

ΤΕΧΝΗΤΗ ΝΟΗΜΟΣΥΝΗ

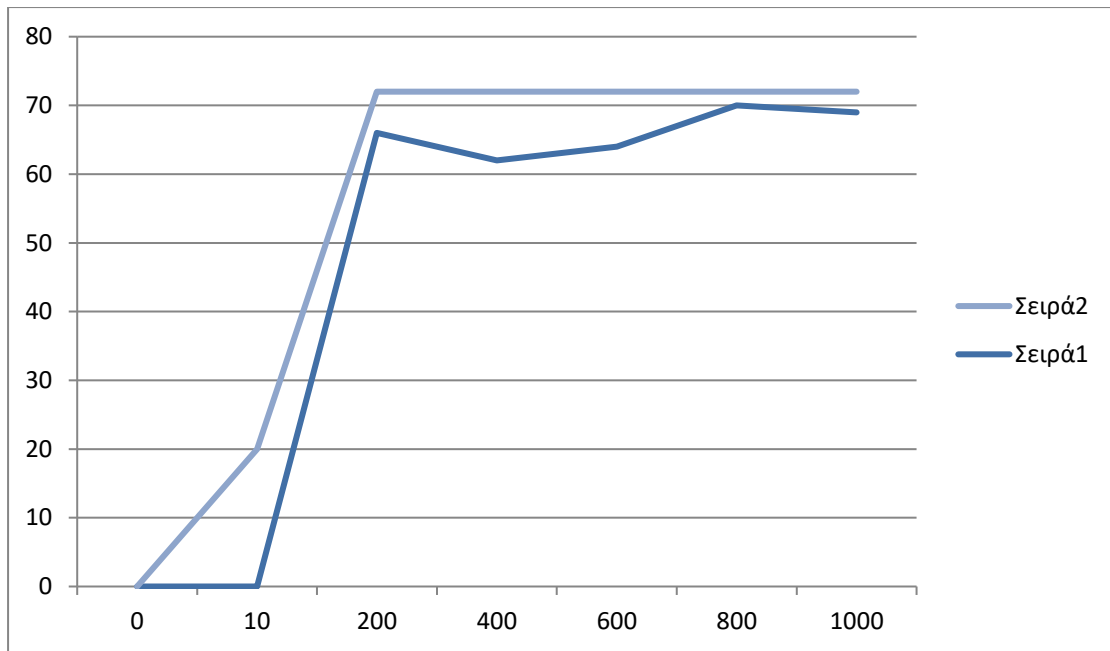
ΕΡΓΑΣΙΑ 3

ΙΑΚΩΒΟΣ ΕΥΔΑΙΜΩΝ 3130059

Στην εργασία αυτή πρέπει να αναπτύξουμε έναν αλγόριθμο μηχανικής μάθησης. Επέλεξα τον ID3. Στην υλοποίηση μου παρέχονται 8 κλάσεις. Μία για την δημιουργία των κόμβων του δέντρου απόφασης, μία που δημιουργεί το δέντρο απόφασης, μία κλάση που δημιουργεί τα χαρακτηριστικά κάθε παραδείγματος, άλλη μία που δημιουργεί στιγμιότυπα παραδειγμάτων και άλλη μία που διαβάζει τα αρχεία και μία που εκτελεί τον ID3. Επίσης έχουμε και την main μας που απλά καλεί τον id3 με διαφορετικά παραδείγματα. Χρησιμοποιήθηκαν 2 Dataset που αντλήθηκαν από το <http://archive.ics.uci.edu/ml/>. Ένα με παραδείγματα αμαξιών και ένα με παραδείγματα λουλουδιών. Θα αναλύσουμε παρακάτω τα αποτελέσματα του ID3 για τα παραδείγματα των αυτοκινήτων κυρίως. Ο ID3 παρουσιάζει συνεχώς για τα δοκιμαστικά παραδείγματα Accuracy of DT learner: 32.0% Error of DT learner: 68.0% και διαφαίνεται να δημιουργείται πρόβλημα υπερπροσαρμογής στα παραδείγματα εκπαίδευσης και να μην βελτιώνεται στα δοκιμαστικά τέστ. Το γεγονός αυτό μπορεί να οφείλεται στα «κακά» παραδείγματα εκπαίδευσης όπου έχουμε λίγες ιδιότητες και ασυνεπή παραδείγματα τα οποία μοιάζουν αρκετά μεταξύ τους σε πολλές ιδιότητες και μπορεί να ανήκουν σε άλλες κατηγορίες. Επίσης πρόβλημα σίγουρα δημιουργεί το γεγονός ότι :

class	N	N[%]
-----	-----	-----
unacc	1210	(70.023 %)
acc	384	(22.222 %)
good	69	(3.993 %)
v-good	65	(3.762 %)

Βλέπουμε ότι οι 2 κλάσεις είναι σχεδόν ανύπαρκτες ενώ η άλλη έχει καλύτερο ποσοστό αλλά και πάλι είναι μικρό. Το recall και το precision διαφαίνονται όταν τρέξουμε τον κώδικα.



Series 2 : Συνάρτηση ακρίβειας για τα testset

Series 1: Συνάρτηση ακρίβειας για τα trainingset

Αξονας x: Αριθμός παραδειγμάτων εκπαίδευσης

Άξονας y : Ακρίβεια της %(εκατό).