤΑ ΖΗΤΗΜΑΤΑ ΤΑ ΟΠΟΙΑ ΘΑ ΑΠΟΤΕΛΟΥΝ ΑΝΤΙΚΕΙΜΕΝΑ ΤΗΣ ΜΕΛΕΤΗΣ ΘΑ ΕΙΝΑΙ:

- ΕΞΟΡΥΞΗ ΔΙΑΦΟΡΕΤΙΚΩΝ ΤΥΠΟΥ ΓΝΩΣΕΩΝ ΑΠΟ ΚΟΙΝΑ ΣΥΝΟΛΑ ΔΕΔΟΜΕΝΩΝ
- ΔΙΑΔΡΑΣΤΙΚΟΤΗΤΑ ΣΤΗΝ ΕΞΟΡΥΞΗ ΓΝΩΣΗΣ Η ΟΠΟΙΑ ΘΑ ΜΠΟΡΕΙ ΝΑ ΑΝΑΠΡΟΣΑΡΜΟΖΕΙ ΤΟΝ ΠΡΟΣΑΝΑΤΟΛΙΣΜΟ ΤΗΣ ΑΝΑΒΑΘΜΙΖΟΝΤΑΣ ΤΗΝ ΕΞΟΔΟ ΤΗΣ.
- ΕΠΑΝΑΧΡΗΣΙΜΟΠΟΙΗΣΗ ΤΗΣ ΓΝΩΣΗΣ ΠΟΥ ΑΝΑΠΤΥΧΘΗΚΕ ΓΙΑ ΤΗΝ ΔΗΜΙΟΥΡΓΙΑ ΝΕΑΣ Η ΤΗΝ ΕΠΙΚΑΙΡΟΠΟΙΗΣΗ ΤΗΣ.
- ΔΗΜΙΟΥΡΓΙΑ ΤΥΠΟΠΟΙΗΜΕΝΩΝ ΜΗΧΑΝΙΣΜΩΝ ΠΟΥ ΘΑ ΕΠΙΤΡΕΠΟΥΝ ΚΟΙΝΩΣ ΧΡΗΣΙΜΟΠΟΙΟΥΜΕΝΕΣ ΔΙΑΔΙΚΑΣΙΕΣ ΑΠΟ ΕΤΕΡΟΓΕΝΗ ΣΥΣΤΗΜΑΤΑ.
- ΤΡΟΠΟΙ ΠΑΡΟΥΣΙΑΣΗΣ ΤΗΣ ΓΝΩΣΗΣ
- ΑΛΓΟΡΙΘΜΟΙ ΕΞΟΡΥΞΗΣ ΓΝΩΣΗΣ ΜΕ ΧΑΜΗΛΗ ΠΟΛΥΠΛΟΚΟΤΗΤΑ.
- ΑΣΦΑΛΕΙΑ ΤΩΝ ΔΙΑΔΙΚΑΣΙΩΝ ΚΑΙ ΤΩΝ ΑΠΟΤΕΛΕΣΜΑΤΩΝ
- ΑΞΙΟΛΟΓΗΣΗ ΤΩΝ ΔΙΑΔΙΚΑΣΙΩΝ ΚΑΙ ΤΩΝ ΓΝΩΣΕΩΝ ΜΕ ΑΝΤΙΚΕΙΜΕΝΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ.

