ҒТАМР 28.23.00 DOI: https://doi.org/10.62687/STJ.2.2.2026.26

STEM ЖӘНЕ КОМПЬЮТЕРЛІК ҒЫЛЫМДАРДАҒЫ СТУДЕНТТЕРДІҢ ҮЛГЕРІМІН БОЛЖАУДЫ ИНТЕРПРЕТАЦИЯЛАУҒА АРНАЛҒАН ТҮСІНДІРМЕЛІ ЖАСАНДЫ ИНТЕЛЛЕКТ (ХАІ): ЖҮЙЕЛІ ӘДЕБИЕТТІК ШОЛУ

[1]А.Ә. Әлібай*, **[2]Г.К. Тағанова**
[1,2]Астана халықаралық университеті, Астана қ., Қазақстан Республикасы
*e-mail: aidanaalibai9@gmail.com

А.Ә. Әлібай – ВШИТИИ Бакалавриант студенті, Астана халықаралық университеті, Астана қ., Қазақстан, e-mail: aidanaalibai9@gmail.com, https://orcid.org/0009-0004-2908-3222
Г.К. Тағанова – аға оқытушы, ғылыми жетекші, Астана халықаралық университеті, Астана қ., Қазақстан, e-mail: guldana.kileuzhanova@gmail.com, https://orcid.org/0000-0001-6139-6270

Аңдатпа. Білім беру деректерін интеллектуалды талдау (Educational Data Mining, EDM) оқушылардың үлгерімін болжау арқылы оқудағы қиындықтарды ерте анықтауға мүмкіндік береді. Дегенмен, Машиналық оқыту үлгілері көбінесе "қара жәшік" принципі бойынша жұмыс істейді. Түсіндірілетін жасанды интеллект (Explainable Artificial Intelligence, XAI) неліктен "қара жәшік" модельдері белгілі бір болжамдарды беретінін түсінуге көмектеседі. Бұл мақалада информатика және жаратылыстану ғылымдары бойынша оқушылардың үлгерімін болжау үшін түсіндірмелі жасанды интеллектті қолдану саласындағы соңғы бес жылдағы зерттеулерге жүйелі шолу берілген. Біз ең жиі қолданылатын сипаттамалар оқушылардың мінез-құлқы мен үлгерімі туралы деректер екенін және болжаудың негізгі мақсаттары пән бойынша үлгермеу қаупімен немесе бағалаулармен байланысты екенін анықтадық. Зерттеу сонымен қатар XAI қолданбаларын қарастырды, нәтижесінде ең көп таралған қолданбалар белгілердің маңыздылығын жаһандық талдау, жеке болжамдарды түсіндіру және араласулар мен шешім қабылдауды қолдау болып табылады. Сонымен қатар, біз Шеплидің аддитивті түсіндірмелері (SHAP) жеке деңгейде шектеулі қолданумен, негізінен жаһандық деңгейде ең көп қолданылатын XAI әдісі екенін анықтадық. Сонымен қатар, курстарды жақсарту, визуализация параметрлері және жекелендірілген ұсыныстарды қалыптастыру үшін түсіндірілетін жасанды интеллектті қолдануға қатысты зерттеулердегі олқылық анықталды. Бұл олқылықты жою мұғалімдерге жеке оқушыларға жақсы қолдау көрсету үшін деректерге негізделген жеке ұсыныстар беруге мүмкіндік береді.

Түйінді сөздер: Түсіндірілетін Жасанды Интеллект, XAI, Білім Беру Деректерін Өндіру, EDM, Өнімділікті Болжау, Информатика Білімі, STEM Білімі, Жүйелі Әдебиеттерге Шолу.

ОБЪЯСНИМЫЙ ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ (XAI) ДЛЯ ИНТЕРПРЕТАЦИИ ПРОГНОЗОВ УСПЕВАЕМОСТИ СТУДЕНТОВ В ОБЛАСТИ STEM И КОМПЬЮТЕРНЫХ НАУК: СИСТЕМАТИЧЕСКИЙ ОБЗОР ЛИТЕРАТУРЫ

[1]А.А. Алибай*, **[2]Г.К. Таганова**
[1,2]Международный университет Астана, г. Астана, Республика Казахстан
*e-mail: aidanaalibai9@gmail.com

А.А. Алибай – студент бакалавриата ВШИТИИ, Международный университет Астана, г. Астана, Казахстан, e-mail: aidanaalibai9@gmail.com, https://orcid.org/0009-0004-2908-3222
Г.К. Таганова – старший преподаватель, научный руководитель, Международный университет Астана, г. Астана, Казахстан, e-mail: guldana.kileuzhanova@gmail.com, https://orcid.org/0000-0001-6139-6270

Аннотация. Интеллектуальный анализ образовательных данных (Educational Data Mining, EDM) позволяет выявлять трудности в обучении на ранних этапах, прогнозируя успеваемость учащихся. Однако модели машинного обучения часто работают по принципу «черного ящика». Объяснимый искусственный интеллект (Explainable Artificial Intelligence, XAI) помогает понять, почему модели «черного ящика» выдают те или иные прогнозы. В этой статье представлен систематический обзор исследований, проведенных за последние пять лет в области применения объяснимого искусственного интеллекта для прогнозирования успеваемости учащихся в сфере компьютерных наук и естественно-научного образования. Мы выяснили, что наиболее часто используемыми характеристиками являются данные о поведении и успеваемости учащихся, а основные цели прогнозирования связаны с риском неуспеваемости по предмету или с оценками. В исследовании также рассматривались области применения XAI, в результате чего выяснилось, что наиболее распространенными областями применения являются глобальный анализ важности признаков, объяснение индивидуальных прогнозов и поддержка вмешательств и принятия решений. Более того, мы обнаружили, что аддитивные объяснения Шепли (SHAP) были наиболее часто используемой методикой XAI, применяемой преимущественно на глобальном уровне, с ограниченным использованием на индивидуальном уровне. Кроме того, был выявлен пробел в исследованиях, связанных с использованием объяснимого искусственного интеллекта для улучшения курсов, настройки визуализаций и формирования персонализированных рекомендаций. Устранение этого пробела позволит преподавателям предоставлять персонализированные рекомендации на основе данных, чтобы лучше поддерживать отдельных учащихся.

Ключевые слова: Объяснимый искусственный интеллект, XAI, Интеллектуальный анализ образовательных данных, EDM, прогнозирование производительности, Образование в области компьютерных наук, STEM-образование, Систематический обзор литературы.

EXPLAINABLE ARTIFICIAL INTELLIGENCE (XAI) FOR INTERPRETING STUDENT PERFORMANCE PREDICTION IN STEM AND COMPUTER SCIENCE: A SYSTEMATIC LITERATURE REVIEW

[1]A.A. Alibai*, **[2]G.K. Taganova**
[1,2]Astana International University, Astana, Republic of Kazakhstan
*e-mail: aidanaalibai9@gmail.com

A.A. Alibai – Bachelor student of the Higher School of Information Technology and Engineering, Astana International University, Astana, Kazakhstan, e-mail: aidanaalibai9@gmail.com, https://orcid.org/0009-0004-2908-3222

G.K. Taganova – Senior Lecturer, Research Supervisor, Astana International University, Astana, Kazakhstan, e-mail: guldana.kileuzhanova@gmail.com, https://orcid.org/0000-0001-6139-6270

Annotation. Educational Data Mining (EDM) enables the early identification of learning difficulties through the prediction of students' academic performance. However, machine learning models often operate as "black boxes," making their decision-making processes difficult to interpret. Explainable Artificial Intelligence (XAI) addresses this challenge by providing insights into why these models generate specific predictions. This article presents a systematic literature review of research published over the past five years concerning the application of Explainable Artificial Intelligence (XAI) to predict student performance in computer science and STEM education. The findings indicate that the most frequently used features were students' behavioral data and academic records, while the primary prediction tasks focused on grade prediction and identifying students at risk of academic failure. The study also analyzed the application domains of XAI, identifying global feature importance analysis, the explanation of individual predictions, and support for pedagogical interventions as the most prevalent areas. Furthermore, SHapley Additive exPlanations (SHAP) emerged as the most

widely used XAI method, applied predominantly for global model interpretation, with limited use for explaining individual predictions. Finally, a significant research gap was identified regarding the use of Explainable Artificial Intelligence (XAI) for course optimization, customized visualizations, and the generation of personalized recommendations. Addressing these gaps could enable educators to provide data-driven, personalized support to improve individual student outcomes.

