

ΠΟΛΥΤΕΧΝΕΙΟ ΚΡΗΤΗΣ
ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ
ΠΑΡΑΓΩΓΗΣ & ΔΙΟΙΚΗΣΗΣ

***ΚΑΤΗΓΟΡΙΟΠΟΙΗΣΗ ΚΑΙ ΚΑΤΑΤΑΞΗ
ΜΕ ΧΡΗΣΗ ΣΤΑΤΙΣΤΙΚΩΝ ΜΕΘΟΔΩΝ
ΚΑΙ ΑΛΓΟΡΙΘΜΩΝ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ***



ΕΦΑΡΜΟΓΗ ΣΕ ΤΡΑΠΕΖΙΚΑ ΥΠΟΚΑΤΑΣΤΗΜΑΤΑ

***ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ ΤΟΥ
ΜΕΛΕΤΙΑΔΗ ΒΑΣΙΛΕΙΟΥ***

***ΕΠΙΒΛΕΠΩΝ ΚΑΘΗΓΗΤΗΣ
ΜΙΧΑΛΟΠΟΥΛΟΣ ΜΙΧ.***

ΧΑΝΙΑ 1997

ΒΙΟΓΡΑΦΙΚΟ ΣΗΜΕΙΩΜΑ

ΟΝΟΜ/ΜΟ: ΜΕΛΕΤΙΑΔΗΣ ΒΑΣΙΛΕΙΟΣ

ΕΤΟΣ ΓΕΝΝΗΣΕΩΣ: 1971

ΤΟΠΟΣ ΓΕΝΝΗΣΕΩΣ: ΝΙΚΑΙΑ

1983-1986: ΟΛΟΚΛΗΡΩΝΩ ΤΙΣ ΓΥΜΝΑΣΙΑΚΕΣ ΣΠΟΥΔΕΣ ΣΤΟ 1ο ΓΥΜΝΑΣΙΟ ΚΟΡΥΔΑΛΛΟΥ ,ΑΠΟΣΠΩΝΤΑΣ ΚΑΘΕ ΧΡΟΝΙΑ ΑΡΙΣΤΕΙΟ. ΤΟ 1984 ΚΑΤΑΛΑΜΒΑΝΩ ΤΗΝ 4η ΘΕΣΗ ΣΤΟ ΠΡΩΤΑΘΛΗΜΑ ΣΤΙΒΟΥ ΠΑΜΠΑΙΔΩΝ ΣΤΑ 60Μ.

1986-1989: ΜΑΘΗΤΕΥΩ ΣΤΟ 3ο ΛΥΚΕΙΟ ΚΟΡ/ΛΟΥ ΚΑΙ ΕΠΙΤΥΓΧΑΝΩ ΤΟΝ 2ο ΚΑΛΥΤΕΡΟ ΒΑΘΜΟ ,ΤΟ 1988 ,ΣΤΑ ΛΥΚΕΙΑ ΤΟΥ ΔΗΜΟΥ. ΤΟ ΙΔΙΟ ΕΤΟΣ Η ΕΚΘΕΣΗ ΜΟΥ ΕΠΙΛΕΓΕΤΑΙ ΝΑ ΕΚΠΡΟΣΩΠΗΣΕΙ ΜΑΖΙ ΜΕ ΑΛΛΕΣ ΤΗΝ Δ.ΑΤΤΙΚΗ ΣΤΟΝ ΠΑΝΕΛΛΗΝΙΟ ΔΙΑΓΩΝΙΣΜΟ. ΤΟ 1987 ΣΤΗΝ ΤΕΛΙΚΗ ΦΑΣΗ ΤΟΥ ΠΑΙΔΙΚΟΥ ΠΡΩΤ/ΤΟΣ ΜΠΑΣΚΕΤ ΑΝΑΚΥΡΗΣΟΜΑΙ 2ος ΚΑΛΥΤΕΡΟΣ ΡΙΜΠΑΟΥΝΤΕΡ.

1990: ΕΙΣΑΓΟΜΑΙ 3ος ΣΤΟ ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΠΑΡΑΓΩΓΗΣ ΚΑΙ ΔΙΟΙΚΗΣΗΣ ΤΟΥ ΠΟΛΥΤΕΧΝΕΙΟΥ ΚΡΗΤΗΣ ,ΤΟ ΟΠΟΙΟ ΑΠΟΤΕΛΕΙ ΚΑΙ ΠΡΩΤΗ ΜΟΥ ΕΠΙΛΟΓΗ.

1991: ΒΑΣΙΚΟ ΜΕΛΟΣ ΤΗΣ ΟΜΑΔΑΣ ΜΠΑΣΚΕΤ ΠΟΥ ΠΕΤΥΧΑΙΝΕΙ ΤΗΝ ΠΡΩΤΗ ΙΣΤΟΡΙΚΗ ΝΙΚΗ ΤΟΥ ΠΟΛ/ΧΝΕΙΟΥ ΚΑΙ ΤΗΣ ΟΜΑΔΑΣ ΥΔΑΤ/ΣΗΣ ΠΟΥ ΛΑΜΒΑΝΕΙ ΤΟ ΧΑΛΚΙΝΟ ΜΕΤΑΛΛΙΟ, ΣΤΗΝ 1η ΠΑΝΕΠ/ΜΙΑΔΑ.

1992: ΕΚΠΟΝΩ ΕΡΓΑΣΙΑ ΣΤΑ ΠΛΑΙΣΙΑ ΤΟΥ ΜΑΘΗΜΑΤΟΣ “ΤΕΧΝΟΛΟΓΙΑ ΠΑΡΑΓΩΓΗΣ “ ΜΕ ΘΕΜΑ “ΥΠΕΡΠΛΗΡΩΜΕΝΟΙ ΚΙΝΗΤΗΡΕΣ (TURBO)”.Η ΕΡΓΑΣΙΑ ΑΥΤΗ ΧΡΗΣΙΜΟΠΟΙΕΙΤΑΙ ΑΡΓΟΤΕΡΑ ΑΠΟ ΤΗΝ FIAT AUTO HELLAS ΣΤΟΝ ΤΟΜΕΑ ΕΚΠΑΙΔΕΥΣΗΣ-ΕΠΙΜΟΡΦΩΣΗΣ.

1993 : ΙΔΡΥΤΙΚΟ ΜΕΛΟΣ ΤΗΣ ΠΟΛΙΤΙΣΤΙΚΗΣ ΟΜΑΔΑΣ ΤΟΥ ΠΟΛΥΤ/ΝΕΙΟΥ. ΒΑΣΙΚΟ ΣΤΕΛΕΧΟΣ ΤΗΣ ΑΠΟΣΤΟΛΗΣ ΤΗΣ 2ΗΣ ΑΘΛΗΤΙΚΗΣ ΠΑΝ/ΜΙΑΔΟΣ .

1995: ΟΛΟΚΛΗΡΩΝΩ ΕΠΙΤΥΧΩΣ ΤΟΝ ΚΥΚΛΟ ΤΩΝ ΜΑΘΗΜΑΤΩΝ ΚΑΙ ΕΚΠΟΝΩ ΤΗΝ ΔΙΠΛΩΜΑΤΙΚΗ ΜΟΥ ΕΡΓΑΣΙΑ ΜΕ ΘΕΜΑ “ΚΑΤΗΓΟΡΙΟΠΟΙΗΣΗ ΚΑΙ ΚΑΤΑΤΑΞΗ ΜΕ ΧΡΗΣΗ ΣΤΑΤΙΣΤΙΚΩΝ ΜΕΘΟΔΩΝ ΚΑΙ ΑΛΓΟΡΙΘΜΩΝ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ “

1995-1997 : ΑΝΑΛΑΜΒΑΝΩ ΤΟ ΟΙΚΟΓΕΝΕΙΑΚΟ ΔΙΑΓΝΩ-ΕΠΙΣΚΕΥΑΣΤΙΚΟ ΚΕΝΤΡΟ ΑΥΤ/ΤΩΝ ΣΥΓΧΡΟΝΗΣ ΤΕΧΝΟΛΟΓΙΑΣ ΣΤΟ ΟΠΟΙΟ ΜΕ ΤΗ ΒΟΗΘΕΙΑ ΤΗΣ ΤΕΧΝΟΓΝΩΣΙΑΣ ΠΟΥ ΕΧΩ ΑΠΟΛΑΒΕΙ ΑΠΟ ΤΟ Π.Κ ΚΑΙ ΤΩΝ ΣΥΓΧΡΟΝΩΝ ΗΛΕΚΤΡΟΝΙΚΩΝ ΜΗΧΑΝΗΜΑΤΩΝ ΠΟΥ ΧΡΗΣΙΜΟΠΟΙΩ ΕΦΑΡΜΟΖΩ ΜΕΘΟΔΟΥΣ ΒΕΛΤΙΣΤΟΠΟΙΗΣΗΣ ΤΗΣ ΠΟΙΟΤΗΤΑΣ ΔΙΑΓΝΩΣΗΣ ΣΕ ΑΥΤ/ΤΑ ΜΟΝΟΥ ΚΑΙ ΠΟΛΑΠΛΟΥ ΣΗΜΕΙΟΥ ΨΕΚΑΣΜΟΥ.

1996-1997: ΣΥΜΒΟΥΛΟΣ ΕΠΙΜΟΡΦΩΣΗΣ ΣΤΗΝ ΝΟΜΑΡΧΙΑ ΠΕΙΡΑΙΑ. ΠΑΡΑΛΛΗΛΑ ΔΙΔΑΣΚΩ ΣΤΑ ΝΟΜΑΡΧΙΑΚΑ ΠΡΟΓΡΑΜΜΑΤΑ ,ΤΑ ΠΑΚΕΤΑ “EXCEL 95” ΚΑΙ “ACCESS 95”.

1996-1997: ΣΥΜΒΟΥΛΟΣ ΟΡΓΑΝΩΣΗΣ ΣΤΗΝ SARMED, ΕΤΑΙΡΕΙΑ Η ΟΠΟΙΑ ΕΚΤΕΛΩΝΙΖΕΙ ΚΑΙ ΠΡΑΓΜΑΤΟΠΟΙΕΙ ΤΟΝ ΠΡΩΤΟ ΤΕΧΝΙΚΟ ΕΛΕΓΧΟ ΣΕ ΟΛΑ ΤΑ ΑΥΤ/ΤΑ ΤΗΣ FIAT,ALFA ROMEO,LANCIA,SUBARU,FORD.

ΞΕΝΕΣ ΓΛΩΣΣΕΣ: ΑΓΓΛΙΚΑ, ΥΨΗΛΟ ΕΠΙΠΕΔΟ
ΓΕΡΜΑΝΙΚΑ,ΙΚΑΝΟΠΟΙΗΤΙΚΟ ΕΠΙΠΕΔΟ

Πίνακας Περιεχομένων

ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ	1
ΠΡΟΛΟΓΟΣ	3
ΕΙΣΑΓΩΓΗ	5
ΕΝΟΤΗΤΑ 1 - CLUSTER ANALYSIS	7
CLUSTER ANALYSIS.....	7
ΟΡΙΣΜΟΙ EVERITT	7
ΙΕΡΑΡΧΙΚΗ ΟΜΑΔΟΠΟΙΗΣΗ	11
ΔΕΙΓΜΑΤΙΚΟΣ ΠΙΝΑΚΑΣ	16
ΠΙΝΑΚΑΣ ΕΓΓΥΤΗΤΑΣ	17
ΔΕΙΚΤΕΣ ΕΓΓΥΤΗΤΑΣ.....	18
ΜΕΘΟΔΟΙ CLUSTER ANALYSIS	24
ΜΑΘΗΜΑΤΙΚΗ ΠΑΡΟΥΣΙΑΣΗ WARD METHOD	24
ΕΦΑΡΜΟΓΗ ΤΗΣ CLUSTER ANALYSIS ΣΤΗΝ ΚΑΤΗΓΟΡΙΟ-ΠΟΙΗΣΗ ΤΩΝ ΤΡΑΠΕΖΙΚΩΝ ΥΠΟΚΑΤΑΣΤΗΜΑΤΩΝ	26
ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ	59
Ο ΑΛΓΟΡΙΘΜΟΣ ID-3.....	61
ΠΟΛΥΠΛΟΚΟΤΗΤΑ ΤΟΥ ΑΛΓΟΡΙΘΜΟΥ	63
NEW ID.....	64

ΜΑΘΗΜΑΤΙΚΗ ΠΑΡΟΥΣΙΑΣΗ ΤΟΥ ΑΛΓΟΡΙΘΜΟΥ	66
ID-3.....	
ΠΑΡΑΔΕΙΓΜΑ	71
ΜΕΤΑΤΡΟΠΗ ΔΕΝΔΡΟΥ ΣΕ ΚΑΝΟΝΕΣ ΑΠΟΦΑΣΗΣ	75
ΕΦΑΡΜΟΓΗ NEW ID-3 ΣΤΟ ΠΡΟΒΛΗΜΑ ΚΑΤΗΓΟΡΙΟΠΟΙΗΣΗΣ.....	76
ΣΧΟΛΙΑΣΜΟΣ ΚΑΝΟΝΩΝ ΑΠΟΦΑΣΗΣ.....	89
 ΑΝΑΛΥΣΗ & ΚΡΙΤΙΚΗ ΤΩΝ ΑΠΟΤΕΛΕΣΜΑΤΩΝ	92
 ΠΑΡΑΡΤΗΜΑΤΑ.....	94
 ΒΙΒΛΙΟΓΡΑΦΙΑ	97

ΠΡΟΛΟΓΟΣ

Μία από τις πρώτες και κυριότερες δραστηριότητες του ανθρώπου ήταν να κατηγοριοποιεί πρόσωπα, αντικείμενα, γεγονότα σε ομάδες. Κάθε ομάδα αποτελείται από στοιχεία που ομοιάζουν, δηλαδή έχουν όμοιες ή συναφείς ιδιότητες. Έτσι, γνωρίζοντας κάποιος την ομάδα που ανήκει κάποιο στοιχείο μπορεί να γνωρίζει και τις ιδιότητες που χαρακτηρίζουν το συγκεκριμένο στοιχείο.

Αν προβάλλουμε τον παραπάνω συλλογισμό στο πεδίο της επιστήμης της Διοίκησης θα διατυπώναμε την άποψη πως η ομαδοποίηση (clustering) αποτελεί σημαντικό εργαλείο στην λήψη αποφάσεων.

Η κατάταξη (classification) είναι η άλλη δραστηριότητα που μας απασχολεί. Έχει ως αντικείμενο να κατατάσει στοιχεία σε κατηγορίες (ομάδες). Ο προβληματισμός που αναδύεται από αυτό το πεδίο της επιστήμης της λήψης αποφάσεων είναι να μπορέσουμε να πετύχουμε την κατάταξη χρησιμοποιώντας όσο το δυνατόν λιγότερα δεδομένα, ώστε η απόφαση να είναι λιγότερο χρονοβόρος και επίπονη.

Το παραπάνω έχει ιδιαίτερη σημασία γιατί στο σύγχρονο επιχειρηματικό περιβάλλον οι αποφάσεις πρέπει να λαμβάνονται σωστά, αλλά και όσο το δυνατόν ταχύτερα. Τα δεδομένα που χρησιμοποιούνται για κατατάξη πρέπει να είναι και τα πιο διαχωριστικά, δηλαδή να πετυχαίνουν τον καλύτερο δυνατό διαχωρισμό μεταξύ των ομάδων.

ΕΙΣΑΓΩΓΗ

Η παρούσα εργασία εξετάζει 27 υποκαταστήματα μίας τράπεζας τα οποία χαρακτηρίζονται από 28 μεταβλητές (τραπεζικές εργασίες).

*Η αναζήτηση μας είναι διττή . Πρώτον, επιδιώκουμε να **κατηγοριοποιήσουμε** τα 27 υποκαταστήματα σε 5 ομάδες (κλάσεις) με την βοήθεια μεθόδων της *cluster* ανάλυσης.*

*Δεύτερον, χρησιμοποιώντας αλγόριθμους μηχανικής μάθησης να εξάγουμε **κανόνες απόφασης**, οι οποίοι θα κατατάσουν μελλοντικά τα υποκαταστήματα σε κλάσεις.*

*Οι κανόνες απόφασης μας δίνουν την δυνατότητα να αποφανθούμε για το αν ένα υποκατάστημα χαρακτηρίζεται ,ως, πολύ καλό, καλό,μέτριο,κακό, εξετάζοντας μόνο τα πιο **διαχωριστικά κριτήρια** (μεταβλητές) από τα οποία αυτοί δομούνται.*

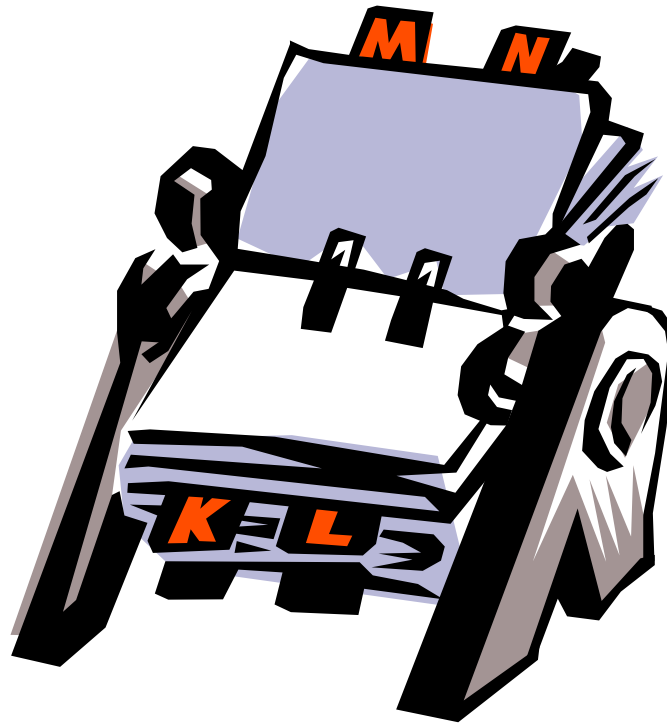
Η πρώτη ενότητα περιλαμβάνει μία θεωρητική προσέγγιση της *cluster* ανάλυσης. Παρουσιάζονται οι μέθοδοι (αλγόριθμοι) που εφαρμόζονται στην ιεραρχική ομαδοποίηση, καθώς και μέτρα ομοιότητας-ανομοιότητας που χρησιμοποιούν οι μέθοδοι. Επίσης παρουσιάζεται το μαθηματικό μοντέλο της μεθόδου του *WARD*, η οποία εφαρμόζεται για να εξάγει τα αποτελέσματα των ομαδοποιήσεων.

