

# Θέματα ομαδικής εργασίας

## Μηχανικής Μάθησης 2022

---

### 7G. Πρόβλεψη μουσικών επιτυχιών με το Spotify

[Million Song Dataset](#) / υποσύνολο 10.000 - 20.000 κομμάτια

[Billboard API](#)

[Spotify API](#)

Χρησιμοποιήστε το API του Spotify για εξαγωγή χαρακτηριστικών από το Million Song Dataset και προβλέψτε τις επιτυχίες του Billboard.

---

### 4Z. Μελέτη κατάταξης αποτελεσμάτων μηχανών αναζήτησης

[Microsoft Learning to Rank Datasets](#)

Βελτιστοποίησης της σειράς εμφάνισης των αποτελεσμάτων ως προς κρίσεις συνάφειας των χρηστών.

---

### BC. Αναγνώριση ζωγράφου από πίνακες

[Painter by Numbers](#) / λάβετε υποσύνολο 20 καλλιτεχνών με κάποιες εκατοντάδες πίνακες για τον καθένα

Συγκρίνετε CNN και κλασικούς ταξινομητές. Για τους κλασικούς εξάγετε διάφορα χαρακτηριστικά εικόνας.

---

### JI. Εκτίμηση τιμών Airbnb στην Αθήνα

[Inside Airbnb](#) / δεδομένα της Αθήνας

Χρησιμοποιήστε και ανάλυση συναισθήματος στα reviews.

---

### XU. Καταμέτρηση αντικειμένων σε κάδους συσκευασιών της Amazon

[Amazon Bin Image Dataset](#) / θα χρειαστείτε το [AWS CLI](#)

Συγκρίνετε CNN και κλασικούς ταξινομητές. Για τους κλασικούς εξάγετε διάφορα χαρακτηριστικά εικόνας.

---

## **OC. Υβριδικό σύστημα συστάσεων συνταγών**

[Food.com Recipes and Interactions](#)

Προτείνετε συνταγές με βάση προηγούμενες επιλογές των χρηστών και τα κείμενα των συνταγών.

---

## **11. Ανίχνευση ανωμαλιών σε υλικό από κλειστό κυκλώμα τηλεόρασης**

[Anomaly Detection and Localization in Crowded Scenes](#)

Ανίχνευση των αποκλίσεων από την κανονική συμπεριφορά του πλήθους όπως καταγράφεται από κάμερες παρακολούθησης.

---

## **6W. Ταξινόμηση μουσικών ειδών**

[GTZAN Genre Collection](#)

[LibROSA](#)

Μετατρέψτε τα κομμάτια σε εικόνες (MFCC) και χρησιμοποιήστε CNN για εξαγωγή χαρακτηριστικών.

---

## **WF. Αναγνώριση σπάνιων ειδών πτηνών**

[Caltech-UCSD Birds-200-2011](#)

Στοχεύστε να βελτιστοποιήσετε το σύστημα σας για τις κατηγορίες πτηνών με τα λιγότερα δείγματα (μη ισορροπημένο σύνολο δεδομένων).

---

## **RZ. Πρόβλεψη δημοφιλίας βίντεο στο YouTube**

## [Trending YouTube Video Statistics](#)

Χρησιμοποιήστε στατιστικά, στοιχεία του βίντεο και κειμενική πληροφορία.

---

## **BK. Ανίχνευση fake news με βάση το Twitter**

[FakeNewsNet](#) / τα tweets θα πρέπει να κατέβουν με το Twitter API

Χρησιμοποιήστε τα URLs των ειδήσεων και τα tweets που τις μοιράζονται για να ανιχνεύσετε fake news.

---

## **7Z. Διάγνωση ακτινογραφιών θώρακα**

[CheXpert](#) / θα χρειαστεί να κάνετε register

Συνδυάστε τις εικόνες με χαρακτηριστικά που έχουν εξαχθεί από γιατρούς για να κάνετε διάγνωση.

---

## **AP. Πρόβλεψη τιμών αεροπορικών εισιτηρίων με ενισχυτική μάθηση**

[Travelpayouts API](#)

Εύρεση της καλύτερης χρονικής στιγμής για την αγορά αεροπορικών εισιτηρίων.