Keywords: Explainable Artificial Intelligence, XAI, Educational Data Mining, EDM, Performance Prediction, Computer Science Education, STEM Education, Systematic Literature Review.

Кіріспе. 21 ғасырда информатика мен STEM-ге қатысты дағдылар маңызды бола бастады. Алайда, бұл пәндер көбінесе оқушыларға қиындық туғызады, өйткені олар математикалық амалдар мен абстрактілі ұғымдардың жиынтығын, бағдарламалау синтаксисін және тіпті тілді түсінуді қажет етеді, бұл жаңадан бастаушыларға қиындық тудырады (Келлехер, Пауш, 2005: 83–137, Чой және т.б., 2024: 1–8).

Білім беру деректерін өндіру (EDM) - машиналық оқытуды, терең оқытуды немесе статистикалық әдістерді пайдалана отырып, білім беру деректерін талдайтын деректерді өндіру бөлімі. EDM қолданудың бірі-оқушылардың үлгерімін болжау (Чой және т.б., 2023). Бұл оқушылардың табысқа жету жолында көмектесетін педагогикалық араласуларға мүмкіндік беру арқылы (Ромеро және т.б., 2008: 368–384) оқу проблемаларын ерте анықтауға ықпал етеді (Бейкер, Йасеф, 2009: 3–17). Дегенмен, машиналық және терең оқытудың дәстүрлі үлгілері әдетте "қара жәшіктер" болып табылады, бұл олардың неліктен нақты болжамдар беретінін түсінуді қиындатады.

Түсіндірілетін жасанды интеллект (XAI) әдістері мұғалімдерге оқушылардың жетістігіне немесе сәтсіздігіне ықпал ететін факторларды түсінуге көмектесу арқылы оқушылардың үлгерімін болжау үлгілерін түсіндіреді (Чой және т.б., 2024ә) (Чой және т.б., 2024a: 1–8). XAI әдістері жасанды интеллект модельдерінің шешімдерін неғұрлым түсіндіруге мүмкіндік береді, модельдердің ашықтығын арттырады және шешім қабылдау процестерін жақсартады (Гилпин және т.б., 2018: 80–89) (Адади, Беррада, 2018: 52138–52160).

Бұл жүйелі әдебиеттерге шолу студенттердің информатика және STEM білім беру саласындағы үлгерімі туралы болжамдарды түсіндірудің заманауи XAI әдістерін қарастырды. Білім берудің осы салаларына ерекше назар аударуымыздың себебі олардың болашақ технологиялық дамуға айтарлықтай әсері бар болып табылады. Зерттеу барысында келесі сұрақтар қойылған болатын.

Осы зерттеуге негізделген сұрақтар:

1. XAI зерттеулерінде оқушылардың үлгерімін түсіндіру үшін қандай мәліметтер жиынтығы, белгілер түрлері, болжау мақсаттары мен алгоритмдер қолданылды?
2. бұл зерттеулерде қандай XAI әдістері, интерпретация деңгейлері, бейнелеу әдістері және практикалық іске асыру қолданылды?

Шолуда әдебиеттегі ықтимал олқылықтарды анықтау және өнімділікті болжау модельдерінің интерпретациясы мен пайдалылығын жақсарту бағыттарын ұсыну үшін осы мәселелер қарастырылды.

Зерттеу материалдары мен әдістері. Қарастырылған әдебиеттерде зерттеушілер әдетте XAI-де "интерпретация" және "түсініктілік" терминдерін бір-бірінің орнына қолданады, бұл адамның жасанды интеллект моделінің неліктен нақты болжам жасағанын түсіну дәрежесін білдіреді (Гилпин және т.б., 2018: 80–89) (Матрани және т.б., 2021: 100060). Алайда, бірнеше зерттеулерде интерпретация мен түсініктілік жеке ұғымдар ретінде қарастырылды (Марцинкевичс, Фогт, 2020), онда интерпретация модельге тән мөлдірлікті білдіреді.

XAI саласындағы алғашқы шолулар интерпретацияның негізгі тұжырымдамаларын анықтауға және XAI әдістерін жіктеуге бағытталған. Адади мен Беррада (Адади, Беррада, 2018: 52138–52160)». 2018 жылы 2004 жылдан 2018 жылға дейін жарияланған 381 мақаланы қамтитын әдебиетке айтарлықтай шолу жасады. Олар XAI әдістерін Машиналық оқыту алгоритмдеріне, интерпретация деңгейіне (жергілікті немесе ғаламдық) және интерпретация уақытына (кірістірілген немесе пост-арнайы) қолданылуына қарай жіктеді.Ішкі интерпретация

олардың арқасында жанама түсіндірілетін модельдерге жатады сызықтық регрессия сияқты салыстырмалы түрде қарапайым құрылым. Фактіден кейінгі интерпретация күрделі "қара жәшік"модельдерінде қабылданған шешімдерді түсіндіру үшін қосымша әдістерді қолдануды қамтиды.

Гидотти және т.б. (Гвидотти және т.б., 2018: 1–42) 2019 жылы XAI әдістерін "қара жәшік" модельдерін түсіндіру, модельдерді тексеру, олардың нәтижелерін түсіндіру және мөлдір модельдерді құру үшін қолданылатын әдістерге бөлу үшін шолу жасады. Олар интерпретацияны бағалаудың ресми көрсеткіштерінің жоқтығын және белгісіз немесе жасырын ерекшеліктерге негізделген модельдерді түсіндірудің қиындығын атап өтті.

Арриета және т.б. (Арриета және т.б., 2020: 82–115) 2020 жылы мөлдір және post-hoc әдістері арасындағы айырмашылыққа және терең оқытуды түсіндіру әдістері үшін нақты Ішкі санаттарды анықтауға бағытталған шолуды ұсынды.

2023 жылы Чауши және оның авторлары (Чауши және т.б., 2023: 48–71) XAI-ге білім беруде шолу жасап, сенімділікті арттыру үшін ашықтық қажеттілігін атап өтті. Олар жасанды интеллекттің күрделі модельдерімен және интерпретацияға қойылатын талаптармен байланысты мәселелерді анықтады және білім берудегі XAI-ге қатысты болашақ зерттеулерге арналған бағыттарды ұсынды.

Алайда, қолданыстағы шолуларда XAI туралы құнды ақпарат болғанымен, олар негізінен XAI-ді білім берудің нақты салаларында қолдануға емес, технологиялық жағына қатысты болды.

Біздің әдебиеттерге жүйелі шолу информатика және STEM білім беру үшін EDM-де XAI әдістерін қолдану бойынша соңғы бес жылдағы (2020-2024) соңғы зерттеулерді талдау арқылы бұл олқылықты жоюға мүмкіндік берді. Біз ақпарат беруге, зерттеулердегі олқылықтарды анықтауға және болашаққа бағыттар ұсынуға тырыстық.

Бұл әдебиетке жүйелі шолу Китченхэм тұжырымдамасына сәйкес (Китченхем, Чартерс, 2007) инженерлік немесе жаратылыстану ғылымдары бағыты бойынша білім беру саласында жүйелі шолулар жүргізілді.

Іздеу стратегиясы. Біз информатика және инженерлік зерттеулер, ACM Digital Library, IEEE Xplore және Web of Science бойынша ең танымал дерекқорларды пайдалана отырып, әдебиеттерді жан-жақты іздедік.

Іздеу жолағы төрт негізгі тұжырымдамаға негізделген: білім беру деректерін өндіру, XAI әдістері, өнімділікті болжау және бағдарламалауды оқыту немесе STEM. Нақтылаудың бірнеше қайталануынан кейін біз іздеу жолағын тұжырымдадық:

("жасанды интеллект" немесе "AI" немесе "машиналық оқыту" немесе "терең оқыту" немесе "білім беру деректерін іздеу" немесе"EDM") және ("SHAP*" немесе "LIME"немесе" модельге тәуелді емес жергілікті түсіндірмелер", Немесе "PDP" немесе "ішінара тәуелділік графигі" немесе "Якорь "немесе" мұз "немесе" жеке шартты күту "немесе" але "немесе" жинақталған жергілікті әсер") және ("түсіндіру* "немесе" түсіндіру* "немесе" транспарациялау*") және ("өнімділік") және ("болжау*") Және ("бағдарламалау*" немесе "CS1" немесе "компьютер" немесе "STEM") және ("білім "немесе" оқыту* "немесе" мектеп "немесе " оқушы" немесе "сынып *" немесе "академиялық" немесе "курс" немесе "оқу бағдарламасы").