Η δεύτερη ενότητα επιχειρεί μία διεξοδική ανάλυση της φιλοσοφίας των αλγορίθμων της μηχανικής μάθησης *ID-3* και *NEWID*. Έννοιες, όπως, τα δένδρα απόφασης και οι κανόνες απόφασης αναλύονται, ενώ παραθέτονται και στοιχεία της θεωρίας πιθανοτήτων και θεωρίας πληροφοριών που αποτελούν το θεωρητικό υπόβαθρο του μαθηματικού μοντέλου του *ID3*. Έπειτα, εφαρμόζουμε τον αλγόριθμο *NEWID* για να εξάγουμε το δένδρο κατάταξης και τους επιθυμητούς κανόνες κατάταξης.

Η τρίτη ενότητα αποτελεί πεδίο γενικότερων συμπερασμάτων και κριτικής ανάλυσης των αποτελεσμάτων της εργασίας.

ΕΝΟΤΗΤΑ

1



CLUSTER ANALYSIS

CLUSTER ANALYSIS

Το αντικείμενο της cluster analysis είναι να βρει ένα κατάλληλο και έγγυρο τρόπο οργάνωσης των δεδομένων και όχι να δημιουργήσει κανόνες για διαχωρισμό των μελλοντικών δεδομένων σε κατηγορίες.

ΟΡΙΣΜΟΙ EVERITT (1974)

1. Μια ομάδα είναι ένα σύνολο από αντικείμενα, τα οποία ομοιάζουν, ενώ τα αντικείμενα διαφορετικών ομάδων δεν ομοιάζουν.
2. Μια ομάδα είναι μια συνάθροιση σημείων στο εξεταζόμενο διάστημα, τέτοια ώστε η απόσταση μεταξύ δύο σημείων στην ομάδα είναι μικρότερη από την απόσταση μεταξύ κάθε σημείου στην ομάδα και σε κάθε σημείου εκτός αυτής.
3. Ομάδες μπορούν να περιγραφούν σαν συγκροτημένες περιοχές ενός πολυδιάστατου διαστήματος περιέχοντας μια σχετικά υψηλή πυκνότητα των στοιχείων, διαχωριζόμενες από άλλες περιοχές που περιέχουν σχετικά χαμηλή πυκνότητα.

Η CLUSTER ANALYSIS γνώρισε μεγάλη ανάπτυξη μετά την έκρηξη των υπολογιστικών μηχανημάτων. Ερευνητές σχεδόν σε κάθε περιοχή της επιστήμης που διαχειρίζεται δεδομένα έχουν βοηθηθεί, όπως οι ψυχολόγοι, βιολόγοι, στατιστικοί, κοινωνικοί επιστήμονες και μηχανικοί. Ο Good πρότεινε τον όρο **βοτρυολογία (botryology)** για αυτόν τον κλάδο της επιστήμης.

Αν θελήσουμε να δούμε την cluster analysis από μια άλλη άποψη, θα λέγαμε ότι είναι μια **διαδικασία ομαδοποίησης** αντικειμένων σε **υποσύνολα**, τα οποία έχουν νόημα στο πλαίσιο ενός ιδιαίτερου προβλήματος.

Η σχέση ανάμεσα στα αντικείμενα απεικονίζεται σε έναν **πίνακα εγγύτητας** στον οποίο γραμμές και στήλες αντιστοιχούν στα αντικείμενα. Αν τα αντικείμενα χαρακτηρίζονται ως δείγματα, ή σημεία στο d-διάστατο μετρικό διάστημα, οι **εγγύτητες** μπορεί να **είναι η απόσταση μεταξύ του ζεύγους των σημείων**, όπως η ευκλείδεια απόσταση. Αναγκαία συνθήκη για να δώσει η cluster analysis ουσιαστικά αποτελέσματα είναι η επιλογή κάποιου κατάλληλου μέτρου απόστασης ή εγγύτητας.

Ο πίνακας εγγύτητας είναι το μόνο δεδομένο που χρειάζεται ο αλγόριθμος ομαδοποίησης.

Μπορούμε να κατατάξουμε την ομαδοποίηση ακολούθως (11) :

I. ΑΠΟΚΛΕΙΣΤΙΚΗ & ΜΗ-ΑΠΟΚΛΕΙΣΤΙΚΗ

Η αποκλειστική διαχωρίζει το σύνολο των αντικειμένων, ώστε καθένα να ανήκει σε ένα και μόνο υποσύνολο ή ομάδα. Η μη-αποκλειστική ή επικαλυπτόμενη κατάταξη μπορεί να αποδώσει ένα αντικείμενο σε αρκετές κλάσεις (ομάδες). Για παράδειγμα μια ομαδοποίηση των ανθρώπων σύμφωνα με την ηλικία ή το φύλλο είναι αποκλειστική. Αντιθέτως, μια ομαδοποίηση σύμφωνα με τις ασθένειες που μπορεί να έχει ένας είναι επικαλυπτόμενη, αφού κάποιος μπορεί να έχει αρκετές ασθένειες. Εμείς θα ασχοληθούμε με αποκλειστικές κατατάξεις.

II. ΕΣΩΓΕΝΗΣ & ΕΞΩΓΕΝΗΣ

Η εσωγενής ομαδοποίηση χρησιμοποιεί μόνο τον πίνακα εγγύτητας για να εκτελέσει την ομαδοποίηση. Η εξωγενής χρησιμοποιεί ετικέτες κατηγοριών στα αντικείμενα, καθώς και τον πίνακα εγγύτητας.

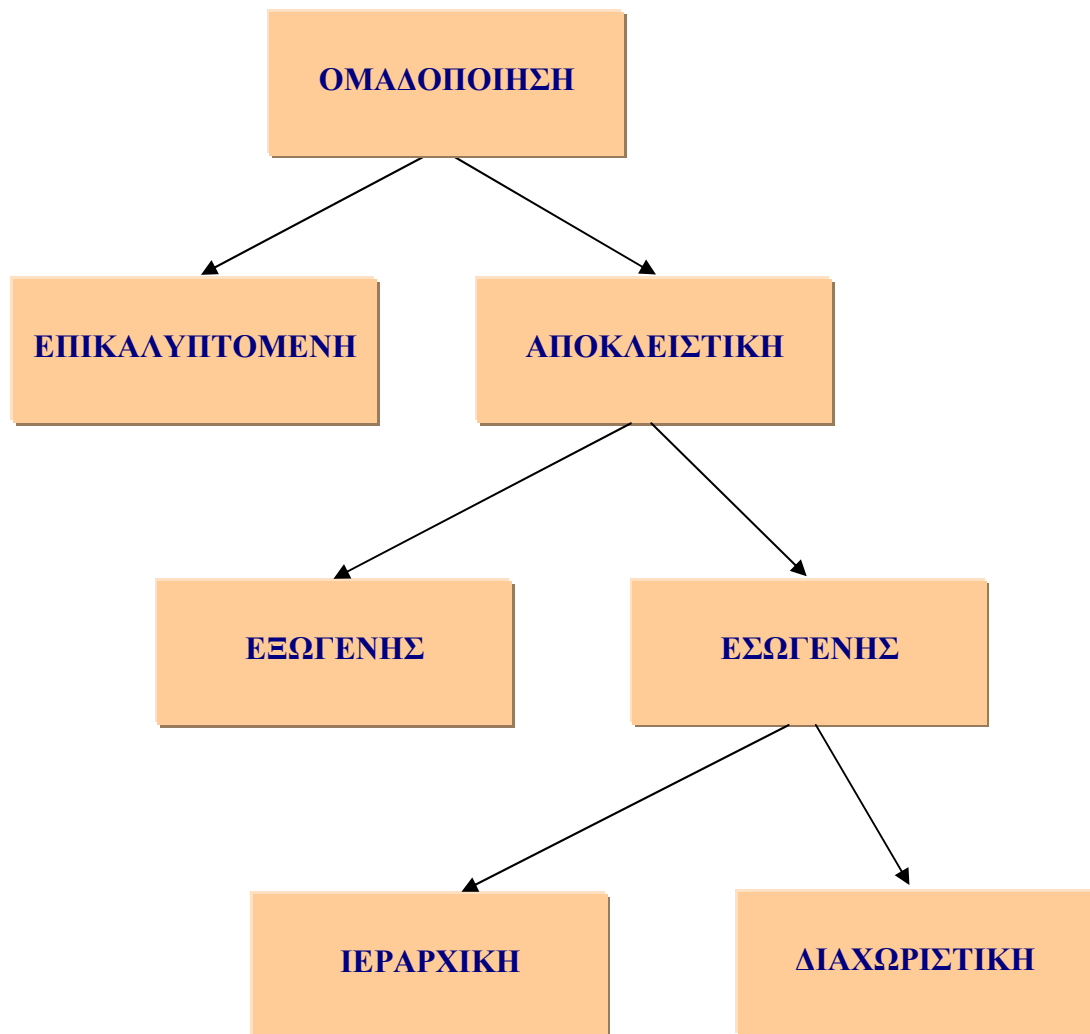
π.χ.

Υποθέστε ότι διάφοροι δείκτες της προσωπικής υγείας συλλέγησαν από καπνιστές και μη-καπνιστές. Μια εσωγενής ομαδοποίηση θα ομαδοποιούσε τα άτομα βασισμένη στις ομοιότητες ανάμεσα στους δείκτες υγείας και μετά θα προσπαθούσε να αποφασίσει αν το κάπνισμα ήταν ένας παράγοντας ροπής των ατόμων σε επικείμενο πλήθος ασθενειών. Ενώ αντιθέτως η εξωγενής θα επιχειρούσε να εξάγει τρόπους διαχωρισμού καπνιστών από μη-καπνιστές που να βασίζονται στους δείκτες υγείας. Εμείς θα ασχοληθούμε με τους εσωγενείς.

III. ΙΕΡΑΡΧΙΚΕΣ & ΔΙΑΧΩΡΙΣΤΙΚΕΣ

Οι εσωγενείς και εξωγενείς υποδιαιρούνται σε ιεραρχικές και διαχωριστικές σύμφωνα με την μορφή διαχείρισης των δεδομένων. Η ιεραρχική ομαδοποίηση είναι μια φωλιασμένη ακολουθία διαχωρισμών. Η ιεραρχική ομαδοποίηση θα εξεταστεί παρακάτω διεξοδικότερα. Οι διαχωριστικές προτείνουν έναν μόνο διαχωρισμό.

Η ομαδοποίηση που μας ενδιαφέρει σε αυτήν την εργασία είναι **η αποκλειστική, εσωγενής, ιεραρχική**. Το 1973 οι Sokal και Sneath πρότειναν το ακρωνύμιο SAHN (Sequential Agglomerative Hierarchical Nonoverlapping) για τους αποκλειστικούς, εσωγενείς, ιεραρχικούς, επαναληπτικούς αλγόριθμους.



Σχήμα 1 - Κατατάξεις ομαδοποίησης

ΙΕΡΑΡΧΙΚΗ ΟΜΑΔΟΠΟΙΗΣΗ

Μια ιεραρχική μέθοδος είναι μια διαδικασία για μετατροπή ενός πίνακα εγγύτητας σε μια **ακολουθία φωλιασμένων διαχωρισμών**. Ένας αλγόριθμος ιεραρχικής ομαδοποίησης είναι ο καθορισμός των βημάτων για την εκτέλεση της ιεραρχικής ομαδοποίησης.

(12),(13),(14)

Ας υποθέσουμε ότι έχουμε N αντικείμενα να ομαδοποιήσουμε τα οποία αποτελούν το σύνολο X .

$$X = \{x^1, x^2, \dots, x^N\}$$

όπου x^i είναι το i -οστό αντικείμενο. Ένας διαχωρισμός B του X **διαμελίζει το X σε υποσύνολα** $\{C_1, C_2, \dots, C_m\}$ ικανοποιώντας τα παρακάτω:

$$C_i \cap C_j = \emptyset \quad \text{για κάθε } i \text{ και } j \text{ από } 1 \text{ έως } m, i \neq j$$

$$C_1 \cup C_2 \cup \dots \cup C_m = X$$

το " \cap " σημαίνει το σύμβολο της τομής, ενώ το " \cup " το σύμβολο της ένωσης και το " \emptyset " σημαίνει το σύμβολο του κενού συνόλου. Μια ομαδοποίηση είναι ένας διαχωρισμός. Οι συνιστώσες του διαχωρισμού ονομάζονται κλάσεις. Ο διαχωρισμός B λέμε ότι είναι

φωλιασμένος στο διαχωρισμό H αν κάθε συνιστώσα του \mathbf{B} αποτελεί υποσύνολο συνιστώσας του H . Το H δημιουργείται από συγχώνευση των συνιστωσών του \mathbf{B} .

π.χ.

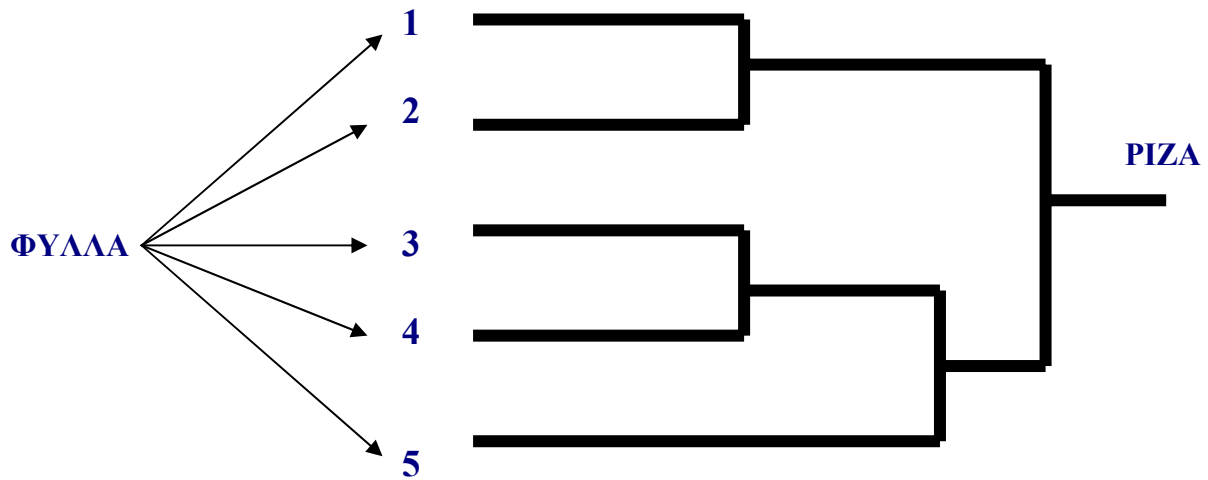
Τα \mathbf{B} και \mathbf{H} αποτελούν ομαδοποιήσεις του συνόλου των αντικειμένων $\{\chi_1, \chi_2, \dots, \chi_{10}\}$

$$\mathbf{H} = \{(\chi_1, \chi_3, \chi_5, \chi_7), (\chi_2, \chi_4, \chi_6, \chi_8), (\chi_9, \chi_{10})\}$$

$$\mathbf{B} = \{(\chi_1, \chi_3), (\chi_5, \chi_7), (\chi_2), (\chi_4, \chi_6, \chi_8), (\chi_9, \chi_{10})\}$$

Ένας επαναληπτικός αλγόριθμος για ιεραρχική ομαδοποίηση ξεκινά τοποθετώντας κάθε ένα από τα n αντικείμενα του συνόλου σε μια **ξεχωριστή κλάση**. Στην συνέχεια επεξεργαζόμενος τον πίνακα εγγύτητας σε κάθε βήμα του **συγχωνεύει κλάσεις**, ωστόσο να σχηματιστεί μια κλάση που να περιέχει όλα τα αντικείμενα του συνόλου.

Ακολουθώντας αυτή την πορεία έχουμε την διαδικασία της **συνθετικής ομαδοποίησης**, ενώ αν ακολουθήσουμε την αντίστροφη πορεία, δηλαδή από μια ομάδα που περιέχει όλα τα αντικείμενα καταλήξουμε σε ατομικές ομάδες θα έχουμε εφαρμόσει **την αναλυτική ομαδοποίηση**.

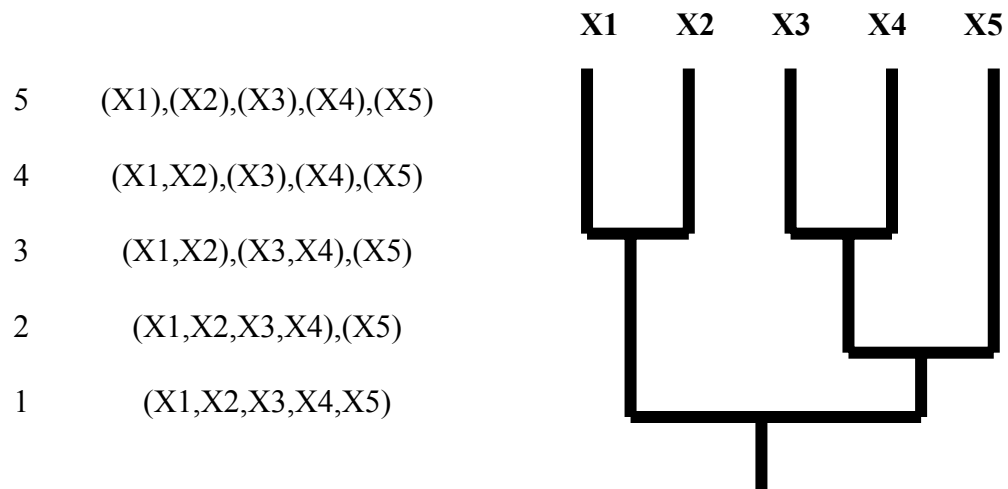


Σχήμα 2 - Δενδροδιάγραμμα διαδικασίας ομαδοποίησης

Τα φύλλα αντιπροσωπεύουν μια μονάδα, ενώ η ρίζα το σύνολο των μονάδων. Η μετακίνηση από τα φύλλα στην ρίζα περιγράφει την αυξανόμενη συνάθροιση των μονάδων σε κλάσεις. Οι ιεραρχικές μέθοδοι που χτίζουν ένα δένδρο από τα φύλλα στη ρίζα ονομάζονται συγχωνευτικές μέθοδοι.

