

ΕΡΓΑΣΤΗΡΙΑΚΗ ΕΡΓΑΣΙΑ

ΦΑΣΗ 3: ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΑΞΙΟΛΟΓΗΣΗ ΜΟΝΤΕΛΩΝ ΠΡΟΒΛΕΨΗΣ

Καλείστε να εκπαιδεύσετε μοντέλα πάνω στο επιλεγμένο dataset σας και να αξιολογήσετε την προβλεπτική τους ικανότητα. Η ανάλυση θα περιλαμβάνει **1) ένα γραμμικό μοντέλο** και **2) ένα μη γραμμικό μοντέλο**.

0) Προεπεξεργασία δεδομένων

Βεβαιωθείτε ότι τα δεδομένα είναι σε κατάλληλη μορφή:

1. Να μην υπάρχουν ελλείπουσες τιμές
2. Να έχουν κωδικοποιηθεί κατάλληλα όλες οι κατηγορικές μεταβλητές
3. Να έχουν κλιμακωθεί αν αυτό απαιτείται από το μοντέλο

Στη συνέχεια χωρίστε τα δεδομένα σε Training set και Test set με αναλογία 80%-20%.

1) Γραμμικό μοντέλο

Εκπαιδεύστε το κατάλληλο γραμμικό μοντέλο για το πρόβλημά σας: Γραμμική Παλινδρόμηση για προβλήματα παλινδρόμησης, Λογιστική Παλινδρόμηση για προβλήματα ταξινόμησης.

Αξιολογήστε την επίδοση του μοντέλου με τα κατάλληλα κριτήρια:

Παλινδρόμηση: R2, RMSE, MAE, διάγραμμα Predicted vs Actual, διάγραμμα Residuals vs Fitted

Ταξινόμηση: Confusion Matrix. Precision, Recall, F1-score. Καμπύλη ROC, μετρική AUC

Και για τα δυο μοντέλα, εξετάστε για πολυσυγγραμικότητα. Υπάρχει ένδειξη ότι κάποιες μεταβλητές θα είχε νόημα να αφαιρεθούν; Αν ναι, δοκιμάστε να τις αφαιρέσετε και επαναλάβετε τη διαδικασία εκπαίδευσης-αξιολόγησης.

2) Μη γραμμικό μοντέλο

Εκπαιδεύστε ένα μοντέλο kNN classification ή regression, θέτοντας την υπερπαράμετρο k σε μια τιμή της επιλογής σας. Υπολογίστε εκ νέου τις αντίστοιχες μετρικές αξιολόγησης και συγκρίνετέ τις με αυτές του Ερωτήματος 1).

Με βάση τις μετρικές, το φαινόμενο φαίνεται να περιγράφεται καλύτερα από το γραμμικό ή το μη γραμμικό μοντέλο;