

**UNIVERSITETI I PRISHTINËS**  
**FAKULTETI I SHKENCAVE MATEMATIKE - NATYRORE**  
**DEPARTAMENTI I MATEMATIKËS**  
**PROGRAMI: SHKENCË KOMPJUTERIKE**



# Teza master

*Analizimi dhe parashikimi i ndikimit të faktorëve të ndryshëm në Performancën Akademike të Studentëve*

**Mentori:**

*Prof. Asoc. Dr. Ermir Rogova*

**Kandidati:**

*Ilir Këçmari*

*Tetor 2024, Prishtinë*

## Abstrakti i zgjeruar

Parashikimi i Performancës Akademike të Studentëve (PAS) dhe identifikimi i faktorëve që ndikojnë ua mundësojnë stafit akademik dhe institucioneve relevante të ndërhyjnë herët për të përmirësuar rezultatet dhe për të rritur cilësinë e studimeve. Numri i studentëve, veçanërisht në universitetet publike, siç raportojnë Agjencia e Statistikave të Kosovës (ASK) dhe Instituti i Statistikave të Shqipërisë (INSTAT) në raportet e tyre të vazhdueshme për numrin e studentëve, ka pësuar rënie vazhdimisht viteve të fundit në Kosovë dhe Shqipëri përkatësisht. Prandaj, është thelbësore të parashikohet PAS-ja, si dhe të identifikohen faktorët kryesorë që ndikojnë në të, me qëllim që të zhvillohen politika dhe strategji që përmirësojnë cilësinë e përgjithshme të arsimit.

Ky hulumtim synon të identifikojë faktorët demografikë, socio-ekonomikë, psikologjikë dhe akademikë që ndikojnë në PAS dhe të përdorë algoritme të të Mësuarit të Makinës (ang. Machine Learning [ML]) për të parashikuar atë në dy raste të studimit: atë të Kosovës dhe atë të Shqipërisë. Përdorimi i këtyre teknikave ka treguar rezultate premtuese në literaturën ekzistuese, ku faktorë të ndryshëm dhe metoda të ndryshme të ML-së janë përdorur për të bërë parashikime të PAS-së.

Metodologjia e ndjekur nga ky hulumtim është Procesi Standard Ndëri-industrial për Mihjen e të Dhënave (ang. Cross-Industry Standard Process for Data Mining [CRISP-DM]) për mbledhjen, përpunimin, analizimin e të dhënave dhe krijimin dhe vlerësimin e modeleve parashikuese. Fillimisht, u analizua gjendja në terren dhe realizuam një shqyrtim të gjerë të literaturës për të identifikuar faktorët që ndikojnë në PAS dhe algoritmet më të përshtatshme të ML-së për parashikimin e saj. Më pas, përmes pyetësorëve të strukturuar, u mblodhën të dhëna nga 943 studentë (497 nga Kosova dhe 446 nga Shqipëria). Pas paraprosimit, për analizë u përdorën 469 përgjigje nga Kosova dhe 413 nga Shqipëria.

Procesi i paraprosimit përfshinte eliminimin e të dhënave të pavlefshme, trajtimin e anomalive dhe standardizimin e të dhënave të ndryshore të grupit të të dhënave. Të dhënat u analizuan me Testin Kruskal-Wallis (TKW) dhe përmes korrelacionit, më pas u aplikuan modele të ML-së, ku përfshihen: Modelet lineare të përgjithësuara (ang. General Linear Models - GLM), Mësimi i thellë (ang. Deep Learning-DL), Trungjet e Vendimmarrjes (ang. Decision Tree [DT]), Pyllëzimi i Rëndomtë (ang. Random Forest [RF]), Trungjet e Forcimit të Gradientit (ang. Gradient Boosting Trees [GBT]) dhe Makina Mbështetëse e Vektorit (ang. Support Vector Machine [SVM]). Përzgjedhja e algoritmeve u bazua në rekomandimet e literaturës dhe në veglën Mjeti i Optimizimit

të Pipeline-ve të Bazuara në Trungje (ang. Tree-based Pipeline Optimization Tool [TPOT]), e cila automatikisht identifikon modelet më të përshtatshme për një grup të caktuar të dhënash.

Analiza tregoi se disa faktorë kanë ndikim të rëndësishëm në PAS, duke përfshirë zonën e banimit para studimeve, suksesin në shkollën e mesme, nivelin arsimor të prindërve, notën në lëndën më të rëndësishme të programit të studimeve, numrin maksimal të rihyrjeve në një provim për të marrë notë kaluese, qasjen e mësimdhënësve, vetëbesimin dhe shumë faktorë të tjerë të diskutuar në hulumtim. Algoritmi GBM, i rekomanduar edhe nga TPOT-ja, performoi më mirë në të dhënat nga rasti i studimit në Kosovë, duke arritur një saktësi të lartë si para, ashtu edhe pas optimizimit të hipërparametrave. Pas optimizimit të hipërparametrave kishim MAE 0.334 dhe RMSE 0.398. Disa nga faktorët që u vlerësuan më së larti nga ky model ishin: sukcesi në shkollën e mesme, koha e marrë për të plotësuar pyetëtorin, perceptimi i studentëve mbi intensitetin e programit dhe nota në lëndën kryesore të studimeve. Ngjashëm ishte edhe në rastin e studimit në Shqipëri, vetëm se në këtë rast studimi algoritmi SVM dha rezultate më të sakta pas optimizimit të hipërparametrave me MAE 0.352 dhe RMSE 0.426. Disa nga faktorët që u vlerësuan më së larti nga ky model ishin: nota në lëndën kryesore, sukcesi në shkollën e mesme, qyteti i banimit para fillimit të studimeve, programi i studimeve, koha e marrë për të plotësuar pyetëtorin dhe besimi në aftësitë e tyre për të përfunduar studimet me sukses. Optimizimi i hipërparametrave luajti rol kyç në rritjen e saktësisë së modeleve të aplikuara, duke përmirësuar parashikimet për PAS-në.

Ky hulumtim ka identifikuar faktorët kryesorë demografikë, socio-ekonomikë, psikologjikë dhe akademikë që ndikojnë në PAS në rastet e studimit në Kosovë dhe Shqipëri. Duke përdorur algoritme të ML-së dhe duke optimizuar hipërparametrat, hulumtimi ka demonstruar se është e mundur të parashikohet me saktësi PAS-ja, duke ofruar kështu një mjet të vlefshëm për institucionet arsimore në zhvillimin e politikave dhe të strategjive që përmirësojnë cilësinë e përgjithshme të arsimit dhe mbështesin studentët në arritjen e rezultateve më të mira.

**Fjalët kyçe:** Performanca Akadematike e Studentëve, faktorët ndikues, Mësimi i Makinës, optimizimi i hipërparametrave.