Το παραπάνω δενδροδιάγραμμα αποτελεί ειδικό τύπο δομής δένδρου. Στο δενδροδιάγραμμα απεικονίζεται η διαδικασία της ομαδοποίησης βήμα-βήμα. Ανάλογα με τον αριθμό των ομάδων που επιδιώκουμε να λάβουμε, κόβουμε στο κατάλληλο επίπεδο.

ΑΡΙΘΜΟΣ ΟΜΑΔΕΣ



Σχήμα 3

Πριν προχωρήσουμε στην παρουσίαση αλγορίθμων ιεραρχικής ομαδοποίησης, θεωρείται επιβεβλημένο να αναφερθούμε στον πίνακα των δεδομένων που επεξεργάζονται αυτοί οι αλγόριθμοι.

ΔΕΙΓΜΑΤΙΚΟΣ ΠΙΝΑΚΑΣ

Έστω ότι έχουμε ένα σύνολο N αντικειμένων. Αν κάθε αντικείμενο παριστάνεται από ένα σύνολο d μέτρων (ή χαρακτηριστικών ή αποτελεσμάτων), τότε κάθε αντικείμενο εκφράζεται από ένα d -διάστατο διάνυσμα, το οποίο καλείται δείγμα. Το σύνολο των αντικειμένων εξετάζεται σαν ένας $n \times d$ δειγματικός πίνακας. Κάθε γραμμή του πίνακα καθορίζει ένα δείγμα, ενώ κάθε στήλη καθορίζει ένα χαρακτηριστικό ή μέτρο.

ΠΙΝΑΚΑΣ ΕΓΓΥΤΗΤΑΣ

Οι μέθοδοι ομαδοποιήσεων απαιτούν να προσδωθεί ένας δείκτης εγγύτητας ή ομοιότητας ή συγγένειας ή σχέσης ανάμεσα στα ζευγάρια των δειγμάτων. Ένας πίνακας εγγύτητας μπορεί να προκύψει από έναν δειγματικό πίνακα ή να ληφθεί απ'ευθείας από ειδικά διαμορφωμένα δεδομένα.

π.χ.

Αν κάποια άτομα ερωτηθούν να εκφράσουν την προτίμησή τους σε κρασιά, τότε η εγγύτητα μεταξύ δύο κρασιών μπορεί να ληφθεί από την πρόσθεση των ατομικών προτιμήσεων. Επίσης κάλλιστα θα μπορούσε κάποιος να δώσει την προσωπική του άποψη για την ομοιότητα μεταξύ δύο κρασιών με κλίμακα από 1 έως 10. Σε έναν πίνακα εγγύτητας τόσο οι γραμμές όσο και οι στήλες αντιπροσωπεύουν δείγματα, ενώ τα στοιχεία του πίνακα εκφράζουν δείκτες **ομοιότητας** ή **ανομοιότητας**.

ΔΕΙΚΤΕΣ ΕΓΓΥΤΗΤΑΣ

Ένας δείκτης εγγύτητας ανάμεσα στο i και στο k δείγμα συμβολίζεται ως $d(i,k)$ και ικανοποιεί τις παρακάτω τρεις ιδιότητες (15) :

1. (α). Για την ανομοιότητα: $d(i,i) = 0$, για όλα τα i
(β). Για την ομοιότητα: $d(i,i) \geq \max_k d(i,k)$, για όλα τα i
2. $d(i,k) = d(k,i)$, για όλα τα (i,k)
3. $d^*(i,k) \geq 0$, για όλα τα (i,k)

Ο πιο γνωστός δείκτης είναι του **Minkowski**, ο οποίος μετρά ανομοιότητες. Το i -στό δείγμα το οποίο αντιστοιχεί στην i -στή γραμμή του δειγματικού πίνακα, συμβολίζεται με το διάνυσμα $x_i = (x_{i1}, x_{i2}, \dots, x_{id})^T$, $i = 1, 2, 3, \dots, n$.

Το d είναι ο αριθμός των χαρακτηριστικών το n ο αριθμός των δειγμάτων και το T συμβολίζει το ανάστροφο διάνυσμα.

Το μέτρο του Minkowski συμβολίζεται ως: $d(i,k) = (\sum |x_{ij} - x_{kj}|^r)^{1/r}$ όπου $r \geq 1$

Όλα τα μέτρα του Minkowski ικανοποιούν εκτός από τις μετρικές ιδιότητες που ήδη έχουμε αναφέρει και άλλες δύο επιπρόσθετες ιδιότητες.

4. $d(i,k) = 0$, μόνο αν $x_i = x_k$
5. $d(i,k) \leq d(i,m) + d(m,k)$, για όλα τα (i,k,m)

Τα πιο γνωστά μέτρα του Minkowski είναι:

a. Για $r = 2$ (Ευκλείδεια απόσταση)

$$d(i,k) = [\sum (x_{ij} - x_{kj})^2]^{1/2} = [(x_i - x_k)^T (x_i - x_k)]^{1/2}$$

b. Για $r = 1$ (απόσταση του Manhattan)

$$d(i,k) = \sum |x_{ij} - x_{kj}|$$

c. Για $r \rightarrow \infty$ ("υπερ"απόσταση)

$$d(i,k) = \max |x_{ij} - x_{kj}|$$

Η Ευκλείδεια απόσταση είναι η πιο γνωστή από τα μέτρα του Minkowski.

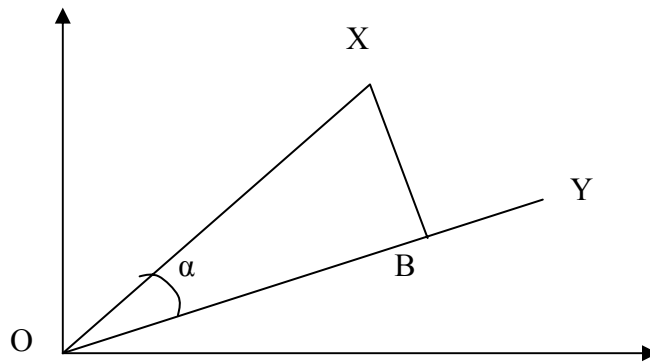
Άλλα γνωστά μέτρα είναι:

a. Pearson correlation (συσχέτιση)

Η συσχέτιση του Pearson είναι ένα μέτρο που λαμβάνει τιμές στο διάστημα $[-1, 1]$ ανάλογα με την γραμμική εξάρτηση των διανυσμάτων των τιμών. Δέχεται την τιμή 0 όταν δεν υπάρχει γραμμική εξάρτηση.

β. Cosine similarity measure

Επίσης λαμβάνει τιμές στο διάστημα $[-1, 1]$ δεχόμενο την τιμή 0 σε ορθογώνια διαστήματα. Έστω X και Y στοιχεία που αντιπροσωπεύονται από τα διανύσματα $X^T = (x_1, \dots, x_N)$ $Y^T = (y_1, \dots, y_N)$ όπου $X_i \quad i=1 \dots N$ οι τιμές που λαμβάνει το στοιχείο στις αντίστοιχες μεταβλητές.



Σχήμα 4

Στο χώρο των δύο διαστάσεων έχουμε την παραπάνω απεικόνιση.

$$X^T Y = |X| |Y| \cos \alpha$$

Το συνημίτονο της γωνίας α είναι ένα μέτρο ομοιότητας μεταξύ X και Y . Όσο περισσότερο παράλληλα είναι τα διανύσματα τόσο μεγαλύτερο είναι το συνημίτονο, με μέγιστη τιμή 1.

ΜΕΘΟΔΟΙ CLUSTER ANALYSIS

Στη συνέχεια παραθέτουμε τις βασικότερες μεθόδους (διαδικασίες) της cluster analysis. (16)

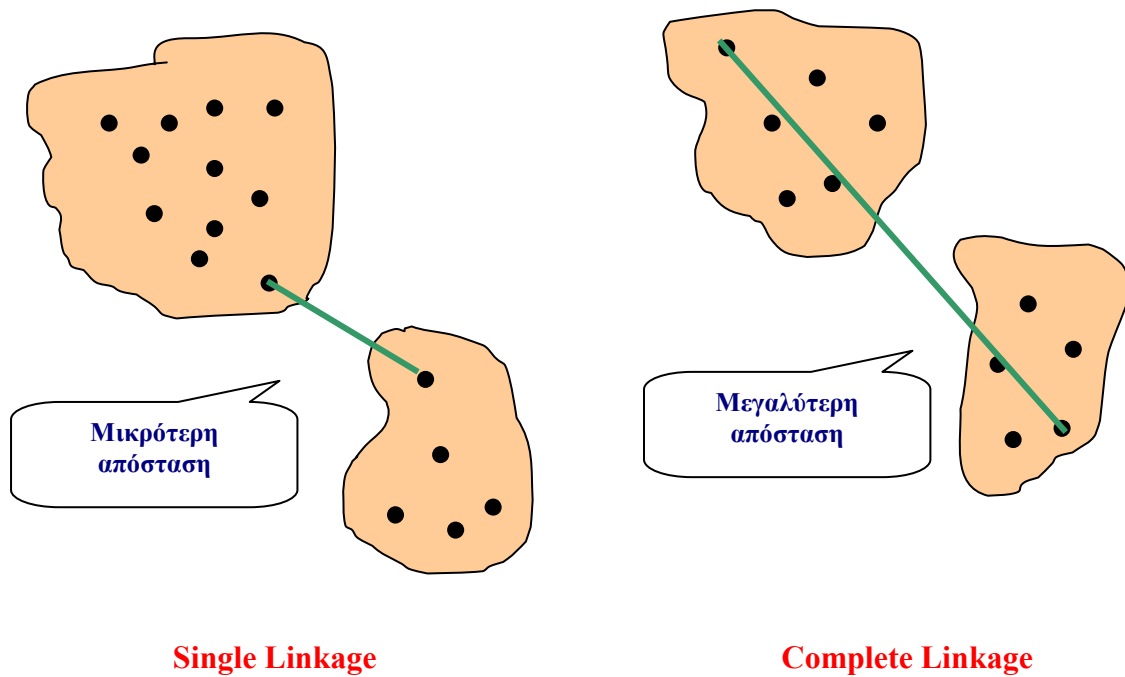
1. SINGLE LINKAGE

Η SINGLE LINKAGE είναι μία διαδικασία η οποία βασίζεται στην **μικρότερη απόσταση**. Τα δύο στοιχεία που χωρίζονται από την μικρότερη απόσταση αποτελούν την πρώτη ομάδα. Στη συνέχεια εξετάζεται η επόμενη κοντινότερη απόσταση, και ή ένα νέο στοιχείο συνδέεται με τα πρώτα δύο για να αποτελέσει ομάδα ή μαζί με κάποιο άλλο αποτελεί μία νέα ομάδα. Αυτή η διαδικασία ονομάζεται επίσης και nearest-neighbor .

Η απόσταση μεταξύ δυο ομάδων είναι η κοντινότερη απόσταση μεταξύ οποιουδήποτε σημείου της μίας ομάδας και οποιουδήποτε σημείου μεταξύ της άλλης ομάδας.

Δύο ομάδες συγχωνεύονται σε κάθε στάδιο όταν έχουν την μικρότερη απόσταση από κάθε άλλο ζεύγος.

Προβλήματα ανακύπτουν με αυτή τη μέθοδο ,όταν οι αποστάσεις μεταξύ των ομάδων δεν είναι σημαντικά μεγάλες. Σε αυτή την περίπτωση, στοιχεία που βρίσκονται στα αντίθετα άκρα μίας ομάδας μπορεί να είναι ανόμοια.



Σχήμα 5

2. COMPLETE LINKAGE

Η COMPLETE LINKAGE είναι μία διαδικασία η οποία παρουσιάζει συνάφεια με την προηγούμενη. Η διαφορά τους βρίσκεται στο γεγονός ,ότι βασίζεται στην **μεγαλύτερη απόσταση**. Για αυτό το λόγο ονομάζεται και furthest-neighbor διαδικασία.

Πιο συγκεκριμένα ,για να συγχωνεύσουμε δύο ομάδες ,εξετάζουμε τις μεγαλύτερες αποστάσεις μεταξύ σημείων των ομάδων και από αυτές επιλέγουμε την μικρότερη για συγχώνευση. Με αυτόν τον τρόπο εξαλείφουμε το πρόβλημα που παρουσίαζε η προηγούμενη μέθοδος.

3. AVERAGE LINKAGE

Η AVERAGE LINKAGE ξεκινά όπως και οι δύο προηγούμενες ,αλλά το κριτήριο που χρησιμοποιεί για συγχώνευση ομάδων ο μέσος όρος των αποστάσεων όλων των στοιχείων μίας ομάδας σε σχέση με κάθε στοιχείο της άλλης ομάδας. Το πλεονέκτημα αυτής της μεθόδου είναι ότι δεν εξαρτάται από ακραίες τιμές ,αλλά από την συμμετοχή όλων των μελών μίας ομάδας. Στόχος αυτής της διαδικασίας είναι να συνενώσει ομάδες που έχουν μικρές αποκλίσεις.

4. WARDS' METHOD

Η WARDS' METHOD ξεκινά με την εύρεση των μέσων των τιμών ,για όλες τις μεταβλητές ,των στοιχείων της ομάδας. Έπειτα , υπολογίζει την τετραγωνισμένη ευκλείδια απόσταση μεταξύ κάθε στοιχείου και του μέσου. Αυτές οι αποστάσεις στη συνέχεια προσθέτονται για όλα τα στοιχεία . Οι ομάδες

που συγχωνεύονται σε αυτήν την μέθοδο είναι αυτές που ελαχιστοποιούν αυτό το άθροισμα.

5. CENTROID METHOD

Το κριτήριο συγχώνευσης αυτής της μεθόδου είναι η απόσταση μεταξύ των κέντρων δύο ομάδων. Ως κέντρο λαμβάνουμε τις μέσες τιμές των στοιχείων της ομάδας για όλες τις μεταβλητές. Συγχωνεύουμε τις ομάδες με την μικρότερη απόσταση μεταξύ των κέντρων τους.

Μαθηματική παρουσίαση του WARD METHOD

Ο Ward (1963) και ο Wourd Hook (1963) ανέπτυξαν αυτόν τον αλγόριθμο.

Καθορίζουμε τις επόμενες ποσότητες:

- X_{ijk} : τιμή της i από τις h μεταβλητές, για το j από τα m στοιχεία, στην k από τις h ομάδες.

- $\bar{X}_{ik} = \sum_{j=1}^{J=mk} X_{ijk} / mk$: ο μέσος της I μεταβλητής για τα στοιχεία της k ομάδας.

- $E_k = \sum_{I=1}^{I=n} \sum_{J=1}^{J=mk} (X_{ijk} - \bar{X}_{ik})^2 = \sum_{I=1}^{I=n} \sum_{J=1}^{J=mk} X_{ijk}^2 - mk \sum_{I=1}^{I=n} \bar{X}_{ik}^2$

Λάθος του αθροίσματος των τετραγώνων για την k . : άθροισμα ευκλείδειων αποστάσεων από κάθε στοιχείο στην ομάδα και προς το διάνυσμα μέσω των τιμών της ομάδας k .

- $E = \sum E_k$: Σύνολο των αθροισμάτων για όλες τις ομάδες.

Ξεκινώντας τον αλγόριθμο, κάθε στοιχείο αποτελεί μια ομάδα από μόνο του. Έτσι το $E_k=0$ για όλες τις ομάδες. Στόχος του αλγορίθμου είναι να βρει σε κάθε στάδιο αυτές τις δύο ομάδες των οποίων η συγχώνευση δίνει την μικρότερη αύξηση του E .

ΕΦΑΡΜΟΓΗ ΤΗΣ CLUSTER ΑΝΑΛΥΣΗΣ ΣΤΗ ΚΑΤΗΓΟΡΙΟΠΟΙΗΣΗ ΤΩΝ ΤΡΑΠΕΖΙΚΩΝ ΥΠΟΚΑΤΑΣΤΗΜΑΤΩΝ

Στο συγκεκριμένο πρόβλημα θέλουμε να κατηγοριοποιήσουμε τα υπκ/τα μας σε 5 ομάδες. Στην πρώτη ομάδα θα βρίσκονται τα καλύτερα, ενώ στην τελευταία τα χειρότερα.

Εφαρμόζουμε την μέθοδο του Ward με μέτρο εγγύτητας την ευκλίδεια απόσταση. Θεωρήσαμε σκόπιμο και σύμφωνα με την φύση του προβλήματος να εφαρμόσουμε 28 κατηγοριοποιήσεις μία για κάθε χαρακτηριστικό.

Στην συνέχεια παρατηρούμε την συχνότητα εμφάνισης των υποκ/των σε κάθε ομάδα. Με την βοήθεια αυτών των μετρήσεων εξετάζουμε τις αποδόσεις κάθε υποκ/τος και αποφαινόμαστε σε ποιά ομάδα ανήκει.