ΑΠΟ ΤΗΝ ΠΑΡΟΥΣΙΑΣΗ ΤΩΝ ΔΕΝΔΡΩΝ ΑΠΟΦΑΣΗΣ ΩΣ ΜΙΑ ΤΕΧΝΙΚΗ ΕΞΟΡΥΞΗΣ ΔΕΔΟΜΕΝΩΝ ΠΡΟΚΥΠΤΕΙ ΟΤΙ ΤΟ ΤΕΛΙΚΟ ΑΠΟΤΕΛΕΣΜΑ ΑΠΟΤΕΛΕΙ ΜΙΑ ΕΥΚΟΛΟΝΟΗΤΗ ΑΠΕΙΚΟΝΙΣΗ ΤΗΣ ΚΑΤΗΓΟΡΙΟΠΟΙΗΣΗΣ ΜΕ ΒΑΣΗ ΤΙΣ ΤΙΜΕΣ ΤΩΝ ΠΑΡΑΜΕΤΡΩΝ ΕΝΟΣ ΜΕΓΑΛΟΥ ΣΥΝΟΛΟΥ ΔΕΟΜΕΝΩΝ. ΕΙΝΑΙ ΕΝΑ ΕΡΓΑΛΕΙΟ ΤΟ ΟΠΟΙΟ ΜΠΟΡΕΙ ΝΑ ΔΩΣΕΙ ΜΙΑ ΣΑΦΗ ΕΙΚΟΝΑ ΤΟΥ ΑΠΟΤΕΛΕΣΜΑΤΟΣ ΤΗΣ ΑΝΑΛΥΣΗΣ ΣΕ ΔΙΕΥΘΥΝΤΙΚΑ ΣΤΕΛΕΧΗ ΜΕ ΛΙΓΕΣ Η ΚΑΘΟΛΟΥ ΓΝΩΣΗ ΠΕΡΙ ΜΗΧΑΝΙΣΜΩΝ ΕΞΟΡΥΞΗΣ ΔΕΔΟΜΕΝΩΝ. ΟΙ ΣΥΝΕΠΕΙΕΣ ΤΗΣ ΕΠΙΛΟΓΗΣ ΤΩΝ ΕΝΑΛΛΑΚΤΙΚΩΝ ΔΙΑΘΕΣΙΜΩΝ ΔΡΑΣΕΩΝ ΕΙΝΑΙ ΠΡΟΦΑΝΕΙΣ ΚΑΙ ΕΥΑΝΑΓΝΩΣΤΕΣ. ΈΧΟΥΝ ΕΠΙΠΛΕΟΝ ΤΗΝ ΔΥΝΑΤΟΤΗΤΑ ΝΑ ΑΠΟΤΥΠΩΝΟΥΝ, ΕΠΙΣΗΣ ΜΕ ΠΑΡΑΣΤΑΤΙΚΟ ΤΡΟΠΟ ΤΙΣ ΣΥΝΕΠΕΙΕΣ ΜΙΑΣ ΣΕΙΡΑΣ ΔΙΑΔΟΧΙΚΩΝ ΑΠΟΦΑΣΕΩΝ ΣΤΗΝ ΠΟΡΕΙΑ ΕΝΟΣ ΟΡΓΑΝΙΣΜΟΥ. ΜΕ ΛΙΓΑ ΛΟΓΙΑ ΠΑΡΕΧΟΥΝ ΕΝΑ ΜΟΝΤΕΛΟ ΜΙΑΣ ΑΠΛΟΥΣΤΕΥΜΕΝΗΣ ΑΠΕΙΚΟΝΙΣΗΣ ΑΙΤΙΟΥ ΚΑΙ ΑΠΟΤΕΛΕΣΜΑΤΟΣ ΑΚΟΜΑ ΚΑΙ ΑΝ Η ΕΝΝΟΙΑ ΤΟΥ ΑΙΤΙΟΥ ΕΜΠΕΡΙΕΧΕΙ ΜΕΓΑΛΗ ΠΟΛΥΠΛΟΚΟΤΗΤΑ, ΑΠΟΡΡΟΙΑ ΤΗΣ ΦΥΣΗΣ ΤΩΝ ΕΞΕΤΑΖΟΜΕΝΩΝ ΟΝΤΟΤΗΤΩΝ ΚΑΙ ΠΑΡΑΜΕΤΡΩΝ. ΕΙΝΑΙ ΕΞΙΣΟΥ ΑΠΟΤΕΛΕΣΜΑΤΙΚΑ ΚΑΙ ΓΙΑ ΥΠΕΡΒΟΛΙΚΑ ΜΕΓΑΛΑ ΣΥΝΟΛΑ ΔΕΔΟΜΕΝΩΝ ΑΦΟΥ Η ΠΟΛΥΠΛΟΚΟΤΗΤΑ ΤΟΥ ΑΠΟΤΕΛΕΣΜΑΤΟΣ ΔΕΝ ΕΞΑΡΤΑΤΑΙ ΑΠΟ ΤΟ ΠΛΗΘΟΣ ΤΩΝ ΕΞΕΤΑΖΟΜΕΝΩΝ ΣΤΙΓΜΙΟΤΥΠΩΝ ΑΛΛΑ ΑΠΟ ΑΥΤΟ ΤΩΝ ΠΑΡΑΜΕΤΡΩΝ. ΣΤΑ ΘΕΤΙΚΑ ΣΤΟΙΧΕΙΑ ΤΟΥΣ ΣΥΜΠΕΡΙΛΑΜΒΑΝΕΤΑΙ ΚΑΙ Η ΠΡΟΒΛΕΨΗ ΑΓΝΩΣΤΩΝ ΤΙΜΩΝ ΤΙΣ ΟΠΟΙΕΣ ΚΑΙ ΔΙΑΧΕΙΡΙΖΟΝΤΑΙ.

