

MASHINAVIY O‘RGANISH ALGORITMLARI YORDAMIDA OLIY TA’LIMDA TALABALAR AKADEMIK O‘ZLASHTIRISHINI BASHORATLASH

Abbosjon To‘ychiboyev Erali o‘g‘li

*Qo‘qon universiteti Raqamli texnologiyalar va
matematika kafedrası o‘qituvchisi
toychiboyevabbosbek@gmail.com*

Annotatsiya

Ushbu maqolada oliy ta‘lim muassasalarida talabalar akademik o‘zlashtirishini mashinaviy o‘rganish algoritmlari yordamida bashoratlash masalalari tadqiq etiladi. Tadqiqot obyekti sifatida talabalarning davomati oraliq baholari topshiriqlarni bajarish faolligi va o‘quv platformasidagi harakatlari kabi ko‘rsatkichlar tanlandi. An‘anaviy baholashda talabaning qiyinchiliklari ko‘pincha kech ya‘ni yakuniy nazorat bosqichida ma‘lum bo‘ladi. Mashinaviy o‘rganishga asoslangan bashoratlash modellari mavjud ma‘lumotlardan qonuniyatlarni o‘rganib talabaning kelgusi natijasini yoki xavf darajasini erta aniqlash imkonini beradi. Maqolada belgilarni tanlash model turlari baholash mezonlari (aniqlik precision recall F1-ko‘rsatkich) model talqini hamda ma‘lumotlar maxfiyligi va algoritmik adolat masalalari tahlil qilindi. Natijalar shuni ko‘rsatadiki bashoratlash tizimi xavf ostidagi talabalarni erta aniqlab o‘z vaqtida yordam ko‘rsatishga xizmat qiladi biroq u talabaga qat‘iy yorliq qo‘ymasdan o‘qituvchining qaror qabul qilishini qo‘llab-quvvatlovchi vosita sifatida qo‘llanilishi maqsadga muvofiqdir.

Kalit so‘zlar: *oliy ta‘lim mashinaviy o‘rganish ta‘lim ma‘lumotlarini qazib olish o‘quv analitikasi bashoratlash akademik o‘zlashtirish tasodifiy o‘rmon gradient busting neyron tarmoq erta ogohlantirish tizimi F1-ko‘rsatkich.*

Abstract

This article investigates the prediction of students' academic performance in higher education using machine learning algorithms. Indicators such as attendance midterm grades assignment completion activity and learning-platform behavior were selected as the research object. In traditional assessment a student's difficulties often become apparent late — at the final examination stage. Machine learning prediction models learn patterns from existing data and make it possible to identify a student's future result or risk level early. The article analyzes feature selection model types evaluation metrics (accuracy precision recall F1-score) model interpretability as well as data privacy and algorithmic fairness. The results show that a prediction system helps identify at-risk students early and provide timely support;

however rather than assigning a rigid label to a student it should serve as a tool supporting the instructor’s decision-making.

Keywords: *higher education machine learning educational data mining learning analytics prediction academic performance random forest gradient boosting neural network early warning system F1-score.*

Kirish

Talabalar akademik o‘zlashtirishi oliy ta’lim sifatining markaziy ko‘rsatkichlaridan biridir. Talabaning bilimini erta va aniq baholash ayniqsa qiyinchilikka uchragan talabalarni o‘z vaqtida aniqlash ta’lim natijalarini yaxshilash va talabalar tomonidan ta’lim tark etilishini kamaytirish imkonini beradi. Biroq an’anaviy baholash tizimida talabaning ortda qolayotgani ko‘pincha kech bilinadi va yordam ko‘rsatish uchun vaqt qoladi. Oliy ta’lim jarayonida katta hajmdagi ma’lumotlar to‘planadi: davomat oraliq va joriy baholar topshiriqlarni topshirish vaqti o‘quv platformasidagi faollik va boshqalar. Bu ma’lumotlar qo‘lda tahlil qilinganda ko‘pincha yetarlicha foydalanilmaydi. Mashinaviy o‘rganish esa ana shu ma’lumotlardan yashirin qonuniyatlarni topib talabaning natijasini bashoratlash imkonini beradi. Ta’lim ma’lumotlarini qazib olish (educational data mining) va o‘quv analitikasi (learning analytics) so‘nggi yillarda alohida ilmiy yo‘nalishga aylandi. Bu sohada bashoratlash modellari talabani “xavf ostida” yoki “barqaror” toifalarga ajratish yakuniy bahoni taxminlash yoki ta’lim tark etish ehtimolini baholash uchun qo‘llaniladi. Bunday tizimlar erta ogohlantirish tizimi (early warning system) sifatida amal qiladi.

