ΕΡΓΑΣΙΑ ΕΞΟΡΥΞΗΣ ΔΕΔΟΜΕΝΩΝ (2026-2027) - Εναλλακτική

Αναγνώριση τύπου δαπέδου με το FOSSBot

Σενάριο και στόχος

Ένα κινητό ρομπότ FOSSBot κινείται σε διαφορετικές επιφάνειες. Για την εργασία θα πρέπει να συλλέξετε πραγματικές καταγραφές από το περιβάλλον προσομοίωσης που υπάρχει στο fossbot.gr, να αναλύσετε τις μετρήσεις αισθητήρων και να αναπτύξετε μοντέλα για την αναγνώριση του τύπου δαπέδου στο οποίο κινείται κάθε φορά το ρομπότ.

Στόχος είναι η τεκμηριωμένη εκπαίδευση, σύγκριση και χρήση ταξινομητών. Ως πρόσθετο σκέλος του έργου FOSSBot προβλέπεται η μεταφορά του μοντέλου στο πραγματικό ρομπότ και, ο έλεγχος λειτουργίας του σε πραγματικές συνθήκες, με προσαρμογή όπου απαιτείται.

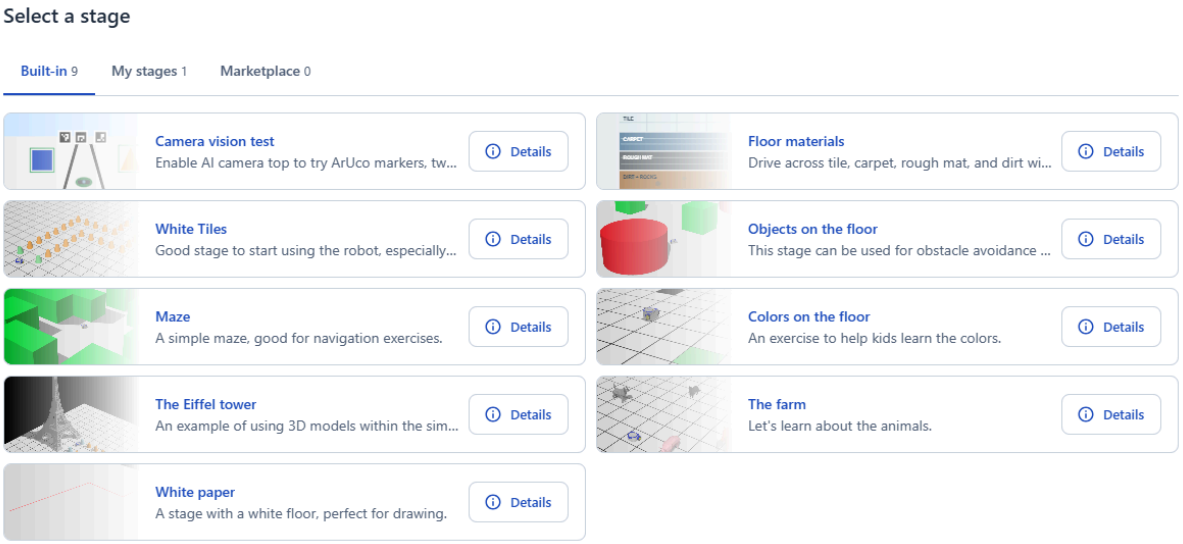
Δεδομένα και πρόσβαση

Για την εργασία θα πρέπει να φτιάξετε ένα σύνολο δεδομένων που θα συλλέξετε μέσω του https://fossbot.gr προσομοιωτή.

