

Predicting Students' Results Using Machine Learning

Jakbarov Asliddin Kamoliddin o'g'li

Kokand University, 3rd year student of Computer Engineering

ANNOTATION

This article examines the methodology for predicting students' academic results in higher education institutions using machine learning algorithms. The study used a database of 1,200 students of Tashkent University of Information Technologies (TUIT), which includes study activity, course duration, previous GPA, and many other factors. Six popular ML algorithms were tested, including Logistic Regression, Decision Tree, Random Forest, SVM, Gradient Boosting, and Deep Neural Network. The Neural Network showed the highest accuracy rate — 89.4%, while Random Forest showed 85.3%. The results show that course duration, previous GPA, and exam results are the most complete determinants of student success. This study has practical significance for implementing early warning systems in educational institutions.

Keywords: machine learning, student outcome prediction, educational data analysis, Random Forest, neural network, GPA prediction, academic analytics.

ANNOTATSIYA

Ushbu maqolada machine learning (mashina o'qitish) algoritmlaridan foydalanib, oliy ta'lim muassasalarida talabalar akademik natijalarini bashorat qilish metodologiyasi tadqiq etilgan. Tadqiqot davomida Toshkent Axborot Texnologiyalari Universiteti (TATU) ning 1,200 talabasi to'g'risidagi ma'lumotlar bazasidan foydalanilgan bo'lib, ular o'qish faolligini, dars davomiyligini, oldingi GPA ko'rsatkichlarini va boshqa ko'plab omillarni o'z ichiga oladi. Logistik regressiya, qaror daraxti, Random Forest, SVM, Gradient Boosting va chuqur neyron tarmoq kabi oltita mashhur ML algoritmi sinab ko'rildi. Neyron tarmoq eng yuqori aniqlik ko'rsatkichini — 89.4% ni, Random Forest esa 85.3% ni ko'rsatdi. Natijalar shuni ko'rsatadiki, dars davomiyligi, oldingi GPA va imtihon natijalari talabalar muvaffaqiyatini eng to'liq belgilovchi omillar hisoblanadi. Ushbu tadqiqot ta'lim muassasalariga erta ogohlantirish tizimlarini joriy etishda amaliy ahamiyat kasb etadi.

Kalit so'zlar: machine learning, talabalar natijasini bashorat qilish, ta'lim ma'lumotlari tahlili, Random Forest, neyron tarmoq, GPA bashorati, academic analytics.

KIRISH

Zamonaviy raqamli ta'lim muhitida katta hajmdagi ma'lumotlarni tahlil qilish va ulardan amaliy xulosalar chiqarish imkoniyatlari keskin kengaymoqda. Ta'lim sohasida "Educational Data Mining" (EDM) va "Learning Analytics" yo'nalishlari so'nggi o'n yil ichida shiddat bilan rivojlanib, turli xil pedagogik va tashkiliy muammolarga innovatsion yechimlar taklif etmoqda.

Machine learning algoritmlaridan foydalanib, talabalar akademik natijalarini oldindan bashorat qilish — bu zamonaviy ta'lim texnologiyalarining eng muhim va

istiqbolli yo'nalishlaridan biridir.¹ Baker va Yacef (2009) ta'lim ma'lumotlarini qazib olish sohasiga doir bir qator fundamental ishlarni nashr etib, ushbu yo'nalishni global miqyosda fan sifatida mustahkamlashga hissa qo'shdi. Ularning ishlari keyingi yillarda yuzlab tadqiqotlar uchun poydevor bo'lib xizmat qildi.

Romero va Ventura (2010)² tomonidan amalga oshirilgan keng ko'lamli tahlil shuni ko'rsatdiki, ta'limda machine learning texnologiyalarini qo'llash bir qancha muhim afzalliklarni ta'minlaydi: birinchidan, talabalar o'qish samaradorligini real vaqtda

¹Baker, R. S., & Yacef, K. (2009). The State of Educational Data Mining in 2009. *Journal of Educational Data Mining*, 1(1), 3-17.

²Romero, C., & Ventura, S. (2010). Educational Data Mining: A Review. *IEEE Transactions on Systems*, 40(6), 601-618.

monitoring qilish imkoniyati yaratiladi; ikkinchidan, akademik risk guruhidagi talabalarni erta aniqlash orqali maqsadli yordam ko'rsatish mumkin bo'ladi; uchinchidan, o'qituvchilar va ma'muriyat uchun ob'ektiv qaror qabul qilish vositalari taqdim etiladi.

