



ΠΑΝΕΠΙΣΤΗΜΙΟ
ΠΑΤΡΩΝ
UNIVERSITY OF PATRAS

Τμήμα Ηλεκτρολόγων Μηχανικών
& Τεχνολογίας Υπολογιστών

Εξατομικευμένα Βιοϊατρικά και Τηλεϊατρικά Συστήματα

Ατομική Προγραμματιστική Εργασία

Ημερομηνία παράδοσης: **13/01/2023**

Σκοπός: Στο πλαίσιο της εργασίας αυτής, θα υλοποιήσετε βασικούς αλγόριθμους μηχανικής μάθησης και μεθόδους ερμηνευσιμότητας με τη χρήση της γλώσσας προγραμματισμού python για τη διάγνωση της νόσου του Πάρκινσον από ηχητικές καταγραφές ομιλίας. Η νόσος του Πάρκινσον είναι μία προοδευτική νευροεκφυλιστική διαταραχή που χαρακτηρίζεται από κινητικές και μη κινητικές εκδηλώσεις. Ο επιπολασμός της νόσου αφορά στις 93 περιπτώσεις ανά 100.000 σε ηλικίες 70-79 ετών. Η σημασία της έγκαιρης διάγνωσης της νόσου έγκειται στο ότι μπορεί να συμβάλει στην καλύτερη διαχείρισή της και ακόμα και στην επιβράδυνσή της.

Περιγραφή: Με τη χρήση της γλώσσας προγραμματισμού python:

- Να μεταφορτώσετε το σύνολο δεδομένων Parkinson's Data Set από τον ιστότοπο του UCI Machine Learning Repository (<https://archive.ics.uci.edu/ml/datasets/Parkinsons>) σε ένα αρχείο csv. Σημείωση: η πρώτη στήλη των δεδομένων δεν είναι χρήσιμη για την ταξινόμηση των στιγμιότυπων, καθώς αποτελεί ένα id για κάθε περιστατικό, οπότε μπορείτε να την αγνοήσετε.
- Να κανονικοποιήσετε κατάλληλα τα χαρακτηριστικά του συνόλου δεδομένων πριν τα εισάγετε ως είσοδο στους αλγόριθμους μηχανικής μάθησης.
- Να ταξινομήσετε τα στιγμιότυπα του συνόλου δεδομένων σε υγιή (στήλη status 0) και σε άτομα με τη νόσο του Πάρκινσον (στήλη status 1) με τη χρήση των ακόλουθων ταξινομητών:
 - Πολυεπίπεδο Νευρωνικό Δίκτυο Πρόσθιας Τροφοδότησης δύο ή περισσότερων κρυφών επιπέδων (Multilayer Perceptron)
 - Βαθύ Νευρωνικό Δίκτυο αρχιτεκτονικής που θα επιλέξετε εσείς από τη διεθνή βιβλιογραφία (deep learning algorithm for tabular data)
 - Ταξινομητή K-Κοντινότερων Γειτόνων (K-Nearest Neighbor)
 - Μηχανή Διανυσμάτων Υποστήριξης (Support Vector Machines)

- Δένδρο Απόφασης (Decision Tree)
- Τυχαίο Δάσος (Random Forest)
- Για κάθε ταξινομητή να πειραματιστείτε με τουλάχιστον δύο διαφορετικούς συνδυασμούς των υπερπαραμέτρων (hyperparameter tuning/ gridsearch) σχεδίασής του. Σημείωση: Ο πειραματισμός με τις υπερπαραμέτρους ενός ταξινομητή πραγματοποιείται πάντα στο σύνολο εκπαίδευσης (training set) και όχι στο σύνολο ελέγχου (test set).
- Να αξιολογήσετε τους ταξινομητές ανά άτομο (subject) με τη μέθοδο Leave-One-Subject-Out με βάση τα κριτήρια της Ισορροπημένης Ακρίβειας (Balanced Accuracy), Ευαισθησίας (Sensitivity), Ειδικότητας (Specificity), AUC και F1 score.
- Να χρησιμοποιήσετε τη μέθοδο ερμηνευσιμότητας Shapley για να καταγράψετε τα πιο σημαντικά χαρακτηριστικά της εισόδου στην πρόβλεψη για ένα υγιές άτομο και για ένα άτομο με τη νόσο του Πάρκινσον με κάποιον από τους ταξινομητές που είναι μη εγγενώς ερμηνεύσιμοι (interpretable) από αυτούς που υλοποιήσατε.

Παραδοτέο: Το παραδοτέο της συγκεκριμένης εργασίας θα είναι ένας συμπιεσμένος φάκελος, τον οποίο θα ανεβάσετε στο e-class, ο οποίος θα περιέχει τα ακόλουθα αρχεία:

- Μία γραπτή αναφορά στην οποία θα περιγράψετε συνοπτικά το σύνολο δεδομένων που χρησιμοποιήσατε και τα βήματα υλοποίησης που ακολουθήσατε. Θα παραθέσετε, επίσης, σε έναν πίνακα τις τιμές των βέλτιστων υπερπαραμέτρων και τις τιμές των κριτηρίων αξιολόγησης (Ισορροπημένη Ακρίβεια, Ευαισθησία, Ειδικότητα, AUC, F1 score) για κάθε ταξινομητή που υλοποιήσατε και θα αιτιολογήσετε την επιλογή του καλύτερου ταξινομητή από αυτούς που υλοποιήθηκαν με βάση αυτά τα αποτελέσματα. Ακόμα θα παραθέσετε το διάγραμμα με τα πιο σημαντικά χαρακτηριστικά από τη μέθοδο ερμηνευσιμότητας Shapley για ένα υγιές άτομο και για ένα άτομο με τη νόσο του Πάρκινσον με κάποιον από τους ταξινομητές που υλοποιήσατε.
- Αρχεία με τον κώδικα python για την υλοποίηση των ερωτημάτων της εργασίας. Στη συγγραφή του κώδικα καλό θα ήταν να συμπεριλάβετε σχόλια με λατινικούς χαρακτήρες, τα οποία να εξηγούν τη λειτουργία των βασικών του μερών.

Καλή επιτυχία!