Қосу Және Алып Тастау Критерийлері. Біз таңдалған мақалалардың зерттеу міндеттерімізге сәйкес келуін қамтамасыз ету үшін қосу және алып тастау критерийлерін белгіледік.

Мақаланы қосу үшін мынадай критерийлер айқындалды: i) Машиналық оқыту немесе терең оқыту тәсілдерін пайдалана отырып, Информатика немесе STEM-білім беру саласындағы оқушылардың үлгерімін болжайтын зерттеулер; ii) болжау модельдерін түсіндіру үшін XAI әдістері анық қолданылатын зерттеулер; iii) XAI іске асыру әдістерінің нақты сипаттамаларын қамтитын зерттеулер; IV) рецензияланатын журналдарда немесе конференцияларда жарияланған зерттеулер.

Алып тастау критерийлері: i) студенттердің үлгерімін болжаусыз мінез-құлқын талдауға

бағытталған зерттеулер; ii) информатика немесе STEM білімімен байланысты емес зерттеулер; iii) ағылшын тілінде емес мақалалар; iv) эмпирикалық нәтижелерсіз аналитикалық мақалалар, шолулар немесе теориялық жұмыстар.

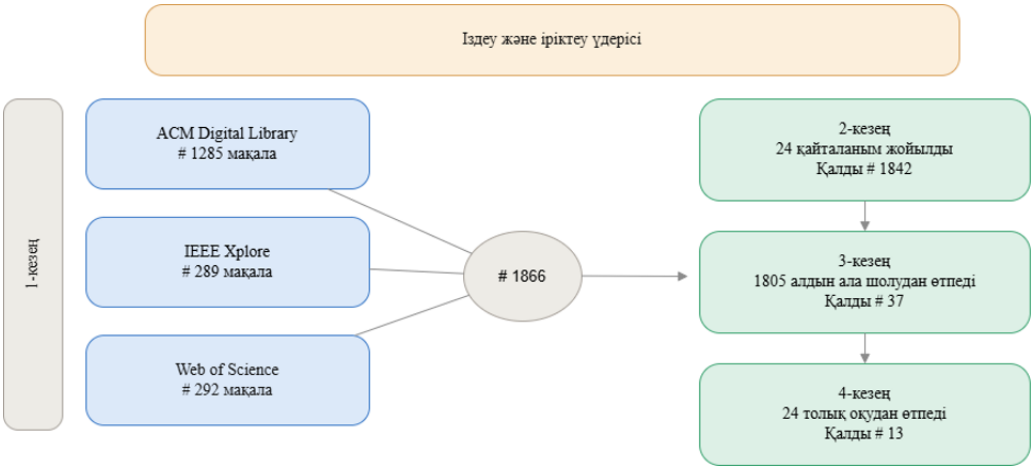
Іздеу және таңдау процесі. Суретте көрсетілгендей, біз іздеу және іріктеу процесінің төрт кезеңін қолданып, кейінгі талдау үшін ең маңызды әдебиеттерді анықтадық.

1 кезең: 2024 жылдың шілдесінде біз дерекқордағы тезистер мен толық мәтіндерге алдын ала анықталған іздеу жолын қолдандық, нәтижесінде 1866 мақала пайда болды.

2 кезең: тақырыпқа, авторларға және DOI-ге негізделген 24 телнұсқаны Алып тастау 1842 мақаланы қалдырды.

3 кезең: біз қосу және алып тастау критерийлерін тексеру үшін тақырыптарды, тезистерді және кілт сөздерді қолмен қарап шықтық, содан кейін кілт сөз тіркестері бойынша мәтіндік іздеу жүргіздік. Бұл тиісті абзацтар мақалалардағы XAI рөлін анықтау үшін оқылды. Бұл қадам бізге XAI-ді тек байланысты жұмыс ретінде атап өткен, бірақ оны оқушылардың үлгерімін болжау үшін арнайы қарастырмаған зерттеулерді алып тастауға мүмкіндік берді. Содан кейін біз 1805 зерттеуді алып тастадық, нәтижесінде 37 зерттеу қалды.

4-қадам: Қалған мақалалардың толық мәтінін іріктеу нәтижесінде егжей-тегжейлі метадеректер (мысалы, атауы, мақсаттары, қорытындылары, білім беру контексті) алынып, өзектілігіне негізделген қосымша алып тастаулар жасалды. Бұл соңғы қадам талдау үшін 13 мақала қалдырды.



Сурет 1. Іздеу және таңдау процесі [Search and selection process]

Зерттеулердің сапасын бағалау (Quality Assessment). Таңдалған зерттеулердің сапасы Kitchenham ұсынған жүйелі әдеби шолуларды жүргізу жөніндегі әдістемелік ұсынымдарға сәйкес бағаланды. Әрбір мақала зерттеу мақсаттарының айқындығы, қолданылған машиналық оқыту алгоритмдерінің негізділігі, Explainable Artificial Intelligence (XAI) әдістерінің сипатталу толықтығы және алынған нәтижелердің негізділігі сияқты критерийлер бойынша сарапталды. Осы талаптарға сәйкес келетін және STEM білімі саласында практикалық маңызы жоғары зерттеулер ғана қорытынды талдауға енгізілді.

Жүйелі қателіктер қаупін талдау (Risk of Bias). Шолу нәтижелерінің объективтілігін қамтамасыз ету мақсатында ықтимал жүйелі қателіктердің көздері талданды. Біріншіден, тек ағылшын тіліндегі жарияланымдардың енгізілуіне байланысты тілдік ауытқу (language bias) болуы мүмкін. Екіншіден, мақалаларды іріктеу қолмен жүргізілгендіктен, таңдау ауытқуының (selection bias) ықтималдығын азайту үшін алдын ала анықталған қосу және алып тастау критерийлері қатаң сақталды. Сонымен қатар, тек рецензияланатын ғылыми жарияланымдарды қамту арқылы жарияланымдық ауытқу (publication bias) ескерілді.

Нәтижелер мен талқылау. Rq1: деректер жиынтығы, сипаттамалары, болжау мақсаты және алгоритмдер. Деректер жиынына келетін болсақ, 1-кестеде көрсетілгендей, зерттеулердің 69,2% - ы өздігінен жиналған деректер жиынын пайдаланды, ал 30,8% -

ы STEM-мен байланысты Open University Learning Analytics (OULAD) деректер жиынын пайдаланды (Кузилек және т.б., 2017: 1–8).

Кесте 1. Деректер жиынтығы мен деректер көздерін тарату [Distribution of datasets and data sources]

Дерек көзі	Қолжетімділік	N	Пайыз
LMS	Өзі жинаған	4	30,8%
LMS	Жалпыға қолжетімді	4	30,8%
AJS	Өзі жинаған	2	15,4%
SIS	Өзі жинаған	2	15,4%
Басқалар	Өзі жинаған	1	7,7%

Деректер көздеріне келетін болсақ, ең көп таралған деректер көзі зерттеудің 61,5%-ында қолданылған оқуды басқару жүйелері (LMS) болды, одан кейін автоматтандырылған бағалау жүйелері (AJS) және оқушылардың ақпараттық жүйелері (SIS) әрқайсысы 15,4%-дан, ал 7,7%-ы оқытушылардың сауалнамалары мен бақылауларына сүйенді (Джаннакас және т.б., 2021).

Іріктеме өлшемдері 100-ден кем емес және 30 000-нан астам оқушыға дейін болды, OULAD сияқты үлкен деректер жиынтығы жоғары болжам дәлдігін қамтамасыз етті. Мысалы Аднан және басқалары (Аднан және т.б., 2022: 129843–129864) жүргізген зерттеу бойынша 91,9% құрады. Кішігірім деректер жиынтығын да құнды ақпарат берді. Мысалы, Рико-Хуан және т.б. (Рико-Хуан және т.б., 2023: 955–969) 90 оқушыдан алынған деректерді пайдалана отырып, AUC 0,70-ке жетті.

Нысан түрлері және кіріс айнымалылары. Болжалды модельдер кіріс ретінде қызмет ететін әртүрлі нысандар негізінде құрылады. 1-кестеде көрсетілгендей, әрбір нысан түрі белгілі бір кіріс айнымалыларын қамтиды.