Ομαδοποίηση στο κριτήριο **K1 : ΠΡΟΣΩΠΙΚΟ**

Cluster Membership of Cases using Ward Method

Number of Clusters		
Label	Case	5
Case 1	1	5
Case 2	2	4
Case 3	3	5
Case 4	4	2
Case 5	5	5
Case 6	6	2
Case 7	7	3
Case 8	8	2
Case 9	9	3
Case 10	10	3
Case 11	11	5
Case 12	12	4
Case 13	13	2
Case 14	14	4
Case 15	15	4
Case 16	16	2
Case 17	17	4
Case 18	18	4
Case 19	19	2
Case 20	20	5
Case 21	21	4
Case 22	22	4
Case 23	23	4
Case 24	24	5
Case 25	25	2
Case 26	26	1
Case 27	27	1

Ομαδοποίηση στο κριτήριο **K2 : ΤΑΜΕΙΑΚΕΣ ΠΡΑΞΕΙΣ ΕΙΣΠΡ.**

Cluster Membership of Cases using Ward Method

Number of Clusters		
Label	Case	5
Case 1	1	3
Case 2	2	4
Case 3	3	5
Case 4	4	3
Case 5	5	3
Case 6	6	3
Case 7	7	3
Case 8	8	3
Case 9	9	4
Case 10	10	3
Case 11	11	4
Case 12	12	3
Case 13	13	2
Case 14	14	4
Case 15	15	4
Case 16	16	3
Case 17	17	5
Case 18	18	4
Case 19	19	2
Case 20	20	5
Case 21	21	4
Case 22	22	5
Case 23	23	4
Case 24	24	2
Case 25	25	3
Case 26	26	2
Case 27	27	1

Ομαδοποίηση στο κριτήριο **K3 : TAM. ΠΡΑΞΕΙΣ ΕΝΤΟΛΕΣ**

Cluster Membership of Cases using Ward Method

Number of Clusters		
Label	Case	5
Case 1	1	4
Case 2	2	5
Case 3	3	4
Case 4	4	4
Case 5	5	4
Case 6	6	2
Case 7	7	2
Case 8	8	4
Case 9	9	4
Case 10	10	3
Case 11	11	4
Case 12	12	5
Case 13	13	2
Case 14	14	4
Case 15	15	4
Case 16	16	4
Case 17	17	4
Case 18	18	4
Case 19	19	3
Case 20	20	4
Case 21	21	4
Case 22	22	4
Case 23	23	3
Case 24	24	1
Case 25	25	3
Case 26	26	2
Case 27	27	1

Ομαδοποίηση στο κριτήριο **K4 : ΕΠΙΤ. ΕΝΤΟΛ. ΤΡΑΠ. ΠΟΥ ΕΚΔ.**

Cluster Membership of Cases using Ward Method

Number of Clusters		
Label	Case	5
Case 1	1	3
Case 2	2	2
Case 3	3	4
Case 4	4	2
Case 5	5	5
Case 6	6	4
Case 7	7	3
Case 8	8	2
Case 9	9	3
Case 10	10	3
Case 11	11	5
Case 12	12	3
Case 13	13	2
Case 14	14	4
Case 15	15	5
Case 16	16	2
Case 17	17	2
Case 18	18	4
Case 19	19	3
Case 20	20	4
Case 21	21	1
Case 22	22	4
Case 23	23	3
Case 24	24	4
Case 25	25	3
Case 26	26	2
Case 27	27	1

Ομαδ. στο κριτήριο K5 : ΕΠΙΤ. ΕΝΤΟΛ. ΤΡΑΠ. ΠΟΥ ΠΛΗΡΩΘ.

Cluster Membership of Cases using Ward Method

Number of Clusters		
Label	Case	5
Case 1	1	3
Case 2	2	4
Case 3	3	4
Case 4	4	5
Case 5	5	3
Case 6	6	5
Case 7	7	4
Case 8	8	5
Case 9	9	5
Case 10	10	5
Case 11	11	5
Case 12	12	4
Case 13	13	3
Case 14	14	4
Case 15	15	2
Case 16	16	3
Case 17	17	3
Case 18	18	5
Case 19	19	4
Case 20	20	5
Case 21	21	1
Case 22	22	5
Case 23	23	5
Case 24	24	3
Case 25	25	2
Case 26	26	3
Case 27	27	2

Ομαδοποίηση στο κριτήριο **K6 : ΧΡΗ/ΛΕΣ ΞΕΝ. ΤΡΑΠ. ΕΞΩΤ.**

Cluster Membership of Cases using Ward Method

Number of Clusters		
Label	Case	5
Case 1	1	4
Case 2	2	4
Case 3	3	4
Case 4	4	2
Case 5	5	5
Case 6	6	5
Case 7	7	3
Case 8	8	5
Case 9	9	4
Case 10	10	1
Case 11	11	5
Case 12	12	4
Case 13	13	4
Case 14	14	3
Case 15	15	5
Case 16	16	2
Case 17	17	4
Case 18	18	5
Case 19	19	5
Case 20	20	2
Case 21	21	5
Case 22	22	2
Case 23	23	4
Case 24	24	5
Case 25	25	4
Case 26	26	5
Case 27	27	5

Ομάδ. στο κριτ. K7 : ΧΡΗ/ΛΕΣ ΞΕΝ. ΤΡΑΠ. ΕΞΩΤ. ΜΕΣΩ Κ.Κ

Cluster Membership of Cases using Ward Method

Number of Clusters		
Label	Case	5
Case 1	1	5
Case 2	2	5
Case 3	3	5
Case 4	4	2
Case 5	5	5
Case 6	6	3
Case 7	7	3
Case 8	8	3
Case 9	9	3
Case 10	10	3
Case 11	11	5
Case 12	12	3
Case 13	13	3
Case 14	14	5
Case 15	15	5
Case 16	16	5
Case 17	17	3
Case 18	18	2
Case 19	19	1
Case 20	20	2
Case 21	21	5
Case 22	22	3
Case 23	23	3
Case 24	24	5
Case 25	25	3
Case 26	26	3
Case 27	27	5

Ομαδοποίηση στο κριτήριο **K8 : ΧΡΗ/ΛΕΣ ΞΕΝ. ΤΡΑΠ. ΠΡΟΣ Κ.Κ**

Cluster Membership of Cases using Ward Method

Number of Clusters		
Label	Case	5
Case 1	1	2
Case 2	2	3
Case 3	3	2
Case 4	4	5
Case 5	5	5
Case 6	6	2
Case 7	7	3
Case 8	8	4
Case 9	9	4
Case 10	10	5
Case 11	11	5
Case 12	12	5
Case 13	13	5
Case 14	14	3
Case 15	15	5
Case 16	16	1
Case 17	17	4
Case 18	18	3
Case 19	19	4
Case 20	20	5
Case 21	21	2
Case 22	22	5
Case 23	23	5
Case 24	24	2
Case 25	25	5
Case 26	26	5
Case 27	27	5

Ομαδοποίηση στο κριτήριο **K9 : ΑΓΟΡΑ ΞΕΝ. ΤΡΑΠΕΖΟΓΡΑΜΜ.**

Cluster Membership of Cases using Ward Method

Number of Clusters		
Label	Case	5
Case 1	1	4
Case 2	2	4
Case 3	3	3
Case 4	4	5
Case 5	5	4
Case 6	6	2
Case 7	7	3
Case 8	8	3
Case 9	9	4
Case 10	10	3
Case 11	11	5
Case 12	12	5
Case 13	13	4
Case 14	14	4
Case 15	15	5
Case 16	16	3
Case 17	17	4
Case 18	18	2
Case 19	19	1
Case 20	20	4
Case 21	21	5
Case 22	22	4
Case 23	23	5
Case 24	24	3
Case 25	25	3
Case 26	26	5
Case 27	27	2

Ομαδ. στο κριτήριο **K10 : ΠΩΛ. ΞΕΝ. ΤΡΑΠ/ΤΙΩΝ ΣΕ ΠΕΛΑΤ.**

Cluster Membership of Cases using Ward Method

Number of Clusters		
Label	Case	5
Case 1	1	4
Case 2	2	4
Case 3	3	5
Case 4	4	4
Case 5	5	3
Case 6	6	3
Case 7	7	3
Case 8	8	3
Case 9	9	4
Case 10	10	5
Case 11	11	2
Case 12	12	4
Case 13	13	2
Case 14	14	4
Case 15	15	3
Case 16	16	3
Case 17	17	5
Case 18	18	5
Case 19	19	3
Case 20	20	5
Case 21	21	3
Case 22	22	5
Case 23	23	5
Case 24	24	3
Case 25	25	3
Case 26	26	5
Case 27	27	1

**Ομαδ. στο κριτήριο K11 : ΠΩΛ. ΞΕΝ. ΤΡΑΠ/ΤΙΩΝ ΣΕ ΠΕΛ. &
ΧΡΗ/ΛΕΣ ΠΡΟΣ Κ.Κ**

Cluster Membership of Cases using Ward Method

Number of Clusters		
Label	Case	5
Case 1	1	3
Case 2	2	4
Case 3	3	3
Case 4	4	5
Case 5	5	4
Case 6	6	5
Case 7	7	3
Case 8	8	3
Case 9	9	3
Case 10	10	3
Case 11	11	5
Case 12	12	2
Case 13	13	2
Case 14	14	5
Case 15	15	5
Case 16	16	1
Case 17	17	3
Case 18	18	2
Case 19	19	1
Case 20	20	2
Case 21	21	4
Case 22	22	3
Case 23	23	3
Case 24	24	2
Case 25	25	3
Case 26	26	3
Case 27	27	5

Ομαδ. στο κριτήριο K12 : ΑΓΟΡΑ ΕΠΙΤΑΓΩΝ

Cluster Membership of Cases using Ward Method

Number of Clusters		
Label	Case	5
Case 1	1	5
Case 2	2	5
Case 3	3	3
Case 4	4	5
Case 5	5	5
Case 6	6	2
Case 7	7	5
Case 8	8	4
Case 9	9	4
Case 10	10	2
Case 11	11	5
Case 12	12	4
Case 13	13	5
Case 14	14	4
Case 15	15	5
Case 16	16	3
Case 17	17	5
Case 18	18	4
Case 19	19	1
Case 20	20	4
Case 21	21	5
Case 22	22	5
Case 23	23	5
Case 24	24	5
Case 25	25	3
Case 26	26	1
Case 27	27	3

Ομάδ. στο κριτήριο **K13 : ΕΚΔΟΣΗ ΕΠΙΤΑΓΩΝ**

Cluster Membership of Cases using Ward Method

Number of Clusters		
Label	Case	5
Case 1	1	1
Case 2	2	1
Case 3	3	1
Case 4	4	3
Case 5	5	5
Case 6	6	2
Case 7	7	2
Case 8	8	4
Case 9	9	4
Case 10	10	5
Case 11	11	5
Case 12	12	3
Case 13	13	3
Case 14	14	4
Case 15	15	5
Case 16	16	2
Case 17	17	5
Case 18	18	3
Case 19	19	3
Case 20	20	5
Case 21	21	5
Case 22	22	5
Case 23	23	5
Case 24	24	5
Case 25	25	3
Case 26	26	1
Case 27	27	1

Ομάδ. στο κριτήριο K14 : ΕΙΣΑΓΩΓΗ ΕΝΤΟΛΩΝ

Cluster Membership of Cases using Ward Method

Number of Clusters		
Label	Case	5
Case 1	1	5
Case 2	2	5
Case 3	3	5
Case 4	4	3
Case 5	5	5
Case 6	6	2
Case 7	7	3
Case 8	8	3
Case 9	9	5
Case 10	10	3
Case 11	11	5
Case 12	12	5
Case 13	13	5
Case 14	14	5
Case 15	15	5
Case 16	16	4
Case 17	17	5
Case 18	18	5
Case 19	19	2
Case 20	20	5
Case 21	21	5
Case 22	22	5
Case 23	23	5
Case 24	24	5
Case 25	25	3
Case 26	26	1
Case 27	27	1

Ομάδ. στο κριτήριο K15 : ΕΚΔΟΣΗ ΕΝΤΟΛΩΝ

Cluster Membership of Cases using Ward Method

Number of Clusters		
Label	Case	5
Case 1	1	5
Case 2	2	5
Case 3	3	4
Case 4	4	4
Case 5	5	5
Case 6	6	5
Case 7	7	5
Case 8	8	3
Case 9	9	4
Case 10	10	3
Case 11	11	5
Case 12	12	5
Case 13	13	5
Case 14	14	5
Case 15	15	5
Case 16	16	3
Case 17	17	5
Case 18	18	5
Case 19	19	3
Case 20	20	5
Case 21	21	5
Case 22	22	5
Case 23	23	5
Case 24	24	5
Case 25	25	2
Case 26	26	1
Case 27	27	1

Ομάδ. στο κριτήριο **K16 : ΣΥΝΑΛ/ΜΑ ΓΙΑ ΛΟΙΠΕΣ ΑΝΑΓΚΕΣ**

Cluster Membership of Cases using Ward Method

Number of Clusters		
Label	Case	5
Case 1	1	5
Case 2	2	5
Case 3	3	5
Case 4	4	2
Case 5	5	5
Case 6	6	2
Case 7	7	2
Case 8	8	2
Case 9	9	4
Case 10	10	4
Case 11	11	4
Case 12	12	5
Case 13	13	4
Case 14	14	4
Case 15	15	4
Case 16	16	2
Case 17	17	5
Case 18	18	4
Case 19	19	2
Case 20	20	5
Case 21	21	5
Case 22	22	5
Case 23	23	5
Case 24	24	5
Case 25	25	3
Case 26	26	1
Case 27	27	1

Ομαδ. στο κριτήριο K17 : ΕΙΣΑΓ. ΑΞΙΩΝ ΕΞΩΤ. ΕΣΩΤ.

Cluster Membership of Cases using Ward Method

Number of Clusters		
Label	Case	5
Case 1	1	5
Case 2	2	5
Case 3	3	2
Case 4	4	2
Case 5	5	5
Case 6	6	2
Case 7	7	4
Case 8	8	5
Case 9	9	4
Case 10	10	4
Case 11	11	5
Case 12	12	5
Case 13	13	4
Case 14	14	4
Case 15	15	5
Case 16	16	3
Case 17	17	5
Case 18	18	5
Case 19	19	1
Case 20	20	5
Case 21	21	5
Case 22	22	5
Case 23	23	5
Case 24	24	5
Case 25	25	4
Case 26	26	4
Case 27	27	2

Ομαδ. στο κριτήριο **K18 : ΕΚΚΡΕΜ. ΑΞΙΕΣ ΣΕ ΤΡΙΤΟΥΣ**

Cluster Membership of Cases using Ward Method

Number of Clusters		
Label	Case	5
Case 1	1	5
Case 2	2	5
Case 3	3	4
Case 4	4	5
Case 5	5	5
Case 6	6	3
Case 7	7	5
Case 8	8	4
Case 9	9	3
Case 10	10	5
Case 11	11	5
Case 12	12	5
Case 13	13	5
Case 14	14	5
Case 15	15	5
Case 16	16	4
Case 17	17	5
Case 18	18	5
Case 19	19	1
Case 20	20	4
Case 21	21	5
Case 22	22	5
Case 23	23	5
Case 24	24	5
Case 25	25	5
Case 26	26	5
Case 27	27	2

Ομάδ. στο κριτήριο K19 : ENTOΛΕΣ ΣΥΝ/ΤΟΣ ΚΑΗΡΙΠΓΚ

Cluster Membership of Cases using Ward Method

Number of Clusters		
Label	Case	5
Case 1	1	5
Case 2	2	5
Case 3	3	5
Case 4	4	5
Case 5	5	5
Case 6	6	5
Case 7	7	5
Case 8	8	4
Case 9	9	5
Case 10	10	5
Case 11	11	5
Case 12	12	3
Case 13	13	5
Case 14	14	5
Case 15	15	5
Case 16	16	1
Case 17	17	5
Case 18	18	5
Case 19	19	5
Case 20	20	5
Case 21	21	5
Case 22	22	5
Case 23	23	5
Case 24	24	5
Case 25	25	5
Case 26	26	2
Case 27	27	1

Ομαδ. στο κριτήριο K20 : ENTOΛΕΣ ΣΥΝ/ΤΟΣ ΠΟΥ ΔΙΑΒΙΒ.

Cluster Membership of Cases using Ward Method

Number of Clusters		
Label	Case	5
Case 1	1	3
Case 2	2	4
Case 3	3	5
Case 4	4	3
Case 5	5	4
Case 6	6	2
Case 7	7	3
Case 8	8	3
Case 9	9	3
Case 10	10	3
Case 11	11	5
Case 12	12	5
Case 13	13	3
Case 14	14	4
Case 15	15	5
Case 16	16	5
Case 17	17	3
Case 18	18	4
Case 19	19	3
Case 20	20	4
Case 21	21	3
Case 22	22	5
Case 23	23	5
Case 24	24	5
Case 25	25	3
Case 26	26	1
Case 27	27	1

Ομαδ. στο κριτήριο **K21 : ΕΓΚΡ. ΕΙΣΑΓ. ΠΟΥ ΕΚΔΟΘΗΚΑΝ**

Cluster Membership of Cases using Ward Method

Number of Clusters		
Label	Case	5
Case 1	1	3
Case 2	2	5
Case 3	3	5
Case 4	4	5
Case 5	5	5
Case 6	6	5
Case 7	7	5
Case 8	8	5
Case 9	9	5
Case 10	10	5
Case 11	11	3
Case 12	12	5
Case 13	13	2
Case 14	14	5
Case 15	15	5
Case 16	16	2
Case 17	17	4
Case 18	18	4
Case 19	19	2
Case 20	20	5
Case 21	21	3
Case 22	22	4
Case 23	23	5
Case 24	24	2
Case 25	25	3
Case 26	26	1
Case 27	27	1

Ομαδ. στο κριτήριο K22 : ΑΔΕΙΕΣ ΕΚΤΕΛ. ΠΟΥ ΕΚΔΟΘ.

Cluster Membership of Cases using Ward Method

Number of Clusters		
Label	Case	5
Case 1	1	5
Case 2	2	5
Case 3	3	5
Case 4	4	5
Case 5	5	5
Case 6	6	5
Case 7	7	5
Case 8	8	5
Case 9	9	5
Case 10	10	3
Case 11	11	5
Case 12	12	5
Case 13	13	5
Case 14	14	5
Case 15	15	5
Case 16	16	5
Case 17	17	4
Case 18	18	4
Case 19	19	2
Case 20	20	5
Case 21	21	5
Case 22	22	5
Case 23	23	5
Case 24	24	5
Case 25	25	3
Case 26	26	1
Case 27	27	4

Ομάδ. στο κριτήριο K23 : ΑΔΕΙΕΣ ΕΚΤΕΛΩΝ. ΕΚΚΡ. ΤΡΑΠ.

Cluster Membership of Cases using Ward Method

Number of Clusters		
Label	Case	5
Case 1	1	3
Case 2	2	4
Case 3	3	2
Case 4	4	5
Case 5	5	4
Case 6	6	1
Case 7	7	5
Case 8	8	4
Case 9	9	4
Case 10	10	5
Case 11	11	2
Case 12	12	5
Case 13	13	3
Case 14	14	4
Case 15	15	5
Case 16	16	2
Case 17	17	3
Case 18	18	5
Case 19	19	2
Case 20	20	5
Case 21	21	5
Case 22	22	4
Case 23	23	5
Case 24	24	5
Case 25	25	4
Case 26	26	5
Case 27	27	5

Ομ. στο κρ. K24 : ΑΔ.ΕΚΤΕΛ.ΕΚΚΡ.ΧΩΡΙΣ ΣΥΝΑΛ.ΔΙΑΤΥΠ.