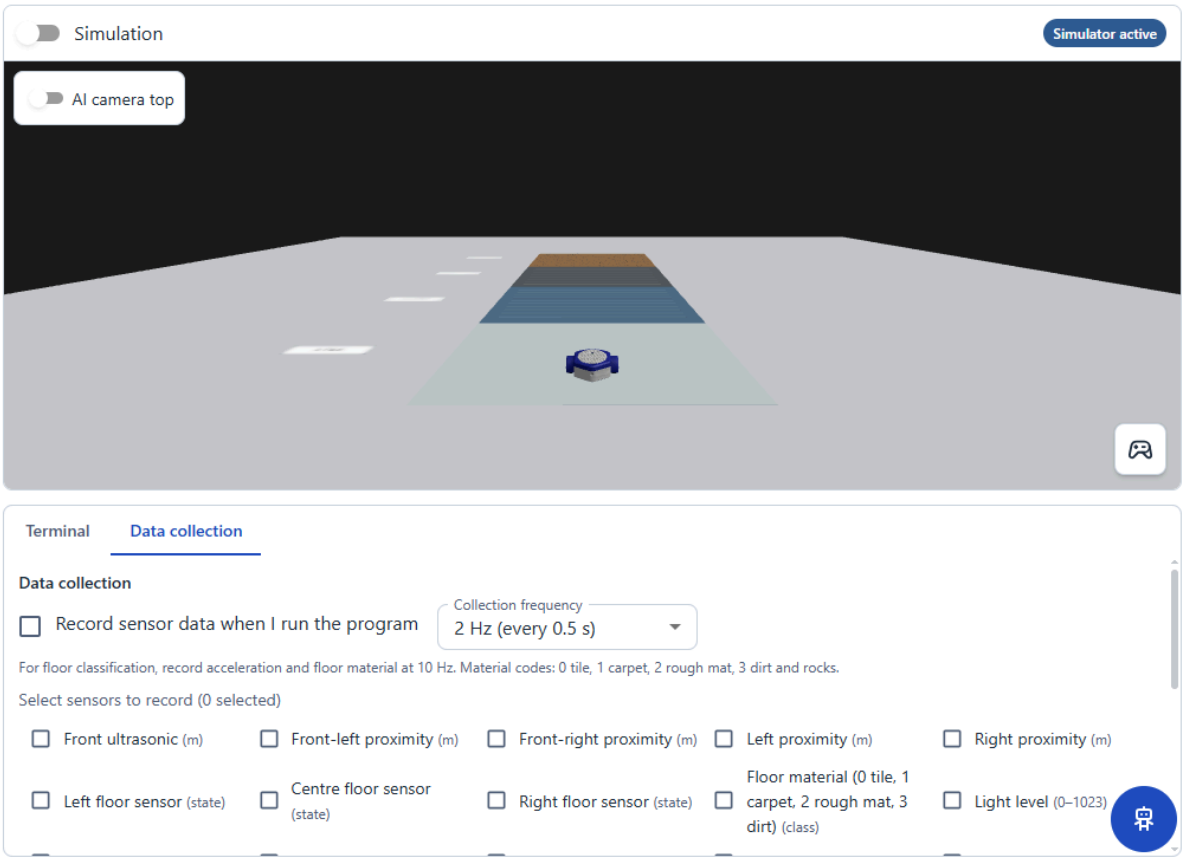
Θα πρέπει να πάτε στην επιλογή Python Editor και από εκεί στην επιλογή Change stage.



