

UNIVERSITETI I PRISHTINËS “HASAN PRISHTINA”

FAKULTETI I SHKENCAVE MATEMATIKE-NATYRORE

DEPARTAMENTI I MATEMATIKËS

PROGRAMI SHKENCË KOMPJUTERIKE



TEZA MASTER

**PARASHIKIMI I RREZIKUT TË MIGRIMIT TË
PROFESIONISTËVE NGA KOSOVA ME MODELE TË
INTERPRETUESHME DHE ME VLERËSIM TË PASIGURISË TË TË
MËSUARIT TË MAKINËS**

Mentori:

Prof. Ass. Dr. Eliot Bytyçi

Kandidati:

Enda Alidema

Maj 2026, Prishtinë

Abstrakt i zgjeruar

Migrimi i profesionistëve dhe i të rinjve të kualifikuar nga Kosova përbën një sfidë me ndikim të drejtpërdrejtë në zhvillimin ekonomik, në kapacitetet e tregut të punës dhe në qëndrueshmërinë afatgjatë të sektorëve strategjikë, si shëndetësia, arsimi dhe teknologjia. Humbja e kapitalit njerëzor nuk është thjesht çështje demografike, por lidhet me uljen e produktivitetit, dobësimin e shërbimeve publike dhe vështirësimin e ndërtimit të një ekonomie konkurruese. Përballë këtij realiteti, kuptimi i faktorëve që ndikojnë në prirjen për migrim dhe identifikimi i grupeve më të ekspozuara ndaj këtij rreziku janë të rëndësishme si nga aspekti akademik, ashtu edhe nga ai praktik dhe politikëbërës.

Ky punim trajton parashikimin e rrezikut të migrimit të profesionistëve dhe të rinjve të kualifikuar nga Kosova përmes modeleve të interpretueshme të të mësuarit të makinës, duke përfshirë edhe vlerësimin e pasigurisë së rezultateve. Punimi ndërthur dy nivele analitike: analizën makro të migrimit të dokumentuar dhe modelimin mikro të synimit për migrim në nivel individ. Në nivel makro, përdoren të dhënat e Eurostat-it mbi lejet e para të qëndrimit për shtetasit e Kosovës gjatë periudhës 2013–2024, të ndara sipas arsyes së lëshimit, kohëzgjatjes së vlefshmërisë dhe vendeve pritëse. Këto të dhëna shërbejnë për të ndërtuar kontekstin empirik të migrimit të rregullt dhe për të identifikuar trendet kryesore të lëvizjes së shtetasve të Kosovës drejt vendeve evropiane.

Megjithatë, të dhënat administrative nuk e shpjegojnë plotësisht vendimmarrjen individuale për migrim, sepse motivet reale shpesh lidhen me faktorë socio-ekonomikë, profesionalë dhe perceptime personale. Prandaj, analiza makro plotësohet me një bashkësi të dhënash të ndërtuar nga pyetësi me 310 të anketuar. Në këtë bashkësi të dhënash, variabla e varur tregon nëse individ planifikon të migrojë brenda tri viteve të ardhshme, ndërsa variablat shpjeguese përfshijnë karakteristika demografike, arsimore, profesionale, ekonomike dhe sociale, si dhe qëndrime e perceptime lidhur me punën, pagën, zhvillimin profesional dhe lidhjet jashtë vendit.

Nga pikëpamja metodologjike, problemi formulohet si klasifikim binar, ku modeli synon të parashikojë nëse individ i përket klasës “Po”, pra planifikon të migrojë, apo klasës “Jo”, pra nuk planifikon të migrojë brenda periudhës së përcaktuar. Për të shmangur rrjedhjen e informacionit, nga modelimi janë përjashtuar variablat që lidhen drejtpërdrejt me planin, arsyet, destinacionin ose përgatitjen konkrete për migrim, në mënyrë që vlerësimi i modeleve të jetë më realist. Në kuadër të eksperimenteve janë trajnuar dhe krahasuar dhjetë modele klasifikimi: Logistic Regression, Random Forest, Extra Trees, XGBoost, CatBoost, Support Vector Machine, Gradient Boosting, Decision Tree, K-Nearest Neighbors dhe Gaussian Naive Bayes. Përveç tyre, janë shqyrtuar edhe qasje plotësuese, si ansamblet me votim të peshuar (ang. weighted voting ensembles), modelet me njohuri paraprake, përfshirë BERT dhe TabPFN, si dhe konfigurime me regularizim.