Cluster Membership of Cases using Ward Method

Number of Clusters		
Label	Case	5
Case 1	1	2
Case 2	2	5
Case 3	3	5
Case 4	4	5
Case 5	5	5
Case 6	6	5
Case 7	7	5
Case 8	8	5
Case 9	9	5
Case 10	10	5
Case 11	11	5
Case 12	12	5
Case 13	13	5
Case 14	14	5
Case 15	15	5
Case 16	16	5
Case 17	17	5
Case 18	18	5
Case 19	19	5
Case 20	20	5
Case 21	21	5
Case 22	22	5
Case 23	23	5
Case 24	24	5
Case 25	25	5
Case 26	26	5
Case 27	27	1

Ομ. στο κρ. K25 : ΑΗΦΘΕΙΣΕΣ ΕΚΔΟΣΕΙΣ ΕΞΑΓΩΓΗΣ

Cluster Membership of Cases using Ward Method

Number of Clusters		
Label	Case	
Case 1	1	5
Case 2	2	5
Case 3	3	5
Case 4	4	5
Case 5	5	3
Case 6	6	4
Case 7	7	5
Case 8	8	2
Case 9	9	4
Case 10	10	5
Case 11	11	5
Case 12	12	5
Case 13	13	5
Case 14	14	5
Case 15	15	5
Case 16	16	5
Case 17	17	5
Case 18	18	5
Case 19	19	5
Case 20	20	5
Case 21	21	5
Case 22	22	5
Case 23	23	5
Case 24	24	5
Case 25	25	5
Case 26	26	5
Case 27	27	1

Ομ. στο κρ. K26: ΑΔΕΙΕΣ ΕΞΑΓΩΓΗΣ ΠΟΥ ΕΚΔΟΘΗΚΑΝ

Cluster Membership of Cases using Ward Method

Number of Clusters		
Label	Case	
Case 1	1	5
Case 2	2	5
Case 3	3	5
Case 4	4	5
Case 5	5	4
Case 6	6	4
Case 7	7	4
Case 8	8	2
Case 9	9	4
Case 10	10	5
Case 11	11	4
Case 12	12	5
Case 13	13	5
Case 14	14	5
Case 15	15	3
Case 16	16	3
Case 17	17	5
Case 18	18	5
Case 19	19	5
Case 20	20	5
Case 21	21	5
Case 22	22	5
Case 23	23	5
Case 24	24	2
Case 25	25	5
Case 26	26	5
Case 27	27	1

Ομ. στο κρ. K27: ΕΚΚΡ. ΔΗΛΩΣ. ΚΑΙ ΑΔΕΙΕΣ ΕΞΑΓΩΓ.

Cluster Membership of Cases using Ward Method

Number of Clusters

Label	Case	5
Case 1	1	4
Case 2	2	5
Case 3	3	5
Case 4	4	4
Case 5	5	4
Case 6	6	4
Case 7	7	5
Case 8	8	3
Case 9	9	4
Case 10	10	5
Case 11	11	4
Case 12	12	5
Case 13	13	5
Case 14	14	5
Case 15	15	4
Case 16	16	3
Case 17	17	4
Case 18	18	5
Case 19	19	4
Case 20	20	5
Case 21	21	5
Case 22	22	5
Case 23	23	5
Case 24	24	4
Case 25	25	5
Case 26	26	2
Case 27	27	1

Ομ. στο κρ. K28: ΟΜΟΛΟΓ. Κ ΤΟΚΟΜΕΡ. ΠΟΥ ΕΞΟΦΛΗΘ.

Cluster Membership of Cases using Ward Method

Number of Clusters

Label	Case	5
Case 1	1	3
Case 2	2	4
Case 3	3	4
Case 4	4	3
Case 5	5	5
Case 6	6	3
Case 7	7	3
Case 8	8	4
Case 9	9	4
Case 10	10	4
Case 11	11	3
Case 12	12	4
Case 13	13	3
Case 14	14	3
Case 15	15	4
Case 16	16	2
Case 17	17	4
Case 18	18	4
Case 19	19	5
Case 20	20	5
Case 21	21	3
Case 22	22	5
Case 23	23	5
Case 24	24	5
Case 25	25	5
Case 26	26	2
Case 27	27	1

	1	2	3	4	5
27	16	6	1	2	3
26	9	6	3	2	7
25		2	15	1	7
24		4	3	3	18
23			4	3	21
22		1	2	4	21
21	3	1	5	4	14
20		3		6	19
19	6	7	6	3	6
18		3	3	10	12
17		1	5	9	13
16	3	8	9	3	5
15		1	1	6	20
14			3	13	12
13		8	6	7	7
12		2	5	6	16
11		2	2	5	18
10	1	1	11	5	10
9			6	14	8
8		3	9	8	8
7		3	12	3	10
6	1	10	5	4	8
5			3	6	19
4		7	6	6	9
3		3	3	6	16
2		1	1	12	14
1		2	9	6	11

Τα παραπάνω αποτελέσματα σε συνδυασμό με την κρίση του αποφασίζοντα θα μας οδηγήσουν στον σχηματισμό των ομάδων.

Ας ξεκινήσουμε από την πρώτη ομάδα. Παρατηρούμε ότι τα υποκαταστήματα Y27,Y26 παρουσιάζουν τις περισσότερες συμμετοχές τους σε αυτή την ομάδα , έχοντας μάλιστα μεγάλη διαφορά όγκου εργασιών από τα άλλα υποκ/τα.

Τα υποκ/τα 6,19,13 εμφανίζονται πιο συχνά στην δεύτερη ομάδα .Το υποκ/μα 16 το τοποθετούμε στην δεύτερη ομάδα αν και παρουσιάζεται περισσότερες φορές στην τρίτη ομάδα γιατί αναδεικνύεται πρώτο σε 3 χαρακτηριστικά , στα οποία τα περισσότερα υποκ/τα παρουσιάζουν λιγιστές ή και μηδαμινές εργασίες.

Επίσης το υπ/μα 4 το τοποθετούμε στη δεύτερη ομάδα λόγω του , ότι , βρίσκεται στο γκρουπ των καλών υποκ/των σε τραπεζικές εργασίες που απαιτούν μεγάλη χρονική διάρκεια.

Τα υποκ/τα 7,8,10,25 εμφανίζονται σταθερά στην ομάδα 3 .Το υποκ/μα 1 εισέρχεται στην ομάδα από τη συμμετοχή του στις πρώτες θέσεις χαρακτηριστικών όπως το K24 όπου εργασίες παρουσιάζουν μόνο το Y1 και το Y27.

Το Y21 απουσιάζει από αρκετές εργασίες στις οποίες απουσιάζει και η πλειονότητα των υποκ/των. Όμως, πρωτεύει σε κάποια χαρακτηριστικά συγκεντρώνοντας μεγάλο όγκο εργασιών ,εισερχόμενο έτσι στην 3η ομάδα.

Τα Y12,Y2,Y17,Y18 επειδή πετυχαίνουν καλά αποτελέσματα σε ορισμένα χαρακτηριστικά , αποφεύγουν την τελευταία ομάδα και αποτελούν μαζί με τα Y9,Y14 την 4η ομάδα.

Τα υπόλοιπα υποκ/τα με χαμηλές επιδόσεις αποτελούν την 5η ομάδα.

Στις ομάδες που καταλήγουμε είναι :

ΟΜΑΔΑ 1

27, 26

ΟΜΑΔΑ 2

19, 16, 6, 4, 13

ΟΜΑΔΑ 3

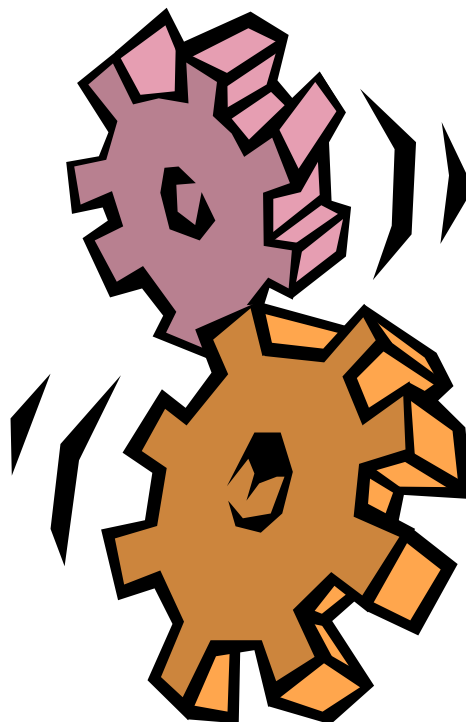
1, 10, 25, 7, 8, 21

ΟΜΑΔΑ 4

14, 17, 18, 9, 2, 12

ΟΜΑΔΑ 5

5, 3, 20, 22, 23, 24, 11, 15



ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Είναι κλάδος της τεχνητής νοημοσύνης, ο οποίος μελετά τη φύση της ανθρώπινης μάθησης και προσπαθεί να υλοποιήσει τους μηχανισμούς της στα υπολογιστικά συστήματα. Η μηχανική μάθηση έχει τρεις βασικές ερευνητικές κατευθύνσεις: (3)

1. **ΠΡΟΣΑΝΑΤΟΛΙΣΜΟΣ ΩΣ ΠΡΟΣ ΤΟΝ ΣΤΟΧΟ (TASK-ORIENTED STUDIES)** που αφορά στην ανάπτυξη συστημάτων για την επίλυση συγκεκριμένων προβλημάτων.
2. **ΠΡΟΣΟΜΟΙΩΣΗ ΓΝΩΣΤΙΚΩΝ ΛΕΙΤΟΥΡΓΙΩΝ (COGNITIVE SIMULATION)** που αφορά στην έρευνα και προσομοίωση του ανθρώπινου τρόπου μάθησης.
3. **ΘΕΩΡΗΤΙΚΗ ΑΝΑΛΥΣΗ (THEORETICAL ANALYSIS)** που αφορά στην έρευνα των δυνατών μηχανισμών μάθησης, ανεξαρτήτως πεδίου εφαρμογής.

Μπορούμε να διακρίνουμε στον μηχανισμό ανθρώπινης μάθησης δύο βασικές μορφές που είναι:

- α. **Απόκτηση γνώσης (Knowledge acquisition)**
- β. **Αναπροσαρμογή χειρισμών (Skill refinement)**

Η πρώτη μπορεί να είναι και υποσυνείδητη ενώ η δεύτερη είναι, βεβαίως, συνειδητή προσπάθεια του ανθρώπου να βελτιώσει τις ικανότητες του στον χειρισμό κάποιων καταστάσεων με βάση της εξάσκηση.

Ο συγκερασμός των δύο αυτών μορφών μάθησης αποτελεί την ανθρώπινη νοημοσύνη. Ο άνθρωπος προβαίνει σε κάποιους χειρισμούς, κατευθυνόμενους απ'την προσλαμβάνουσα γνώση, τους οποίους αναπροσαρμόζει αν αποδειχθούν αναποτελεσματικοί, οδηγούμενος έτσι σε απόκτηση νέας γνώσης.

Ο χώρος της Μηχανικής Μάθησης διαιρείται σε δύο βασικούς τομείς:

α. Την Συμπερασματική

β. Την Επαγωγική

Στην **Συμπερασματική** δημιουργούμε υποθέσεις με την βοήθεια συμπερασματικής λογικής πάνω στην πληροφορία που κατέχουμε.

Στην **Επαγωγική** δημιουργούμε με βάση την γνώση που έχουμε καινούργιες υποθέσεις για **έννοιες (concepts)**, εξαιρέσεις, επεξηγήσεις ή υψηλού επιπέδου χαρακτηρισμοί της πληροφορίας εισόδου.

Ανάλογα με το επίπεδο της πληροφορίας που κατέχουμε, μπορούμε να λάβουμε τέσσερα είδη υποθέσεων.

1. ΜΑΘΗΣΗ ΧΩΡΙΣ ΥΠΟΘΕΣΕΙΣ (ROTE LEARNING)

Το σύστημα διαθέτει ακριβώς την πληροφορία που του χρειάζεται και δεν είναι αναγκαία η δημιουργία υποθέσεων.

2. ΜΑΘΗΣΗ ΜΕΣΩ ΟΔΗΓΙΩΝ (LEARNING BY BEING TOLD)

Η πληροφορία είναι γενική και το σύστημα πρέπει να δημιουργήσει υποθέσεις για την πληροφορία που λείπει.

3. ΜΑΘΗΣΗ ΜΕΣΩ ΠΑΡΑΔΕΙΓΜΑΤΩΝ (LEARNING FROM EXAMPLES)

Η πληροφορία είναι πολύ εξειδικευμένη και το σύστημα πρέπει να δημιουργήσει υποθέσεις για πιο γενικούς κανόνες.

4. ΜΑΘΗΣΗ ΜΕΣΩ ΑΝΑΛΟΓΙΩΝ (LEARNING BY ANALOGY)

Η πληροφορία σχετίζεται με ανάλογα προβλήματα, οπότε το σύστημα πρέπει να εντοπίζει τις αναλογίες και να δημιουργήσει ανάλογες υποθέσεις για το τρέχον πρόβλημα.

Ο ΑΛΓΟΡΙΘΜΟΣ ID3

Ο ID3 είναι ένας αλγόριθμος κατάταξης απλός αλλά ταυτόχρονα πολύ αποτελεσματικός. Χρησιμοποιεί μια ευριστική διαδικασία κατασκευής κανόνων, τέτοιων που να πετυχαίνουν διαχωρισμό μεταξύ των κλάσεων. Για να μπορέσουμε όμως να εξάγουμε τέτοιους κανόνες, θεωρείται απαραίτητη η είσοδος του συστήματος. (1),(2)

- η περιγραφή του πεδίου εφαρμογής
- ένα σύνολο παραδειγμάτων

Το πεδίο εφαρμογής περιγράφεται από ένα σύνολο **ιδιοτήτων (attributies)**, καθώς και από τις τιμές (**attribute values**) που λαμβάνει κάθε μια απ' αυτές.

Το σύνολο παραδειγμάτων εξυπηρετεί εκπαιδευτικούς σκοπούς και για αυτό συνήθως ονομάζεται σύνολο παραδειγμάτων διδασκαλίας (**training set**).

Κάθε διάνυσμα τιμών των ιδιοτήτων μαζί με την πληροφορία για την συγκεκριμένη κλάση στην οποία ανήκει αυτό το διάνυσμα, αποτελεί ένα **παράδειγμα**.

Το σύνολο των κανόνων που παράγονται, μας δίνονται με την μορφή ενός **δένδρου απόφασης**. Οι κόμβοι του δένδρου αντιστοιχούν σε ελέγχους των τιμών κάποιων ιδιοτήτων, ενώ τα κλαδιά είναι οι δυνατές τιμές των αντίστοιχων ιδιοτήτων.

Κάθε φύλλο αναπαριστά μια συγκεκριμένη κλάση. Κάθε μονοπάτι του δένδρου από την ρίζα του μέχρι κάποιο συγκεκριμένο φύλλο αποτελεί ένα διαφορετικό κανόνα κατάταξης

παραδειγμάτων σε μια από τις κλάσεις.

Ο ID3 είναι ένας **αναδρομικός** αλγόριθμος, ο οποίος σε κάθε βήμα του επιλέγει την ιδιότητα εκείνη η οποία πετυχαίνει τον καλύτερο διαχωρισμό μεταξύ παραδειγμάτων διαφορετικών κλάσεων. Η ιδιότητα που επιλέγεται κάθε φορά είναι εκείνη με την μεγαλύτερη **πληροφοριακή ισχύ**.

Πρακτικά, μια ιδιότητα έχει μεγάλη πληροφοριακή ισχύ αν οι τιμές της στο σύνολο των παραδειγμάτων είναι τέτοιες, ώστε να πετυχαίνεται όσο το δυνατόν μεγαλύτερος διαχωρισμός μεταξύ των παραδειγμάτων που ανήκουν σε διαφορετική κλάση.

Η ιδιότητα με την μεγαλύτερη πληροφοριακή ισχύ γίνεται κόμβος του δένδρου και δημιουργούνται τόσα κλαδιά, όσες και οι τιμές της.

Το σύνολο των παραδειγμάτων χωρίζεται σε υποσύνολα, έτσι ώστε κάθε υποσύνολο να περιέχει εκείνα τα παραδείγματα που έχουν μια συγκεκριμένη τιμή στην ιδιότητα που επιλέχθηκε. Κάθε τέτοιο υποσύνολο αποτελεί το σύνολο παραδειγμάτων για το αντίστοιχο κλαδί.

Ο αλγόριθμος προσπαθεί σε κάθε βήμα να καλύψει τα παραδείγματα, δηλαδή να δημιουργήσει υποσύνολα του αρχικού συνόλου παραδειγμάτων, έτσι ώστε τα παραδείγματα κάθε υποσυνόλου να ανήκουν σε μια και μόνο κλάση.

ΠΟΛΥΠΛΟΚΟΤΗΤΑ ΤΟΥ ΑΛΓΟΡΙΘΜΟΥ

Η πολυπλοκότητα εύρεσης κάθε φύλλου είναι ανάλογη του αριθμού των παραδειγμάτων (EX) και του αριθμού ιδιοτήτων (A) $O([EX] X[A])$.

Η ολική πολυπλοκότητα του αλγόριθμου είναι ανάλογη του γινομένου του αριθμού παραδειγμάτων, του αριθμού ιδιοτήτων και του αριθμού των κόμβων του δένδρου που είναι τα φύλλα.

Μέχρι τώρα τα πειράματα έχουν δείξει ότι ο ID3 δεν παρουσιάζει εκθετική αύξηση χώρου ή χρόνου που απαιτεί για πολύ μεγάλες εφαρμογές (πολλά παραδείγματα και ιδιότητες).

Δημιουργήθηκε στο *Turing Institute* στην *Γλασκώβη*, στα πλαίσια του *MLT (Machine Learning Toolbox)*. Αποτελεί μια μετεξελιγμένη έκδοση του ID3. (4)

Στην συνέχεια θα αναφέρουμε τα νέα στοιχεία του NEW ID σε σχέση με τον ID3, καθώς και τις ιδιαιτερότητές του. Η βασική διαφορά αφορά στον τύπο των ιδιοτήτων που οι δύο αλγόριθμοι χειρίζονται. Έτσι, ενώ ο ID3 χειρίζεται μόνο έναν τύπο ιδιοτήτων, τις **κατηγορηματικές (nominal)**, ο NEW ID μπορεί να χειρίζεται και **αριθμητικές ιδιότητες (linear)**, που παίρνουν ακέραιες ή και πραγματικές τιμές.

