

А.Е. СҰЛТАНӘЛІ¹, Г.Н. КАЗБЕКОВА²

Қожа Ахмет Ясауи атындағы Халықаралық қазақ-түрік университеті, Түркістан қ.,
Қазақстан

¹ E-mail: aigerim.sultanali@ayu.edu.kz*

² E-mail: gulnur.kazbekova@ayu.edu.kz

БІЛІМ АЛУШЫЛАРДЫҢ ОҚУ ҮЛГЕРІМІНЕ ӘСЕР ЕТЕТІН ФАКТОРЛАРДЫ МАШИНАЛЫҚ ОҚЫТУ ӘДІСТЕРІМЕН ТАЛДАУ ЖӘНЕ БОЛЖАУ

Аңдатпа. Зерттеудің мақсаты білім алушылардың қорытынды оқу нәтижесіне әсер ететін негізгі академиялық және оқу мінез-құлық факторларын анықтау және машиналық оқыту әдістерінің көмегімен қорытынды емтихан бағасын болжау дәлдігін салыстырмалы бағалау болып табылады. Экспериментте 2000 білім алушы туралы жазбадан тұратын ашық деректер жиыны пайдаланылды. Кіріс белгілері ретінде сабаққа қатысу, бірінші және екінші аралық тест нәтижелері, тапсырма бағасы және күнделікті оқу уақыты алынды, ал мақсатты айнымалы ретінде 100 балдық шкаладағы қорытынды емтихан бағасы таңдалды. Деректерде бос мәндер анықталмады, студент идентификаторы модельдеуге енгізілмеді. Linear Regression, Decision Tree, Random Forest және Gradient Boosting регрессиялық модельдері 80/20 оқыту-тестілеу бөлуі арқылы салыстырылды. Модельдер MAE, RMSE және R^2 көрсеткіштерімен бағаланды және нәтижелер 5-қабатты кросс-валидациямен тексерілді. Linear Regression тест жиынында $R^2=0.830$, MAE=3.675 және RMSE=4.618 нәтижелерін көрсетіп, зерттелген модельдердің ішіндегі ең жоғары болжамдық сапаға қол жеткізді. Кросс-валидациядағы орташа R^2 0.836 ± 0.008 болды. Корреляциялық талдау сабаққа қатысудың қорытынды бағамен ең жоғары байланысын көрсетті ($r=0.726$), ал Random Forest бойынша оның маңыздылығы 41.49% болды. Алынған нәтижелер салыстырмалы түрде қарапайым және интерпретацияланатын модельдердің білім беру деректерінде тиімді бола алатынын көрсетеді және ерте академиялық мониторинг жүйелерін әзірлеуге негіз бола алады.

Түйін сөздер: машиналық оқыту, оқу үлгерімі, білім алушы, академиялық болжау, білім беру деректері, регрессиялық модель, факторлық талдау.

А.Е. СУЛТАНАЛИ¹, Г.Н. КАЗБЕКОВА²

Международный казахско-турецкий университет имени Ходжи Ахмеда Ясави, Туркестан,
Казахстан

¹ E-mail: aigerim.sultanali@ayu.edu.kz*

² E-mail: gulnur.kazbekova@ayu.edu.kz

АНАЛИЗ И ПРОГНОЗИРОВАНИЕ ФАКТОРОВ, ВЛИЯЮЩИХ НА УСПЕВАЕМОСТЬ ОБУЧАЮЩИХСЯ, МЕТОДАМИ МАШИННОГО ОБУЧЕНИЯ

Аннотация. Цель исследования заключается в выявлении основных академических и поведенческих факторов, связанных с итоговой успеваемостью обучающихся, и сравнительной оценке точности методов машинного обучения при прогнозировании итоговой экзаменационной оценки. В эксперименте использован открытый набор данных, содержащий 2000 записей. В качестве входных признаков использованы посещаемость, результаты двух промежуточных тестов, оценка за задания и ежедневная

