

## Lecture 4: Instance-based Learning

### Project 3

Να φορτώσετε το συνημμένο dataset **diabetes** και να εφαρμόσετε τη μέθοδο **K-Nearest Neighbors** (συνιστάται η χρήση της βιβλιοθήκης **sci-kit** για Python).

Κατά τη διάρκεια της εφαρμογής του αλγορίθμου, να γίνουν πειράματα τα οποία θα αντιστοιχούν στο συνδυασμό παραμέτρων που βρίσκονται στο αρχείο **KNN\_Results.xlsx**, για πλήθος γειτόνων  $k=1:200$ .

Για κάθε πείραμα, να αξιολογηθούν οι συνδυασμοί των παραμέτρων και να παραδοθούν τα αποτελέσματα που θα αντιστοιχούν στην περίπτωση με τη μέγιστη τιμή της μετρικής **F1**. Επίσης, μαζί με τα αποτελέσματα, να παραδοθεί και το πλήθος γειτόνων  $k$  για το οποίο παίρνει τη μέγιστη τιμή της η **F1**.

Επιπλέον, να δημιουργηθούν και να παραδοθούν τα γραφήματα του κάθε πειράματος, τα οποία θα περιέχουν στον άξονα  $x$  το (αυξανόμενο) πλήθος γειτόνων, και στον άξονα  $y$  τις τιμές των μετρικών **Recall**, **Precision** και **F1**.

Η συμπλήρωση των αποτελεσμάτων να γίνει στο αρχείο **KNN\_Results.xlsx**.

**Σημείωση 1:** Κατά την υποβολή της εργασίας, είναι απαραίτητο να συμπεριληφθεί και ο κώδικας που χρησιμοποιήθηκε, πέραν των αποτελεσμάτων και των γραφημάτων. Για διευκόλυνση, το αρχείο **KNN\_Template.py** πάνω στο οποίο μπορεί να συμπληρωθεί ο απαραίτητος κώδικας σε Python.