Біз әртүрлі рецензияланған мақалалардың сипаттамалары әртүрлі екенін анықтадық, бұл мұғалімдерге "қара жәшік" үлгілері туралы нақты түсінік алуды қиындатады. XAI әдістері сипаттамалардың үлесін нақтылау және мағыналы интерпретацияны қолдау үшін қажет.

Зерттеулердің 29,7% пайдаланылған мінез-құлық деректері басу және қатысу деректері сияқты айнымалыларды қамтиды. Мысалы, Swamy және басқалары (Суэми және т.б., 2023: 345–356) Scala бағдарламалау курсының сәтсіздікке ұшырау қаупін бейне мен викторинаны көруді қадағалау арқылы болжады. Вахид және басқалары (Уахид және т.б., 2023) Moodle-мен өзара әрекеттесу негізінде STEM курсынан өту немесе сәтсіздік нәтижелерін болжады.

Зерттеулердің 27,0% қолданылатын академиялық үлгерім туралы мәліметтер, мысалы, бағалау нәтижелері және курстың аяқталу мәртебесі нақты жетістік көрсеткіштерін береді. Мысалы, Аднан және басқалары (Аднан және т.б., 2022: 129843–129864) оқушылардың нәтижелерін келесідей жіктеу үшін бағаларды қадағалады: ерекшелену, емтиханнан өту, сәтсіздікке ұшырау немесе кету.

Демографиялық деректер (16,2%) жас және жыныс сияқты айнымалылармен кеңірек контекст береді. Бағдарламалау деректері (13,5%) код пен синтаксистегі қателер сияқты айнымалыларды қамтиды. Мәтінмәндік деректер (13,5%) курстың күрделілігі, ұпайлар және сабақ схемалары сияқты айнымалыларды қамтиды.

Олардың ішінде мінез-құлық сипаттамалары және өнімділік көрсеткіштері ең көп қолданылатын болып ғана емес, сонымен қатар XAI қосымшаларында түсіндіру үшін ең маңызды болып табылады, өйткені олар оқушылардың қатысуы мен жетістіктерін тікелей көрсетеді.

Кесте 2. Кіріс айнымалыларының түрлері [Types of input variables]

Түрі	**Кіріс айнымалылары**	**Пайыз**	**Авторлар**
Мінез-құлықтық	Курс бойынша кликстрим деректері, мәтіндік пікірлер, қатысу, қатысушылық	29,7%	Swamy және т.б., Alrajhi және т.б., Rico-Juan және т.б.
Академиялық үлгерім	ОБК, курсты аяқтау мәртебесі, бағалау балдары	27,0%	Lu және т.б., Giannakas және т.б.
Демографиялық	Жынысы, жасы, ең жоғары білім деңгейі	16,2%	Baranyi және т.б., Mendez және т.б.
Бағдарламалау деректері	Код, пернелерді басу кідірісі, бағдарламалау конструкциялары, синтаксис қателері	13,5%	Pereira және т.б., Waheed және т.б.
Контексттік	Курстың күрделілігі, кредиттер, белсенділік үлгілері	13,5%	Rahul және Katarya, Giannakas және т.б.

Болжау мақсаттары мен алгоритмдері. Болжау мақсаттарын түсіну өте маңызды, өйткені олар XAI әдістері түсіндіруге тырысатын контекстті қамтамасыз етеді. 2-кестеде көрсетілгендей, мақсаттар мен болжау алгоритмдері арасында нақты байланыс бар.

Ең көп таралған мақсат - курстың сәтсіздікке ұшырау қаупі, өйткені зерттеулердің 61,5% студенттердің емтихан тапсыратынын, сәтсіздікке ұшырайтынын немесе оқудан бас тартатынын болжайды. Random Forest (Ригатти, 2017: 31–39), XGBoost (Чен, Гестрин, 2016: 785–794) және LSTM (Грефф және тб., 2016: 2222–2232) сияқты жіктеу алгоритмдері ең көп қолданылады. Мысалы, Лу және оның авторлары (Лу және тб., 2023: 1–9) STEM курсында xgboost көмегімен оқуды болжауда 90% дәлдікке қол жеткізді.

Кесте 3. Студенттердің үлгерімін болжауда қолданылған XAI техникаларына шолу [Review of XAI techniques used in student performance prediction]

Авторлар	Болжау мақсаты	Мақсатты айнымалылар	Үздік алгоритмдер	XAI техникасы	Деңгей	Визуализация әдістері
Swamy және т.б.	Курстан қалу қаупі	Scala бағдарламалауда: өтті/өтпеді	BiLSTM, дәлдігі 90%	SHAP, LIME	Жаһандық	XAI ұқсастық картасы; белгілер маңыздылығының картасы
Baranyi және т.б.	Курстан қалу қаупі	STEM курсын бітірді/шықты	FCNN, дәлдігі 72,4%	SHAP, Permutation Importance	Жаһандық	SHAP жиынтық диаграммасы; маңыздылық кестесі
Alrajhi және т.б.	Студент белсенділігі	Big Data курсындағы пікір кезектілігі	BERT, дәлдігі 92%	Integrated Gradients	Жеке	VisualizationDataRecord (Captum [36])
Mendez және т.б.	Баға немесе ұпай	CS курсындағы баға	GBT (дәлдік көрсетілмеген)	SHAP	Жеке	Дербес дөңгелек диаграммалар (SHAP мәндері)
Adnan және т.б.	Курстан қалу қаупі	STEM: үздік/өтті/ өтпеді/шықты	Random Forest, дәлдігі 91,9%	SHAP	Жаһандық, Жеке	Жаһандық: жиынтық, тәуелділік диаграммасы; Жеке: жергілікті түсіндіру
Pereira және т.б.	Курстан қалу қаупі	Python курсында: өтті/өтпеді	XGBoost, дәлдігі 81,3%	SHAP	Жаһандық, Жеке	Жаһандық: бағаналы, жиынтық; Жеке: шешім, күш диаграммасы
Rahul және Katarya	Курстан қалу қаупі	STEM курсында: өтті/өтпеді	EDL_KDF, дәлдігі 99,6%	SHAP	Жаһандық	Бағаналы диаграмма, жиынтық диаграмма
Rico-Juan және т.б.	Курстан қалу қаупі	Бағдарламалау тапсырмасы: өтті/өтпеді	Random Forest, AUC 0,70	SHAP	Жаһандық	Жиынтық диаграмма, тәуелділік диаграммасы
Lu және т.б.	Курстан қалу қаупі	STEM курсынан шықты/шықпады	XGBoost, дәлдігі 90%, AUC-ROC 0,89	SHAP	Жаһандық, Жеке	Жаһандық: жиынтық, бағаналы, тәуелділік, өзара әрекет, SHAP картасы; Жеке: сарқырама диаграммасы
Giannakas және т.б.	Баға немесе ұпай	Бағдарламалық инженерия: A немесе F	DNN, дәлдігі 82,39%	SHAP	Жаһандық	Бағаналы диаграмма, жиынтық диаграмма
Waheed және т.б.	Курстан қалу қаупі	STEM курсында: өтті/өтпеді	LSTM, дәлдігі 84,57%, AUC 0,82	SHAP	Жаһандық	Жиынтық диаграмма, тәуелділік диаграммасы
Effenberger және Pelánek	Кодты түсіну	Python курсындағы код кластерлері	Дербес кластерлеу алгоритмі	Түсіндірілетін кластерлеу алгоритмі	Жаһандық	Кластерленген белгілер матрицасы, кластер жиынтық диаграммасы
Afzaal және т.б.	Баға немесе ұпай	Бағдарламалау тапсырмаларының нәтижелері	ANN, дәлдігі 90%	Counterfactual Explanations	Жеке	Кері байланыс бақылау тақтасы (DICE)

ISSN (print) 3135-1840 eISSN (online) 3135-1859 Smart Technologies Journal, 2(2) · 2026

3-кестеде зерттеулердің 23,1% нәтижелерін болжау үшін gradientboosting Trees (GBT) (Фридман, 2001: 1189–1232), жасанды нейрондық желілер (ANN) (Егнанараяна, 2009) және терең нейрондық желілер (DNN) (Ше және т.б., 2017: 2295–2329) сияқты күшейтетін алгоритмдер немесе нейрондық желілер қолданылды. Мысалы, Мендес және басқалар. (Мендес және т.б., 2021) информатика курсының тұрақты бағаларын болжау үшін регрессия GBT қолданды. "Кодты түсіну" санатында Эфенбергер мен Пеланек (Эффенбергер және Пеланек, 2021: 101–112) Python курсында код ерекшеліктеріне негізделген студенттердің бағдарламалық шешімдерін топтастыру үшін кластерлеуді қолданды. Студенттерді Тарту үшін Алраджи және т.б. (Әлражхи және т.б., 2023) трансформаторлардан Екі Бағытты Кодтаушы Көріністері бар қолданбалы мәтінді өндіру (Девлин, 2018) (BERT негізіндегі нейрондық желі) жеделдігін жіктеу үшін студент Big Data курсында түсініктеме беріп, 92% жоғары дәлдікке қол жеткізді.