продолжительность самостоятельного обучения; целевой переменной выступила итоговая экзаменационная оценка по 100-балльной шкале. Пропущенные значения отсутствовали, идентификатор студента исключался из моделирования. Сравнивались Linear Regression, Decision Tree, Random Forest и Gradient Boosting при разделении данных 80/20. Качество оценивалось по MAE, RMSE и R^2 , устойчивость результатов дополнительно проверялась пятикратной кросс-валидацией. Linear Regression показала на тестовой выборке $R^2=0,830$, $MAE=3,675$ и $RMSE=4,618$, превзойдя более сложные ансамблевые модели. Среднее значение R^2 при кросс-валидации составило $0,836\pm0,008$. Наиболее сильная корреляция с итоговой оценкой выявлена для посещаемости ($r=0,726$), а ее важность в Random Forest составила 41,49%. Результаты подтверждают целесообразность применения интерпретируемых моделей для анализа образовательных данных и поддержки раннего академического мониторинга.

Ключевые слова: машинное обучение, академическая успеваемость, обучающийся, прогнозирование, образовательные данные, регрессия, анализ факторов.

A.Ye. SULTANALI¹, G.N. KAZBEKOVA²

Khoja Akhmet Yassawi International Kazakh-Turkish University, Turkestan, Kazakhstan

¹ E-mail: aigerim.sultanali@ayu.edu.kz*

² E-mail: gulnur.kazbekova@ayu.edu.kz

ANALYSIS AND PREDICTION OF FACTORS AFFECTING STUDENT ACADEMIC PERFORMANCE USING MACHINE LEARNING METHODS

Abstract. This study aims to identify the main academic and learning-behaviour factors associated with students' final academic outcomes and to compare machine learning methods for predicting final examination marks. An open dataset containing 2,000 student records was used. Attendance, two internal test scores, assignment score, and daily study hours were selected as predictors, while the final examination mark on a 100-point scale was used as the target variable. No missing values were detected, and Student_ID was excluded because it does not represent an educational characteristic. Linear Regression, Decision Tree, Random Forest, and Gradient Boosting regression models were compared using an 80/20 train-test split. Predictive performance was evaluated with mean absolute error (MAE), root mean squared error (RMSE), and coefficient of determination (R^2), and robustness was additionally assessed using five-fold cross-validation. Linear Regression achieved the best test performance with $R^2=0.830$, $MAE=3.675$, and $RMSE=4.618$. Its mean cross-validated R^2 was 0.836 ± 0.008 . Pearson correlation analysis showed that attendance had the strongest relationship with final examination marks ($r=0.726$). Random Forest feature importance also ranked attendance first, with an importance of 41.49%. The results demonstrate that a relatively simple and interpretable regression model can outperform more complex tree-based models on this dataset. The proposed experimental workflow can support data-driven academic monitoring and provide a reproducible basis for further development of intelligent student assessment systems.

Keywords: machine learning, academic performance, student, performance prediction, educational data, regression model, factor analysis.

ҚЫСҚАША

Білім беру жүйелерінің цифрлануы білім алушының оқу әрекеті туралы құрылымдалған деректер көлемін арттырып, оларды академиялық шешім қабылдауда пайдалануға мүмкіндік берді. Educational Data Mining және Learning Analytics бағыттарында машиналық оқыту әдістері студенттің үлгерімін болжау, тәуекел тобын ерте анықтау және оқытуды дербестендіру үшін кеңінен қолданылады (Romero & Ventura, 2024). Мұндай тәсілдер

дәстүрлі қорытынды бағаға ғана сүйенбей, оқу барысындағы бірнеше көрсеткіштің бірлескен ықпалын талдауға мүмкіндік береді.

Соңғы зерттеулерде студент үлгерімін болжау үшін сызықтық модельдер, шешім ағаштары, Random Forest, Gradient Boosting, нейрондық желілер және гибриді ансамбльдер қолданылуда (Abougalala et al., 2025; Baashar et al., 2022; Chen & Ding, 2023; Hegazi et al., 2023; Huynh-Cam et al., 2022; Ji et al., 2024; C. Liu et al., 2022; Nnamoko et al., 2025; Olabanjo et al., 2022; Pallathadka et al., 2022; Poudyal et al., 2022). Zhao et al. (2023) машиналық оқыту негізінде академиялық нәтижелерді сандық талдау білім сапасын бағалауда маңызды екенін көрсетсе, Guanin-Fajardo et al. (2024) жоғары білімдегі академиялық табысты бірнеше ML әдісімен болжаған. Онлайн ортадағы мінез-құлық деректерінің де болжамдық мәні бар екені clickstream және оқу белсенділігіне негізделген зерттеулерде көрсетілген (Y. Liu et al., 2023; Wen & Juan, 2023).

