

Τεχνητή Νοημοσύνη

3η Εργασία

Γραμμένος Θεόδωρος

A.E.M.: 3294

Άσκηση 1

Ερώτημα α

Ένα χαρακτηριστικό που είναι χρήσιμο είναι το `director_name`. Ο σκηνοθέτης μιας ταινίας επηρεάζει σε μεγάλο βαθμό το τελικό αποτέλεσμα. Συνεπώς είναι λογικό ένας καλός σκηνοθέτης να βγάζει ταινίες με υψηλή βαθμολογία.

Ένα χαρακτηριστικό που δεν είναι χρήσιμο είναι το `facenumber_in_poster`. Ο αριθμός των προσώπων ίσως παίζει κάποιο ρόλο στο εάν κάποιος θα επιλέξει να δει την ταινία, αλλά σίγουρα δεν επηρεάζει τη βαθμολογία που θα της δώσει.

Ερώτημα β

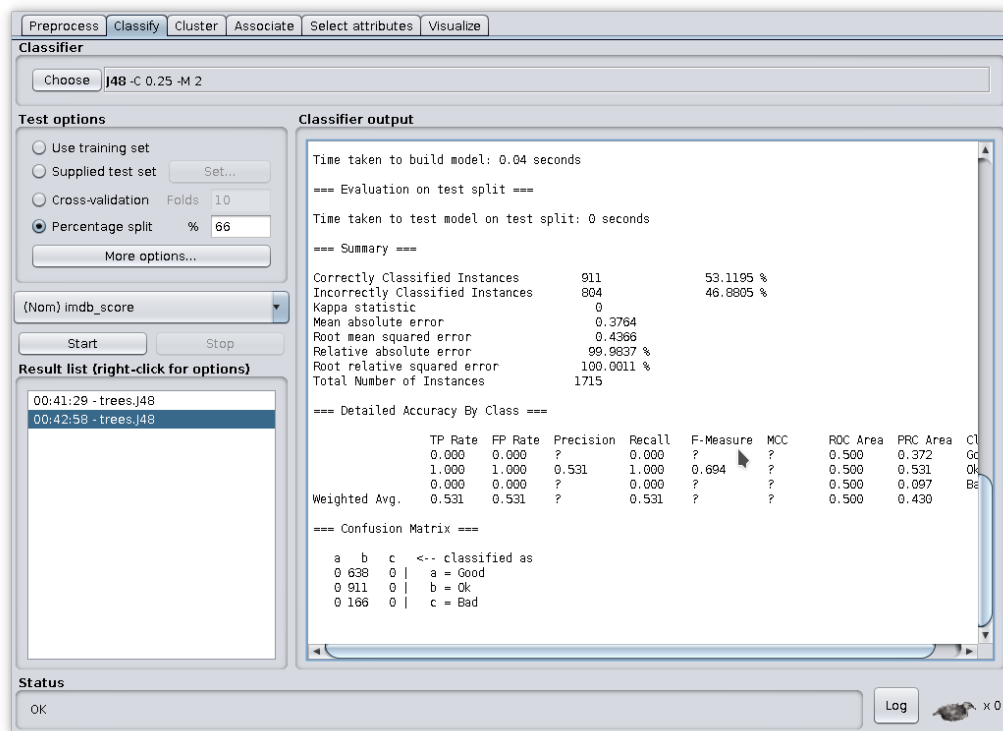
Ένα χρήσιμο χαρακτηριστικό θα ήταν ο αριθμός των βραβείων που έχει πάρει. Ταινίες που έχουν πολλά βραβεία, είναι φυσικό να έχουν μεγαλύτερη βαθμολογία.

Θα μπορούσε να προστεθεί ο συνθέτης του soundtrack της ταινίας. Βραβευμένοι συνθέτες μπορούν να δημιουργήσουν κομμάτια που βελτιώνουν μια ταινία και συνδέονται μαζί της. Για παράδειγμα ο John Williams έχει δημιουργήσει τη μουσική για πολλές διάσημες ταινίες, όπως η σειρά Star Wars, Η Λίστα του Σίντλερ, Σούπερμαν κ.α..

Μία ακόμα πληροφορία είναι η εταιρεία που διανέμει την ταινία. Διαφορετικές εταιρείες επικεντρώνονται σε ταινίες διαφορετικού είδους και αυτό μπορεί να επηρεάσει τη βαθμολογία.