Rq2: XAI әдістері, интерпретация деңгейлері, бейнелеу әдістері және практикалық іске асыру. Болжамды модельдердің интерпретациясын жақсарту үшін әртүрлі XAI әдістері қолданылды. Ең көп қолданылатын әдіс ретінде Шеплидің қосымша түсіндірмелері (SHAP) (Лундберг, Ли, 2017) 11 зерттеуде келтірілген. Мысалы, Перейра (Перейра және т.б., 2021: 117097–117119) XGBOOST моделімен SHAP-ті оқушылардың негізгі мінез-құлқын анықтау үшін қолданды, мұғалімдерге табысқа қандай факторлар ықпал ететінін түсінуге көмектесті және Python курсында тәуекел тобындағы студенттерге мақсатты қолдау көрсетті. SHAP-тің танымалдылығын оның ойын теориясындағы берік теориялық негізімен (Шепли, 1953), жергілікті және жаһандық түсініктемелер беру қабілетімен, машиналық оқыту модельдерінің көпшілігімен жұмыс істеу икемділігімен және ашық бастапқы кітапханалармен қол жетімділігімен түсіндіруге болады.

LIME. Тек бір зерттеуде жергілікті түсіндірілетін модельдік диагностикалық түсініктемелер (LIME) қолданылды (Рибейро и др., 2016: 1135–1144), әр жағдайға жеңілдетілген модельдер жасау арқылы жеке болжамдарды түсіндіретін кеңінен қолданылатын әдіс. LIME нақты болжамдар үшін күрделі модельдің мінез-құлқын жуықтау үшін сызықтық регрессиялар немесе шешім ағаштары сияқты жергілікті суррогат модельдерін жасайды. Атап айтқанда, Swamy (Суэми және т.б., 2023: 345–356) викториналар жіберу және бейнелерге қатысу сияқты сәттілікке әсер ететін маңызды факторларды анықтау үшін LIME қолданды.

Ауыстырудың маңыздылығы. Зерттеулердің бірінде Permutation Importance (Альтман және т.б., 2010: 1340–1347) әдісі қолданылды, ол әр белгінің мәндерін араластыру және модельдің дәлдігіне әсерін өлшеу арқылы белгілердің маңыздылығын бағалауға мүмкіндік береді. Дәлдіктің айтарлықтай төмендеуі маңызды белгіні көрсетеді. Бараньи және басқалар (Бараньи және т.б., 2020: 13–19) оны STEM курсының маңызды көрсеткіштері ретінде университетке түсу кезінде балл сияқты факторларды растай отырып, бітіруді болжаушыларды тексеру үшін пайдаланды.

Біріктірілген градиенттер. Альраджи және басқалар (Альраджхи және т.б., 2023) оқушылардың шұғыл түсініктемелерін алу үшін Bert моделін түсіндіру үшін объектілерге апаратын жолдар бойындағы градиенттерді есептейтін (Сундарараджан және т.б., 2017: 3319–3328) интеграцияланған градиенттерді қолданды. Бұл әдіс мұғалімдерге жедел жауап беруді қажет ететін аймақтарды анықтауға және модельдің ашықтығын арттыруға көмектесетін кілт сөздерді бөліп көрсетті.

Контрафактілік Түсініктемелер. Афзаалда және басқалар (Афзаал және т.б., 2021: 37–42), контрафактілік түсініктемелер (Вахтер және т.б., 2017: 841) мүмкіндік мәндерін реттеу арқылы "what-if" сценарийлерін жасау үшін пайдаланылды. Бұл тәсіл енгізу мүмкіндіктеріндегі ең аз өзгерістердің модельдік болжамдарға қалай әсер ететінін зерттейді. Бұл зерттеу бағдарламалау курсында студенттердің мүмкіндіктерін кеңейту үшін жекелендірілген ұсыныстарды пайдалануға бағытталған, бұл олардың үлгерімін жақсарту үшін мақсатты әрекеттерді орындауға мүмкіндік береді.

Пайдаланушы түсіндіретін кластерлеу алгоритмі. Эфенбергер және Пеланек (Эффенбергер, Пеланек, 2021: 101–112) Python курсында оқушылар ұсынған кодты

жіктеу үшін пайдаланушының интерпретацияланған кластерлеу алгоритмін цикл және шартты операторлар сияқты стандартты бағдарламалау үлгілерінде пайдаланды. Бұл тәсіл мұғалімдерге жалпы кодтау әдістерін жылдам анықтауға мүмкіндік береді, бұл мақсатты кері байланысты жеңілдетеді және бағдарламалау тапсырмаларындағы өнімділік деректерінің интерпретациясын арттырады. Әр түрлі XAI әдістері қолданылғанымен, қарастырылған зерттеулерде жинақталған жергілікті эффекттерде (ALE) (Апли, Жу, 2020: 1059–1086) байланыстыру (Рибейро және т.б., 2018) және терең оқытудың маңызды функциялары (DeepLIFT) (Шрикумар және т.б., 2017: 3145–3153) сияқты әдістер айтарлықтай жетіспеді. Бұл алшақтық болашақ зерттеулерге осы әдістерді оқушылардың үлгерімін болжау модельдеріне енгізу мүмкіндігін ашады. Біз XAI-дің информатика және STEM білім берудегі практикалық қосымшаларына арналған әдебиеттерге бірде-бір шолуды таппадық. Бұл оқылықты жою үшін біз 1-кестеде білім беру нәтижелеріне ықпал ететін тиісті қосымшаларды қорытындыладық.

Кесте 4. CS және STEM білімінде XAI қолдану [Application of XAI in CS and STEM education]

Санат	**Пайыз**
Жаһандық белгілердің маңыздылығын талдау	33,3%
Жеке болжамды түсіндіру	20%
Араласуларды және шешім қабылдауды қолдау	26,7%
Белгілердің өзара әрекеттесуін зерттеу	13,3%
Курсты жақсартуға бағытталған іс-шаралар	3,3%
Ұсынымдар жасау	3,3%

Жаһандық белгілердің маңыздылығын талдау. Қарастырылған зерттеулердің 33,3% құрайтын негізгі қолдану жаһандық белгілердің маңыздылығын талдау болып табылады, оның мақсаты оқушылардың жалпы үлгерімі үшін ең маңызды белгілерді анықтау болып табылады. Мысалы, Перейра және бірлескен авторлар (Перейра және т.б., 2021: 117097–117119) студенттердің үлгеріміне әсер ететін маңызды факторларды талдау үшін SHAP жиынтық кестелерін қолданды және бірінші емтихан бағасы мен жүйеге кіруге кететін уақыт CS1 курсының оқу үлгерімінің негізгі факторлары екенін анықтады.

Жеке болжамды түсіндіру. Зерттеулердің 20% - ы жеке болжамды түсіндіру арқылы жеке оқушылардың мінез-құлық үлгілеріне әсер ететін, мақсатты ақпараттарды бөліп көрсетеді. Мысалы, Лу (Лу және т.б., 2023: 1–9) ШАП сарқырамасының кестесін (SHAP Waterfall Plot) ресурстар мен тапсырмаларды басу сияқты функциялардың әсерін көрсету үшін және бұл факторлар оқушының STEM курсында оқуын жалғастыру немесе оны тастап кету ықтималдығына қалай әсер ететінін түсіндіру үшін пайдаланды.

Қолдау Шаралары Және Шешім қабылдау. Зерттеулердің 26,7% - ында XAI қосымшалары оқытушыларға оқушылардың үлгерімін қолдау үшін уақытылы араласуға көмектеседі. Мысалы, Алраджи және басқалар (Альраджхи және т.б., 2023) Captum визуализациясын (Кохликян және т.б., 2020) мұғалімдерге шұғыл назар аударуды қажет ететін түсініктемелерді анықтауға көмектесу үшін, қажет болған жағдайда мақсатты қолдау көрсетуге мүмкіндік беру үшін қолданды.