Сонымен бірге күрделі модельдің қолданылуы автоматты түрде жоғары дәлдікке кепілдік бермейді. Нәтиже деректер құрылымына, белгілердің табиғатына және мақсатты айнымалының байланыс сипатына тәуелді. Сондықтан бір деректер жиынында әртүрлі күрделіліктегі алгоритмдерді бірыңғай эксперименттік хаттамамен салыстыру және нәтижені интерпретациялау маңызды. Әсіресе білім беру саласында болжамның түсіндірмелілігі педагогке немесе әкімшілікке қандай фактордың нәтижеге көбірек әсер ететінін түсінуге көмектеседі (Borna et al., 2024; Jin, 2026).

Осы зерттеудің мақсаты – сабаққа қатысу, аралық тест нәтижелері, тапсырма бағасы және күнделікті оқу уақыты негізінде қорытынды емтихан нәтижесін болжау және төрт регрессиялық модельдің тиімділігін салыстыру. Зерттеу міндеттері: деректердің сапасын тексеру; факторлар мен қорытынды баға арасындағы корреляцияны бағалау; Linear Regression, Decision Tree, Random Forest және Gradient Boosting модельдерін оқыту; MAE, RMSE және R^2 арқылы салыстыру; 5-fold cross-validation көмегімен тұрақтылығын тексеру; Random Forest негізінде белгілер маңыздылығын талдау.

ЗЕРТТЕУ МАТЕРИАЛДАРЫ МЕН ӘДІСТЕРІ

Эксперименттік зерттеу үшін Kaggle платформасынан алынған Student Academic Performance типіндегі ашық CSV деректер жиыны пайдаланылды. Файлда 2000 бақылау және 7 баған бар. Бос мәндер анықталған жоқ. Student_ID тек бірегей идентификатор болғандықтан, болжамдық модельдерге енгізілген жоқ. Модельдің мақсатты айнымалысы Final Exam Marks (out of 100) болды. Деректер жиынының құрылымы 1-кестеде берілген.

Кесте 1. Экспериментте пайдаланылған белгілердің сипаттамасы.

Белгі	Мазмұны	Рөлі
Attendance (%)	Сабаққа қатысу, %	Кіріс
Internal Test 1 (out of 40)	1-аралық тест, 40 балл	Кіріс
Internal Test 2 (out of 40)	2-аралық тест, 40 балл	Кіріс
Assignment Score (out of 10)	Тапсырма бағасы, 10 балл	Кіріс
Daily Study Hours	Күнделікті оқу уақыты	Кіріс
Final Exam Marks (out of 100)	Қорытынды емтихан, 100 балл	Мақсат

Деректер train_test_split процедурасымен 80% оқыту және 20% тест жиынына бөлінді; кездейсоқтықты қайталау үшін random_state=42 белгіленді. Нәтижесінде 1600 жазба модельдерді оқытуға, 400 жазба тәуелсіз тестілеуге пайдаланылды. Эксперимент Python ортасында pandas және scikit-learn кітапханалары арқылы орындалды. Салыстыруға Linear Regression, Decision Tree Regressor, Random Forest Regressor (100 ағаш) және Gradient Boosting Regressor алынды. Ағаштық модельдер үшін random_state=42 пайдаланылды.

Регрессиялық сапаны бағалау үшін орташа абсолюттік қате (MAE), орташа квадраттық қатенің түбірі (RMSE) және детерминация коэффициенті (R^2) қолданылды. MAE мен RMSE неғұрлым аз, ал R^2 бірлікке неғұрлым жақын болса, модельдің болжамдық сапасы соғұрлым

жоғары деп есептелді. Нәтиженің бір ғана train/test бөлінісіне тәуелді болмауын тексеру үшін деректер араластырылып, 5-қабатты KFold кросс-валидация жүргізілді.