ΤΑ ΔΕΝΔΡΑ ΑΠΟΦΑΣΕΩΝ ΕΙΝΑΙ ΚΑΤΑΛΛΗΛΑ ΓΙΑ ΤΗΝ ΑΠΕΙΚΟΝΙΣΗ ΚΑΤΑΣΤΑΣΕΩΝ ΠΟΥ ΔΙΑΜΟΡΦΩΝΟΝΤΑΙ ΑΠΟ ΤΙΣ ΤΙΜΕΣ ΟΣΟ ΤΟ ΔΥΝΑΤΟΝ ΛΙΓΟΤΕΡΩΝ ΠΑΡΑΜΕΤΡΩΝ. ΣΕ ΑΝΤΙΘΕΤΗ ΠΕΡΙΠΤΩΣΗ ΠΡΟΚΥΠΤΟΥΝ ΑΡΚΕΤΑ «ΨΗΛΑ» ΔΕΝΔΡΑ ΜΕ ΑΡΚΕΤΕΣ ΔΙΑΚΛΑΔΩΣΕΙΣ ΠΟΥ ΕΝΔΕΧΟΜΕΝΩΣ ΝΑ ΑΜΒΛΥΝΟΥΝ ΤΟ ΠΛΕΟΝΕΚΤΗΜΑ ΤΗΣ ΠΑΡΑΣΤΑΤΙΚΗΣ ΑΠΕΙΚΟΝΙΣΗΣ ΑΛΛΑ ΚΑΙ ΠΟΥ ΑΠΑΙΤΟΥΝ ΑΥΞΗΜΕΝΟ ΥΠΟΛΟΓΙΣΤΙΚΟ ΚΑΙ ΕΠΕΞΕΡΓΑΣΤΙΚΟ ΚΟΣΤΟΣ. Η ΓΕΝΙΚΕΥΣΗ ΤΟΥ ΜΟΝΤΕΛΟΥ ΔΕΝ ΜΠΟΡΕΙ ΝΑ ΑΠΟΤΕΛΕΙ ΠΑΝΑΚΕΙΑ ΔΙΟΤΙ ΕΧΕΙ ΣΑΝ ΣΥΝΕΠΕΙΑ ΑΠΟΔΟΧΗ ΣΦΑΛΜΑΤΟΣ ΤΟ ΟΠΟΙΟ ΕΝΔΕΧΕΤΑΙ ΝΑ ΕΙΝΑΙ ΣΗΜΑΝΤΙΚΟ ΩΣ ΠΡΟΣ ΤΟ ΑΠΟΤΕΛΕΣΜΑ ΠΟΥ ΠΑΡΕΧΟΥΝ. ΈΝΑ ΑΛΛΟ ΜΕΙΟΝΕΚΤΗΜΑ ΤΩΝ ΔΕΝΔΡΩΝ ΑΠΟΦΑΣΗΣ

ΕΙΝΑΙ ΟΤΙ ΔΕΝ ΔΙΑΧΕΙΡΙΖΟΝΤΑΙ ΣΥΝΕΧΕΙΣ ΤΙΜΕΣ. Η ΑΠΟΤΕΛΕΣΜΑΤΙΚΟΤΗΤΑ ΤΗΣ ΔΙΑΚΡΙΤΟΠΟΙΗΣΗΣ ΕΙΝΑΙ ΕΚΕΙΝΗ ΠΟΥ ΠΡΟΣΔΙΟΡΙΖΕΙ ΤΗΝ ΚΑΤΑΛΛΗΛΟΛΗΤΑ ΤΗΣ ΧΡΗΣΗΣ ΤΟΥ ΚΑΙ ΓΙΑ ΠΑΡΑΜΕΤΡΟΥΣ ΓΙΑ ΣΥΝΕΧΕΙΣ ΤΙΜΕΣ.