Mazkur tadqiqotda mashinaviy o‘rganish asosida talabalar o‘zlashtirishini bashoratlashning ilmiy-amaliy asoslari o‘rganildi. Maqolaning asosiy g‘oyasi shundan iboratki bashoratlash modeli talabaga qat’iy hukm chiqaruvchi vosita emas balki o‘qituvchiga qaysi talabalarga e’tibor qaratish kerakligini ko‘rsatuvchi qaror qo‘llab-quvvatlash vositasidir. Shu bilan birga model ma’lumotlar maxfiyligi va adolat talablariga rioya qilishi shart. Tadqiqotning maqsadi — talabalar akademik o‘zlashtirishini bashoratlashda qo‘llaniladigan mashinaviy o‘rganish algoritmlarini o‘rganish ularning amaliy ahamiyatini asoslash hamda aniqlik talqin qilinish va axloqiy talablar o‘rtasidagi muvozanatni ko‘rsatishdan iborat.

Adabiyotlar tahlili

Mashinaviy o‘rganish nazariyasi T. Hastie va hammualliflarning hamda C. Bishopning klassik ishlarida tizimli yoritilgan. Ular nazorat ostida o‘rganish (supervised learning) regressiya va tasniflash masalalari model murakkabligi hamda haddan tashqari moslashish (overfitting) muammosini ilmiy asosda izohlaydi. Bu tushunchalar ta’lim ma’lumotlarini bashoratlashning nazariy poydevorini tashkil etadi. Tasniflash va regressiya masalalarida bir

qancha algoritmlar keng qo‘llaniladi. Logistik regressiya oddiy va talqin qilish oson bo‘lgani uchun bazaviy model sifatida ishlatiladi. Qaror daraxtlari qoidaga asoslangan tushunarli natija beradi biroq alohida olinganda barqaror emas. L. Breymanning tasodifiy o‘rmon (random forest) usuli ko‘plab daraxtlarni birlashtirib aniqlik va barqarorlikni oshiradi. Gradient busting yo‘nalishidagi modellar jumladan T. Chen va C. Guestrin taklif etgan XGBoost tabaqaviy (tabular) ma‘lumotlarda yuqori aniqlik bilan ajralib turadi. Ular ketma-ket qurilgan zaif modellarni birlashtirib murakkab qonuniyatlarni o‘rganadi. Neyron tarmoqlar esa katta hajmdagi ma‘lumotlarda kuchli natija beradi ammo talqin qilish qiyinligi va ko‘proq ma‘lumot talab etishi bilan farqlanadi. Ta‘lim sohasiga oid ishlarda C. Romero va S. Ventura ta‘lim ma‘lumotlarini qazib olishning holatini umumlashtirgan S. Kotsiantis esa talabalar baholarini bashoratlovchi qaror qo‘llab-quvvatlash tizimini taklif etgan. Boshqa tadqiqotlarda davomat joriy baholar va platformadagi faollik talaba xavfini erta aniqlashda eng ahamiyatli belgilar ekanligi qayd etiladi. Adabiyotlarda axloqiy va maxfiylik jihatlari alohida ta‘kidlanadi. A. Pardo va G. Siemens o‘quv analitikasi uchun maxfiylik va axloq tamoyillarini ishlab chiqqan. Bashoratlash modeli talabaga “muvaqqiyatsiz” degan qat’iy yorliq qo‘yishi ma‘lumotlarning noto‘g‘ri ishlatilishi yoki ayrim guruhlariga nisbatan adolatsizlik kabi xavflarni keltirib chiqarishi mumkin. Shu bois yakuniy qaror inson zimmasida qolishi zarur. Adabiyotlar tahlili shuni ko‘rsatadiki bashoratlash tizimi sifatini uchta mezon belgilaydi: bashorat aniqligi model talqin qilinishi (qaysi belgi qaror beradi) va axloqiy ishonchlilik (maxfiylik hamda adolat). Ushbu uch mezon muvozanati tizimning ta‘lim jarayonidagi amaliy qiymatini ta‘minlaydi.