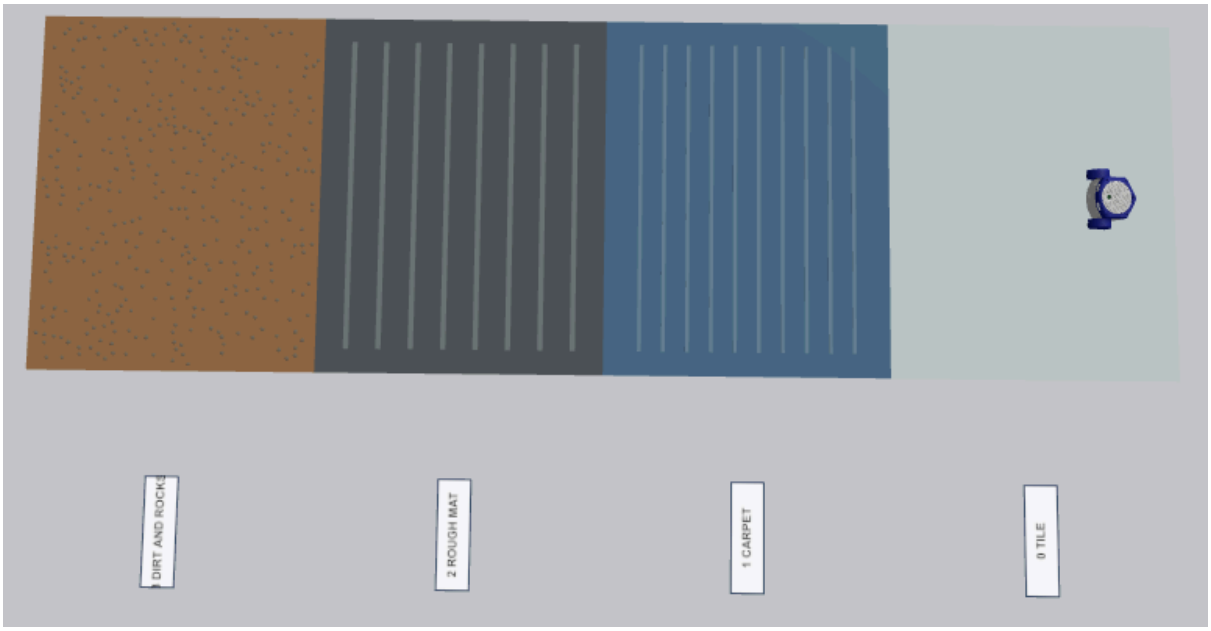
Στην οθόνη που ανοίγει επιλέξτε την πίστα Floor materials



Μόλις επιλέξετε την πίστα θα πρέπει να ρυθμίσετε τους αισθητήρες με τους οποίους θα καταγράψετε δεδομένα.



Ενδεικτικά μπορείτε να καταγράψτε επιτάχυνση, γυροσκόπιο, τις καταστάσεις των τριών αισθητήρων δαπέδου, οδομετρία κλπ. Το γνώρισμα floor_label είναι η μεταβλητή-κλάσης y. Υπάρχουν 4 κλάσεις (πλακάκι, χαλί, τραχύ πατάκι, χώμα).



Πραγματοποιήστε πιλοτικές διαδρομές με διαφορετικές ταχύτητες και δειγματοληψίες ώστε να συλλέξετε δεδομένα ανά κλάση και να βρείτε τα δεδομένα που εκπαιδεύουν καλύτερα τα μοντέλα σας.

Αν καταγράψετε πολλαπλά αρχεία CSV (ένα για κάθε διαδρομή) συγχωνεύστε τα σε ένα ενιαίο αρχείο (ανά ρυθμό δειγματοληψίας). Κάθε έγκυρη γραμμή αποτελεί ανεξάρτητο παράδειγμα στην αναπαράσταση. Μπορείτε να χρησιμοποιήσετε χρονικά παράθυρα αλλά δεν είναι υποχρεωτικό.

ΕΡΓΑΣΙΕΣ

Α) Προετοιμασία και αρχική ανάλυση

Α1. Φόρτωση και έλεγχος ποιότητας

- Φτιάξτε το ενιαίο αρχείο και κρατήστε τα γνωρίσματα που συμβάλουν στην πρόβλεψη. Καταγράψτε τα τελικά ονόματα/τύπους μεταβλητών και το πλήθος αρχείων και γραμμών πριν και μετά τον καθαρισμό.
- Παρουσιάστε κάποια στατιστικά δεδομένα όπως: τις συχνότητες των τύπων δαπέδου, ελλιπείς τιμές ανά αισθητήρα κλπ
- Εντοπίστε διπλότυπα, σταθερές στήλες, περίεργες τιμές και ρυθμούς δειγματοληψίας.
- Διερευνήστε θόρυβο και ακραίες τιμές. Ξεχωρίστε σφάλματα αισθητήρα από έγκυρες έντονες κινήσεις. Εξετάστε δεδομένα ακινησίας και την επίδραση της βαρύτητας στην επιτάχυνση.

