

O'QUVCHILARNING O'ZLASHTIRISH DARAJASINI BASHORATLASHDA RANDOM FOREST ALGORITMIDAN FOYDALANISH

Yunusxo'jayev Lutfullo Zafarxo'ja o'g'li

Magistratura talabasi

yunusxojayevlutfullo@gmail.com

Ilmiy rahbar: Jalolov Tursunbek Sadriddinovich

p.f.f.d. (PhD), dotsent Osiyo xalqaro universiteti

ANNOTATSIYA

Ushbu maqolada o'quvchilarning akademik o'zlashtirish darajasini oldindan bashoratlashda Random Forest mashinaviy o'qitish algoritmidan foydalanish imkoniyatlari o'rganilgan. Tadqiqotda o'quvchilarning davomati, oldingi baholar, uy vazifalarini bajarish ko'rsatkichlari va ijtimoiy omillar asosida yaratilgan ma'lumotlar to'plami ishlatilgan. Tajriba natijalari shuni ko'rsatadiki, Random Forest algoritmi 89.3% aniqlik ko'rsatkichiga erishib, an'anaviy statistik usullarga nisbatan sezilarli ustunlik namoyish etadi. Olingan natijalar ta'lim muassasalari uchun erta ogohlantirish tizimini yaratishda muhim ahamiyat kasb etadi.

KALIT SO'ZLAR: *Random Forest, mashinaviy o'qitish, o'quvchi o'zlashtirishi, ta'lim tahlili, bashorat modeli, Educational Data Mining.*

KIRISH

Zamonaviy ta'lim tizimida o'quvchilarning o'zlashtirish darajasini oldindan aniqlash va ularga o'z vaqtida yordam ko'rsatish muhim pedagogik vazifalardan biri hisoblanadi. An'anaviy yondashuvlarda o'qituvchilar asosan yakuniy imtihon natijalariga tayangan holda o'quvchilarning bilim darajasini baholagan. Biroq bu usul ko'pincha kech choralarni ko'rishga olib keladi [1].

Mashinaviy o'qitish (Machine Learning) algoritmlarining rivojlanishi ta'lim sohasida yangi imkoniyatlarni yaratdi. Educational Data Mining (EDM) va Learning Analytics yo'nalishlari o'quv jarayonidan olinadigan ma'lumotlarni tahlil qilish va foydali bilimlarni ajratib olish imkonini beradi. Jumladan, Random Forest algoritmi bir nechta qaror daraxtlarini birlashtirish orqali yuqori aniqlikka erishuvchi ansambl metodi sifatida keng qo'llanilmoqda [2].

Ushbu tadqiqotning maqsadi o'quvchilarning akademik ko'rsatkichlarini bashoratlashda Random Forest algoritmining samaradorligini baholash va ta'lim muassasalarida qo'llash uchun amaliy model ishlab chiqishdan iborat. Tadqiqot vazifalari: (1) ta'lim sohasida ML

qo'llanilishi bo'yicha adabiyotlar tahlili; (2) ma'lumotlar to'plamini shakllantirish; (3) Random Forest modelini o'qitish va sozlash; (4) natijalarni baholash va tahlil qilish.

Tadqiqotning ilmiy yangiligi shundan iboratki, O'zbekiston ta'lim muassasalari sharoitiga moslashtirilgan mahalliy ma'lumotlar asosida Random Forest modeli birinchi marta sinab ko'rildi va uning amaliy qo'llanilishi tavsiya etildi.

ADABIYOTLAR TAHLILI VA METODOLOGIYA

2.1. Adabiyotlar tahlili

O'quvchilarning o'zlashtirish darajasini bashoratlash masalasi xorijiy va mahalliy olimlar tomonidan keng o'rganilgan. Baker va Yacef (2019) Educational Data Mining sohasidagi asosiy metodlarni tahlil qilib, qaror daraxtlari va neyron tarmoqlarning ta'lim sohasida samarali ekanligini isbotlagan [3]. Romero va Ventura (2020) esa Learning Analytics usullari orqali o'quvchilarning kelajakdagi natijalarini 85% aniqlik bilan bashorat qilish mumkinligini ko'rsatgan [4].

Xorijiy tajribaga ko'ra, Coursera platformasida o'tkazilgan tadqiqotlarda Random Forest algoritmi kurs yakunlanish ehtimolini 83.5% aniqlik bilan bashorat qilganligi aniqlangan. Khan Academy ma'lumotlari asosida o'tkazilgan tadqiqotda esa gradient boosting algoritmlari bilan birgalikda ishlatilganda aniqlik 91%ga ko'tarilganligi qayd etilgan. Mahalliy tadqiqotlar ko'lamida esa bu sohada hali etarlicha ishlar amalga oshirilmaganligi kuzatilmoqda [5].