Metodologiyasi

Tadqiqot metodologiyasi ma‘lumotlarni to‘plash tozalash va belgilarga aylantirish modelni o‘rgatish baholash hamda natijalarni erta ogohlantirish tizimiga bog‘lash bosqichlariga asoslandi. Tadqiqot obyekti sifatida talabalarining o‘quv jarayonidagi ko‘rsatkichlari kiruvchi ma‘lumotlar manbasi sifatida qaraldi. Metodologik yondashuvda har bir talaba belgilar vektori bilan ifodalandi. Belgilar sifatida davomat ulushi oraliq baholar topshiriqlarni o‘z vaqtida topshirish darajasi o‘quv platformasidagi faollik va o‘tgan davrlardagi natijalar olindi. Maqsad o‘zgaruvchisi sifatida talabaning kelgusi yakuniy natijasi yoki “xavf ostida / barqaror” toifasi belgilandi. Modellar nazorat ostida o‘rganish doirasida o‘rgatildi. Ma‘lumotlar o‘rgatish va test qismlariga ajratildi model haddan tashqari moslashmasligi uchun kross-validatsiya qo‘llanildi. Bashoratlash masalasi vazifaga qarab tasniflash (xavf toifasi) yoki regressiya (taxminiy ball) ko‘rinishida qo‘yildi. Modellar baholashda quyidagi mezonlar belgilandi: aniqlik (accuracy) precision recall F1-ko‘rsatkich va ROC-AUC. Ta‘limda xavf ostidagi talabani o‘tkazib yubormaslik muhim bo‘lgani uchun recall ko‘rsatkichiga alohida e‘tibor qaratildi. Keltirilgan sonli natijalar laboratoriya sharoitidagi

umumlashtirilgan baholash ko‘rinishida berildi va amaliyotda ma’lumotlar hajmi hamda sifatiga qarab o‘zgaradi.

1-jadval.

Akademik o‘zlashtirishni bashoratlash quvurining bosqichlari

Bosqich	Vazifa	Algoritmik vosita	Asosiy talab
1	Ma’lumot to‘plash	O‘quv platformasi baholar bazasi	Ishonchli manba
2	Tozalash va belgilar	Belgilarni shakllantirish	Sifatli belgilar
3	Modelni o‘rgatish	RF gradient busting NN	Overfittingdan saqlanish
4	Baholash	Accuracy F1 ROC-AUC	Recallga e’tibor
5	Talqin qilish	Belgi ahamiyati tahlili	Tushunarliklik
6	Erta ogohlantirish	Hisobot o‘qituvchi qarori	Adolat va maxfiylik

Natijalar

Talabalar o‘zlashtirishini bashoratlash bo‘yicha o‘tkazilgan baholash shuni ko‘rsatdiki mashinaviy o‘rganish modellari mavjud ma’lumotlardan qonuniyatlarni o‘rganib xavf ostidagi talabalarni erta aniqlay oldi. Oddiy modellar tez o‘rgatildi va talqin qilish oson bo‘ldi murakkabroq modellar esa aniqlikni oshirdi ammo tushunarliklik bilan resurs talabini birgalikda baholashni taqozo etdi. Logistik regressiya bazaviy model sifatida barqaror va talqin qilinadigan natija berdi: u qaysi belgilar natijaga ta’sir qilishini aniq ko‘rsatdi. Biroq murakkab chiziqsiz bog‘lanishlarni to‘liq qamrab ololmadi. Yakka qaror daraxti tushunarli ammo ma’lumotdagi kichik o‘zgarishlarga sezgir bo‘ldi. Tasodifiy o‘rmon ko‘plab daraxtlarni birlashtirgani uchun aniqlik va barqarorlikni oshirdi hamda belgilar ahamiyatini baholash imkonini berdi. Gradient busting (XGBoost kabi) tabaqaviy ma’lumotlarda eng yuqori aniqlik ko‘rsatkichlaridan birini berdi lekin parametrlarni ehtiyotkor sozlashni talab qildi. Neyron tarmoq katta ma’lumotda kuchli bo‘lsa-da kichik ma’lumotda ustunlik bermadi va talqini qiyin bo‘ldi. Belgilar tahlili davomat joriy baholar va topshiriqlarni o‘z vaqtida topshirish darajasi bashorat uchun eng ahamiyatli ko‘rsatkichlar ekanligini ko‘rsatdi. Bu natija o‘qituvchiga aynan qaysi ko‘rsatkichlarni kuzatish kerakligini aniqlashda amaliy yordam beradi.