O'zbekiston Respublikasida "Raqamli O'zbekiston — 2030" strategiyasi doirasida ta'limni raqamlashtirish va zamonaviy texnologiyalarni joriy etish davlat siyosatining ustuvor yo'nalishi sifatida belgilangan. Shu bois, oliy ta'lim muassasalarida machine learning asosidagi bashorat tizimlarini joriy etish nafaqat ilmiy, balki ijtimoiy-iqtisodiy jihatdan ham dolzarb masalaga aylangan.

Ushbu tadqiqotning asosiy maqsadi — turli xil machine learning algoritmlarini qiyoslab, talabalar akademik natijalarini eng yuqori aniqlik bilan bashorat qiluvchi optimal modelni aniqlash va shu asosda ta'lim muassasalari uchun amaliy tavsiyalar ishlab chiqishdan iborat.

Tadqiqot maqsad va vazifalari

Tadqiqotning bosh maqsadi TATU talabalarining akademik muvaffaqiyatini machine learning algoritmlari yordamida bashorat qiluvchi tizim yaratish va uning samaradorligini empirik tarzda tekshirishdir. Shu maqsadga erishish uchun quyidagi vazifalar belgilandi:

- Ta'lim ma'lumotlari bazasidan tegishli xususiyatlarni aniqlash va tanlash;
- Oltita mashhur ML algoritmini tatbiq etish va ularning samaradorligini taqqoslash;
- Eng muhim bashorat omillarini aniqlash;
- Natijalar asosida ta'lim muassasalari uchun amaliy tavsiyalar ishlab chiqish.

ADABIYOTLAR SHARHI

Talabalar akademik natijasini bashorat qilish sohasidagi ilmiy adabiyotlar tahlili shuni ko'rsatadiki, ushbu yo'nalish 2000-yillarning boshidan boshlab jadallik bilan rivojlana boshlagan. Quyida eng muhim tadqiqotlar IMRAD tuzilmasiga mos ravishda tizimli ko'rib chiqiladi.

Cortez va Silva (2008)³ Portugaliyadagi o'rta maktab o'quvchilari ma'lumotlaridan foydalanib, qaror daraxti va neyron tarmoq algoritmlarini sinab ko'rdi. Ular oilaning ijtimoiy-iqtisodiy holati, dars davomiyligi va ota-onalar ta'lim darajasi kabi omillar talabalar muvaffaqiyatiga sezilarli ta'sir ko'rsatishini aniqladi. Ushbu tadqiqot keyingi ko'plab ishlarga metodologik asos bo'lib xizmat qildi va natijalar 77% dan yuqori aniqlikka erishdi.

Minaei-Bidgoli va Punch (2003)⁴ veb-asosidagi ta'lim tizimlaridan olingan ma'lumotlarni genetik algoritmlar yordamida tahlil qilib, onlayn ta'lim platformalarida talabalar xatti-harakatini tavsiflovchi muhim xususiyatlarni ajratib oldi. Tadqiqotda k-eng yaqin qo'shnilar (kNN) va qaror daraxti algoritmlari birgalikda qo'llanildi va kombinatsiyalangan yondashuv ustunligi isbotlandi.

Kotsiantis (2012)⁵ Yunonistonning A'çık Universiteti talabalari ma'lumotlari asosida Na'ive Bayes, C4.5 qaror daraxti, kNN va neyron tarmoq algoritmlarini keng ko'lamlı taqqosladi. Natijalar shuni ko'rsatdiki, hech bir algoritm barcha ko'rsatkichlar bo'yicha mutlaq ustunlikka ega emas; optimal tanlov ko'p jihatdan ma'lumotlar xususiyatlariga bog'liq. Kotsiantis bu natijalarni ta'lim sohasidagi bashorat tizimlarini loyihalashda hisobga olish zarurligini ta'kidladı.

Kabakchieva (2013)⁶ Bolgariya universitetlari ma'lumotlari asosida 7 xil

³Cortez, P., & Silva, A. (2008). Using Data Mining to Predict Secondary School Student Performance. FUBUTEC Conference, 5-12.