Α2. Διερευνητική ανάλυση

- Εξετάστε κατανομές μετρήσεων ανά δάπεδο και ταχύτητα. Παρουσιάστε 4–6 στοχευμένα διαγράμματα με άξονες, μονάδες και σύντομη ερμηνεία.
- Εντοπίστε συσχετισμένα ή πλεονάζοντα γνωρίσματα με πίνακα/χάρτη συσχετίσεων και σχολιάστε επιλεγμένα ζεύγη.
- Κατατάξτε τα χαρακτηριστικά ως προς τη χρησιμότητα για αναγνώριση δαπέδου με τεκμηριωμένη μέθοδο. Ελέγξτε μήπως η ταχύτητα ή η κάποιο άλλο γνώρισμα οδηγεί σε αποφάσεις που δε γενικεύουν.

Α3. Προεπεξεργασία και μείωση διαστάσεων

- Σχεδιάστε αντιμετώπιση κενών, διαφορετικών κλιμάκων και προβληματικών τιμών. Συγκρίνετε εναλλακτικές επιλογές και αιτιολογήστε την τελική.
- Συγκρίνετε την αρχική έγκυρη αναπαράσταση με τουλάχιστον μία μειωμένη, π.χ. PCA. Αιτιολογήστε τον αριθμό διαστάσεων και αξιολογήστε την επίδραση σε ταξινόμηση και συσταδοποίηση.

Β) Κατηγοριοποίηση/ταξινόμηση

Β1. Πειραματική οργάνωση

Χωρίστε ολόκληρα αρχεία σε training, validation και ανεξάρτητο test (ενδεικτικά 80/10/10 ανά κλάση), με κάλυψη διαφόρων ταχυτήτων. Αναφέρετε το μέγεθος και τη σύνθεση των συνόλων που χρησιμοποιείτε. Επιλέξτε μια συνεπή διαδικασία ώστε οι συγκρίσεις των μοντέλων να είναι ουσιαστικές.

Β2. Μοντέλα και μετρικές

- Συγκρίνετε τουλάχιστον πέντε ταξινομητές από διαφορετικές οικογένειες. Τεκμηριώστε γιατί επιλέχθηκαν.
- Διερευνήστε εναλλακτικά σύνολα υπερπαραμέτρων, καταγράψτε τις τιμές και την τελική επιλογή. Χρησιμοποιήστε macro-F1 ως κύρια μετρική και επιπλέον

balanced accuracy, precision/recall ανά κλάση, πίνακα σύγχυσης και χρόνους εκπαίδευσης/πρόβλεψης.

Β3. Τι βελτίωσε τη λύση;

- Για τουλάχιστον έναν ταξινομητή συγκρίνετε αρχική και μειωμένη αναπαράσταση ως προς macro-F1, χρόνο και πλήθος γνωρισμάτων.
- Κατασκευάστε καμπύλη μάθησης με 10%, 30% και 100% του training, επιλέγοντας ολόκληρες διαδρομές ή δηλώνοντας τον περιορισμό της δειγματοληψίας. Χρησιμοποιήστε κοινό σύνολο αξιολόγησης.
- Αναλύστε δύο συχνές συγχύσεις κλάσεων και την επίδοση σε λιγότερο συχνές κατηγορίες.