---

## **TR. Κατηγοριοποίηση αποικιών κοραλλιογενών υφάλων**

[Data from: Identifying multiple coral reef regimes and their drivers across the Hawaiian archipelago](#)

Ταξινόμηση κοραλλιογενών υφάλων με βάση περιβαλλοντικούς και ανθρώπινους παράγοντες.

---

## **70. Εκτίμηση επιβίωσης ασθενών με καρκίνο ωοθηκών**

[Integrated genomic analyses of ovarian carcinoma \(Supplementary Tables\)](#)

Θα χρησιμοποιηθούν κάποια από τα χαρακτηριστικά του άτλαντα του γονιδιώματος του καρκίνου.

---

#### **4C. Συστήματα συστάσεων και αραιότητα χαρακτηριστικών**

[MovieLens](#)

Μελέτη του προβλήματος cold start.

---

#### **WC. Αναγνώριση ανθρώπινων δραστηριοτήτων στα δάση του Αμαζονίου**

[planetlabs planet-amazon-deforestation](#)

Εντοπίστε ανθρώπινες δραστηριότητες που οδηγούν στην αποψίλωση των δασών.

---

#### **ED. Πρόβλεψη αστεριών εστιατορίων από reviews**

[Yelp](#)

Ανάλυση της κειμενικής περιγραφής των reviews εστιατορίων για πρόβλεψη των αστεριών τους.

---

#### **U6. Κατάταξη ξενοδοχείων του Expedia**

[Personalize Expedia Hotel Searches - ICDM 2013](#)

Βελτιστοποίηση της σειράς εμφάνισης των ξενοδοχείων με στόχο την επίτευξη πώλησης.

---

#### **OJ. Χρηματιστηριακές προβλέψεις με βάση ειδήσεις**

[Yahoo! Finance's API](#)

[GoogleNews API](#)

[NYTimes API](#)

Προβλέψτε τιμές χρηματιστηριακών δεικτών ή μετοχών με βάση τις ειδήσεις.

---

## **QN. Εξερεύνηση του παιχνιδιού Rogue (NetHack) με βαθιά ενισχυτική μάθηση**

[NetHack Learning Environment \(NLE\)](#)

Το παιχνίδι είναι πολύπλοκο, οπότε θα εξεταστεί μόνο η εξερεύνηση όπως [εδώ](#).

---

## **PC. Πρόβλεψη απώλειας βάρους από το MyFitnessPal**

[MyFitnessPal Dataset](#)

Χρησιμοποιήστε τα ημερολόγια χρηστών του MyFitnessPal για να προβλέψετε την απώλεια βάρους τους.

## Μηχανική Μάθηση

[Ταμπλό](#) ➤ [Μαθήματα](#) ➤ [Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών](#) ➤

[ΔΠΜΣ Επιστήμη Δεδομένων και Μηχανική Μάθηση](#) ➤ [Μηχανική Μάθηση](#) ➤ [Εργαστήριο](#) ➤

[Ομαδική εργασία εξαμήνου \(Εκφώνηση Εργασίας 2\)](#)

## Ομαδική εργασία εξαμήνου (Εκφώνηση Εργασίας 2)

Άνοιξε: Τετάρτη, 19 Ιανουαρίου 2022, 12:00 πμ

Λήγει: Κυριακή, 13 Μαρτίου 2022, 11:59 μμ

## Ομαδική εργασία Μηχανικής Μάθησης 2022

### Ομάδες

Η εργασία είναι ομαδική σε ομάδες δύο ή τριών ατόμων. Αν δεν είστε σε ομάδα δεν μπορείτε να επιλέξετε θέμα.

Μπορείτε να σχηματίσετε ομάδα από διαφορετικά προγράμματα σπουδών, διδακτορικές σπουδές ή ΔΠΜΣ.

Τυχόν μεμονωμένες συμμετοχές στο τέλος της περιόδου επιλογής ομάδας θα τοποθετηθούν τυχαία σε ομάδες.