Όταν μια αριθμητική ιδιότητα επιλέγεται σαν κόμβος του δένδρου, υπολογίζεται στατιστικά από τα παραδείγματα εκείνη η τιμή που διαχωρίζει καλύτερα τις κλάσεις. Η τιμή αυτή ονομάζεται **σημείο διαχωρισμού (split point)**. Δημιουργείται τότε ένα κλαδί για τα παραδείγματα με τιμές ιδιότητας μεγαλύτερες του σημείου διαχωρισμού και ένα δεύτερο κλαδί με τιμές μικρότερες του σημείου διαχωρισμού και ένα δεύτερο κλαδί με τιμές μικρότερες του σημείου διαχωρισμού.

Πιο συγκεκριμένα υποθέτουμε ότι A μια ιδιότητα και οι διατεταγμένες τιμές που λαμβάνει στο σύνολο των παραδειγμάτων είναι $\{v_1, v_2, \dots, v_3\}$. Άρα τα πιθανά σημεία διαχωρισμού είναι $m-1$. Συνηθίζεται να λαμβάνεται ως πιθανό σημείο διαχωρισμού το μέσο του διαστήματος $\{v_i, v_i + 1\}$, δηλαδή $v_i + v_i + 1/2$ το οποίο διαχωρίζει τα παραδείγματα σε δύο υποσύνολα.

Η διαδικασία για την εύρεση της πληροφοριακής ισχύς παραμένει η ίδια.

➤ Επιπλέον ο NEW ID μπορεί να χειριστεί διατεταγμένες κατηγορηματικές ιδιότητες.

π.χ. Για την ιδιότητα επάγγελμα έχουμε τις τιμές {υπάλληλος, προϊστάμενος, διευθυντής}. Αν θέλουμε να δηλώσουμε πως ο διευθυντής έχει υψηλότερη θέση, τότε δεν έχουμε παρά να παραθέσουμε τις τιμές με τη σειρά στην οποία διατάσσονται και να ορίσουμε ότι το επάγγελμα είναι διατεταγμένη ιδιότητα:

επάγγελμα : (ORDERED) διευθυντής, προϊστάμενος, υπάλληλος.

➤ Επίσης μπορεί να λάβει ιεραρχικές ιδιότητες

π.χ. το χαρακτηριστικό "σχήμα" μπορεί να πάρει τιμές {τετράγωνο, τρίγωνο} και επιπλέον το τρίγωνο να αποτελεί ξεχωριστό χαρακτηριστικό χαμηλότερης ιεραρχίας με τιμές {σκαληνό, ισοσκελές, ισόπλευρο}.

➤ Τέλος διαχειρίζεται και αδιάφορες τιμές ιδιοτήτων.

π.χ. όταν ένα χαρακτηριστικό λαμβάνει μια συγκεκριμένη τιμή, τέτοια ώστε καθίσταται αδιάφορη η τιμή ενός άλλου. Θα δώσουμε ένα εύγωττο παράδειγμα. Έχουμε μια ιατρική εφαρμογή για την οποία πρέπει να καταγραφούν όλες οι δυνατές εξετάσεις για να καταλήξουμε σε διάγνωση. Αν όμως αυτή η διάγνωση αποφασιστεί με μερικά στοιχεία μόνο δεν είναι βεβαίως απαραίτητο ο ασθενής να υποβληθεί σε όλες τις εξετάσεις. Οι τιμές των άλλων ιδιοτήτων είναι μεν άγνωστες, αλλά όχι γιατί τα στοιχεία δεν είναι διαθέσιμα, αλλά γιατί δεν είναι σημαντικά.

ΜΑΘΗΜΑΤΙΚΗ ΠΑΡΟΥΣΙΑΣΗ ΑΛΓΟΡΙΘΜΟΥ

Πριν προχωρήσουμε στην μαθηματική παρουσίαση του αλγόριθμου, κρίνεται σκόπιμο να επισημάνουμε εν συντομία κάποια βασικά στοιχεία της θεωρίας πιθανοτήτων, αλλά και της θεωρίας πληροφοριών.

ΔΙΑΜΕΡΙΣΜΟΣ

Διαμερισμός \mathcal{U} ενός συνόλου \mathcal{X} είναι μια ομάδα ασυμβίβαστων υποσυνόλων A_i του \mathcal{X} των οποίων η ένωση ισούται με \mathcal{X} .

$$A_1 + \dots + A_n = \mathcal{X} \quad A_i \cap A_j = \{\emptyset\} \quad i \neq j$$

ΠΙΘΑΝΟΤΗΤΑ

Πιθανότητα ενός ενδεχομένου A μπορεί να ερμηνευτεί ως μέτρο της αβεβαιότητάς μας για την πραγματοποίηση ή μη A .

ΕΝΤΡΟΠΙΑ

Είναι το μέτρο αβεβαιότητας στην πραγματοποίηση ή μη-πραγματοποίηση, όχι ενός μοναδικού ενδεχομένου του \mathcal{X} , αλλά οποιουδήποτε ενδεχομένου A_i ενός διαμερισμού \mathcal{U} του \mathcal{X} .

Επίσης μπορεί να διατυπωθεί ότι η εντροπία είναι:

- Μέτρο αταξίας σε μια τυχαία κατάσταση
- Μέτρο επιλογής σε μια τυχαία κατάσταση

ΠΛΗΡΟΦΟΡΙΑ

Είναι ένα μέτρο επιλογής που έχει κάποιος όταν επιλέγει ένα μήνυμα. Κατά κάποιο τρόπο, ένα μέτρο έκπληξης όταν πάρουμε ένα μήνυμα. Έστω ένα τυχαίο γεγονός E με πιθανότητα $P(E)$. Αν το E συμβεί τότε μας έδωσε $I = -\log p(e)$ μονάδες πληροφορίας.

ΑΒΕΒΑΙΩΤΗΤΑ ΚΑΙ ΠΛΗΡΟΦΟΡΙΑ

Στην εμπειρική ερμηνεία της εντροπίας, ο αριθμός $H(u)$ είναι ένα μέτρο της αβεβαιότητας μας για τα ενδεχόμενα A_i του διαμερισμού u , πριν την εκτέλεση του πειράματος. Αν εκτελεστεί το πείραμα και γίνουν γνωστά τα αποτελέσματα που σχετίζονται με τα A_i , η αβεβαιότητα αναιρείται. Έτσι μπορούμε να πούμε ότι το πείραμα παρέχει πληροφορία για τα ενδεχόμενα A_i , ίση με την εντροπία του διαμερισμού τους. Έτσι η αβεβαιότητα ισούται με πληροφορία.

Θα θεωρήσουμε ότι έχουμε μόνο δύο κλάσεις και ότι το σύνολο των παραδειγμάτων Y παρέχει p παραδείγματα της κλάσης P και n παραδείγματα της κλάσης N . (5),(6),(7),(8),(9)

Οι παραδοχές μας είναι:

- Κάθε σωστό δένδρο απόφασης κατατάσσει τα παραδείγματα με αναλογία ίδια με αυτής του συνόλου εκπαίδευσης. Ένα τυχαίο παράδειγμα ανήκει P με πιθανότητα $p/(p+n)$ και στην κλάση N με πιθανότητα $n/(p+n)$.
- Το δένδρο απόφασης χρησιμοποιείται για να κατατάξει παραδείγματα στην κλάση P ή N . Έτσι θεωρείται ως γεννήτρια P και N και η απαιτούμενη πληροφορία για την δημιουργία αυτών των μηνυμάτων είναι:

$$I(p,n) = -(p/(p+n) \cdot \log_2(p/(p+n))) - (n/(p+n) \cdot \log_2(n/(p+n))) \quad (1)$$

Ο παραπάνω τύπος γενικεύεται για περισσότερες κλάσεις. Αν C_1, \dots, C_K οι δυνατές κλάσεις και e_1, \dots, e_K το πλήθος των παραδειγμάτων κάθε κλάσης, τότε:

$$I(C1, ..., CK) = -\sum (e_i/\Sigma e_i) \log_2(e_i/\Sigma e_i) \quad (2)$$

Ας υποθέσουμε πως η ιδιότητα A με τιμές A1, A2, ... An αποτελεί την ρίζα του δένδρου. Τότε το σύνολο των εκπαιδευτικών παραδειγμάτων Y διαιρείται αντίστοιχα στα υποσύνολα Y1, Y2, ..., YU, όπου το Yi περιέχει εκείνα τα παραδείγματα του Y με τιμή Ai για την ιδιότητα A. Υποθέτουμε ότι το Yi περιέχει ri παραδείγματα κλάσης P και ni παραδείγματα της κλάσης N.

Η πληροφορία που απαιτείται για την κάλυψη του Yi είναι **I(pi, ni)**.

Για ένα δένδρο με ρίζα A, χρησιμοποιείται η παρακάτω συνάρτηση εντροπίας για την μέτρηση της αβεβαιότητας, δηλαδή έλλειψη πληροφορίας:

$$E(A) = \sum (p_i + n_i / p + n) I(p_i, n_i) \quad (3)$$

Η **πληροφοριακή ισχύς** του χαρακτηριστικού είναι:

$$\text{GAIN}(A) = I(p_i, n_i) - E(A) \quad (4)$$

Στην σχέση αυτή, η ποσότητα I(p, n) είναι σταθερή. Επομένως μεγάλη πληροφοριακή ισχύ έχει η ιδιότητα με την ελάχιστη εντροπία.

Ο αλγόριθμος υπολογίζει την πληροφοριακή ισχύ κάθε ιδιότητας και επιλέγει εκείνη με την μεγαλύτερη πληροφοριακή ισχύ. Η ιδιότητα αυτή ονομάζεται πιο διαχωριστική ή πιο πληροφοριακή.

Η ιδιότητα αυτή γίνεται κόμβος του δένδρου, δημιουργούνται δε τόσα κλαδιά, όσες είναι οι δυνατές τιμές της και τα παραδείγματα χωρίζονται στα αντίστοιχα υποσύνολα.

Η παραπάνω διαδικασία επιλογής της ιδιότητας - κόμβος επαναλαμβάνεται αναδρομικά για κάθε υποσύνολο παραδειγμάτων. Όταν το σύνολο αυτό περιέχει παραδείγματα μιας μόνο κλάσης, ή είναι κενό, ικανοποιείται η συνθήκη τερματισμού και δημιουργείται ένα φύλλο στο οποίο αντιστοιχίζεται η κλάση, ή το ειδικό σύμβολο κενό (null) αντίστοιχα. Οι τυχόν εμφανίσεις της κενής κλάσης οφείλονται στο γεγονός ότι ο ID3 δημιουργεί κλαδιά για όλες τις τιμές της ιδιότητας που επιλέχθηκε σαν κόμβος και όχι μόνο για αυτές που εμφανίζονται στα παραδείγματα.

ΠΑΡΑΔΕΙΓΜΑ

Στην συνέχεια παραθέτουμε ένα παράδειγμα για την καλύτερη κατανόηση των παραπάνω (8),(1) . Ο Μίκι και ο Γκούφι που έχουν προγραμματίσει να κάνουν περίπατο στην Λιμνούπολη θέλουν να αποφασίσουν αν θα πάρουν ομπρέλα μαζί τους. Η απόφαση αυτή θα καθοριστεί από τις ενδείξεις του καιρού.

Ο καιρός περιγράφεται από τα εξής χαρακτηριστικά:

όψη καιρού {λιακάδα, συννεφιά, βροχερός}

υγρασία {υψηλή, χαμηλή}

θερμοκρασία {χαμηλή, μέτρια, υψηλή}

αέρας {ναι, όχι}

Έχουμε δύο κλάσεις {NAI, OXI}

a/a	Όψη καιρού	Θερμοκρασία	Υγρασία	Αέρας	Κλάση
1	ΛΙΑΚΑΔΑ	ΥΨΗΛΗ	ΥΨΗΛΗ	ΟΧΙ	ΟΧΙ
2	«	«	«	ΝΑΙ	«
3	ΣΥΝΝΕΦΙΑ	«	«	ΟΧΙ	ΝΑΙ
4	ΒΡΟΧΕΡΟΣ	ΜΕΤΡΙΑ	«	«	«
5	«	ΧΑΜΗΛΗ	ΚΑΝΟΝΙΚΗ	«	«
6	«	«	«	ΝΑΙ	ΟΧΙ
7	ΣΥΝΝΕΦΙΑ	«	«	«	ΝΑΙ
8	ΛΙΑΚΑΔΑ	ΜΕΤΡΙΑ	ΥΨΗΛΗ	ΟΧΙ	ΟΧΙ
9	«	ΥΨΗΛΗ	ΚΑΝΟΝΙΚΗ	«	ΝΑΙ
10	ΒΡΟΧΕΡΟΣ	ΜΕΤΡΙΑ	«	«	«
11	ΛΙΑΚΑΔΑ	«	«	ΝΑΙ	«
12	ΣΥΝΝΕΦΙΑ	«	ΥΨΗΛΗ	«	«
13	«	ΥΨΗΛΗ	ΚΑΝΟΝΙΚΗ	ΟΧΙ	«
14	ΒΡΟΧΕΡΟΣ	ΜΕΤΡΙΑ	ΥΨΗΛΗ	ΝΑΙ	ΟΧΙ

Από τα 14 παραδείγματα, 9 ανήκουν στην κλάση N και 5 στην κλάση O. Άρα η απαιτούμενη πληροφορία για την κατάταξη είναι:

$$I(N,O) = -9/14 \log_2(9/14) - 5/14 \log_2(5/14) = 0.94 \text{ bits}$$

➤ Εξετάζουμε το χαρακτηριστικό όψη του καιρού:

Πέντε από τα παραδείγματα λαμβάνουν την τιμή (λιακάδα), 2 από την κλάση N και 3 από την κλάση O.

$$\begin{aligned} \text{Επομένως, } N1=2, O1=3 & \quad I(N1,O1) = 0.971 & \quad I(N2,O2) = 0 \\ N2=4, O2=0 & \end{aligned}$$

$$N3=3, O3=3$$

$$I(N3,O3)=0.971$$

Άρα η αναμενόμενη πληροφορία γι' αυτό το χαρακτηριστικό είναι:

$$E(\text{όψη καιρού}) = 5/14 I(N1,O1) + 4/14 I(N2,O2) + 5/14 I(N3,O3) = 0.694 \text{ bits}$$

Η πληροφοριακή ισχύς του χαρακτηριστικού αυτού είναι τότε:

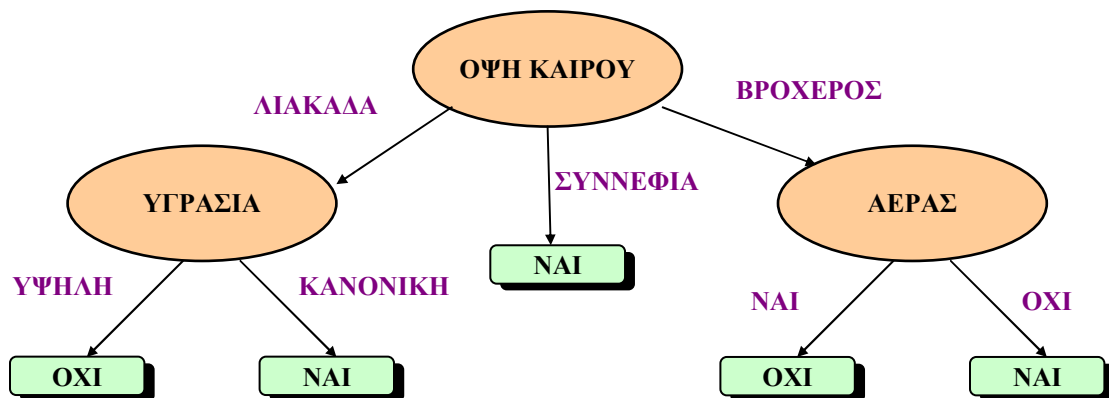
$$\text{gain}(\text{όψη καιρού}) = 0.94 - E(\text{όψη καιρού}) = 0.246 \text{ bits}$$

$$\text{gain}(\text{θερμοκρασία}) = 0.029 \text{ bits}$$

$$\text{gain}(\text{υγρασία}) = 0.151 \text{ bits}$$

$$\text{gain}(\text{αέρας}) = 0.048 \text{ bits}$$

Η όψη καιρού είναι το χαρακτηριστικό με την μεγαλύτερη πληροφοριακή ισχύ και άρα αυτό επιλέγεται ως ρίζα του δένδρου απόφασης.



Σχήμα 5 - Δένδρο απόφασης παραδείγματος

Λαμβάνουμε το παραπάνω δένδρο, το οποίο είναι και το οικονομικότερο. Υπάρχουν περισσότερα από ένα σωστά δένδρα απόφασης, τα οποία είναι πιο πολύπλοκα απ' το

παραπάνω.

Αν τώρα θεωρήσουμε ότι το συγκεκριμένο απόγευμα του περιπάτου τα χαρακτηριστικά του καιρού έχουν τις εξής τιμές:

όψη καιρού	{λιακάδα}
θερμοκρασία	{υψηλή}
υγρασία	{κανονική}
αέρας	{όχι}

Τότε διαπιστώνουμε απ' το δένδρο απόφασης ότι ο Μίκι και ο Γκούφι πρέπει να εφοδιαστούν με μια ομπρέλα.

ΜΕΤΑΤΡΟΠΗ ΔΕΝΔΡΟΥ ΣΕ ΚΑΝΟΝΕΣ ΑΠΟΦΑΣΗΣ

Όπως καταλαβαίνουμε, ένα δένδρο απόφασης δεν μπορεί να είναι εύκολα κατανοητό από έναν ειδικό. Θεωρείται αναγκαία η μετατροπή του σε κανόνες παραγωγής (production rules), οι οποίοι είναι και ο πιο διαδεδομένος τρόπος αναπαράστασης της γνώσης στα έμπειρα συστήματα. (10) Η μορφή των κανόνων παραγωγής είναι:

AN (προϋποθέσεις) TOTE (κλάση) (μέτρο βεβαιότητας)

Επίσης πρέπει να σημειωθεί πως η μετατροπή αυτή μπορεί να αυξήσει την ακρίβεια κατάταξης των ιδιοτήτων με την αφαίρεση από το δένδρο ιδιοτήτων που ανταποκρίνονται σε ιδιαιτερότητες του εκπαιδευτικού συνόλου.