Сызықтық регрессия моделі жалпы түрде келесі өрнекпен сипатталды:

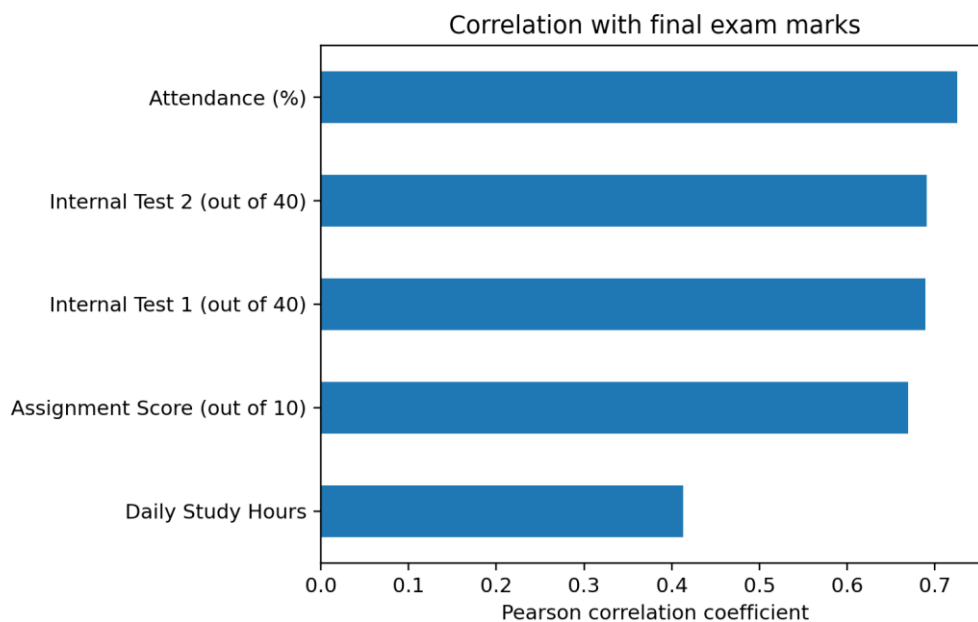
$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5, \quad (1)$$

мұнда $\hat{y}$ – болжанған қорытынды баға; x_1 – сабаққа қатысу; x_2 және x_3 – екі аралық тест; x_4 – тапсырма бағасы; x_5 – күнделікті оқу уақыты; $\beta_0 \dots \beta_5$ – модель параметрлері.

Факторлардың мақсатты айнымалымен сызықтық байланысын Pearson корреляция коэффициенті арқылы, ал сызықтық емес және өзара әрекеттесетін әсерлерді қосымша интерпретациялау үшін Random Forest feature importance көрсеткіші арқылы бағаладық. Мұндай салыстырмалы әдіс студент үлгерімін әртүрлі алгоритмдермен болжауға арналған қазіргі зерттеулердің тәжірибесіне сәйкес келеді (Holicza & Kiss, 2023; Tang et al., 2024; Zhao et al., 2023).

НӘТИЖЕЛЕР МЕН ТАЛҚЫЛАУ

Алдымен кіріс белгілерінің қорытынды емтихан бағасымен байланысы есептелді. Барлық бес белгі оң корреляция көрсетті. Ең жоғары коэффициент сабаққа қатысуда байқалды ($r=0.726$). Екінші және бірінші аралық тесттердің корреляциясы тиісінше 0.691 және 0.689 болды. Тапсырма бағасы үшін $r=0.669$, ал күнделікті оқу уақыты үшін $r=0.413$ алынды. Бұл деректерде жүйелі қатысу мен аралық бақылау нәтижелері қорытынды емтиханмен ең тығыз байланысты екенін көрсетеді (1-сурет).