Άσκηση 2

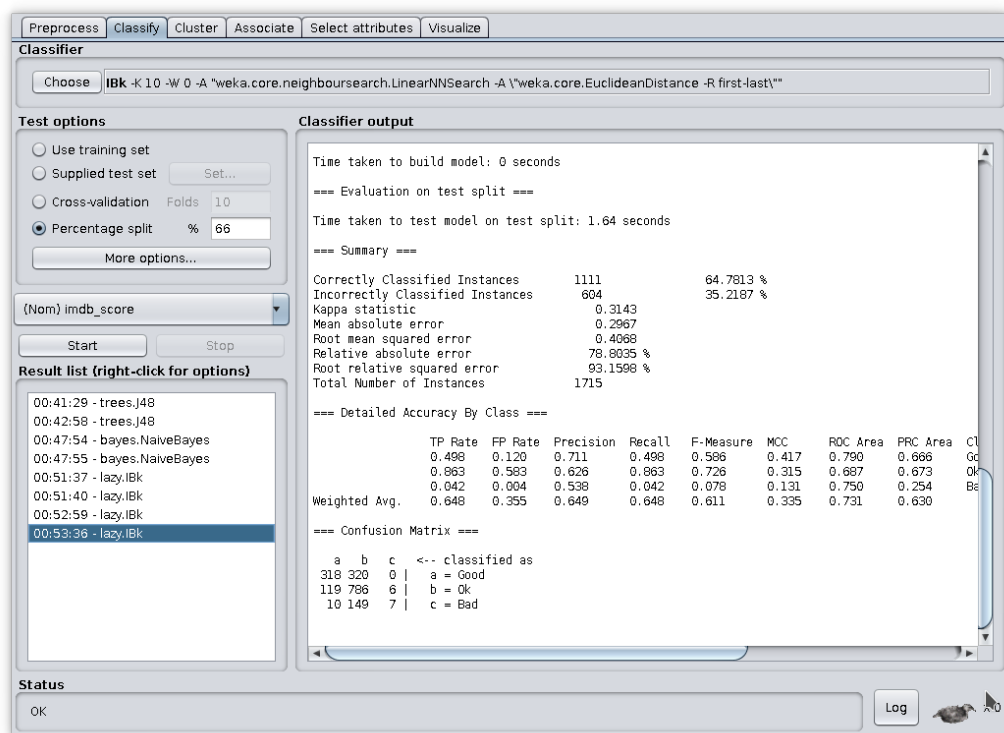
Ο πρώτος αλγόριθμος που δοκιμάστηκε είναι ο J48. Με 10-fold cross validation ο αλγόριθμος έχει ακρίβεια 55%, ενώ όταν επιλέγεται 66% split έχει ακρίβεια 53%.



Σχήμα 1: Τα αποτελέσματα του J48

Ο δεύτερος αλγόριθμος που δοκιμάστηκε είναι ο NaiveBayes. Ο αλγόριθμος με cross validation είχε ακρίβεια 43% και με 66% split ακρίβεια 40%. Τα αποτελέσματα αυτού του αλγορίθμου είναι χειρότερα από του J48.

Ο τρίτος αλγόριθμος που δοκιμάστηκε είναι ο k-nearest-neighbour. Δοκιμάστηκε με 5 και με 10 γείτονες. Με 5 γείτονες είχε ακρίβεια 62% στο percentage split και 65% στο cross validation. Με 10 γείτονες είχε ακρίβεια 64% στο percentage split και 65% στο cross validation. Συνεπώς η αύξηση του αριθμού των γειτόνων που λαμβάνονται υπόψη δεν επηρέασε την ακρίβεια του μοντέλου.



