

Ατομική Εργασία #2

Προθεσμία υποβολής: 5 Ιουνίου 2022

Μηχανική Μάθηση

Χρησιμοποιείτε έναν αλγόριθμο της επιλογής σας (π.χ., ID3, νευρωνικά δίκτυα, Naïve Bayes) και εφαρμόστε τον σε ένα από τα παρακάτω σύνολα δεδομένων (datasets). Το μοντέλο που θα αναπτύξετε θα πρέπει να υποστηρίζει τη μηχανική μάθηση με επίβλεψη (supervised learning).

Διαθέσιμα σύνολα δεδομένων:

ΕΥΡΕΣΗ ΤΥΠΟΥ ΙΡΙΔΑΣ (IRIS)	
Χαρακτηριστικά	Κλάση (κατηγοριοποίηση)
1. Μήκος σεφάλου (sepal length) σε εκατοστά 2. Πλάτος σεφάλου (sepal width) σε εκατοστά 3. Μήκος πετάλου (petal length) σε εκατοστά 4. Πλάτος πετάλου (petal width) σε εκατοστά	<ul style="list-style-type: none">• Τύπος<ul style="list-style-type: none">○ Iris Setosa○ Iris Versicolour○ Iris Virginica

ΑΞΙΟΛΟΓΗΣΗ ΑΥΤΟΚΙΝΗΤΩΝ (CAR EVALUATION)	
Χαρακτηριστικά	Κλάση (κατηγοριοποίηση)
1. Κόστος αγοράς (buying): <ul style="list-style-type: none">a. Πολύ υψηλό (vhigh)b. Υψηλό (high)c. Μέτριο (med)d. Χαμηλό (low) 2. Κόστος συντήρησης (maint): <ul style="list-style-type: none">a. Πολύ υψηλό (vhigh)b. Υψηλό (high)c. Μέτριο (med)d. Χαμηλό (low) 3. Αριθμός θυρών (doors): <ul style="list-style-type: none">a. 2b. 3c. 4d. 5+ (5more) 4. Αριθμός ατόμων (persons): <ul style="list-style-type: none">a. 2b. 4c. 4+ (more) 5. Χώρος αποσκευών (lug_boot): <ul style="list-style-type: none">a. Μικρός (small)b. Μεσαίος (med)c. Μεγάλος (big) 6. Ασφάλεια (safety): <ul style="list-style-type: none">a. Χαμηλή (low)b. Μέτρια (med)c. Υψηλή (high)	<ul style="list-style-type: none">• Κατάσταση<ul style="list-style-type: none">○ Μη ικανοποιητική (unacc)○ Ικανοποιητική (acc)○ Καλή (good)○ Άριστη (very good)

ΕΥΡΕΣΗ ΤΥΠΟΥ ΣΤΑΦΙΔΑΣ (RAISIN)	
Χαρακτηριστικά	Κλάση (κατηγοριοποίηση)
<ol style="list-style-type: none"> 1. Περιοχή (Area): Αριθμός των πίξελ που περιέχεται η σταφίδα. 2. Περίμετρος (Perimeter): Αριθμός των πίξελ που οριοθετούν την περίμετρο γύρω από την σταφίδα. 3. Μέγιστο μήκος (MajorAxisLength): Το μέγιστο μήκος (σε πίξελ) που μπορεί να φανεί σε μια σταφίδα. 4. Ελάχιστο μήκος (MinorAxisLength): Το ελάχιστο μήκος (σε πίξελ) που μπορεί να φανεί σε μια σταφίδα. 5. Εκκεντρότητα (Eccentricity): Το μέσο μέγεθος της εκκεντρότητας για τις σταφίδες. 6. Κυρτότητα (ConvexArea): Η ελάχιστη περιοχή κυρτότητας (σε πίξελ) για τις σταφίδες. 7. Έκταση (Extent): Το ποσοστό έκτασης μεταξύ των σταφίδων και του υπόλοιπου χώρου. 	<ul style="list-style-type: none"> • Τύπος <ul style="list-style-type: none"> ο Kecimen ο Besni

Εκτός από τα παραπάνω σύνολα δεδομένων, μπορείτε να χρησιμοποιήσετε οποιοδήποτε άλλο θέλετε. Στο διαδίκτυο υπάρχει πληθώρα δεδομένων που διανέμονται ελεύθερα. Ενδεικτικοί ιστότοποι με σύνολα δεδομένων:

- UCI Machine Learning Repository: <https://archive.ics.uci.edu/ml/index.php>
- Google Dataset Search: <https://datasetsearch.research.google.com/>
- Kaggle: <https://www.kaggle.com/datasets>
- Datahub: <https://datahub.io/collections>

Παράδοση:

- Το παραδοτέο θα είναι η εφαρμογή αλγόριθμου της επιλογής σας σε ένα από τα παραπάνω σύνολα δεδομένα.
- Μπορείτε να εφαρμόσετε τον αλγόριθμο χρησιμοποιώντας κάποια γλώσσα προγραμματισμού (π.χ., Python) ή να τον παρουσιάσετε σε κείμενο (π.χ., αρχείο PDF)
- Η προθεσμία παράδοσης είναι η 5^η Ιουνίου 2022.