### Επιλογή θέματος

Είναι διαθέσιμα 22 θέματα που αντιστοιχούν σε διάφορα tasks μηχανικής μάθησης του πραγματικού κόσμου. Επιλέξτε σύμφωνα με τα ενδιαφέροντά σας και όχι με κριτήριο αν μοιάζει εύκολο ή δύσκολο. Όλα τα προτεινόμενα θέματα έχουν μεγάλο βάθος για έρευνα.

Η επιλογή γίνεται μέσα από τη [φόρμα επιλογής θέματος](#) ομαδικής εργασίας στο Helios. Μόνο ένα μέλος της ομάδας χρειάζεται να υποβάλει και μπορεί να υποβάλει μόνο μία φορά. Θα χρειαστεί να κάνει sign-in με ένα λογαριασμό Google.

Κάθε θέμα μπορεί να επιλεγεί το μέγιστο από δύο ομάδες, στη συνέχεια η επιλογή του θα παύει να είναι διαθέσιμη στη φόρμα.

### Παρατηρήσεις για την εργασία

Το ζητούμενο για κάθε θέμα και dataset διατυπώνεται συνοπτικά. Εντός του ζητούμενου μπορείτε εφαρμόσετε αλγόριθμους μηχανικής μάθησης όπως κρίνετε εσείς. Μελετήστε τη βιβλιογραφία για τις προσεγγίσεις που υπάρχουν στο ζητούμενο σας.

Κάποια dataset είναι πολύ μεγάλα. Μπορείτε να χρησιμοποιήσετε υποσύνολά τους, εφόσον γίνει με σωστό τρόπο η κατασκευή τους.

Σε κάποια θέματα χρειάζεται να χρησιμοποιήσετε API ή το AWS CLI. Αυτό δεν τα κάνει ούτε δυσκολότερα ούτε ευκολότερα.

Θέλουμε να δούμε σύγκριση μεταξύ διάφορων μεθόδων, χαρακτηριστικών, ταξινομητών κλπ.

Από τη βαθιά μάθηση, μπορείτε να χρησιμοποιήσετε μόνο Feed-Forward Networks και CNN εφόσον έχουν εφαρμογή στο πρόβλημά σας. Ο βασικός κορμός της εργασίας πρέπει να αποτελείται από τις μεθοδολογίες της κλασικής μηχανικής μάθησης (≈ scikit-learn).

Εξαίρεση στο προηγούμενο αποτελούν τα θέματα ενισχυτικής μάθησης όπου χρησιμοποιούμε μόνο μοντέλα Deep Reinforcement Learning.

### Παραδοτέο

Το παραδοτέο της εργασίας είναι μια γραπτή αναφορά παρόμοια στη σύνταξη με ένα σύντομο paper για συνέδριο.



Για το περιεχόμενό της σας προτείνουμε να ακολουθήσετε το έγγραφ ["Final Report Guidelines."](#)

Για το formatting, θα θέλαμε αναφορές σε A4 που ακολουθούν το [conference template της IEEE](#) για Word ή *L<sup>A</sup>T<sub>E</sub>X*.

Αν τυχόν χρειαστεί να μοιράσετε επιπλέον υλικό (κώδικα, γραφήματα κλπ) θα χρησιμοποιήσετε το github.

Για οποιαδήποτε απορία είμαστε πάντοτε στη διάθεσή σας μέσω email ή Helios.

## Κατάσταση Υποβολής

Ομάδα	Group 12
Κατάσταση Υποβολής	Υποβλήθηκε για βαθμολόγηση
Κατάσταση βαθμολόγησης	Χωρίς βαθμό
Χρόνος που απομένει	7 ημέρες 3 ώρες
Τελευταία τροποποίηση	Παρασκευή, 4 Μαρτίου 2022, 12:32 μμ
Υποβολές αρχείων	<div> <a href="#">Group_12_Report.pdf</a> 4 Μαρτίου 2022, 12:32 μμ</div>
Σχόλια υποβολής	<div> <a href="#">Σχόλια (1)</a></div>