ΑΝΑΦΟΡΕΣ

[1] MOHAMMED J. ZAKI, WAGNER MEIRA, DATA MINING AND ANALYSIS, 2014, <HTTP://WWW.DATAMININGBOOK.INFO/UPLOADS/BOOK.PDF>

[2] ΓΕΩΡΓΙΟΣ ΠΑΛΙΟΥΡΑΣ, ΕΞΟΡΥΞΗ ΕΞΟΡΥΞΗ ΓΝΩΣΗΣ ΑΠΟ ΔΕΔΟΜΕΝΑ (ACTA MINING), 2011, HTTP://OLDPORTAL.DEMOKRITOS.GR/PAROUSIASEIS/PALIOURAS_140705.PDF

[3] TAN, STEINBACH, KUMAR, DATA MINING: INTRODUCTION, 2004, HTTP://WWW-USERS.CS.UMN.EDU/~KUMAR/DMBOOK/DMSLIDES/CHAP1_INTRO.PDF

[4] JIAWEI HAN, MICHELINE KAMBER, DATA MINING: CONCEPTS AND TECHNIQUES, 2010, <HTTP://WEB.ENGR.ILLINOIS.EDU/~HANJ/BK2/TOC.PDF>

[5] DAVID KRIESEL, NEURAL NETWORKS, 2005, <HTTP://WWW.DKRIESEL.COM/MEDIA/SCIENCE/NEURONALENETZEN-ZETA2-2COL-DKRIESEL.COM.PDF>

[6] RAUL ROJAS, NEURAL NETWORKS, A SYSTEMATIC INTRODUCTION, 1996, <HTTP://PAGE.MI.FU-BERLIN.DE/ROJAS/NEURAL/NEURON.PDF>

[7] TOM M. MITCHELL, DECISION TREE LEARNING, 1997, <HTTP://WWW.CS.CMU.EDU/AFS/CS/PROJECT/THEO-20/WWW/MLBOOK/CH3.PDF>

[8] INTRODUCTION TO DECISION TREES, <HTTP://WWW.TREEPLAN.COM/CHAPTERS/INTRODUCTION-TO-DECISION-TREES.PDF>

[9] DECISION TREE LEARNING, <HTTP://WWW.TREEPLAN.COM/CHAPTERS/INTRODUCTION-TO-DECISION-TREES.PDF>

[10] LIORROKACH, ODEDMAIMON, DECISION TREES, 2002, <HTTP://WWW.ISE.BGU.AC.IL/FACULTY/LIORR/HBCHAP9.PDF>

[11] XINDONG WU, VIPIN KUMAR, J. ROSS QUINLAN, JOYDEEP GHOSH, QIANGYANG, HIROSHIMOTODA, GEOFFREY J. MCLACHLAN, ANGUS NG, BING LIU, PHILIP S. YU, ZHI-HUA ZHOU, MICHAEL STEINBACH, DAVID J. HAND, DAN STEINBERG, TOP 10 ALGORITHMS IN DATA MINING, 2007, <HTTP://WWW.CS.UMD.EDU/~SAMIR/498/10ALGORITHMS-08.PDF>

[12] MAX KUHN, CLASSIFICATION USING C5.0, 2013, HTTP://STATIC1.SQUARESPACE.COM/STATIC/51156277E4B0B8B2FFE11C00/T/51E7E42CE4B0FD2E32684BCA/1374151724529/USER_C5.0.PDF