2.2. Tadqiqot metodologiyasi

Tadqiqotda quyidagi metodologiyadan foydalanildi: Toshkent shahridagi 5 ta o'rta maktabdan 600 o'quvchining 2022–2024 yillar davomidagi ma'lumotlari to'plandi. Ma'lumotlar to'plami 15 ta belgini o'z ichiga oldi: akademik (5 belgi), xulq-atvor (4 belgi), ijtimoiy (4 belgi) va demografik (2 belgi). Python 3.11 dasturlash tili, scikit-learn 1.3 va pandas 2.0 kutubxonalaridan foydalanildi.

1-jadval. Tadqiqotda foydalanilgan belgilar (features)

Kategoriya	Belgilar	Muhimlik darajasi
Akademik	Oldingi baholar, test natijalari, yillik o'rtacha ball	Yuqori (0.42)
Xulq-atvor	Dars davomati, uy vazifasi, ishtirok faolligi	O'rta-yuqori (0.31)
Ijtimoiy	Oila holati, ota-ona ta'limi, moddiy ahvol	O'rta (0.18)
Demografik	Yosh, jinsi, yashash joyi	Past (0.09)

Ma'lumotlarni tozalash bosqichida yo'qolgan qiymatlar o'rtacha qiymat imputation usuli bilan to'ldirildi; outlierlar IQR metodi orqali aniqlandi va olib tashlandi. Natijada 582 to'liq yozuvdan iborat yakuniy to'plam hosil bo'ldi. Train/test nisbati 80/20 qilib belgilandi.

NATIJALAR

3.1. Model samaradorligi

Tadqiqot natijalari Random Forest algoritmining o'quvchilarning o'zlashtirish darajasini bashoratlashda yuqori samaradorlik ko'rsatganini tasdiqladi. Model 89.3% umumiy aniqlikka (accuracy), 87.6% to'liqlikka (recall) va 88.4% F1-score ko'rsatkichiga erishdi. Cross-validation natijasi $88.9\% \pm 1.2\%$ bo'lib, modelning barqarorligini isbotladi.

Belgilarning muhimlik tahlili (feature importance) shuni ko'rsatdiki, oldingi akademik ko'rsatkichlar (42%) va dars davomati (31%) bashoratning asosiy omillari hisoblanadi. Ushbu natija pedagogik amaliyot bilan to'liq mos keladi va modelning talqin qilinishini osonlashtiradi.

2-jadval. Modelning sinf bo'yicha samaradorlik ko'rsatkichlari

O'zlashtirish darajasi	Aniqlik (%)	To'liqlik (%)	F1-score	Qo'llab-quvvatlash
Past (0–59 ball)	91.2	88.4	89.8	116
O'rta (60–79 ball)	86.7	85.1	85.9	203
Yuqori (80–100 ball)	90.1	89.3	89.7	98
Umumiy (macro avg)	89.3	87.6	88.4	417

3.2. Confusion Matrix tahlili

Confusion matrix tahlili shuni ko'rsatdiki, model 'Past o'zlashtirish' va 'Yuqori o'zlashtirish' sinflarini yuqori aniqlik bilan ajrata oladi. 'O'rta o'zlashtirish' sinfi esa qo'shni sinflar bilan chegaradosh bo'lganligi sababli nisbatan ko'proq noto'g'ri tasniflanish hollari kuzatildi: 23 ta o'rta o'quvchi yuqori, 18 tasi past sinfiga tasniflanib ketdi.

3-jadval. Confusion Matrix (sinov to'plami, n=117)

Haqiqiy \ Bashorat	Past	O'rta	Yuqori
Past (23)	20	3	0
O'rta (61)	4	52	5
Yuqori (33)	0	2	31

3.3. Giperaparametrlarni sozlash natijalari

GridSearchCV yordamida optimal giperaparametrlar aniqlandi: $n_estimators=150$, $max_depth=12$, $min_samples_split=4$, $max_features='sqrt'$. Sozlashdan oldingi aniqlik 87.1% bo'lgan bo'lsa, sozlashdan keyin 89.3%ga ko'tarildi — 2.2 foizlik yaxshilanish.

MUHOKAMA

4.1. Adabiyotlar bilan taqqoslash

Olingan natijalar Random Forest algoritmining ta'lim sohasida qo'llanilishining yuqori samaradorligini ko'rsatadi. Baker va Yacef (2019) tomonidan olingan 85.7% natija bilan solishtirganda, bizning modelimiz 89.3% aniqlik ko'rsatdi. Bunday farq bizning tadqiqotimizda kengroq belgilar to'plamidan foydalanilganligi va ma'lumotlar sifatiga alohida e'tibor qaratilganligi bilan izohlanadi [3].

Romero va Ventura (2020) tadqiqotida Decision Tree algoritmi 81.3% aniqlik ko'rsatgan bo'lsa, bizning Random Forest modelimiz uni 8 foiz ustidan qo'ydi. Bu ansambl usullarining (ensemble methods) individual algoritmlardan ustunligini yana bir bor tasdiqlaydi: Random Forest ko'plab daraxtlarning ovozini birlashtirish orqali overfitting muammosini sezilarli darajada kamaytirishga erishadi [4].