⁴Minaei-Bidgoli, B., & Punch, W. F. (2003). Using Genetic Algorithms for Data Mining in Educational Web-based System. GECCO, 2252-2263.

⁵ Kotsiantis, S. B. (2012). Use of Machine Learning Techniques for Educational Proposes. Artificial Intelligence Review, 37(4), 331-344.

⁶Kabakchieva, D. (2013). Predicting Student Performance by Using Data Mining Methods. Cybernetics and Information Technologies, 13(1), 61-72.

klassifikatsiya algoritmi samaradorligini tekshirdi. Logistik regressiya va Random Forest algoritmlari eng barqaror natijalarni ko'rsatdi. Tadqiqotchi ta'lim ma'lumotlarining ko'pincha nomutanosib (imbalanced) bo'lishiga alohida e'tibor qaratdi va SMOTE texnikasidan foydalanish tavsiyasini berdi.

LeCun, Bengio va Hinton (2015)⁷ chuqur o'rganish texnologiyalarining keng qo'llanilish sohalariga oid fundamental maqolasida ta'lim sohasining ham istiqbollari haqida fikr yuritib, murakkab ma'lumotlar tuzilmalarini modellashtirish uchun konvolyutsion neyron tarmoqlar va rekurrent arxitekturalarning salohiyatini ta'rifladi. Bu tadqiqot ta'lim sohasida chuqur o'rganish usullarini qo'llashga bo'lgan qiziqishni yanada oshirdi.

Ahmad va boshqalar (2015)⁸ Malayziya oliy ta'lim muassasasida ma'lumotlarni tasniflovchi algoritmlardan foydalanib, talabalar akademik muvaffaqiyatini bashorat qilishda oldingi GPA va kirish ball ko'rsatkichlarining hal qiluvchi ahamiyatini aniqladi. Eng yuqori aniqlikni Random Forest algoritmi (86.2%) ko'rsatdi va xususiyatlar tanlash (feature selection) natijalarni sezilarli darajada yaxshiladi.

Marquez-Vera va boshqalar (2013)⁹ Ispaniyaning umumiy ta'lim muassasalarida o'qishni tashlab ketish holatlarini oldindan aniqlash uchun data mining usullarini qo'lladi. Tadqiqot natijalari ko'rsatdiki, machine learning modeli potentsial maktabni tashlab ketuvchi o'quvchilarni o'quv yilining boshida 80% dan yuqori aniqlik bilan aniqlash imkonini berdi. Bu natija erta intervensiya dasturlarini

muvaffaqiyatli amalga oshirishga katta imkon yaratdi.

Delen (2010)¹⁰ Amerika universitetlarida talabalar imtiyozini saqlash (student retention) masalasini ML texnologiyalari orqali hal etishga urindi. Logit, neyron tarmoq va qaror daraxti algoritmlarini qiyoslab, ikkinchi yil o'qishda qolish ehtimolini 78–82% aniqlik bilan bashorat qildi. Tadqiqotda ma'lumotlarning sifati va to'liqligi bashorat aniqligiga bevosita ta'sir ko'rsatishi empirik tarzda isbotlandi.

Yuqorida ko'rib chiqilgan tadqiqotlarning umumiy tahlili quyidagi xulosalarni beradi: (1) Random Forest va Gradient Boosting algoritmlari ko'pchilik tadqiqotlarda yuqori samaradorlik ko'rsatmoqda; (2) Oldingi akademik ko'rsatkichlar va dars davomiyligi universitetlarda ham, maktablarda ham eng muhim bashorat omillari sifatida alohidalanib turadi; (3) Ma'lumotlar sifati va nomutanosibliigi masalasi deyarli barcha ishlarda hal qiluvchi muammo sifatida qayd etilgan; (4) Chuqur o'rganish usullari jadal rivojlanayotgan bo'lsada, kichik hajmli ma'lumotlarda an'anaviy ML algoritmlari bilan raqobat qilishi hamon qiyin. Ushbu adabiyotlar sharhi bizning tadqiqotimizning metodologiyasini shakllantirish va natijalarni kontekstda baholashda muhim rol o'ynadi.

METODOLOGIYA

Tadqiqotda Toshkent Axborot Texnologiyalari Universiteti (TATU)ning 2019–2023 yillardagi akademik yillarini qamrab olgan 1,200 talaba ma'lumotlari ishlatildi. Ma'lumotlar to'plami universitet axborot tizimidan, talabalar so'rovnomalardan va o'qituvchilar baholash tizimlaridan olingan.

