



ΠΑΝΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΑΤΤΙΚΗΣ

ΣΧΟΛΗ ΜΗΧΑΝΙΚΩΝ

ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΒΙΟΪΑΤΡΙΚΗΣ

Τίτλος εκπονούμενης διατριβής:

**Πρόβλεψη της ακαδημαϊκής επίδοσης φοιτητών με τη χρήση αλγορίθμων
μηχανικής μάθησης**

1^η Έκθεση προόδου

1/12/2023 - 30/11/2024

Ονοματεπώνυμο Υποψήφιου Διδάκτορα: **ΣΓΟΥΡΑΚΗ ΓΕΩΡΓΙΑ**

Αριθμός Μητρώου: **2303**

Επιβλέπουσα:

ΣΚΟΥΡΟΛΙΑΚΟΥ ΑΙΚΑΤΕΡΙΝΗ

Μέλη Συμβουλευτικής Επιτροπής:

**ΚΩΣΤΟΠΟΥΛΟΣ ΣΠΥΡΙΔΩΝ
ΚΑΒΟΥΡΑΣ ΔΙΟΝΥΣΙΟΣ**

Το αντικείμενο της παρούσης διδακτορικής διατριβής είναι να αναγνωριστεί η ακαδημαϊκή επίδοση των προπτυχιακών φοιτητών, δηλαδή η πρόοδος, τα επιτεύγματα και γενικότερα η συνολική επιτυχία στη μάθηση ενός φοιτητή. Με τον όρο της ακαδημαϊκής επίδοσης δε μετράμε μόνο τα επιτεύγματα των φοιτητών, αλλά επίσης και την ποιότητα εκπαίδευσης του εκπαιδευτικού ιδρύματος φοίτησης, η οποία είναι ένας από τους πιο σημαντικούς δείκτες ποιότητας για κάθε πανεπιστημιακό ίδρυμα. Σκοπός είναι η διαρκής βελτίωση της μάθησης και η αναβάθμιση της ποιότητας της προσφερόμενης εκπαίδευσης.

Κατά το πρώτο έτος πραγματοποιήθηκε μια βιβλιογραφική ανασκόπηση μελετών, στις οποίες έχουν δημιουργηθεί διάφορα μοντέλα πρόβλεψης που έχουν ως στόχο να βοηθήσουν τα ανώτατα εκπαιδευτικά ιδρύματα να επιτύχουν το σκοπό τους. Είναι σημαντικό για τα πανεπιστήμια να προσπαθήσουν να εξασφαλίσουν σε πρώιμο στάδιο, ότι οι φοιτητές δε θα σπαταλήσουν χρόνο και δε θα χάσουν το ενδιαφέρον τους για τις σπουδές τους, βελτιώνοντας τη μάθηση και αναβαθμίζοντας την ποιότητα της προσφερόμενης εκπαίδευσης με την εφαρμογή κατάλληλων στρατηγικών σχεδίων.

Κατά τη διάρκεια της ανασκόπησης, εντοπίστηκαν περισσότερες μελέτες να ασχολούνται με την πρόβλεψη της ακαδημαϊκής επίδοσης φοιτητών [1] [2] [3] [4] [5] [6], ενώ υπάρχουν μελέτες που ασχολούνται με το χρόνο αποφοίτησης [7] [8] [9] [10] χρησιμοποιώντας ως εργαλείο την εξόρυξη (data mining) ακαδημαϊκών ή/και προ-ακαδημαϊκών, κοινωνικών, δημογραφικών, οικονομικών δεδομένων.