Ерекшеліктердің Өзара Әрекеттесуін Зерттеу. Зерттеулердің 13,3% -ын құрайтын функционалдық өзара әрекеттесуді зерттеу әртүрлі факторлардың оқушылардың үлгеріміне қалай әсер ететінін зерттейді. Мысалы, Вахид және басқалар (Уахид және т.б., 2023) викторинаға қатысу мен курстық тапсырмалар арасындағы байланысты зерттеу үшін SHAP тәуелділік сюжеттерін пайдаланды. Олар екі іс-шараға да дәйекті қатысу сәтсіздік қаупі төменгі деңгеймен байланысты екенін анықтады, бұл жиынтық оқушылардың үлгеріміне жиынтық оң әсерін көрсететінін анықтады.

Курсты практикалық жетілдіру. Қарастырылған зерттеулердің біреуінде ғана ХАІ курстың жақсаруын іздеу үшін пайдаланылды. Свами және басқалары (Суэми және т.б., 2023: 345–356) сапалы сауалнамалар жүргізіп, оқытушылардың 85% - дан астамы викториналық тапсырмаларды түзету сияқты бағдарламалау курсының элементтерін нақтылау үшін ХАІ идеяларын бағалағанын анықтады.

Ұсыныстар Жасау. Сондай-ақ, қарастырылған зерттеулердің бірінде жекелендірілген кері байланысқа бағытталған ұсыныстар жасалды. Афзаал және басқалары (Афзаал және т.б., 2021: 37–42) арнайы кеңестер ұсынатын бақылау тақтасын жасады. Мысалы, егер викториналық ұпайлар жақсартуды ұсынса, онда тақтасы студенттерге тиісті оқу материалдары бойынша ұсыныстар жасайды, бұл студенттерге бағдарламалау дағдыларын жетілдірудің практикалық стратегияларына назар аударуға көмектеседі.

Осылайша, жекелендірілген ұсыныстар жасау үшін ХАІ пайдалануда айтарлықтай алшақтық бар. Бұл олқылықты жою оқытушыларға жеке студенттерді жақсырақ қолдау үшін жекелендірілген, деректерге негізделген ұсыныстар ұсынуға мүмкіндік береді. Сонымен қатар, ХАІ курсын жетілдіру бойынша ұсыныстар шектеулі болып қала береді, бұл оқытуды жақсарту және оқу нәтижелерін оңтайландыру үшін қосымша мүмкіндіктер ашады.

Түсіндіру деңгейлері. Оқушылардың оқу үлгерімін болжау үшін ХАІ әдістерін зерттеген кезде, түсіндіру әдістері жаһандық және жеке (жергілікті) деңгейлерге жіктеледі, олардың әрқайсысы ерекше түсініктер ұсынады. Қарастырылған жеті зерттеу жаһандық деңгейдегі түсіндірмелерге, үшеуі жеке деңгейдегі түсіндірмелерге бағытталған, ал үшеуі екеуін де қамтиды.

Әлемдік деңгей. Әлемдік түсіндірме барлық болжамдар бойынша мүмкіндіктердің маңыздылығы туралы түсінік береді, бұл деректер жиынтығы бойынша үлгілерді бақылауды жеңілдетеді. Қарастырылған зерттеулерде SHAP әдетте мүмкіндіктердің маңыздылығы мен тәуелділіктерін әлемдік деңгейде анықтау үшін қолданылады. Ол жан-жақты түсініктер бергенімен, SHAP жоғары есептеу ресурстарын, әсіресе жоғары өлшемді деректер үшін қажет, бұл өңдеу шығындарын арттыруы мүмкін.

Жеке деңгей. Жеке түсіндіру мүмкіндігі жеке болжамдарға ерекшеліктердің әсеріне бағытталған, жеке оқушыларға арналған жекелендірілген талдауларды қолдайды. Интеграцияланған градиенттер (Альраджхи и др., 2023) және контрафактілік түсіндірмелер (Афзаал және т.б., 2021: 37–42) сияқты әдістер ерекшеліктерге негізделген үлестер мен нәтижеге әсер ететін минималды өзгерістерді қамтамасыз етеді, бұл жеке араласуларға мүмкіндік береді.

Екінші жағынан, LIME шолудан өткен зерттеулердің бірінде пайда болды (Суэми және т.б., 2023: 345–356). LIME болжамды жақындату үшін жергілікті суррогат моделін жасау арқылы SHAP-қа қарағанда жеке түсіндірмелердің интуитивті визуализациясын ұсынады. Бұл пайдаланушыларға әрбір мүмкіндіктің белгілі бір жағдай үшін болжамды нәтижеге қалай үлес қосатынын жақсы түсінуге көмектеседі.

Қарастырылған зерттеулердің көпшілігі жаһандық түсіндірулерге баса назар аударады, дегенмен жеке деңгейдегі талдау жеке тұлғалар, білім беру араласулары үшін шешуші болып қала береді. Сонымен қатар, Свами және басқалар (Суэми және т.б., 2023: 345–356) атап өткендей SHAP және LIME сияқты әртүрлі ХАІ әдістері бір деректер жиынтығында әртүрлі түсініктемелер бере алады. Бұл мәселелер зерттеушілер неліктен қолдау үшін бір ХАІ әдісін қолданғанын түсіндіре алады түсініктемелердің дәйектілігі. Бұл мәселелер зерттеушілердің түсіндірмелердің дәйектілігін сақтау үшін ХАІ әдісін не үшін қолданғанын түсіндіруі мүмкін. Өнімділікті болжау модельдерін жасау кезінде болжау дәлдігі мен интерпретация арасындағы тепе-теңдікті қамтамасыз ету маңызды.

Кесте 5. Машиналық оқыту және терең оқыту [Machine learning and deep learning]

Түрі	Пайыз	Алгоритм	Модель дәлдігі	Үздік алгоритм
Дәстүрлі машиналық оқыту	53,8%	GBT, ANN, Random Forest, XGBoost, Custom clustering	88,3% ± 4,8	Random Forest, дәлдігі 91,9%
Терең оқыту	46,2%	BiLSTM, FCNN, BERT, EDL_KDF, DNN, LSTM	86,8% ± 9,3	EDL_KDF, дәлдігі 99,6%

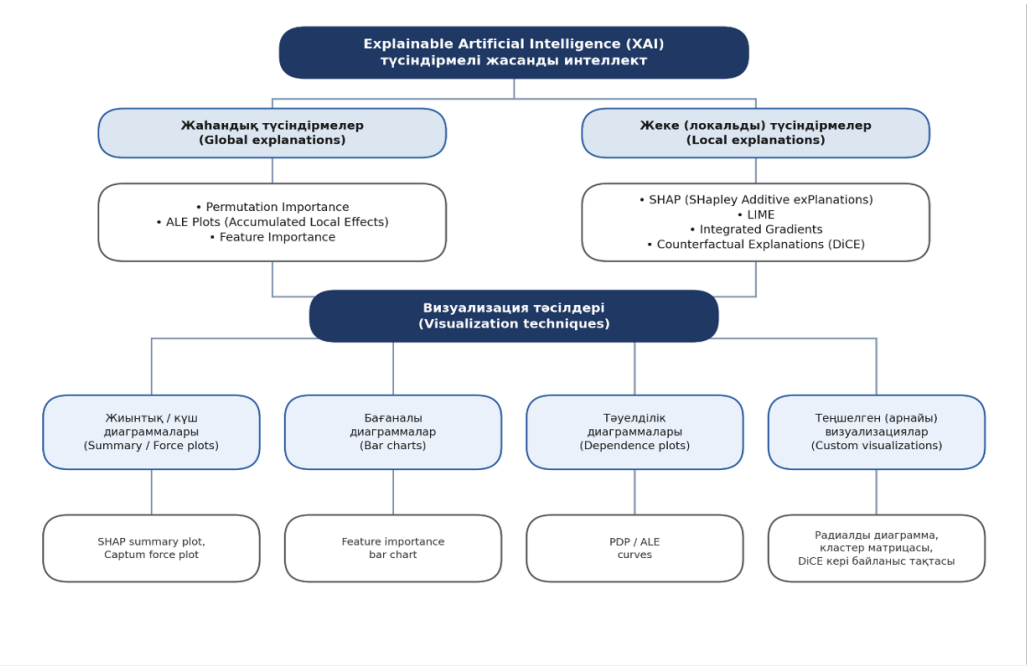
Дәстүрлі машиналық оқыту. 5-кестеде көрсетілгендей, GBT, Random Forest, XGBoost, ANN және арнайы кластерлеу алгоритмдері қарастырылып отырған тәсілдердің 53,8%-ын құрайды. Бұл дәстүрлі машиналық оқыту модельдері орташа дәлдікті 88,3 ± 4,8% құрайды, ал кездейсоқ орман 91,9% ең жақсы нәтижелерге қол жеткізеді (Аднан және т.б., 2022: 129843–129864). Бұл модельдер түсіндіруді жақсарту үшін SHAP және контрафактілік түсіндірмелер сияқты XAI әдістерін пайдаланады.