Σχήμα 2: Τα αποτελέσματα του 10-nearest neighbour

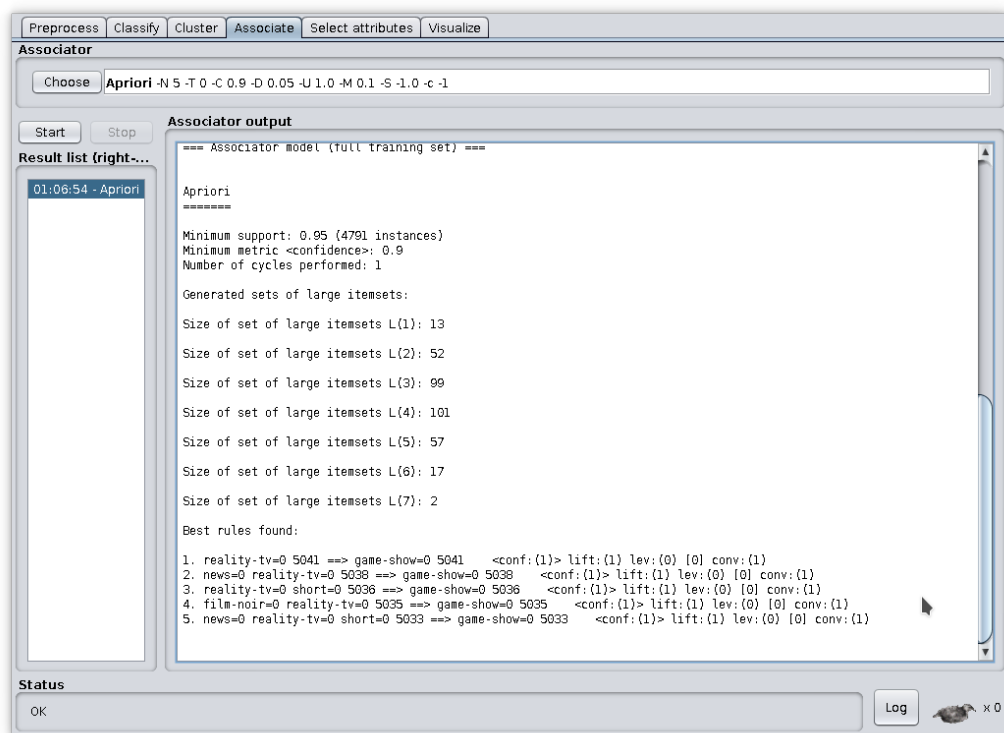
Άσκηση 3

Εκτελώντας τον αλγόριθμο SimpleKMeans με $k = 3$ παίρνουμε 3 ομάδες με μέγεθος: 1475, 2270 και 1298.

Άσκηση 4

Οι 5 κανόνες με $support \leq 0.1$ και το μεγαλύτερο confidence είναι:

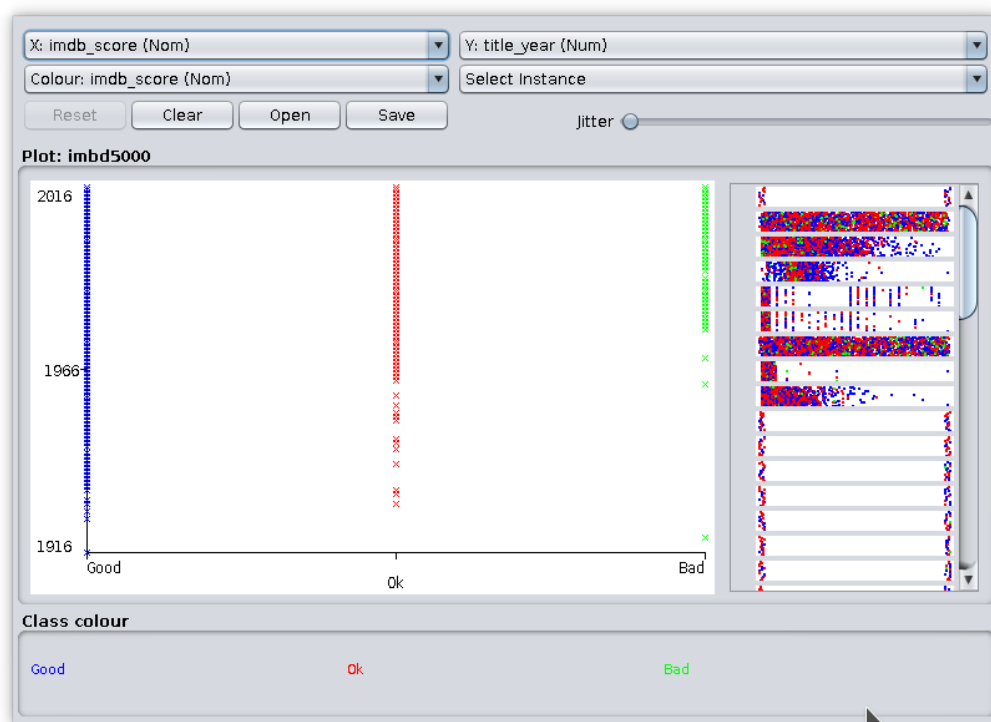
- Αν reality-tv=0 τότε game-show=0
- Αν news=0 και reality-tv=0 τότε game-show=0
- Αν reality-tv=0 και short=0 τότε game-show=0
- Αν film-noir=0 και reality-tv=0 τότε game-show=0
- Αν news=0, reality-tv=0 και short=0 τότε game-show=0



Σχήμα 3: Τα αποτελέσματα του Association

Άσκηση 5

Στο διάγραμμα φαίνεται ότι οι ταινίες με χαρακτηρισμό bad περιορίζονται στα τέλη του 20ου και στον 21ο αιώνα.



Σχήμα 4: Τα αποτελέσματα του Visualisation