ΕΦΑΡΜΟΓΗ NEW ID-3 ΣΤΟ ΠΡΟΒΛΗΜΑ ΜΑΣ

Στο πρόβλημα που εμείς εξετάζουμε, πεδίο εφαρμογής είναι το σύνολο των 28 ιδιοτήτων (attributes) καθώς και οι τιμές τους.

Κλάσεις θεωρούνται οι ομάδες Group 1, Group 2, Group 3, Group 4, Group 5 στις οποίες διαμοιράστηκαν τα 27 υποκαταστήματα.

Έτσι παραδείγματα του προβλήματος αποτελούν τα 27 διανύσματα που περιέχουν τιμές 28 ιδιοτήτων, καθώς και την κλάση (ομάδα) του υποκαταστήματος.

Έτσι ως είσοδο του συστήματος έχουμε:

	K1	K2	K3	K4	K5	K6	K7	K8	K9	K10	K11	K12	K12	K13
Y1	301	13705	17915	1946	8895	1	0	83	1050	1182	83	765	24	2
Y2	311	7618	12882	4826	5302	3	0	58	2037	988	49	1479	70	3
Y3	248	4136	20550	1827	4771	3	0	78	4473	398	82	24976	43	5
Y4	610	11607	18620	3938	909	12	151	0	956	1519	15	2903	430	14
Y5	276	23034	29106	823	9395	0	26	9	1709	205	35	190	50	6
Y6	835	18042	71357	1713	747	0	42	90	9734	3250	2	87353	958	70
Y7	453	21704	75988	3055	1918	5	60	45	2964	2923	106	636	758	10
Y8	536	11820	23540	3925	708	0	44	14	2410	1955	58	5384	144	22
Y9	522	6144	18134	2915	433	4	73	26	1695	1086	89	5631	141	1
Y10	431	16761	15208	3416	962	100	104	9	2617	878	69	81513	43	21
Y11	258	9692	29395	0	0	0	0	3	781	4801	5	13	48	2
Y12	358	10890	13862	3172	34296	1	39	5	389	1580	41	5260	215	3
Y13	532	32226	62462	4077	11333	4	103	4	1437	5257	32	223	280	3
Y14	363	8356	16772	2452	1581	8	9	66	2108	1250	0	6401	146	3
Y15	366	8601	21251	1094	16041	0	0	0	101	3378	0	13	54	2
Y16	975	16748	27032	5558	6247	18	15	246	5253	5414	304	55312	747	22
Y17	367	5290	11389	4393	9039	3	88	25	1691	435	104	2247	52	0
Y18	315	6122	21199	2211	478	0	148	7	8375	783	153	16109	273	4
Y19	642	27668	42718	3136	3840	0	213	17	26758	2807	233	164680	554	64
Y20	240	3624	13044	1585	250	19	137	0	2036	959	137	13853	59	6
Y21	342	8719	15835	7329	30349	0	0	76	937	1948	73	46	25	0
Y22	312	3191	10026	1915	221	13	88	6	1182	105	93	2877	23	0
Y23	321	10079	33835	3476	243	3	0	0	817	343	100	526	16	0
Y24	286	5946	16021	1566	6895	0	110	110	2868	2119	123	1025	9	3
Y25	492	12525	47192	2759	1846	3	0	0	5822	2270	81	34400	413	21
Y26	2344	30779	51401	5152	2324	0	0	0	990	0	81	9338	4824	158
Y27	2971	221325	210195	9967	20477	0	0	0	12133	21958	16	34311	4340	190

Y : ΥΠΟΚΑΤΑΣΤΗΜΑΤΑ

K : ΜΕΤΑΒΛΗΤΕΣ (ΒΛ. ΣΕΛ. ...)

K14	K15	K16	K17	K18	K19	K20	K21	K22	K23	K24	K25	K26	K27	CLASS
0	70	1	0	0	145	710	3	124	71	0	1	67	21141	GROUP 3
5	13	0	2	0	54	0	2	2	0	0	2	2	11475	GROUP 4
37	22	63	18	0	28	56	56	413	0	1	1	3	10897	GROUP 5
33	421	68	6	0	19	39	51	0	0	0	22	71	36485	GROUP 2
2	43	5	0	0	52	0	5	28	0	11	11	136	3582	GROUP 5
12	625	89	84	0	231	0	1	833	0	7	15	53	49739	GROUP 2
0	564	10	0	0	137	52	80	0	0	0	9	0	62833	GROUP 3
69	451	4	24	1	138	0	7	11	0	17	139	392	20025	GROUP 3
15	214	19	80	0	73	2	12	17	0	8	9	171	9960	GROUP 4
47	171	13	1	0	72	0	708	0	0	0	0	0	14997	GROUP 3
0	179	0	0	0	15	909	0	359	0	0	8	88	21717	GROUP 5
6	39	1	0	14	0	0	0	0	0	0	0	0	9067	GROUP 4
1	271	22	0	0	190	3517	40	11	0	0	3	24	21276	GROUP 2
1	116	33	2	0	64	30	25	51	0	0	0	0	31945	GROUP 4
0	113	0	0	0	2	7	6	0	0	0	127	53	9732	GROUP 5
49	1076	55	36	39	34	5464	113	590	0	0	123	570	100321	GROUP 2
0	39	0	0	0	72	357	401	140	0	0	0	112	11611	GROUP 4
7	143	0	0	0	36	318	292	0	0	0	0	0	13630	GROUP 4
63	685	338	495	0	94	5675	6016	385	0	0	2	132	7997	GROUP 2
0	33	0	14	0	42	0	0	0	0	0	3	22	5640	GROUP 5
0	61	0	9	0	111	1051	0	0	0	0	0	0	29229	GROUP 3
13	29	0	0	0	5	101	115	45	0	0	0	0	6391	GROUP 5
0	13	0	0	0	0	0	0	0	0	0	0	0	0	GROUP 5
0	42	0	0	0	0	2846	71	0	0	0	13	86	5548	GROUP 5
108	438	14	7	0	121	1163	1247	66	0	0	3	2	9445	GROUP 3
776	1398	39	10	17	863	2993	37335	0	0	0	0	10774	92111	GROUP 1
566	2877	178	204	39	1273	63856	463	0	246	5271	1093	48650	825566	GROUP 1

ΚΡΙΤΗΡΙΑ (ΜΕΤΑΒΛΗΤΕΣ)

- K1: ΠΡΟΣΩΠΙΚΟ
- K2: ΤΑΜΕΙΑΚΕΣ ΠΡΑΞΕΙΣ ΕΙΣΠΡΑΞΕΙΣ
- K3: ΤΑΜΕΙΑΚΕΣ ΠΡΑΞΕΙΣ ΕΝΤΟΛΕΣ
- K4: ΕΠΙΤΑΓΕΣ ΕΝΤΟΛΩΝ ΤΡΑΠΕΖΩΝ ΠΟΥ ΕΚΔΟΘΗΚΑΝ
- K5: ΕΠΙΤΑΓΕΣ ΕΝΤΟΛΩΝ ΤΡΑΠΕΖΩΝ ΠΟΥ ΠΛΗΡΩΘΗΚΑΝ
- K6: ΧΡΗΜ/ΛΕΣ ΞΕΝΩΝ ΤΡΑΠΕΖΩΝ ΕΞΩΤΕΡΙΚΟΥ
- K7: ΧΡΗΜ/ΛΕΣ ΞΕΝΩΝ ΤΡΑΠΕΖΩΝ ΕΞΩΤΕΡΙΚΟΥ ΜΕΣΩ Κ.Κ
- K8: ΧΡΗΜ/ΛΕΣ ΞΕΝΩΝ ΤΡΑΠΕΖΩΝ ΠΡΟΣ Κ.Κ
- K9: ΑΓΟΡΑ ΞΕΝΩΝ ΤΡΑΠΕΖΟΓΡΑΜΜΑΤΙΩΝ
- K10: ΠΩΛΗΣΗ ΞΕΝΩΝ ΤΡΑΠΕΖΟΓΡΑΜΜΑΤΙΩΝ ΣΕ ΠΕΛΑΤΕΣ
- K11: ΠΩΛΗΣΗ ΞΕΝΩΝ ΤΡΑΠ/ΤΙΩΝ ΣΕ ΠΕΛΑΤΕΣ Κ ΧΡΗΜ/ΛΕΣ ΠΡΟΣ Κ.Κ
- K12: ΑΓΟΡΑ ΕΠΙΤΑΓΩΝ
- K13: ΕΚΔΟΣΗ ΕΠΙΤΑΓΩΝ
- K14: ΕΙΣΑΓΩΓΗ ΕΝΤΟΛΩΝ
- K15: ΕΚΔΟΣΗ ΕΝΤΟΛΩΝ
- K16: ΣΥΝΑΛ/ΜΑ ΓΙΑ ΛΟΙΠΕΣ ΑΝΑΓΚΕΣ
- K17: ΕΙΣΑΓΩΓΗ ΑΞΙΩΝ ΕΞΩΤ. ΕΣΩΤ.
- K18: ΕΚΚΡΕΜΕΙΣ ΑΞΙΕΣ ΣΕ ΤΡΙΤΟΥΣ
- K19: ΕΝΤΟΛΕΣ ΣΥΝ/ΤΟΣ ΚΛΗΡΙΝΓΚ
- K20: ΕΝΤΟΛΕΣ ΣΥΝ/ΤΟΣ ΠΟΥ ΔΙΑΒΙΒ.
- K21: ΕΓΚΡΙΣΕΙΣ ΕΙΣΑΓ. ΠΟΥ ΕΚΔΟΘΗΚΑΝ
- K22: ΑΔΕΙΕΣ ΕΚΤΕΛ. ΠΟΥ ΕΚΔΟΘΗΚΑΝ
- K23: ΑΔΕΙΕΣ ΕΚΤΕΛ. ΕΓΚΡ. ΤΡΑΠΕΖΩΝ
- K24: ΑΔΕΙΕΣ ΕΚΤΕΛ. ΕΓΚΡ. ΧΩΡΙΣ ΣΥΝΑΛ. ΔΙΑΤΥΠ.
- K25: ΛΗΦΘΕΙΣΕΣ ΔΗΛΩΣΕΙΣ ΕΞΑΓΩΓΗΣ
- K26: ΑΔΕΙΕΣ ΕΞΑΓ. ΠΟΥ ΕΚΔΟΘΗΚΑΝ
- K27: ΕΓΚΡ. ΔΗΛΩΣΕΙΣ & ΑΔΕΙΕΣ ΕΞΑΓ.
- K28: ΟΜΟΛ. & ΤΟΚΟΜΕΡ. ΠΟΥ ΕΞΟΦΛΗΘΗΚΑΝ

Επίσης απαραίτητος να εισαχθεί θεωρείται και ο καθορισμός των ιδιοτήτων (defclass).

Property	(K1	(range (real)))
Property	(K2	(range (real)))
Property	(K3	(range (real)))
Property	(K4	(range (real)))
Property	(K5	(range (real)))
Property	(K6	(range (real)))
Property	(K7	(range (real)))
Property	(K8	(range (real)))
Property	(K9	(range (real)))
Property	(K10	(range (real)))
Property	(K11	(range (real)))
Property	(K12	(range (real)))
Property	(K13	(range (real)))
Property	(K14	(range (real)))
Property	(K15	(range (real)))
Property	(K16	(range (real)))
Property	(K17	(range (real)))
Property	(K18	(range (real)))
Property	(K19	(range (real)))
Property	(K20	(range (real)))
Property	(K21	(range (real)))
Property	(K22	(range (real)))
Property	(K23	(range (real)))
Property	(K24	(range (real)))
Property	(K25	(range (real)))
Property	(K26	(range (real)))
Property	(K27	(range (real)))
Property	(K28	(range (real)))
Property	(CLASS	(range (nominal (GROUP1, GROUP 2, GROUP3, GROUP4, GROUP5))))

Τα αποτελέσματα που λαμβάνουμε :

K1

```
( <= 527 : K2
  ( <= 8415 : K5
    ( > 17650 {GROUP 3}[00100]
      < 17650 : K4
        ( <= 1930 {GROUP 5}[00006]
          > 1930 {GROUP 4}[00050]
        )

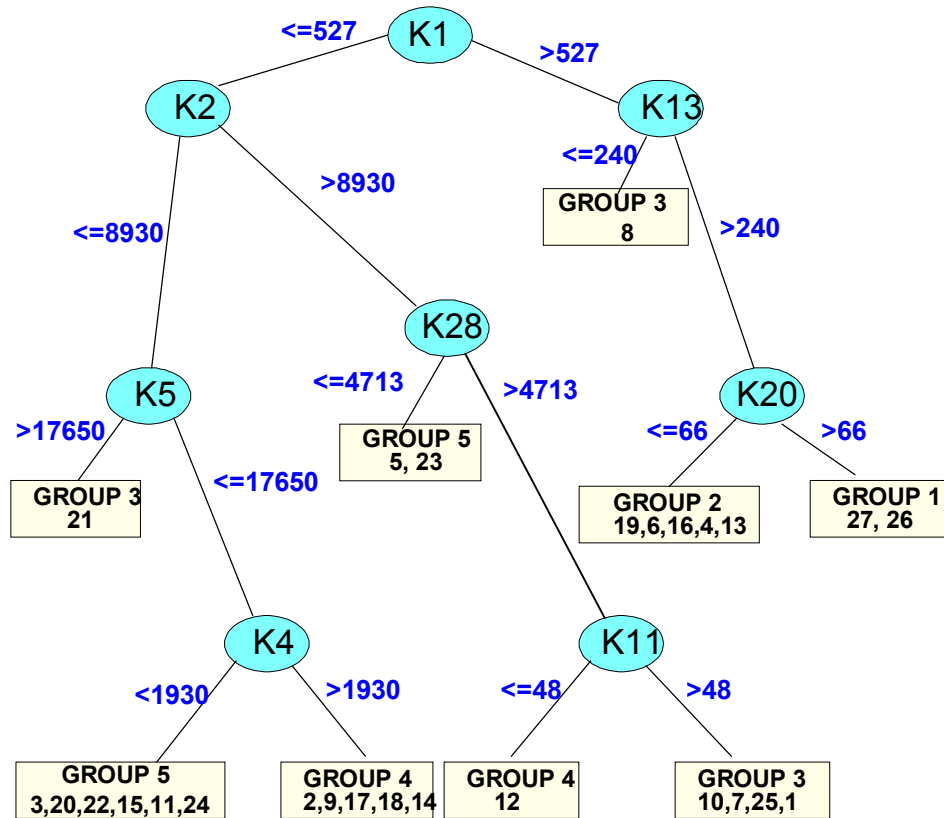
        > 8415 : K28
        ( <= 4713 {GROUP 5}[00008]
          > 4713 : K11
          ( <= 48 {GROUP 4}[00010]
            > 48 {GROUP 3}[00500]
          )
        )
      )
    )
  )
)

> 527 : K13
( <= 240 {GROUP 3}[00100]
  > 240 : K15
  ( <= 66 {GROUP 2}[05000]
    > 66 {GROUP 1}[20000]
  )
)
)
```

Οι κανόνες που παράχθηκαν από τον αλγόριθμο έχουν 100% ακρίβεια αφού κατατάσουν όλα τα υποκαταστήματα στις κλάσεις τους.

	GROUP 2	GROUP 2	GROUP 3	GROUP 4	GROUP 5	ACCURACY
GROUP 1	2.00	0.00	0.00	0.000	0.00	100%
GROUP 2	0	5	0	0	0	100%
GROUP 3	0	0	6	0	0	100%
GROUP 4	0	0	0	6	0	100%
GROUP 5	0	0	0	0	8	100%

Το επόμενο βήμα είναι η κατασκευή του δένδρου



Μετατροπή δέδρου σε κανόνες

ΚΑΝΟΝΑΣ 1

AN

K1 <= 527

& K2 <= 8930

& K5 <= 17650

& K4 < 1930

TOTE (ΟΜΑΔΑ = 5)

ΟΜΑΔΑ 5 : Y3, Y20, Y22, Y15, Y11, Y24

ΚΑΝΟΝΑΣ 2

AN

K1 <= 527

& K2 <= 8930

& K5 <= 17650

& K4 > 1930

TOTE (ΟΜΑΔΑ = 4)

ΟΜΑΔΑ 4 : Y2, Y9, Y17, Y18, Y14

KANONΑΣ 3

AN

K1 \leq 527

& K2 \leq 8930

& K5 $>$ 17650

TOTE (ΟΜΑΔΑ = 3)

ΟΜΑΔΑ 3 : Y21

KANONΑΣ 4

AN

K1 \leq 527

& K2 $>$ 8930

& K28 \leq 4713

TOTE (ΟΜΑΔΑ = 5)

ΟΜΑΔΑ 5 : Y5, Y23

KANONΑΣ 5

AN

K1 <= 527

& K2 > 8930

& K28 > 4713

& K11 < 48

TOTE (ΟΜΑΔΑ = 4)

ΟΜΑΔΑ 4 : Y12

KANONΑΣ 6

AN

K1 <= 527

& K2 > 8930

& K11 > 48

& K28 > 4713

TOTE (ΟΜΑΔΑ = 3)

ΟΜΑΔΑ 3 : Y25, Y1, Y10, Y7

KANONΑΣ 7

AN

$K1 > 527$

& $K13 < 240$

TOTE (ΟΜΑΔΑ = 3)

ΟΜΑΔΑ 3 : Y8

KANONΑΣ 8

AN

$K1 > 527$

& $K13 > 240$

& $K15 < 66$

TOTE (ΟΜΑΔΑ = 2)

ΟΜΑΔΑ 2 : Y19, Y16, Y6, Y13, Y4

KANONΑΣ 9

AN

$K1 > 527$

& $K13 > 240$

& $K15 > 66$

TOTE (ΟΜΑΔΑ = 1)

ΟΜΑΔΑ 1 : Y27, Y26

ΣΧΟΛΙΑΣΜΟΣ ΚΑΝΟΝΩΝ ΑΠΟΦΑΣΗΣ

Από τους 9 παραπάνω κανόνες μπορούμε να εξάγουμε τα παρακάτω συμπεράσματα :

Πιο διαχωριστικό κριτήριο (μεταβλητές) δηλαδή το κριτήριο με την μεγαλύτερη πληροφοριακή ισχύ αποτελεί το προσωπικό των υποκαταστημάτων. Το σημείο διαχωρισμού (split point) αυτού του κριτηρίου είναι το 527, το οποίο διαχωρίζει τα υποκαταστήματα των δύο πρώτων ομάδων με αυτά των υπολοίπων με εξαιρέσεις το υποκατάστημα 8 που ανήκει στην ομάδα 3.