Терең оқыту. Екі бағытты LSTM (BiLSTM), толықтай байланысты нейрондық желілер (FCNN), BERT, айқын терең оқыту білімін бөлу фреймворкі (EDL_KDF), DNN және LSTM сияқты озық архитектуралар әдістердің 46,2%-ын құрайды және орташа дәлдік 86,8% ± 9,3%-ға жетеді. Бұл өзгергіштік терең оқытудың одан да жоғары өнімділікті қамтамасыз ету мүмкіндігі бар екенін көрсетеді. Терең оқыту дәстүрлі түрде онша түсіндірілмейтін болып саналса да (Мин және т.б., 2022: 3503–3568), бұл шолу терең оқыту модельдерін түсіндіру үшін SHAP, LIME және интеграцияланған градиенттер сияқты әртүрлі XAI әдістерін пайдалануға болатынын анықтады. Терең оқыту моделі As DAbDBi-LSTM және ResNet-101 желілерін біріктірген EDL_KDF терең оқыту құрылымын (Рахул, Катарйа, 2024: 1–8) пайдаланып, ең жоғары 99,6% дәлдікке қол жеткізді. Бұл құрылым дәстүрлі машиналық оқыту модельдерінен асып түсті. Сондықтан, егер болжау дәлдігі негізгі мақсат болса және терең оқыту модельдерінің түсіндірілу мәселелерін шешуге болатын болса, терең оқыту алгоритмдері бірінші таңдау болуы керек.

Оқушылардың үлгерімін болжауға арналған XAI визуализация әдістерін екі негізгі санатқа бөлуге болады.

XAI арқылы жасалған графиктер. Бұл тәсіл кітапханалардан алынған стандартты графиктерді пайдаланып, модель болжамдарына ерекшеліктердің әсерін визуализациялады, бұл оқытушылар мен зерттеушілерге күрделі модельдерді түсінуге көмектеседі. Альраджхи және басқалары (Альраджхи және т.б., 2023) Captum интеграцияланған градиенттерін пайдаланып, түсініктемелерде жеделдік жіктемелеріне әсер ететін кілт сөздерді белгіледі. Сонымен қатар, Baranyi және басқалары (Бараньи және т.б., 2020: 13–19) нақты факторлардың бітіру ықтималдығына қалай әсер ететінін көрсету үшін SHAP қорытынды графиктерін пайдаланды. Қарастырылған мақалалар sci-kit-learn (Рико-Хуан және т.б., 2023: 955–969), TensorFlow (Суэми және т.б., 2023: 345–356) және PyTorch (Афзаал және т.б., 2021: 37–42) бағдарламаларындағы XAI кітапханаларын пайдаланып жасалған визуализациялардың түсіндірілетін шешімдерді жылдамырақ әзірлеуге мүмкіндік беретінін көрсетті.

Теңшелген визуализациялар. Бұл тәсіл XAI нәтижелерін қолжетімділікті жақсарту үшін теңшейді, бұл оқытушылар мен оқушылардың түсіндіруін жеңілдетеді. Мендес және т.б. (Мендес және т.б., 2021) ерекшелік үлестерін көрнекі түрде нақтылау үшін SHAP мәндеріне негізделген дөңгелек диаграммалар жасады. Дегенмен, бұл бағытты зерттеген зерттеулер аз, бұл оқытушы мен оқушыға ыңғайлы визуализация құралдарын әзірлеудегі алшақтықты көрсетеді. Мұны шешу білім беру мекемелерінде XAI әдістерінің қолжетімділігін және практикалық құндылығын арттыруы мүмкін.



Сурет 2. Explainable Artificial Intelligence (XAI) әдістері мен визуализация тәсілдерінің жіктелуі [Classification of Explainable Artificial Intelligence (XAI) methods and visualization techniques]

Зерттеудің шектеулері (Limitations). Осы жүйелі шолудың нәтижелерін интерпретациялау кезінде бірнеше әдістемелік шектеулерді ескеру қажет. Біріншіден, әдебиеттерді іздеу тек таңдалған негізгі ғылыми дерекқорлармен (ACM Digital Library, IEEE Xplore және Web of Science) шектелді. Сонымен қатар, талдауға тек ағылшын тіліндегі басылымдар ғана енгізілді, бұл басқа тілдердегі құнды зерттеулердің ескерілмеуіне байланысты «тілдік ауытқуға» (language bias) алып келуі ықтимал. Екіншіден, қосу және алып тастаудың қатаң критерийлеріне сәйкес, шолуға іріктелген зерттеулердің саны шектеулі, бұл алынған қорытындыларды жалпылауға белгілі бір шектеулер қояды. Түсіндірмелі жасанды интеллект (XAI) саласының өте қарқынды дамып жатқанын ескерсек, іздеу процесі аяқталғаннан кейін жарық көрген ең жаңа немесе инновациялық еңбектер осы шолуға енбей қалуы мүмкін. Сонымен қатар, зерттеудің тек STEM және компьютерлік ғылымдар саласына бағытталғандығы алынған қорытындыларды басқа гуманитарлық немесе әлеуметтік пәндерге тікелей қолдану мүмкіндігін шектейді.

Қорытынды. Бұл жүйелі әдебиеттік шолуда STEM пәндері мен компьютерлік ғылымдар аясында үлгерімді болжау модельдерін интерпретациялау үшін түсіндірмелі жасанды интеллект (XAI) құралдарын қолданудың қазіргі жағдайы зерттелді.

Зерттеудің негізгі нәтижелері келесідей тұжырымдалды:

Деректер мен белгілерді талдау (RQ1): Жұмыстардың басым бөлігі тікелей оқытуды басқару жүйелерінен (LMS) алынған авторлық деректер жиынтығына негізделген. Студенттердің мінез-құлық үлгілері мен академиялық көрсеткіштері ең жиі қолданылатын белгілер болып табылады, ал болжамның басты мақсаты — үлгермеушілік қаупін анықтау. Қолданылатын белгілердің әртүрлілігі педагогтар үшін «қара жәшік» модельдерінің логикасын түсінуді қиындатады, бұл XAI әдістерін енгізу қажеттілігін арттырады.

XAI-ды қолдану салалары (RQ2): Әдістер көбінесе белгілердің жаһандық маңыздылығын талдау, жеке болжамдарды түсіндіру және педагогикалық шешімдерді қабылдауды қолдау үшін пайдаланылады. SHAP әдісі ең көп таралған құрал ретінде анықталды, бірақ ол негізінен жаһандық деңгейде қолданылып, жеке жағдайларды талдауда шектеулі болып отыр. Сонымен қатар, ALE, Anchors және DeepLIFT сияқты әдістердің әдебиетте кездеспеуі болашақ зерттеулер үшін жаңа мүмкіндіктерді көрсетеді.

Анықталған олқылықтар: XAI-ды оқу курстарын жетілдіру, бейімделген

визуализацияларды жасау және персоналдандырылған ұсыныстар беру үшін қолдануда айтарлықтай тапшылық байқалады. Болжау дәлдігі басымдыққа ие болған жағдайда, терең оқыту (deep learning) модельдері дәстүрлі алгоритмдерге қарағанда жоғары тиімділік көрсетті. Аталған ғылыми олқылықтарды жою білім берудегі XAI-дың практикалық маңызын арттыруға мүмкіндік береді. Бұл оқытушыларға STEM және компьютерлік ғылымдар саласындағы студенттерге неғұрлым тиімді әрі дербес қолдау көрсетуге жағдай жасайды.

References

Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). IEEE Access, 6, 52138–52160.