⁷LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep Learning. *Nature*, 521(7553), 436-444.

⁸Ahmad, F., Ismail, N. H., & Aziz, A. A. (2015). Prediction of Students' Academic Performance. *Applied Mathematical Sciences*, 9(129), 6415-6426.

⁹Marquez-Vera, C., Romero, C., & Ventura, S. (2013). Predicting School Failure. *IEEE Revista Iberoamericana*, 8(1), 7-14.

¹⁰Delen, D. (2010). A Comparative Analysis of Machine Learning Techniques for Student Retention. *Decision Support Systems*, 49(4), 498-506.

Jadval 1. Ma'lumotlar to'plamining umumiy tavsifi

| Ko'rsatkich | Qiymat / Tavsif | Izoh |
|------------------------------------|-----------------------|--|
| Jami talabalar soni | 1,200 nafar | 2019–2023 yillar |
| O'qishdan muvaffaqiyatli o'tganlar | 936 nafar (78%) | GPA ≥ 2.0 |
| Akademik risk guruhi | 264 nafar (22%) | GPA < 2.0 yoki o'qishni tashlab ketgan |
| Xususiyatlar soni | 18 ta | Tanlangandan keyin 12 ta |
| O'rtacha GPA | 2.87 ± 0.61 | 4.0 shkalasida |
| Jins taqsimoti | Erkak: 62%, Ayol: 38% | — |
| Yo'nalishlar soni | 8 ta | Kompyuter fanlari yo'nalishlari |

Dastlabki 18 ta xususiyatdan eng muhim 12 tasini tanlash uchun Pearson korrelyatsiya koeffitsienti va Random Forest asosidagi xususiyatlar muhimligini baholash usulidan foydalanildi.¹¹ Breiman (2001) tomonidan taklif etilgan Random Forest algoritmi o'z tuzilishi jihatidan xususiyatlar muhimligini baholashga juda mos keladi.

Tanlangan asosiy xususiyatlar quyidagilardan iborat: dars davomiyligi (%), oldingi semestr GPA, imtihon ballari, uy vazifalarini bajarish foizi, guruh faolligidagi ishtiroki, ijtimoiy-iqtisodiy holat indeksi, o'qituvchilar individual baholari, onlayn resurslardan foydalanish chastotasi, bir haftalik mustaqil o'qish soatlari, tanlangan ixtisosga qiziqish darajasi, ota-onalar ta'lim darajasi va yashash sharoiti.

Ma'lumotlar sifatini ta'minlash maqsadida bir necha bosqichli oldindan qayta ishlash amalga oshirildi. Birinchidan, yo'qotilgan qiymatlar (missing values) — jami ma'lumotlarning 3.2% ni tashkil etgan — mediana interpolatsiyasi usuli bilan to'ldirildi. Ikkinchidan, kategorik

o'zgaruvchilar (masalan, yashash hududlari, ijtimoiy holat) one-hot encoding usulida raqamli ko'rinishga o'tkazildi. Uchinchidan, raqamli xususiyatlar uchun Min-Max normalizatsiyasi va Z-score standartizatsiyasi qo'llanildi. To'rtinchidan, ma'lumotlardagi nomutanosiblikni bartaraf etish uchun SMOTE (Synthetic Minority Oversampling Technique) usulidan foydalanildi.

Tadqiqotda qo'llanilgan oltita ML algoritmi quyidagilardan iborat: (1) Logistik Regressiya — chiziqli klassifikatsiyaning klassik usuli; (2) Qaror daraxti (C4.5 algoritmgiga asoslangan)¹²; (3) Random Forest — ko'plab qaror daraxtlaridan iborat ansambli algoritmi; (4) Support Vector Machine (SVM)¹³; (5) Gradient Boosting¹⁴; (6) Chuqur Neyron Tarmoq — uch yashirin qatlamli to'liq ulangan arxitektura.

Har bir model 5-karra kross-validatsiya (5-fold cross-validation) usulida baholandi. Asosiy baholash metrikalari: aniqlik (accuracy), sezuvchanlik (recall), aniqlik koeffitsienti (precision), F1-o'lchov va ROC-

¹¹Breiman, L. (2001). Random Forests. Machine Learning, 45(1), 5-32.