2-jadval.

Mashinaviy o‘rganish algoritmlarining umumlashtirilgan qiyosi

Algoritm	Aniqlik	Talqin qilinishi	Ma'lumot talabi	Izoh
Logistik regressiya	O'rtacha	Yuqori	Past	Yaxshi bazaviy model
Qaror daraxti	O'rtacha	Yuqori	Past	Sezgir beqaror
Tasodifiy o'rmon	Yuqori	O'rtacha	O'rtacha	Barqaror ishonchli
Gradient busting	Juda yuqori	O'rtacha	O'rtacha	Sozlash talab etadi
Neyron tarmoq	Yuqori	Past	Yuqori	Katta ma'lumot uchun

Baholashda recall ko'rsatkichi alohida muhim bo'ldi chunki ta'limda xavf ostidagi talabani o'tkazib yuborish uni yordamsiz qoldiradi. Shu sababli model precision va recall o'rtasidagi muvozanatni ifodalovchi F1-ko'rsatkich asosida baholandi. Eng yuqori aniqlik emas balki amaliy foydali va adolatli model afzal ko'rildi.

Umumlashtirilgan laboratoriya baholashida tasodifiy o'rmon va gradient busting aniqlik bilan barqarorlik o'rtasida maqbul natija berdi; logistik regressiya esa talqin qilinishi muhim bo'lgan hollarda qulay bazaviy model bo'lib qoldi. Belgilar ahamiyati tahlili modelni "qora quti" emas balki tushunarli vositaga aylantirdi.

3-jadval.

Modellar ishlash xususiyatlari bo'yicha namunaviy tahlil

Model	O'rgatish vaqti	Bashorat tezligi	Taxminiy aniqlik	Baholash
Logistik regressiya	Juda tez	Juda tez	O'rtacha	Talqin uchun ideal
Tasodifiy o'rmon	O'rtacha	Tez	Yuqori	Amaliy uchun mos
Gradient busting	O'rtacha-uzoq	Tez	Juda yuqori	Eng aniq sozlanadi
Neyron tarmoq	Uzoq	O'rtacha	Yuqori	Katta ma'lumot kerak

Talabalar o‘zlashtirish ma’lumotlari misolida modellar xavf ostidagi talabalarni yakuniy nazoratdan ancha oldin aniqlashi va o‘qituvchiga aniq asoslangan signal berishi mumkinligi ko‘rsatildi. Bu natija bashoratlash tizimining ta’lim jarayonidagi amaliy ahamiyatini ko‘rsatadi ayni paytda har bir bashorat yakuniy hukm emas balki tekshirilishi lozim bo‘lgan signal ekanligini ta’kidlaydi.

Muhokama

Natijalar talabalar o‘zlashtirishini bashoratlash masalasi faqat texnik aniqlik emas balki talqin qilinish va axloq masalasi ham ekanligini ko‘rsatadi. Aniqligi yuqori ammo tushunarsiz model o‘qituvchining ishonchini qozonmasligi va noto‘g‘ri qarorlarga olib kelishi mumkin. Shu nuqtai nazardan modellarni ikki guruhga ajratish mumkin. Birinchi guruhga talqin qilinishi yuqori ammo aniqligi o‘rtacha modellar (logistik regressiya qaror daraxti) kiradi. Ikkinchi guruhga aniqligi yuqori lekin nisbatan murakkab modellar (tasodifiy o‘rmon gradient busting neyron tarmoq) kiradi. Amaliy tizimda aniqlik va tushunarlik o‘rtasida vazifaga mos tanlov qilinadi. Axloqiy nuqtai nazardan bir necha tamoyilga rioya qilish tavsiya etiladi. Birinchidan bashorat talabaga qat’iy yorliq qo‘yish uchun emas balki yordam ko‘rsatish uchun ishlatiladi. Ikkinchidan model turli guruhlariga nisbatan adolatli ishlashi tekshiriladi. Uchinchidan ma’lumotlar maxfiyligi ta’minlanadi va faqat belgilangan maqsadda foydalaniladi. Ilmiy jihatdan asosiy xulosa shundan iboratki bashoratlash modeli mustaqil hakam emas balki erta ogohlantirish va qaror qo‘llab-quvvatlash vositasidir. U xavf ostidagi talabalarni belgilaydi yakuniy qaror va yordam choralari esa o‘qituvchi hamda muassasa zimmasida qoladi. Shu sababli kelgusida tushunarli (talqin qilinadigan) va adolatli modellar istiqbolli hisoblanadi. Amaliy model Python va Scikit-learn muhitida laboratoriya ishi sifatida bajarilishi mumkin: ma’lumotlar tozalanadi belgilar shakllantiriladi model o‘rgatiladi va baholanadi belgilar ahamiyati tahlil qilinadi va xavf ostidagi talabalar ro‘yxati o‘qituvchiga taqdim etiladi. Bunday yondashuv ma’lumotlarga asoslangan ta’lim qarorlarini qo‘llab-quvvatlaydi.