Σχετικά με τις μελέτες που προβλέπουν την ακαδημαϊκή επίδοση αξίζει να σημειωθεί ότι στις τρεις πρώτες μελέτες εφαρμόζονται 6 διαφορετικοί αλγόριθμοι μηχανικής μάθησης. Στην 1^η [1] μελέτη οι Random Forest(RF), Nearest Neighbor(NN), Support Vector Machines(SVM), Logistic Regression(LR), Naïve Bayes(NB) και K-Nearest Neighbour (KNN) και αξιολογούνται όλοι οι αλγόριθμοι ως ακριβείς. Στη 2^η [2] οι Logistic Regression(LR), Linear Discriminant Analysis(LDA), K-Nearest Neighbour(KNN), Decision Trees(DT), Naïve Bayes(NB) και Support Vector Machines (SVM) και στην 3^η [3] οι Random Forest(RF), Decision Tree(DT), Support Vector Machine(SVM), Naïve Bayes(NB), Bagged Trees(BT) και Adapting Boosting Trees(ABT), ενώ αξιολογείται μόνο ένας που είναι πιο ακριβής ο Support Vector Machines (SVM) και ο Random Forest (RF) αντίστοιχα. Στην 4^η [4] μελέτη εφαρμόζεται ο Random Forest (RF) παρουσιάζοντας ικανοποιητικά αποτελέσματα ως προς την πρόβλεψη. Στην 5^η [5] εφαρμόζονται οι αλγόριθμοι Decision Tree(DT) και Fuzzy Genetic Algorithm(FGA), ενώ στην 6^η [6] ένα γραφικό μοντέλο μηχανικής μάθησης το Bayesian Belief Network (BBN) οι οποίοι εξάγουν τα απαιτούμενα από τη μελέτη αποτελέσματα. Σε όλες τις μελέτες χρησιμοποιούνται μόνο παράμετροι ακαδημαϊκών δεδομένων με εξαίρεση την 3^η μελέτη που χρησιμοποιεί και κοινωνικά, δημογραφικά και οικονομικά δεδομένα.

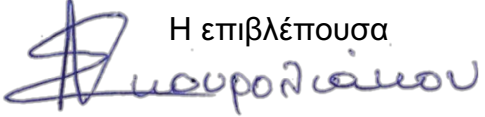
Σχετικά με τις μελέτες που προβλέπουν το χρόνο αποφοίτησης επισημαίνεται ότι επίσης εφαρμόζονται αλγόριθμοι μηχανικής μάθησης και αξιολογείται ο πιο ακριβής. Στην 1^η [7] -οι Random Forest (RF), Neural Networks (NN), Logistic Regression (LR), Naïve Bayes (NB), K-Nearest Neighbor (kNN) - και στην 3^η [9] - οι Decision Tree(DT), Random Forest(RF), Naïve Bayes(NB), Support Vector Machine PolyKernel(SVM PolyKernel) και Support Vector Machine RBFKernel(SVM RBFKernel) - εφαρμόζονται 5 δηλαδή αλγόριθμοι και κρίνονται ως αποτελεσματικοί ο RF και ο NN, και ο SVM PolyKernel αντίστοιχα. Στη 2^η [8] εφαρμόζονται 3 μοντέλα πρόβλεψης, τα Support Vector Machines (SVMs), Gaussian Processes (GPs) και Deep Boltzmann Machines (DBMs), με πιο αποτελεσματικό το DBM. Στην 4^η [10] εφαρμόζεται ο αλγόριθμος Intelligent K-Methods που η απόδοσή του μπορεί να χρησιμοποιηθεί ως σημείο αναφοράς. Τέλος, χρησιμοποιούνται και στις 3 πρώτες μελέτες παράμετροι ακαδημαϊκών και δημογραφικών δεδομένων με εξαίρεση την 4^η μελέτη που χρησιμοποιεί μόνο ακαδημαϊκά δεδομένα.

Στους αλγόριθμους αυτούς εφαρμόζονται διαφορετικές ταξινομήσεις που αποσκοπούν σε συγκεκριμένη πρόβλεψη ακρίβειας (accuracy) η οποία και αξιολογείται σε κάθε μελέτη για την εξαγωγή όσο το δυνατόν ασφαλούς συμπερασμάτων. Για παράδειγμα, χρησιμοποιείται η μήτρα σύγχυσης (confusion matrix) [1] [7] σε διάφορους αλγορίθμους για ακρίβεια ταξινόμησης (Classification Accuracy) προσπαθώντας να προβλεφθεί ο τελικός βαθμός εξετάσεων και ο τελικός χρόνος αποφοίτησης.