Për shkak të pabarazisë ndërmjet klasave, vlerësimi i performancës nuk mbështetet vetëm në saktësinë e përgjithshme, por edhe në metrika më të përshtatshme për të dhëna të pabalancuara, si saktësia për secilën klasë, precizioni, F1-score, saktësia e balancuar dhe ROC-AUC. Për vlerësimin e pasigurisë së këtyre metrikave janë përdorur intervalet e Wilson-it, të cilat mundësojnë paraqitjen e rezultateve jo vetëm si vlera të vetme numerike, por edhe si intervale besueshmërie, veçanërisht në kushte të mostrës relativisht të kufizuar dhe të pabarazisë ndërmjet klasave. Po ashtu, janë testuar strategji të ndryshme të balancimit të klasave, duke

përfshirë peshimin e balancuar të klasave (ang. balanced class weight), mbikampionimin e rastësishëm (ang. random oversampling - ROS), nënkampionimin e rastësishëm (ang. random undersampling - RUS), teknikën sintetike të mbikampionimit të klasës pakicë (ang. Synthetic Minority Over-sampling Technique - SMOTE) dhe SMOTE pranë kufirit ndarës ndërmjet klasave (ang. Borderline-SMOTE). Krahas tyre, janë përdorur edhe kalibrimi i probabiliteteve dhe akordimi i pragut të vendimit, me qëllim përmirësimin e identifikimit të individëve me rrezik më të lartë migrimi.

Rezultatet tregojnë se modeli më i mirë përfundimtar është Gradient Boosting i kalibruar, i kombinuar me Borderline-SMOTE dhe prag të akorduar të vendimit. Ky konfigurim arrin saktësi të përgjithshme 0.823, saktësi 0.800 për klasën “Jo”, saktësi 0.882 për klasën “Po”, saktësi të balancuar 0.841 dhe F1-score 0.732. Këto rezultate tregojnë se modelet e të mësuarit të makinës mund të përdoren me sukses për parashikimin e rrezikut të migrimit në nivel individ, veçanërisht kur trajtohet pabarazia e klasave dhe kur pragu i vendimit përshtatet me qëllimin analitik të studimit.

Përveç performancës parashikuese, punimi i kushton rëndësi edhe interpretueshmërisë së modeleve. Për këtë qëllim janë përdorur rëndësia me permutim (ang. permutation importance), vlerat shpjeguese të Shapley-t (ang. SHAP values) dhe grafikët e varësisë së pjesshme (ang. Partial Dependence Plots - PDP). Këto metoda ndihmojnë në identifikimin e variablave që kontribuojnë më shumë në parashikimin e rrezikut të migrimit dhe e bëjnë modelin më transparent për interpretim akademik dhe politikëbërje. Gjetjet sugjerojnë se prirja për migrim nuk lidhet vetëm me kushtet ekonomike, por edhe me perceptimin për zhvillim profesional, stabilitetin në punë, mundësitë e karrierës dhe praninë e rrjeteve sociale jashtë vendit. Kjo tregon se migrimi i profesionistëve duhet kuptuar si fenomen shumëfaktorial, ku kushtet materiale ndërthuren me pritshmëritë profesionale dhe sociale të individëve.

Në këtë mënyrë, punimi kontribuon në literaturën mbi migrimin në Kosovë duke ofruar një qasje të integruar që kombinon analizën përshkruese të trendeve makro me modelimin parashikues në nivel mikro. Kontributi metodologjik qëndron në përdorimin e modeleve të interpretueshme të të mësuarit të makinës, në trajtimin e pabarazisë së klasave, në kalibrimin e probabiliteteve dhe në vlerësimin e pasigurisë së rezultateve përmes intervaleve të Wilson-it. Megjithatë bashkësia e të dhënave nga pyetësi është relativisht e kufizuar dhe mbështetet në synime të deklaruara për migrim, të cilat nuk përkthehen domosdoshmërisht në migrim real, rezultatet ofrojnë një bazë të dobishme për analiza të mëtejshme, validim me mostra më të mëdha dhe zhvillim të politikave më të targetuara ndaj grupeve me rrezik më të lartë migrimi. Në këtë kuptim, punimi paraqet një hap të rëndësishëm drejt përdorimit të metodave të avancuara, të interpretueshme dhe të besueshme të të mësuarit të makinës për analizimin e migrimit të profesionistëve nga Kosova.

Fjalë kyçe: migrimi i profesionistëve, të mësuarit e makinës, interpretueshmëria, vlerësimi i pasigurisë