Xulosa

Tadqiqot natijalariga ko‘ra mashinaviy o‘rganish algoritmlari oliy ta’limda talabalar akademik o‘zlashtirishini bashoratlashning samarali vositasi hisoblanadi. Ular mavjud ma’lumotlardan qonuniyatlarni o‘rganib xavf ostidagi talabalarni erta aniqlash va o‘z vaqtida yordam ko‘rsatish imkonini beradi. Biroq model tanlashda aniqlik bilan birga talqin qilinishi ham hisobga olinishi kerak. Talabalar o‘zlashtirish ma’lumotlari misolida logistik regressiya tasodifiy o‘rmon gradient busting va neyron tarmoq modellarining xususiyatlari tahlil qilindi. Natijalar tasodifiy o‘rmon va gradient busting aniqlik bilan barqarorlikda ustun logistik regressiya esa talqin qilinishi muhim bo‘lgan hollarda qulay ekanligini ko‘rsatdi. Shu sababli bashoratlash tizimida aniqlik talqin qilinish va axloqiy ishonchlilik muvozanati asosiy mezon

bo‘lishi kerak. Maqolada ko‘rib chiqilgan yondashuv talabalar ma’lumotlarini erta ogohlantirish tizimiga aylantirish xavf ostidagi talabalarni aniqlash va o‘qituvchiga asoslangan signal berishga qaratilgan. Bu yondashuv ta’lim ma’lumotlarini qazib olish va o‘quv analitikasi fanlarini amaliy ta’lim jarayoni bilan bog‘lashga xizmat qiladi. Kelgusida tadqiqotni chuqurlashtirish uchun haqiqiy o‘quv ma’lumotlari asosida turli modellarning precision recall F1-ko‘rsatkich va ROC-AUC qiymatlarini solishtirish hamda model adolatini baholash maqsadga muvofiq. Eng muhimi bunday tizimlar ma’lumotlar maxfiyligi va adolat talablariga rioya qilgan holda o‘qituvchining yordamchi vositasi sifatida joriy etilishi zarur.

Foydalanilgan adabiyotlar

1. Hastie T. Tibshirani R. Friedman J. The Elements of Statistical Learning. 2nd edition. Springer 2009.
2. Bishop C.M. Pattern Recognition and Machine Learning. Springer 2006.
3. Breiman L. Random Forests. Machine Learning 2001.
4. Chen T. Guestrin C. XGBoost: A Scalable Tree Boosting System. Proceedings of KDD 2016.
5. Goodfellow I. Bengio Y. Courville A. Deep Learning. MIT Press 2016.
6. Romero C. Ventura S. Educational Data Mining: A Review of the State of the Art. IEEE Transactions on Systems Man and Cybernetics 2010.
7. Baker R.S.J.d. Inventado P.S. Educational Data Mining and Learning Analytics. Springer 2014.
8. Kotsiantis S.B. Use of Machine Learning Techniques for Educational Purposes: A Decision Support System for Forecasting Students’ Grades. Artificial Intelligence Review 2012.
9. Marbouti F. Diefes-Dux H.A. Madhavan K. Models for Early Prediction of At-Risk Students in a Course Using Standards-Based Grading. Computers & Education 2016.
10. Pardo A. Siemens G. Ethical and Privacy Principles for Learning Analytics. British Journal of Educational Technology 2014.
11. Pedregosa F. va boshqalar. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research 2011.
12. European Union. General Data Protection Regulation (GDPR). Official Journal of the European Union 2016.