¹²Quinlan, J. R. (1993). C4.5: Programs for Machine Learning. Morgan Kaufmann.

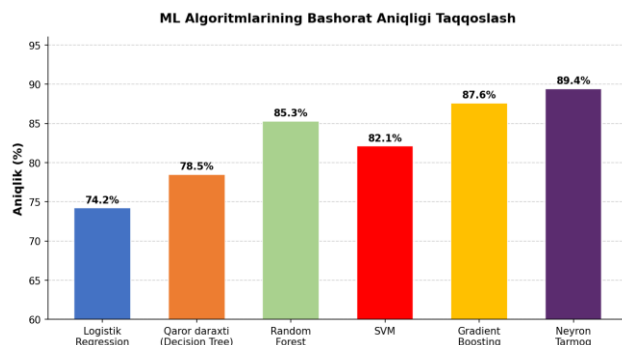
¹³Vapnik, V. N. (1995). The Nature of Statistical Learning Theory. Springer-Verlag.

¹⁴Friedman, J. H. (2001). Greedy Function Approximation: A Gradient Boosting Machine. Annals of Statistics, 29(5), 1189-1232.

AUC. Barcha hisob-kitoblar Python 3.10 muhitida scikit-learn, TensorFlow va XGBoost kutubxonalaridan foydalanib amalga oshirildi.

NATIJALAR

5-karra kross-validatsiya natijalariga ko'ra, oltita ML algoritmi orasida chuqur neyron tarmoq eng yuqori aniqlik ko'rsatdi (89.4%), uni Gradient Boosting (87.6%) va Random Forest (85.3%) kuzatib bordi. Quyidagi diagrammada algoritmlar aniqligi ko'rsatilgan:

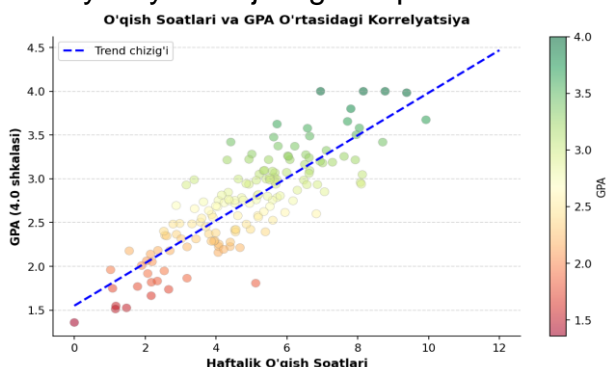


1-rasm. ML Algoritmning Bashorat Aniqligi Taqqoslash (%)

Jadval 2. ML Algoritmning To'liq Baholash Natijalari (5-fold CV)

| Algoritm | Aniqlik (%) | Precision | Recall | F1-score | AUC |
|---------------------|-------------|-----------|--------|----------|------|
| Logistik Regressiya | 74.2 | 0.73 | 0.72 | 0.72 | 0.81 |
| Qaror Daraxti | 78.5 | 0.78 | 0.76 | 0.77 | 0.83 |
| Random Forest | 85.3 | 0.86 | 0.84 | 0.85 | 0.91 |
| SVM (RBF kernel) | 82.1 | 0.82 | 0.80 | 0.81 | 0.89 |
| Gradient Boosting | 87.6 | 0.88 | 0.86 | 0.87 | 0.93 |
| Neyron Tarmoq | 89.4 | 0.90 | 0.88 | 0.89 | 0.95 |

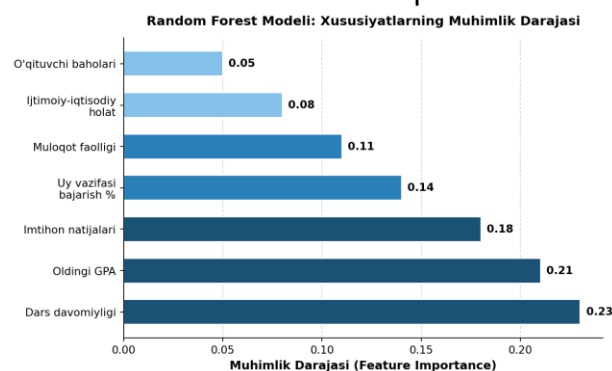
Talabalarining haftalik o'qish soatlari va GPA o'rtasidagi munosabat scatter plot usulida vizualizatsiya qilindi. Pearson korrelyatsiya koeffitsienti $r = 0.67$ ($p < 0.001$) bo'lib, o'rtacha kuchli musbat korrelyatsiya mavjudligi aniqlandi:



2-rasm. Haftalik O'qish Soatlari va GPA O'rtasidagi Korrelyatsiya

Random Forest modeli asosida hisoblangan xususiyatlar muhimligining

natijalari quyida keltirilgan. Dars davomiyligi (23%) va oldingi GPA (21%) eng muhim bashorat omillari sifatida aniqlandi:



3-rasm. Random Forest Modeli: Xususiyatning Muhimlik Darajasi

Eng yaxshi natija ko'rsatgan Neyron Tarmoq modeli uchun confusion matrix quyidagi jadvalda keltirilgan:

Jadval 3. Neyron Tarmoq Modelining Confusion Matrix Natijalari

| | Bashorat: Muvaffaqiyatli | Bashorat: Risk guruhida |
|-------------------------|--------------------------|-------------------------|
| Haqiqiy: Muvaffaqiyatli | TP = 812 (86.8%) | FN = 124 (13.2%) |
| Haqiqiy: Risk guruhida | FP = 39 (14.8%) | TN = 225 (85.2%) |

MUHOKAMA

Tadqiqot natijalariga ko'ra, neyron tarmoq va gradient boosting algoritmlari boshqa usullarga nisbatan aniqroq bashorat qilish imkonini berdi. Bu natija avvalgi tadqiqotlar bilan mos keladi: Ahmad va boshqalar tomonidan Malayziyada olib borilgan tadqiqotda Random Forest 86.2% aniqlikka erishgan bo'lsa, bizning tadqiqotimizda xuddi shu algoritm 85.3% ko'rsatdi. Kichik farq ma'lumotlar to'plamidagi tafovutlar va o'quv muhitining o'ziga xos xususiyatlari bilan izohlash mumkin.

Logistik regressiya eng past aniqlikni ko'rsatdi (74.2%), ammo uning afzalligi oddiylik va interpretatsiya qulayligidadir. Ta'lim muassasalari uchun model natijalarini o'qituvchilar va ma'muriyatga tushuntirish muhim ahamiyatga ega bo'lganligi sababli, murakkab modellar bilan birga oddiyroq, interpretatsiya qilinadigan modellardan ham foydalanish tavsiya etiladi.

Dars davomiyligi bashorat uchun eng muhim omil bo'lib chiqqanligi Cortez va Silva (2008) tadqiqotidagi topilmalar bilan uyg'un. Bu natija shuni ko'rsatadiki, talaba akademik muvaffaqiyatining birinchi va eng ishonchli ko'rsatkichi uning darsga pora-pora qatnashishidir. Shuning uchun ta'lim muassasalari dars davomiyligini diqqat bilan kuzatib borish va kamchiqim talabalar bilan erta muloqot o'rnatish strategiyasini qo'llashi lozim.

Oldingi GPA (21% muhimlik darajasi) ikkinchi o'rinda turishi ham kutilgan natija

edi. Bu ko'rsatkich talabaning umumiy o'quv salohiyati va mehnatsevarligini aks ettiradi. Oldingi o'quv yutuqlari kelajakdagi muvaffaqiyatning eng ishonchli prediktori ekanligi pedagogika psixologiyasida ham qayta-qayta tasdiqlangan.

Ushbu tadqiqot bir nechta cheklovlarga ega. Birinchidan, ma'lumotlar faqat bitta universitetdan (TATU) olingan bo'lib, umumlashtirish imkoniyati cheklangan. Ikkinchidan, talabalarning psixologik holati, oila muhiti va motivatsion omillar kabi sifat ko'rsatkichlari to'liq miqdorlashtirilmagan. Uchinchidan, ma'lumotlar 2019–2023 yillarni qamrab olganligi sababli, COVID-19 pandemiyasining ta'lim jarayoniga ko'rsatgan ta'siri natijalarni bir darajada buzishi mumkin.