Γ) Ανθεκτικότητα και γενίκευση

Γ1 Επαύξηση με αιτιολόγηση

Εφαρμόστε τουλάχιστον μία απλή τεχνική επαύξησης μόνο στις γραμμές εκπαίδευσης, όπως μικρό προσθετικό θόρυβο στις συνεχείς τιμές επιτάχυνσης ή γυροσκοπίου. Δημιουργήστε νέα αντίγραφα γραμμών με μικρές μεταβολές και την ίδια ετικέτα. Αιτιολογήστε την ένταση θορύβου από τις τιμές του training και ελέγξτε ότι τα παραγόμενα δεδομένα είναι εύλογα. Προαιρετικά εξετάστε τυχαία υπερδειγματοληψία όταν υπάρχει ανισορροπία κλάσεων.

- Διατηρήστε αμετάβλητες τις ετικέτες και τις δυαδικές καταστάσεις. Μην προσθέσετε συνεχή θόρυβο στα floor κανάλια και μην αλλάξετε αυθαίρετα τη συνιστώσα βαρύτητας. Κάθε επαυξημένη γραμμή παραμένει στο ίδιο training fold με το αρχικό της αρχείο. Η επαύξηση γίνεται μετά τον διαχωρισμό των αρχείων.
- Για έναν επιλεγμένο ταξινομητή συγκρίνετε εκπαίδευση χωρίς επαύξηση και με επαύξηση, με το ίδιο validation και το ίδιο υπολογιστικό όριο. Η τελική επιλογή αξιολογείται μία φορά στο ανέγγιχτο test. Δεν απαιτείται βελτίωση της επίδοσης· απαιτείται να εξηγήσετε το αποτέλεσμα.

Νέες προσομοιωμένες διαδρομές με μεταβολές ταχύτητας ή θέσης αποτελούν πρόσθετη συλλογή δεδομένων και δηλώνονται χωριστά από την αριθμητική επαύξηση. Η υπερδειγματοληψία δεν αντικαθιστά νέες ανεξάρτητες καταγραφές.

Γ2. Δοκιμή στο πραγματικό FOSSBot

Με υποστήριξη του Χρήστου Χρόνη (chronis@hua.gr) θα οργανωθούν η πρόσβαση στο φυσικό ρομπότ για τον έλεγχο του μοντέλου. Οι δοκιμές θα αφορούν ταξινόμηση της κίνησης σε αντίστοιχες πραγματικές επιφάνειες. Όπου απαιτηθεί μπορείτε να κάνετε fine-tuning του μοντέλου με τα πραγματικά δεδομένα αισθητήρων.

Στοχεύστε σε 6 ανεξάρτητες καταγραφές ανά κλάση: 3 για προσαρμογή και 3 για δοκιμή, με δύο ταχύτητες συνολικά. Συλλέξτε το test σε χωριστή συνεδρία και, αν είναι εφικτό, σε άλλη διάταξη επιφανειών. Προτιμήστε 15–30 s κίνησης όταν επιτρέπεται.

Συγκρίνετε S→S (simulation-train σε simulation-test), S→R (ίδιο παγωμένο μοντέλο σε real-test) και S και R→R (simulation-train + real-adaptation, αξιολόγηση στο ίδιο ανέγγιχτο real-test). Αναφέρετε macro-F1, πίνακες σύγχυσης και περιορισμούς μικρού δείγματος.

ΠΑΡΑΔΟΣΗ

Η ημερομηνία παράδοσης ορίζεται η 15/1/2027 στο eclass.

Κάθε ομάδα παραδίδει ένα ενιαίο πακέτο που συνδέει τα ερωτήματα, τα πειράματα και τα συμπεράσματα. Το πακέτο πρέπει να καλύπτει τις ενότητες Α, Β, Γ.

