

## 16 Πρόβλεψη Downlink Bitrate (Throughput) Με Βάση Μετρικές Ποιότητας

Παρέχεται ένα σύνολο δεδομένων από κινητή συσκευή (κινητό τηλέφωνο) που “κατεβάζει” δεδομένα, συνδεδεμένη σε ασύρματο δίκτυο: <https://www.kaggle.com/datasets/aeryss/lte-dataset/data>. Οι μεταβλητές περιλαμβάνουν τα downlink και uplink bitrates (ρυθμός μετάδοσης δεδομένων), δείκτες ποιότητας της σύνδεσης στο δίκτυο καθώς και δείκτες σχετικούς με την κίνηση της συσκευής. Η συχνότητα μέτρησης των δεδομένων (granularity) είναι 1s, και το dataset περιέχει μετρήσεις για ακίνητη συσκευή καθώς και κινούμενη συσκευή σε διάφορες συνθήκες (περπάτημα, αυτοκίνητο, τρένο, λεωφορείο). Τα ζητούμενα είναι τα εξής:

1. Εξηγήστε την σημασία της πρόβλεψης του throughput (downlink bitrate) των συσκευών στα σύγχρονα δίκτυα 5G.
2. Εκτελέστε την απαραίτητη προεπεξεργασία των δεδομένων (αφαίρεση null τιμών, κανονικοποίηση κ.λπ.) καθώς και σύνδεση των διαφόρων traces για κάθε σενάριο κίνησης σε ένα ενιαίο σύνολο δεδομένων, το οποίο θα χωρίσετε σε train και test datasets επιλογής σας.
3. Εκπαιδεύστε ένα απλό μοντέλο μηχανικής μάθησης για παλινδρόμηση (regression) όπως π.χ. XGBoost regressor, για πρόβλεψη (forecasting) του downlink bitrate με βάση τις προηγούμενες τιμές των μεταβλητών με βάση τα training και testing datasets που έχετε επιλέξει. Υπολογίστε κατάλληλες μετρικές για την αξιολόγηση του μοντέλου, όπως root mean squared error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE).
4. Χρησιμοποιήστε τεχνική explainable AI (XAI) της επιλογής σας (π.χ. XGBoost interpreter, SHAP) για να εντοπίσετε ποιες μεταβλητές επηρεάζουν περισσότερο τις προβλέψεις του μοντέλου. Παρουσιάστε τα αποτελέσματα.
5. Εκπαιδεύστε ένα μοντέλο deep learning για time series forecasting (π.χ. LSTM model) αρχιτεκτονικής της επιλογής σας που θα προβλέπει το downlink bitrate την επόμενη χρονική στιγμή, με βάση τις τιμές των μεταβλητών τις προηγούμενες χρονικές στιγμές (επιλέξτε πόσα προηγούμενα δείγματα θα χρησιμοποιείτε για την πρόβλεψη, π.χ. 64 δείγματα). Διαχειριστείτε το dataset κατάλληλα ώστε να δημιουργήσετε δεδομένα κατάλληλα για training και testing του μοντέλου (δημιουργείτε time-series windows). Υπολογίστε τις απαραίτητες μετρικές αξιολόγησης, παρουσιάστε τα αποτελέσματα και συγκρίνετε με το προηγούμενο μοντέλο.

**Bonus:** Με βάση τα αποτελέσματα της τεχνικής XAI ή άλλων τεχνικών της επιλογής σας, αφαιρέστε σταδιακά μεταβλητές από το training dataset, οι οποίες επηρεάζουν λιγότερο την πρόβλεψη των μοντέλων. Υπολογίστε τις μετρικές και αξιολόγησης και δείξτε πως επηρεάζεται η απόδοση των μοντέλων ανάλογα με τις μεταβλητές που δεν έχουν ληφθεί υπόψη.

**Τελικά παραδοτέα:** Παρουσίαση όλων των παραπάνω σε μορφή PowerPoint, τελική αναλυτική αναφορά καθώς και πηγαίος κώδικας. Για τις παραπάνω εργασίες προτείνεται υλοποίηση στις γλώσσες προγραμματισμού Python ή R. Για την υλοποίηση των απλών μοντέλων machine learning προτείνεται χρήση της βιβλιοθήκης Scikit Learn ενώ για τα deep learning μοντέλα προτείνονται οι βιβλιοθήκες Keras-TensorFlow ή PyTorch.

Υπεύθυνος εκπαιδευτής – σύμβουλος: Παναγιώτης Μαράντης ([pmaradis@isi.gr](mailto:pmaradis@isi.gr)).