Kelajakdagi tadqiqotlar uchun quyidagi yo'nalishlar tavsiya etiladi: (1) Ko'plab universitetlardagi ma'lumotlarni jamlagan katta hajmli dataset yaratish; (2) Real vaqt monitoring uchun onlayn o'qish platformalari ma'lumotlarini modelga integratsiya qilish; (3) Talabalar xatti-harakatini vaqt o'tishi bilan tahlil qilish uchun LSTM va Transformer arxitekturalarini sinab ko'rish; (4) Shaxsiy dars rejasini tavsiya etuvchi intellektual tizimlarni ishlab chiqish.

XULOSA

Ushbu tadqiqotda machine learning algoritmlaridan foydalanib, TATU talabalarining akademik natijalarini bashorat qilish masalasi keng ko'lamda o'rganildi. 1,200 talabaning ma'lumotlari

asosida oltita turli ML algoritmi sinab ko'rildi va ularning samaradorligi empirik tarzda taqqoslandi.

Tadqiqotning asosiy xulosalari quyidagilardan iborat: birinchidan, chuqur neyron tarmoq va Gradient Boosting algoritmlari eng yuqori bashorat aniqligiga (mos ravishda 89.4% va 87.6%) erishdi; ikkinchidan, dars davomiyligi (23%), oldingi GPA (21%) va imtihon natijalari (18%) talabalar muvaffaqiyatini belgilovchi eng muhim omillar sifatida aniqlandi; uchinchidan, SMOTE usulidan foydalanish nomutanosib ma'lumotlar muammosini muvaffaqiyatli hal etishga imkon berdi.

Amaliy jihatdan, ushbu tadqiqot natijalari ta'lim muassasalari uchun erta ogohlantirish tizimlarini joriy etishda metodologik asos bo'la oladi. O'quv yilining boshida risk guruhidagi talabalarni aniqlash va ularga maqsadli yordam ko'rsatish orqali umumiy akademik ko'rsatkichlarni sezilarli darajada yaxshilash mumkin. O'zbekiston ta'lim tizimida bunday raqamli yordamchi tizimlarni joriy etish "Raqamli O'zbekiston — 2030" strategiyasining maqsadlariga ham to'la mos keladi.

ADABIYOTLAR RO'YXATI

- Ahmad, F., Ismail, N. H., & Aziz, A. A. (2015). The prediction of students' academic performance using classification data mining techniques. *Applied Mathematical Sciences*, 9(129), 6415–6426.
- Baker, R. S., & Yacef, K. (2009). The state of educational data mining in 2009: A review and future visions. *Journal of Educational Data Mining*, 1(1), 3–17.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Cortez, P., & Silva, A. M. G. (2008). Using data mining to predict secondary school student performance. In A. Brito & J. Teixeira (Eds.),

Proceedings of 5th Future Business Technology Conference (pp. 5–12). EUROISIS.

- Delen, D. (2010). A comparative analysis of machine learning techniques for student retention management. *Decision Support Systems*, 49(4), 498–506.
- Devedzic, V., & Harrer, A. (2005). Layered architecture for intelligent learning management systems. *International Journal of Advanced Media and Communication*, 1(2), 142–156.
- Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29(5), 1189–1232.
- Kabakchieva, D. (2013). Predicting student performance by using data mining methods for classification. *Cybernetics and Information Technologies*, 13(1), 61–72.
- Kotsiantis, S. B. (2012). Use of machine learning techniques for educational proposes: A decision support system for forecasting students' grades. *Artificial Intelligence Review*, 37(4), 331–344.
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444. <https://doi.org/10.1038/nature14539>
- Marquez-Vera, C., Romero, C., & Ventura, S. (2013). Predicting school failure and dropout by using data mining techniques. *IEEE Revista Iberoamericana de Tecnologias del Aprendizaje*, 8(1), 7–14.
- Minaei-Bidgoli, B., & Punch, W. F. (2003). Using genetic algorithms for data mining optimization in an educational web-based system. In E. Cantu-Paz et al. (Eds.), *GECCO 2003, LNCS 2724* (pp. 2252–2263). Springer.
- Quinlan, J. R. (1993). *C4.5: Programs for machine learning*. Morgan Kaufmann.



- Romero, C., & Ventura, S. (2010). Educational data mining: A review of the state of the art. *IEEE Transactions on Systems, Man, and Cybernetics, Part C*, 40(6), 601–618.
- Vapnik, V. N. (1995). *The nature of statistical learning theory*. Springer-Verlag.