



Heart Disease Classification (2)

Θέλουμε να δημιουργήσουμε ένα σύστημα που να κατηγοριοποιεί ασθενείς με καρδιακές παθήσεις.

Από το παρακάτω link μπορείτε να κατεβάσετε το αρχείο «processed.cleveland.data» το οποίο περιέχει στοιχεία και δεδομένα από 303 ασθενείς.

<http://archive.ics.uci.edu/ml/machine-learning-databases/heart-disease/processed.cleveland.data>

Τα δεδομένα για κάθε ασθενή περιγράφονται με τη βοήθεια 13 στοιχείων-χαρακτηριστικών. Αυτά είναι:

- ~~1. Ηλικία (πραγματικός αριθμός)~~
2. Φύλο (0=γυναίκα, 1=άντρας)
3. Πόνος στο στήθος (τύπος 0=καθόλου, 1, 2, 3 ή 4)
- ~~4. Πίεση αίματος (πραγματικός αριθμός)~~
- ~~5. Χοληστερόλη σε mg/dl (πραγματικός αριθμός)~~
6. Σάκχαρο αίματος > 120 mg/dl (τιμές 0 ή 1)
7. Αποτελέσματα Ηλεκτροκαρδιογραφήματος σε ανάπαυση (τιμές 0=normal, 1, 2)
- ~~8. Μέγιστος αριθμός σφυγμών (πραγματικός αριθμός)~~
9. exercise induced angina (τιμές 0=no ή 1=yes)
- ~~10. Oldpeak = ελάττωση του ST που προκαλείται με την άσκηση σε σχέση με την ηρεμία (πραγματικός αριθμός)~~
11. Η κλίση του μεγίστου ST κατά την άσκηση (τιμές 1, 2, 3)
- ~~12. number of major vessels (0-3) colored by flourescopy (πραγματικός αριθμός)~~
13. thal: (3 = κανονικός, 6 = σταθερή βλάβη, 7 = αναστρέψιμη βλάβη)

Οι κατηγορίες/κλάσεις διάγνωσης που θεωρούμε είναι 5 και εμφανίζονται ως το 14^ο στοιχείο σε κάθε γραμμή στο data-set:

Κλάση 0: Χωρίς καρδιακές παθήσεις

Κλάση 1: Λιγότερο πιθανό να έχει καρδιακή πάθηση

Κλάση 2: Πιθανότερο να έχει καρδιακή πάθηση από την κλάση 1

Κλάση 3: Πιθανότερο να έχει καρδιακή πάθηση από την κλάση 2

Κλάση 4: Η μέγιστη πιθανότητα να έχει καρδιακή πάθηση.

~~Για την επίλυση του προβλήματος αυτού, θα χρησιμοποιήσετε τον αλγόριθμο CART (Classification And Regression Tree).~~