Ερμηνεία Μοντέλων Μηχανικής Μάθησης με Μεθόδους eXplainable AI (XAI) - Εφαρμογή σε πλατφόρμες e-Learning

Η κατηγοριοποίηση ενός εκπαιδευόμενου με βάση την απόδοσή του στα πλαίσια ενός μαθήματος ή μίας σειράς μαθημάτων αποτελεί ένα από τα πλέον σημαντικά ζητήματα που μπορούν να απασχολήσουν έναν εκπαιδευτικό. Η συγκεκριμένη κατηγοριοποίηση είναι εξαιρετικά βοηθητική για τους εκπαιδευτικούς, καθώς μπορούν έγκαιρα να προβούν σε στοχευμένες ενέργειες εφόσον αυτό χρειαστεί.

Για την κατηγοριοποίηση ενός εκπαιδευόμενου έχουν υλοποιηθεί αρκετοί μηχανισμοί βασισμένοι σε διάφορες μεθόδους μηχανικής μάθησης και σε διαφορετικές πλατφόρμες e-learning. Αρκετές από αυτές τις υλοποιήσεις, βασισμένες σε αρκετά διαδεδομένες μεθόδους, φαίνεται να είναι αποτελεσματικές.

Ωστόσο, τα μοντέλα machine learning αντιμετωπίζονται, συνήθως, από τους διαχειριστές αλλά και από τους εκπαιδευτικούς ως black boxes, δηλαδή οι διαχειριστές αδυνατούν να κατανοήσουν την ακριβή λειτουργία τους και, κυρίως, πώς προκύπτουν τα συγκεκριμένα αποτελέσματα [1].

Για το σκοπό έχουν αναπτυχθεί οι τεχνικές eXplainable Artificial Intelligence (XAI) [2] που στοχεύουν στην κατανόηση των μοντέλων machine learning. Οι τεχνικές αυτές εφαρμόζονται είτε για τη συνολική κατανόηση των μοντέλων, δηλαδή πώς οι παράμετροι των μοντέλων επηρεάζουν τις αποφάσεις τους (global explainability) ή την κατανόηση του πώς λαμβάνονται οι αποφάσεις ταξινόμησης (classification) για συγκεκριμένες εισόδους (local explainability) [3]. Παραδείγματα μεθόδων XAI είναι οι LIME, SHAP και LEMNA [3, 4].

Η διπλωματική θα διερευνήσει μεθόδους XAI για την κατανόηση μοντέλων machine, είτε supervised ή unsupervised learning, που εφαρμόζονται για την κατηγοριοποίηση των εκπαιδευόμενων καθώς και την ενδεχόμενη πρόβλεψη της πορείας τους με βάση αυτή. Σκοπός της διπλωματικής θα είναι να διερευνήσει τη συνεισφορά διαφορετικών τύπων features και να συγκρίνει τις διάφορες μεθόδους XAI. Για την πειραματική αξιολόγηση των παραπάνω μοντέλων θα χρησιμοποιηθούν διαθέσιμα σύνολα δεδομένων από πλατφόρμες e-learning.

[1] D. Pantazatos, A. Trilivas, K. Meli, D. Kotsifakos and C. Douligeris, “Machine Learning and Explainable Artificial Intelligence in Education and Training-Status and Trends”, In Maglaras, L.A., Douligeris, C. (eds) Wireless Internet. WiCON 2023. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol 527. Springer, Cham.
[2] Explainable Artificial Intelligence, https://en.wikipedia.org/wiki/Explainable_artificial_intelligence
[3] Interpretable Machine Learning, https://christophm.github.io/interpretable-ml-book/
[4] SHAP, https://github.com/slundberg/shap