Τελικό πακέτο υποβολής

Παραδοτέο	Περιεχόμενο
report.pdf	Τεχνική αναφορά 8–12 σελίδων, χωρίς εξώφυλλο, βιβλιογραφία και προαιρετικά παραρτήματα. Διακριτές ενότητες Α, Β και Γ, συμπεράσματα, περιορισμοί και συνεισφορά μελών.
src/ ή notebooks/	Εκτελέσιμος κώδικας φόρτωσης, προεπεξεργασίας, EDA, ταξινόμησης και φυσικών ελέγχων. Τα notebooks να περιλαμβάνουν τις βασικές εξόδους.
README.md και περιβάλλον	Οδηγίες λήψης/τοποθέτησης δεδομένων και εκτέλεσης, εκδόσεις βιβλιοθηκών, τυχαίοι σπόροι, απαιτήσεις υλικού και ενδεικτικός χρόνος. requirements.txt ή environment.yml.
results/	Πίνακες πειραμάτων σε CSV, βασικά διαγράμματα και ρυθμίσεις που επιτρέπουν την αντιστοίχιση κάθε αποτελέσματος με την εκτέλεσή του.

Υποβάλετε ένα συμπιεσμένο αρχείο ανά ομάδα. Μην συμπεριλάβετε τα αρχικά δεδομένα, αλλά δώστε σαφείς οδηγίες πρόσβασης.

Τελική παρουσίαση

Προτείνεται παρουσίαση 5–7 διαφανειών με έμφαση στις αποφάσεις και στα ευρήματα. Όλα τα μέλη συμμετέχουν στην προετοιμασία και στην παρουσίαση. Κάθε μέλος πρέπει να μπορεί να εξηγήσει τη συνολική προσέγγιση της ομάδας.

Ενδεικτικό περιεχόμενο

- Πρόβλημα, δεδομένα και ερευνητικά ερωτήματα
- Βασικά ευρήματα EDA και αποφάσεις προεπεξεργασίας
- Ταξινόμηση, συγκρίσεις και ανάλυση σφαλμάτων
- Αξιολόγηση σε πραγματικές συνθήκες, ανθεκτικότητα
- Τι βοήθησε, βασικοί περιορισμοί και τελικό συμπέρασμα

Προτιμήστε ευανάγνωστα διαγράμματα και συγκρίσεις αντί για εκτενή αποσπάσματα κώδικα. Να είστε έτοιμοι για σύντομες ερωτήσεις σχετικά με τις επιλογές και τα αποτελέσματά σας.

Αξιολόγηση

Η ποιότητα της τεκμηρίωσης και της ερμηνείας αποτελεί βασικό μέρος της αξιολόγησης. Δεν βαθμολογείται μόνο η υψηλότερη επίδοση. Μια τεκμηριωμένη σύγκριση, μια εύστοχη ανάλυση αποτυχίας ή ένα αρνητικό αποτέλεσμα μπορούν να αποτελούν ουσιαστική συνεισφορά. Δεν απαιτείται κάθε τεχνική να βελτιώνει τις μετρικές· απαιτείται να εξετάζετε και να εξηγείτε την επίδρασή της.

ΚΑΤΑΝΟΜΗ ΒΑΘΜΟΛΟΓΙΑΣ

Κριτήριο	Βαρύτητα	Τι αξιολογείται
Προεπεξεργασία και EDA	20%	Ποιότητα ελέγχων, συσχετίσεις, σημαντικά χαρακτηριστικά, αιτιολόγηση αποφάσεων.
Μείωση διαστάσεων και δειγματοληψία	15%	Ουσιαστικές συγκρίσεις αναπαραστάσεων και μεγεθών δεδομένων, επίδοση και κόστος.
Ταξινόμηση	25%	Πειραματική οργάνωση, baselines, διαφορετικές οικογένειες, μετρικές, και σφάλματα.
Ανθεκτικότητα	25%	Αποτελέσματα στο φυσικό ρομπότ, εκπαίδευση και fine tuning
Αναπαραγωγιμότητα και αναφορά	5%	Εκτελέσιμος κώδικας, πλήρεις οδηγίες, σαφής σύνδεση αποτελεσμάτων και συμπερασμάτων.

Προφορική παρουσίαση	10%	Σαφήνεια, επιλογή ευρημάτων, τήρηση χρόνου, συμμετοχή και κατανόηση της λύσης.

Οι δύο εργασίες με την καλύτερη τεκμηρίωση θα έχουν ένα bonus 10%.