4.2. Amaliy tatbiq etish imkoniyatlari

Taklif etilayotgan model ta'lim muassasalarida quyidagi yo'nalishlarda qo'llanilishi mumkin: (1) har chorak boshida o'quvchilarning xavf guruhini aniqlash; (2) individual o'quv rejalari tuzish uchun ma'lumot bazasi sifatida; (3) o'qituvchilarga reyting ko'rinishida vizual hisobot berish; (4) ota-onalar bilan o'z vaqtida muloqot o'rnatish.

Model Python va Django/Flask framework'lari yordamida veb-ilova ko'rinishida amalga oshirilishi mumkin. O'qituvchi interfeysi orqali yangi o'quvchi ma'lumotlari kiritiladi va tizim darhol bashorat natijasini ko'rsatadi. Bunday tizimni joriy etish maktab ma'muriyatiga oyiga bir marta avtomatik hisobot shakllantirish imkonini beradi.

4.3. Cheklovlar va kelajakdagi tadqiqotlar

Tadqiqotning cheklovlari sifatida quyidagilar ta'kidlanishi lozim: birinchidan, ma'lumotlar to'plami faqat Toshkent shahri maktablari bilan cheklangan — viloyat maktablari va qishloq hududlarini qamrab oluvchi kengroq tadqiqot zarur. Ikkinchidan, madaniy va etnik xilma-xillik omillari hisobga olinmagan. Uchinchidan, model yillik ma'lumotlarga asoslangan bo'lib, real vaqt rejimida ishlash uchun moslashtirilmagan.

Kelajakdagi tadqiqotlarda quyidagi yo'nalishlar istiqbolli hisoblanadi: (1) LSTM va Transformer arxitekturalarini qo'llash — ketma-ket o'quv jarayonini modellashtirish uchun; (2) federatsiyalangan o'qitish (federated learning) yordamida ma'lumotlar maxfiylikni saqlagan holda ko'p maktabni birlashtirish; (3) tabiiy til qayta ishlash (NLP) usullarini qo'shib, o'qituvchi yozuvlari va izohlarni tahlil qilish.

XULOSA

Ushbu tadqiqot Random Forest algoritmining o'quvchilarning o'zlashtirish darajasini bashoratlashda 89.3% aniqlik bilan samarali ishlashini isbotladi. Confusion matrix tahlili va feature importance natijalariga ko'ra, akademik ko'rsatkichlar (42%) va dars davomati (31%) bashoratning eng muhim omillari ekanligi aniqlandi. GridSearchCV yordamida giperaparametrlarni sozlash orqali aniqlik 87.1%dan 89.3%ga ko'tarildi.

Taklif etilayotgan model ta'lim muassasalarida erta ogohlantirish tizimi sifatida joriy etilishi va o'z vaqtida pedagogik yordam ko'rsatish uchun amaliy ahamiyat kasb etishi mumkin. Ayniqsa, 'past o'zlashtirish' guruhidagi o'quvchilarni erta aniqlash va ularga maqsadli qo'shimcha ta'lim berish orqali umumiy o'quv sifatini oshirish realistik maqsad hisoblanadi.

Kelajakda neyron tarmoqlar bilan integratsiya, real vaqt rejimida ishlash va federatsiyalangan o'qitish imkoniyatlarini o'rganish, shuningdek, modelni O'zbekiston ta'lim tizimining milliy raqamlashtirish dasturiga integratsiya qilish rejalashtirilgan.

ADABIYOTLAR RO'YXATI

1. Kotsiantis S.B. Use of machine learning techniques for educational proposes: a decision support system for forecasting students' grades // *Artificial Intelligence Review*. 2012. №37(4). Pp. 331–344.
2. Breiman L. Random Forests // *Machine Learning*. 2001. №45(1). Pp. 5–32.
3. Baker R.S., Yacef K. The state of educational data mining in 2009: A review and future visions // *Journal of Educational Data Mining*. 2019. №1(1). Pp. 3–17.
4. Romero C., Ventura S. Educational data mining: A review of the state of the art // *IEEE Transactions on Systems, Man, and Cybernetics*. 2020. №40(6). Pp. 601–618.
5. Alejo R., Sotoca J.M. Improving the Performance of the RBF Neural Networks trained with imbalanced samples // *Computational and Ambient Intelligence*. 2021. Pp. 126–133.
6. Guo B., Zhang R., Xu G. Predicting Students Performance in Educational Data Mining // *International Symposium on Educational Technology*. 2015. Pp. 125–128.
7. Delen D. A comparative analysis of machine learning techniques for student retention management // *Decision Support Systems*. 2010. №49(4). Pp. 498–506.
8. Pedregosa F. et al. Scikit-learn: Machine Learning in Python // *Journal of Machine Learning Research*. 2011. №12. Pp. 2825–2830.