Στόχος αυτής της διατριβής είναι να χρησιμοποιηθούν και άλλοι αλγόριθμοι μηχανικής μάθησης που σε συνδυασμό με τις απαραίτητες υπολογιστικές τεχνικές μάθησης και εξόρυξης δεδομένων, θα γίνει προσπάθεια να δημιουργηθεί ένα όσο το δυνατόν πιο ακριβές μοντέλο πρόβλεψης της ακαδημαϊκής επίδοσης. Τα δεδομένα (παράμετροι) που θα εισαχθούν δύναται να είναι δημογραφικά όπως η ηλικία και το φύλο, κοινωνικά όπως είναι η οικονομική κατάσταση, καθώς και εκπαιδευτικά προ πανεπιστημίου όπως είναι ο βαθμός απολυτηρίου και η σειρά προτίμησης, και εκπαιδευτικά κατά τη διάρκεια φοίτησης στο πανεπιστήμιο όπως είναι ο αριθμός των προσπαθειών για να περάσουν τα μαθήματα. Ο τελικός σκοπός είναι η υποστήριξη ολόκληρης της διαδικασίας μάθησης σε ένα εκπαιδευτικό ίδρυμα.

Βιβλιογραφία:

- [1] Mustafa Yağcı, “Educational data mining: prediction of students’ academic performance using machine learning algorithms”, <https://doi.org/10.1186/s40561-022-00192-z>, 2022
- [2] Jovana Jović, Emilija Kisić, Miroslava Raspopović Milić, Dragan Domazet, Kavitha Chandra “Prediction of student academic performance using machine learning algorithms”, Proceedings 13th International Conference on eLearning, 2022
- [3] V.L. Miguéisa, Ana Freitasb, Paulo J.V. Garciab, André Silvab, “Early segmentation of students according to their academic performance: A predictive modelling approach”, <https://doi.org/10.1016/j.dss.2018.09.001>, 2018
- [4] M. Nachouki and M. Abou Naaj, “Predicting student performance to improve academic advising using the random forest algorithm,” Int. J. Distance Educ. Technol., vol. 20, no. 1, pp. 1–17, Jan.2022.
- [5] Hashmia Hamsa, Simi Indiradevi, Jubilant J. Kitzhakethottam, “Student academic performance prediction model using decision tree and fuzzy genetic algorithm”, <https://doi.org/10.1016/j.prtcy.2016.08.114>, 2016
- [6] Ahmad Slim, Gregory L. Heileman, Jarred Kozlick and Chaouki T. Abdallah, “Predicting Student Success Based on Prior Performance”, Department of Electrical and Computer Engineering University of New Mexico, 2014
- [7] Agus Santoso, Heri Retnawati, Kartianom, Ezi Apino, Ibnu Rafi, Munaya Nikma Rosyada, “Predicting Time to Graduation of Open University Students: An Educational Data Mining Study”, 2024
- [8] Tushar Ojha, Gregory L. Heileman, Manel Martinez-Ramon, Ahmad Slim “Prediction of graduation delay based on student performance”, IEEE, 2017
- [9] Nurafifah Mohammad Suhaimi, Shuzlina Abdul-Rahman¹, Sofianita Mutalib, Nurzeatul Hamimah Abdul Hamid, and Ariff Md Ab Malik “Predictive Model of Graduate-On-Time Using Machine Learning Algorithms”, https://doi.org/10.1007/978-981-15-0399-3_11, pp. 130–141, 2019
- [10] Leonardo Cahaya, Lely Hiryanto, Teny Handhayani, “Student Graduation Time Prediction Using Intelligent K-Medoids Algorithm”, Faculty of Information Technology Tarumanagara University, 2017

 Η επιβλέπουσα
Α. Σκουρολιάκου