- [13] MELANIE MITCHELL, AN INTRODUCTION TO GENETIC ALGORITHMS, 1999, [HTTP://WWW.BOENTE.ETI.BR/FUZZY/EBOOK-FUZZY-MITCHELL.PDF](http://www.boente.eti.br/fuzzy/ebook-fuzzy-mitchell.pdf)
- [14] PHILIPP KOEHN, COMBINING GENETIC ALGORITHMS AND NEURAL NETWORKS: THE ENCODING PROBLEM, 1994, [HTTP://HOMEPAGES.INF.ED.AC.UK/PKOEHN/PUBLICATIONS/GANN94.PDF](http://homepages.inf.ed.ac.uk/pkoehn/publications/gann94.pdf)
- [15] GENETIC ALGORITHMS, 1997, [HTTP://WWW.CS.CMU.EDU/AFS/CS/PROJECT/THEO-20/WWW/MLBOOK/CH9.PDF](http://www.cs.cmu.edu/afs/cs/project/theo-20/www/mlbook/ch9.pdf)
- [16] OLIVER SUTTON, INTRODUCTION TO K NEAREST NEIGHBOUR CLASSIFICATION AND CONDENSED NEAREST NEIGHBOUR DATA REDUCTION, 2012, [HTTP://WWW.MATH.LE.AC.UK/PEOPLE/AG153/HOMEPAGE/KNN/OLIVER_KNN_TALK.PDF](http://www.math.le.ac.uk/people/ag153/homepage/knn/oliver_knn_talk.pdf)
- [17] SIDDHARTHDEOKAR, WEIGHTED K NEAREST NEIGHBOR, 2009, [HTTP://WWW.CSEE.UMBC.EDU/~TINOOSH/CMPE650/SLIDES/K_NEAREST_NEIGHBOR_ALGORITHM.PDF](http://www.csee.umbc.edu/~tinoosh/cmpe650/slides/k_nearest_neighbor_algorithm.pdf)
- [18] LEROY BESSLER, DATA VISUALIZATION TIPS AND TECHNIQUES FOR EFFECTIVE COMMUNICATION, 2013, [HTTP://PHARMASUG.ORG/PROCEEDINGS/2013/DG/PHARMASUG-2013-DG10.PDF](http://pharmasug.org/proceedings/2013/dg/pharmasug-2013-dg10.pdf)
- [19] MUZAMMIL KHAN, SARWAR SHAH KHAN, DATA AND INFORMATION VISUALIZATION METHODS, AND INTERACTIVE MECHANISMS: A SURVEY, 2011, [HTTP://RESEARCH.IJCAONLINE.ORG/VOLUME34/NUMBER1/PXC3875722.PDF](http://research.ijcaonline.org/volume34/number1/pxc3875722.pdf)
- [20] E.W.T. NGAI, LI XIU, D.C.K. CHAU, APPLICATION OF DATA MINING TECHNIQUES IN CUSTOMER RELATIONSHIP MANAGEMENT: A LITERATURE REVIEW AND CLASSIFICATION, 2008, [HTTP://115.28.209.159/CI/APPLICATION%20OF%20DATA%20MINING%20TECHNIQUES%20IN%20CUSTOMER%20RELATIONSHIP%20MANAGEMENT.PDF](http://115.28.209.159/ci/application%20of%20data%20mining%20techniques%20in%20customer%20relationship%20management.pdf)
- [21] AHMED, KNOWLEDGE DISCOVERY PRACTICES AND EMERGING APPLICATIONS OF DATA MINING, 2004, [HTTPS://BOOKS.GOOGLE.GR/BOOKS?ID=FMJ3ZUYGNHAC&PG=PA239&LPG=PA239&DQ=AHMED+2004+DATA+MINING&SOURCE=BL&OTS=YTFJ-MKUJF&SIG=KCCNOBROMIBIE6LPFW1KIHSVC6S&HL=EL&SA=X&EI=F6VPVZKQJIQKSGHPZYHYBQ&VED=0CF8Q6AEWCA#V=ONEPAGE&Q=AHMED%202004%20DATA%20MINING&F=FALSE](https://books.google.gr/books?id=FMJ3ZUYGNHAC&pg=PA239&lpg=PA239&dq=ahmed+2004+data+mining&source=bl&ots=ytFJ-MKUJF&sig=kccnobromibie6lpfw1kihsvc6s&hl=el&sa=x&ei=f6vpvzkqjIQKSGHPZYHYBQ&ved=0CF8Q6AEWCA#v=onepage&q=ahmed%202004%20data%20mining&f=false)