Adnan, M., Uddin, M. I., Khan, E., Alharithi, F. S., Amin, S., & Alzahrani, A. A. (2022). Earliest possible global and local interpretation of students' performance in virtual learning environment by leveraging explainable AI. IEEE Access, 10, 129843–129864. https://doi.org/10.1109/ACCESS.2022.3227072

Afzaal, M., et al. (2021). Generation of automatic data-driven feedback to students using explainable machine learning. In International Conference on Artificial Intelligence in Education (pp. 37–42). Springer.

Alrajhi, L., Pereira, F. D., Cristea, A. I., & Alamri, A. (2023). Serendipitous gains of explaining a classifier: Artificial versus human performance and annotator support in an urgent instructor-intervention model for MOOCs. In Proceedings of the 6th Workshop on Human Factors in Hypertext. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3603607.3613480

Altmann, A., Toloşi, L., Sander, O., & Lengauer, T. (2010). Permutation importance: A corrected feature importance measure. Bioinformatics, 26(10), 1340–1347.

Apley, D. W., & Zhu, J. (2020). Visualizing the effects of predictor variables in black box supervised learning models. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 82(4), 1059–1086.

Arrieta, A. B., et al. (2020). Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion, 58, 82–115.

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.

Baranyi, M., Nagy, M., & Molontay, R. (2020). Interpretable deep learning for university dropout prediction. In Proceedings of the 21st Annual Conference on Information Technology Education (pp. 13–19). New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3368308.3415382

Chaushi, B. A., Selimi, B., Chaushi, A., & Apostolova, M. (2023). Explainable artificial intelligence in education: A comprehensive review. In World Conference on Explainable Artificial Intelligence (pp. 48–71). Springer.

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785–794).

Choi, W.-C., Lam, C.-T., & Mendes, A. J. (2023). A systematic literature review on performance prediction in learning programming using educational data mining. In 2023 IEEE Frontiers in Education Conference (FIE).

Choi, W.-C., Lam, C.-T., & Mendes, A. J. (2024a). How various educational features influence programming performance in primary school education. In 2024 IEEE Global Engineering Education Conference (EDUCON) (pp. 1–8). IEEE.

Choi, W.-C., Lam, C.-T., & Mendes, A. J. (2024b). Enhance learning performance predictions with explainable machine learning. In 2024 IEEE Frontiers in Education Conference (FIE).

Choi, W.-C., Lam, C.-T., & Mendes, A. J. (2024c). Analyzing the interpretability of machine learning prediction on student performance using SHapley Additive exPlanations. In 2024 IEEE International Conference on Teaching, Assessment and Learning for Engineering (TALE) (pp. 1–8). IEEE.

Devlin, J. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

Effenberger, T., & Pelánek, R. (2021). Interpretable clustering of students' solutions in introductory programming. In International Conference on Artificial Intelligence in Education (pp. 101–112). Springer.

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. The Annals of Statistics, 29(5), 1189–1232.

Giannakas, F., Troussas, C., Voyiatzis, I., & Sgouropoulou, C. (2021). A deep learning classification framework for early prediction of team-based academic performance. Applied Soft Computing, 106, 107355. https://doi.org/10.1016/j.asoc.2021.107355

Gilpin, L. H., Bau, D., Yuan, B. Z., Bajwa, A., Specter, M., & Kagal, L. (2018). Explaining explanations: An overview of interpretability of machine learning. In 2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA) (pp. 80–89). IEEE.

Greff, K., Srivastava, R. K., Koutník, J., Steunebrink, B. R., & Schmidhuber, J. (2016). LSTM: A search space odyssey. IEEE Transactions on Neural Networks and Learning Systems, 28(10), 2222–2232.

Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2018). A survey of methods for explaining black box models. ACM Computing Surveys, 51(5), 1–42.

Kelleher, C., & Pausch, R. (2005). Lowering the barriers to programming: A taxonomy of programming environments and languages for novice programmers. ACM Computing Surveys, 37(2), 83–137.

Kitchenham, B., & Charters, S. (2007). Guidelines for performing systematic literature reviews in software engineering. UK.

Kokhlikyan, N., et al. (2020). Captum: A unified and generic model interpretability library for PyTorch. arXiv preprint arXiv:2009.07896.

Kuzilek, J., Hlosta, M., & Zdrahal, Z. (2017). Open university learning analytics dataset. Scientific Data, 4(1), 1–8.

Lu, J., Mou, J., & Li, P. (2023). Interpretive analyses of learner dropout prediction in online STEM courses. In 2023 5th International Conference on Computer Science and Technologies in Education (CSTE) (pp. 1–9). https://doi.org/10.1109/CSTE59648.2023.00020

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In Advances in Neural Information Processing Systems (Vol. 30).

Marcinkevičs, R., & Vogt, J. E. (2020). Interpretability and explainability: A machine learning zoo mini-tour. arXiv preprint arXiv:2012.01805.

Mathrani, A., Susnjak, T., Ramaswami, G., & Barczak, A. (2021). Perspectives on the challenges of generalizability, transparency and ethics in predictive learning analytics. Computers and Education Open, 2, 100060.

Mendez, G., Galárraga, L., & Chiluiza, K. (2021). Showing academic performance predictions during term planning: Effects on students' decisions, behaviors, and preferences. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3411764.3445718

Minh, D., Wang, H. X., Li, Y. F., & Nguyen, T. N. (2022). Explainable artificial intelligence: A comprehensive review. Artificial Intelligence Review, 55(5), 3503–3568. https://doi.org/10.1007/s10462-021-10088-y

Mothilal, R. K., Sharma, A., & Tan, C. (2020). Explaining machine learning classifiers through diverse counterfactual explanations. In Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (pp. 607–617).

Pereira, F. D., et al. (2021). Explaining individual and collective programming students' behavior by interpreting a black-box predictive model. IEEE Access, 9, 117097–117119. https://doi.org/10.1109/ACCESS.2021.3105956

Rahul, & Katarya, R. (2024). Shapley explainable deep learning based knowledge distillation framework for student's performance prediction. In 2024 Second International Conference on Data Science and Information System (ICDSIS) (pp. 1–8). https://doi.org/10.1109/ICDSIS61070.2024.10594386
Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?" Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 1135–1144).
Ribeiro, M. T., Singh, S., & Guestrin, C. (2018). Anchors: High-precision model-agnostic explanations. In Proceedings of the AAAI Conference on Artificial Intelligence.
Rico-Juan, J. R., Sánchez-Cartagena, V. M., Valero-Mas, J. J., & Gallego, A. J. (2023). Identifying student profiles within online judge systems using explainable artificial intelligence. IEEE Transactions on Learning Technologies, 16(6), 955–969. https://doi.org/10.1109/TLT.2023.3239110
Rigatti, S. J. (2017). Random forest. Journal of Insurance Medicine, 47(1), 31–39.
Romero, C., Ventura, S., & García, E. (2008). Data mining in course management systems: Moodle case study and tutorial. Computers & Education, 51(1), 368–384.
Shapley, L. S. (1953). A value for n-person games. Princeton University Press.
Shrikumar, A., Greenside, P., & Kundaje, A. (2017). Learning important features through propagating activation differences. In Proceedings of the 34th International Conference on Machine Learning (pp. 3145–3153). PMLR.
Sundararajan, M., Taly, A., & Yan, Q. (2017). Axiomatic attribution for deep networks. In Proceedings of the 34th International Conference on Machine Learning (pp. 3319–3328). PMLR.
Swamy, V., Du, S., Marras, M., & Kaser, T. (2023). Trusting the explainers: Teacher validation of explainable artificial intelligence for course design. In LAK23: 13th International Learning Analytics and Knowledge Conference (pp. 345–356). New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3576050.3576147
Sze, V., Chen, Y.-H., Yang, T.-J., & Emer, J. S. (2017). Efficient processing of deep neural networks: A tutorial and survey. Proceedings of the IEEE, 105(12), 2295–2329
Wachter, S., Mittelstadt, B., & Russell, C. (2017). Counterfactual explanations without opening the black box: Automated decisions and the GDPR. Harvard Journal of Law & Technology, 31, 841.
Waheed, H., Hassan, S.-U., Nawaz, R., Aljohani, N. R., Chen, G., & Gasevic, D. (2023). Early prediction of learners at risk in self-paced education: A neural network approach. Expert Systems with Applications, 213(A), 118868. https://doi.org/10.1016/j.eswa.2022.118868
Yegnanarayana, B. (2009). Artificial neural networks. PHI Learning Pvt. Ltd.