Ας προχωρήσουμε όμως σε πιο λεπτομερή ανάλυση των ομάδων.

Η ομάδα 3 αποτελείται από υποκαταστήματα που έχουν προσωπικό κάτω των 527 ατόμων. Τα υποκαταστήματα 3, 11, 15, 24, 20 και 22 έχουν ταμειακές πράξεις εισπράξεις κάτω των 8.930, καθώς και επιταγές εντολών τραπεζών που εκδόθηκαν κάτω των 1.930.

Τα υποκαταστήματα 5, 23 ενώ έχουν ταμειακές πράξεις εισπράξεις άνω των 8.930 τα ομόλογα που εξόφλησαν είναι κάτω των 4.713.

Παρατηρούμε ότι στην ομάδα 5 βρίσκονται υποκαταστήματα με μικρό προσωπικό, μικρό όγκο ταμειακών πράξεων εισπράξεων και εκδόσεων εντολών επιταγών ή μικρό όγκο εξοφλημένων ομολόγων.

Στην ομάδα 4 ανήκουν επίσης υποκαταστήματα με προσωπικό κάτω των 527. Κάποια υποκαταστήματα αυτής της ομάδας ομοιάζουν με κάποιον της ομάδας 5 στον μικρό όγκο ταμειακών πράξεων εισπράξεων, αλλά βεβαίως διαφέρουν στο ότι έχουν μεγαλύτερες εκδόσεις εντολών επιταγών των 1.930.

Στην ίδια ομάδα ανήκει και το Y12 το οποίο αν και έχει όγκο σε ταμειακές πράξεις εισπράξεων μεγαλύτερες των 8.930 και εξοφλημένων ομολόγων μεγαλύτερων των 4.713 έχει χαμηλές πωλήσεις ξένων τρ/τίων και χρημ/λες προς κ.κ. μην μπορώντας έτσι να ανέλθει σε κάποια υψηλότερη ομάδα.

Στην ομάδα 3 ανήκουν κυρίως υποκαταστήματα με προσωπικό επίσης κάτω των 527 ατόμων, αλλά με ικανοποιητικά ποσά στους όγκους των ταμειακών πράξεων εισπράξεων και στην εξόφληση ομολόγων, καθώς και στις πωλήσεις ξένων τρ/τίων και χρημ/λές προς κ.κ.

Επίσης ανήκει και το Y21 το οποίο αν και έχει ταμειακές πράξεις εισπράξεις κάτω των 8930 παρουσιάζει μεγάλο αριθμό επιταγών εντολών που πληρώθηκαν και αναβαθμίζεται ανεβαίνοντας στην ομάδα 3.

Τέλος το υποκατάστημα Y8 αν και έχει προσωπικό μεγαλύτερο των 527, παρουσιάζει μικρό αριθμό στην έκδοση επιταγών (144) και κατατάσσεται και αυτό στην ομάδα 3.

Τα υποκαταστήματα της 2ης ομάδας έχουν προσωπικό μεγαλύτερο των 527 ατόμων, καθώς και εκδόσεως επιταγών μεγαλύτερες των 240 αλλά με εκδόσεις εντολών κάτω των 66.

Η ομάδα 1 αποτελείται από τα υποκαταστήματα Υ26, Υ27 με υψηλό αριθμό προσωπικού, αλλά και μεγάλους όγκους ποσοτήτων εκτός των άλλων και στην έκδοση επιταγών και στην έκδοση εντολών.

Η παραπάνω ανάλυση ανέδειξε τα σημαντικότερα κριτήρια (μεταβλητές) τα οποία θα χρησιμοποιηθούν από τον αναλυτή λήπτη αποφάσεων που θα θελήσει να κατατάξει μελλοντικά κάποιο νέο υποκατάστημα ή τα ίδια υποκαταστήματα με δεδομένα άλλων ετών. Έτσι ανάλογα με τις τιμές που θα λάβουν σε αυτά τα κριτήρια θα μπορεί να τα χαρακτηρίσει ως αποδοτικά ή μη-αποδοτικά ή μέσης απόδοσης.



ΑΝΑΛΥΣΗ & ΚΡΙΤΙΚΗ ΤΩΝ ΑΠΟΤΕΛΕΣΜΑΤΩΝ

ΑΝΑΛΥΣΗ ΤΩΝ ΑΠΟΤΕΛΕΣΜΑΤΩΝ

Η παρούσα εργασία θεωρούμε ,ότι, έδωσε μία πλήρη εικόνα της clustering (ομαδοποίηση) και της classification (κατάταξη) καθώς και της διαφοράς των εννοιών τους.

Αν θελήσουμε να σχολιάσουμε την ολότητα των αποτελέσματος που λήφθηκαν ,θα μπορούσαμε να διατυπώσουμε ,ότι , κρίνονται ικανοποιητικά.

Η CLUSTER καταχώρησε τα υποκαταστήματα σε κλάσεις (ομάδες) οι οποίες έχουν μεγάλη συνάφεια με την διάταξη των υποκ/των στις τραπεζικές εργασίες.

Ο NEWID κατέταξε το 100% των υποκ/των στις ομάδες , τις οποίες ανέδειξε η εργασία της CLUSTER .

Τα διαχωριστικά χαρακτηριστικά που υποδεικνύει θεωρούνται βασικά στο τραπεζικό περιβάλλον.

Αυτές οι εργασίες θα αποτελέσουν την πυξίδα , που θα οδηγήσει τον λήπτη αποφάσεων να αποφανθεί για τον χαρακτηρισμό μελλοντικών στοιχείων.

Τα μελλοντικά αυτά στοιχεία μπορεί να αναφέρονται τόσο στα αποτελέσματα των ήδη εξεταζόμενων υποκ/των σε επερχόμενα έτη, όσο και σε αποτελέσματα κάποιου νέου υποκ/τος που θα δημιουργηθεί στο μέλλον.

Βέβαια πρέπει να αναφερθεί το γεγονός πως οι αλγόριθμοι μάθησης εξάγουν αποτελέσματα από τα εκπαιδευτικά παραδείγματα από τα οποία τροφοδοτούνται .

Από αυτό συνάγεται ,ότι, ίσως ο αριθμός των παραδειγμάτων να μην ήταν ο καλύτερος δυνατός.

Όμως, η φύση του προβλήματος είναι τέτοια που υποχρεωνόμαστε να εξάγουμε αποτελέσματα από αυτό το πεδίο παραδειγμάτων.

Πιστεύουμε πως η παρούσα εργασία αποτελεί πηγή αρωγής για τον αναλυτή-ληπτή αποφάσεων των τραπεζικών δραστηριοτήτων.

Επίσης θεωρούμε σημαντικό το γεγονός ,ότι, μπορέσαμε να αναδείξουμε την φιλοσοφία και τους μηχανισμούς της μηχανικής μάθησης εφαρμόζοντας την σε ένα οικονομικο-διοικητικό πρόβλημα.



ΠΑΡΑΡΤΗΜΑΤΑ

	K1	K2	K3	K4	K5	K6	K7	K8	K9	K10	K11	K12	K12	K13
Y1	301	13705	17915	1946	8895	1	0	83	1050	1182	83	765	24	2
Y2	311	7618	12882	4826	5302	3	0	58	2037	988	49	1479	70	3
Y3	248	4136	20550	1827	4771	3	0	78	4473	398	82	24976	43	5
Y4	610	11607	18620	3938	909	12	151	0	956	1519	15	2903	430	14
Y5	276	23034	29106	823	9395	0	26	9	1709	205	35	190	50	6
Y6	835	18042	71357	1713	747	0	42	90	9734	3250	2	87353	958	70
Y7	453	21704	75988	3055	1918	5	60	45	2964	2923	106	636	758	10
Y8	536	11820	23540	3925	708	0	44	14	2410	1955	58	5384	144	22
Y9	522	6144	18134	2915	433	4	73	26	1695	1086	89	5631	141	1
Y10	431	16761	15208	3416	962	100	104	9	2617	878	69	81513	43	21
Y11	258	9692	29395	0	0	0	0	3	781	4801	5	13	48	2
Y12	358	10890	13862	3172	34296	1	39	5	389	1580	41	5260	215	3
Y13	532	32226	62462	4077	11333	4	103	4	1437	5257	32	223	280	3
Y14	363	8356	16772	2452	1581	8	9	66	2108	1250	0	6401	146	3
Y15	366	8601	21251	1094	16041	0	0	0	101	3378	0	13	54	2
Y16	975	16748	27032	5558	6247	18	15	246	5253	5414	304	55312	747	22
Y17	367	5290	11389	4393	9039	3	88	25	1691	435	104	2247	52	0
Y18	315	6122	21199	2211	478	0	148	7	8375	783	153	16109	273	4
Y19	642	27668	42718	3136	3840	0	213	17	26758	2807	233	164680	554	64
Y20	240	3624	13044	1585	250	19	137	0	2036	959	137	13853	59	6
Y21	342	8719	15835	7329	30349	0	0	76	937	1948	73	46	25	0
Y22	312	3191	10026	1915	221	13	88	6	1182	105	93	2877	23	0
Y23	321	10079	33835	3476	243	3	0	0	817	343	100	526	16	0
Y24	286	5946	16021	1566	6895	0	110	110	2868	2119	123	1025	9	3
Y25	492	12525	47192	2759	1846	3	0	0	5822	2270	81	34400	413	21
Y26	2344	30779	51401	5152	2324	0	0	0	990	0	81	9338	4824	158
Y27	2971	221325	210195	9967	20477	0	0	0	12133	21958	16	34311	4340	190

Y : ΥΠΟΚΑΤΑΣΤΗΜΑΤΑ

K : ΜΕΤΑΒΛΗΤΕΣ (ΒΛ. ΣΕΛ. ...)

K14	K15	K16	K17	K18	K19	K20	K21	K22	K23	K24	K25	K26	K27
0	70	1	0	0	145	710	3	124	71	0	1	67	21141
5	13	0	2	0	54	0	2	2	0	0	2	2	11475
37	22	63	18	0	28	56	56	413	0	1	1	3	10897
33	421	68	6	0	19	39	51	0	0	0	22	71	36485
2	43	5	0	0	52	0	5	28	0	11	11	136	3582
12	625	89	84	0	231	0	1	833	0	7	15	53	49739
0	564	10	0	0	137	52	80	0	0	0	9	0	62833
69	451	4	24	1	138	0	7	11	0	17	139	392	20025
15	214	19	80	0	73	2	12	17	0	8	9	171	9960
47	171	13	1	0	72	0	708	0	0	0	0	0	14997
0	179	0	0	0	15	909	0	359	0	0	8	88	21717
6	39	1	0	14	0	0	0	0	0	0	0	0	9067
1	271	22	0	0	190	3517	40	11	0	0	3	24	21276
1	116	33	2	0	64	30	25	51	0	0	0	0	31945
0	113	0	0	0	2	7	6	0	0	0	127	53	9732
49	1076	55	36	39	34	5464	113	590	0	0	123	570	100321
0	39	0	0	0	72	357	401	140	0	0	0	112	11611
7	143	0	0	0	36	318	292	0	0	0	0	0	13630
63	685	338	495	0	94	5675	6016	385	0	0	2	132	7997
0	33	0	14	0	42	0	0	0	0	0	3	22	5640
0	61	0	9	0	111	1051	0	0	0	0	0	0	29229
13	29	0	0	0	5	101	115	45	0	0	0	0	6391
0	13	0	0	0	0	0	0	0	0	0	0	0	0
0	42	0	0	0	0	2846	71	0	0	0	13	86	5548
108	438	14	7	0	121	1163	1247	66	0	0	3	2	9445
776	1398	39	10	17	863	2993	37335	0	0	0	0	10774	92111
566	2877	178	204	39	1273	63856	463	0	246	5271	1093	48650	825566

ΚΡΙΤΗΡΙΑ (ΜΕΤΑΒΛΗΤΕΣ)

- K1: ΠΡΟΣΩΠΙΚΟ
- K2: ΤΑΜΕΙΑΚΕΣ ΠΡΑΞΕΙΣ ΕΙΣΠΡΑΞΕΙΣ
- K3: ΤΑΜΕΙΑΚΕΣ ΠΡΑΞΕΙΣ ΕΝΤΟΛΕΣ
- K4: ΕΠΙΤΑΓΕΣ ΕΝΤΟΛΩΝ ΤΡΑΠΕΖΩΝ ΠΟΥ ΕΚΔΟΘΗΚΑΝ
- K5: ΕΠΙΤΑΓΕΣ ΕΝΤΟΛΩΝ ΤΡΑΠΕΖΩΝ ΠΟΥ ΠΛΗΡΩΘΗΚΑΝ
- K6: ΧΡΗΜ/ΛΕΣ ΞΕΝΩΝ ΤΡΑΠΕΖΩΝ ΕΞΩΤΕΡΙΚΟΥ
- K7: ΧΡΗΜ/ΛΕΣ ΞΕΝΩΝ ΤΡΑΠΕΖΩΝ ΕΞΩΤΕΡΙΚΟΥ ΜΕΣΩ Κ.Κ
- K8: ΧΡΗΜ/ΛΕΣ ΞΕΝΩΝ ΤΡΑΠΕΖΩΝ ΠΡΟΣ Κ.Κ
- K9: ΑΓΟΡΑ ΞΕΝΩΝ ΤΡΑΠΕΖΟΓΡΑΜΜΑΤΙΩΝ
- K10: ΠΩΛΗΣΗ ΞΕΝΩΝ ΤΡΑΠΕΖΟΓΡΑΜΜΑΤΙΩΝ ΣΕ ΠΕΛΑΤΕΣ
- K11: ΠΩΛΗΣΗ ΞΕΝΩΝ ΤΡΑΠ/ΤΙΩΝ ΣΕ ΠΕΛΑΤΕΣ Κ ΧΡΗΜ/ΛΕΣ ΠΡΟΣ Κ.Κ
- K12: ΑΓΟΡΑ ΕΠΙΤΑΓΩΝ
- K13: ΕΚΔΟΣΗ ΕΠΙΤΑΓΩΝ
- K14: ΕΙΣΑΓΩΓΗ ΕΝΤΟΛΩΝ
- K15: ΕΚΔΟΣΗ ΕΝΤΟΛΩΝ
- K16: ΣΥΝΑΛ/ΜΑ ΓΙΑ ΛΟΙΠΕΣ ΑΝΑΓΚΕΣ
- K17: ΕΙΣΑΓΩΓΗ ΑΞΙΩΝ ΕΞΩΤ. ΕΣΩΤ.
- K18: ΕΚΚΡΕΜΕΙΣ ΑΞΙΕΣ ΣΕ ΤΡΙΤΟΥΣ
- K19: ΕΝΤΟΛΕΣ ΣΥΝ/ΤΟΣ ΚΛΗΡΙΝΓΚ
- K20: ΕΝΤΟΛΕΣ ΣΥΝ/ΤΟΣ ΠΟΥ ΔΙΑΒΙΒ.
- K21: ΕΓΚΡΙΣΕΙΣ ΕΙΣΑΓ. ΠΟΥ ΕΚΔΟΘΗΚΑΝ
- K22: ΑΔΕΙΕΣ ΕΚΤΕΛ. ΠΟΥ ΕΚΔΟΘΗΚΑΝ
- K23: ΑΔΕΙΕΣ ΕΚΤΕΛ. ΕΓΚΡ. ΤΡΑΠΕΖΩΝ
- K24: ΑΔΕΙΕΣ ΕΚΤΕΛ. ΕΓΚΡ. ΧΩΡΙΣ ΣΥΝΑΛ. ΔΙΑΤΥΠ.
- K25: ΛΗΦΘΕΙΣΕΣ ΔΗΛΩΣΕΙΣ ΕΞΑΓΩΓΗΣ
- K26: ΑΔΕΙΕΣ ΕΞΑΓ. ΠΟΥ ΕΚΔΟΘΗΚΑΝ
- K27: ΕΓΚΡ. ΔΗΛΩΣΕΙΣ & ΑΔΕΙΕΣ ΕΞΑΓ.
- K28: ΟΜΟΛ. & ΤΟΚΟΜΕΡ. ΠΟΥ ΕΞΟΦΛΗΘΗΚΑΝ



- [1]. J.R QUINLAN “INDUCTION OF DECISION TREES” 1987
- [2]. J.R QUINLAN “GENERATING PRODUCTION RULES FROM DECISION TREES”. 1986
- [3]. R.S MICHALKSI,J.G GARDONELL, AND T.M MITCELL. “MACHINE LEARNING : AN ARTIFICIAL INTELLIGENCE APPROACH”. 1983
- [4]. R.A BOSWELL “MANUAL FOR NEWID VERSION 2.0. TECHNICAL REPORT” 1990
- [5]. J.R QUINLAN “KNOWLEDGE ACQUISITION FROM STRUCTURED DATA” 1991
- [6]. J.R QUINLAN “LEARNING LOGICAL DEFINITIONS FROM RELATIONS”1990
- [7]. I.KONONENKO AND I.BRATKO “INFORMATION – BASED EVAL
- [8]. J.R QUINLAN “DECISION TREES AND MULTI-VALUED ATTRIBUTES” 1985
- [9]. R.M GOODMAN AND SMYTH “DECISION TREE DESIGN USING INFORMATION THEORY” 1990
- [10]. J.R QUINLAN “GENERATING PRODUCTION RULES FROM DECISION TREES” 1987
- [11]. SNEATH AND SOKAL “ PRINCIPLES OF NUMERICAL TAXONOMY” 1963

- [12]. HARTIGAN J.A “REPRESENTATION OF SIMILARITY MATRICES BY TREEW” 1981
- [13]. FRIEDMAN AND RUBIN “ ON SOME INVARIANT CRITERIA FOR GROUPING DATA” 1981
- [14]. JOHNSON S.C “HIERARCHICAL CLUSTERING SCHEMES” 1967
- [15]. ANDERBERG “CLUSTER ANALYSIS FOR APPLICATIONS” 1973
- [16]. SCHAUM “ΘΕΩΡΙΑ ΠΙΘΑΝΟΤΗΤΩΝ”
- [17]. Ι.ΦΙΛΗΣ “ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ Κ ΘΕΩΡΙΑ ΠΛΗΡΟΦΟΡΙΩΝ"
- [18]. “GUIDE OF SPSS”