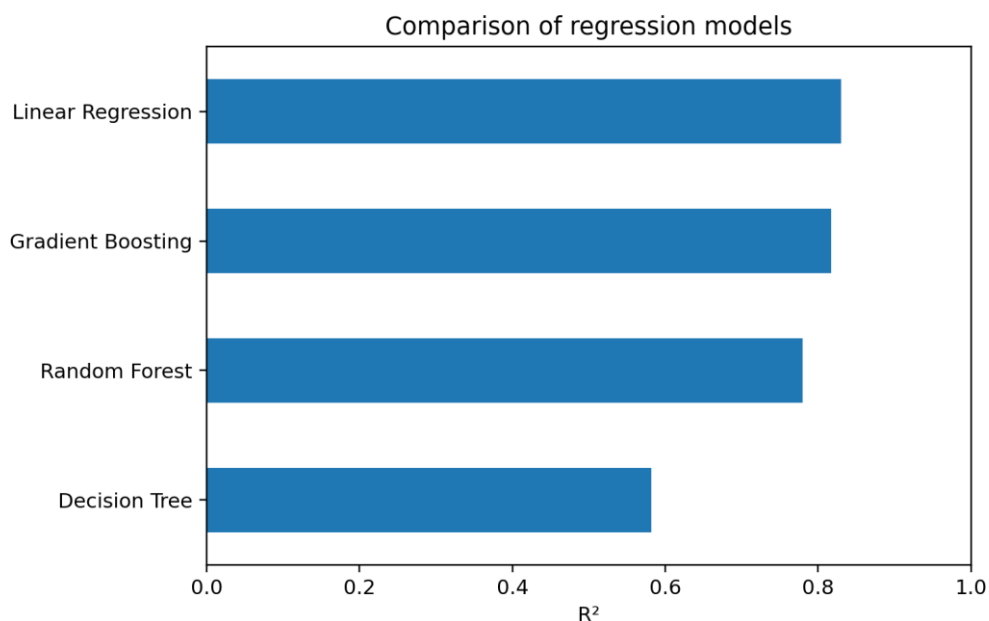
Сурет 1. Кіріс факторларының қорытынды емтихан бағасымен корреляциясы.

Төрт модель бірдей оқыту және тест жиындарында салыстырылды. Нәтижелер 2-кестеде және 2-суретте келтірілген. Linear Regression барлық үш негізгі метрика бойынша ең жақсы нәтиже көрсетті: MAE=3.675, RMSE=4.618 және $R^2=0.830$. Gradient Boosting нәтижесі оған жақын болғанымен ($R^2=0.817$), Random Forest және Decision Tree модельдерінде қате жоғары болды.

Кесте 2. Машиналық оқыту модельдерінің тест жиынындағы нәтижелері.

Модель	MAE	RMSE	R^2
Linear Regression	3.675	4.618	0.830

Модель	MAE	RMSE	R ²
Gradient Boosting	3.737	4.786	0.817
Random Forest	4.160	5.252	0.780
Decision Tree	5.747	7.235	0.582



Сурет 2. Регрессиялық модельдердің R² көрсеткіші бойынша салыстырылуы.

R²=0.830 нәтижесі таңдалған бес белгі арқылы тест жиынындағы қорытынды емтихан бағасының вариациясының шамамен 83%-ы модельмен түсіндірілетінін көрсетеді. MAE=3.675 болғандықтан, Linear Regression болжамы нақты қорытынды бағадан орта есеппен шамамен 3.7 баллға ауытқыды. Нақты және болжанған бағалардың таралуы 3-суретте көрсетілген; нүктелердің диагональға жақын орналасуы модель болжамдарының нақты мәндермен жалпы сәйкестігін визуалды түрде растайды.

Алынған нәтиженің кездейсоқ бөлініске тәуелділігін бағалау үшін 5-fold cross-validation жүргізілді. Linear Regression орташа R²=0.836±0.008 көрсетіп, қайтадан бірінші орын алды. Gradient Boosting үшін орташа R²=0.818, Random Forest үшін 0.790, Decision Tree үшін 0.617 болды (3-кесте). Төмен стандарттық ауытқу сызықтық модельдің әртүрлі деректер бөліктерінде тұрақты нәтиже беретінін көрсетеді.

Кесте 3. 5-fold cross-validation нәтижелері.

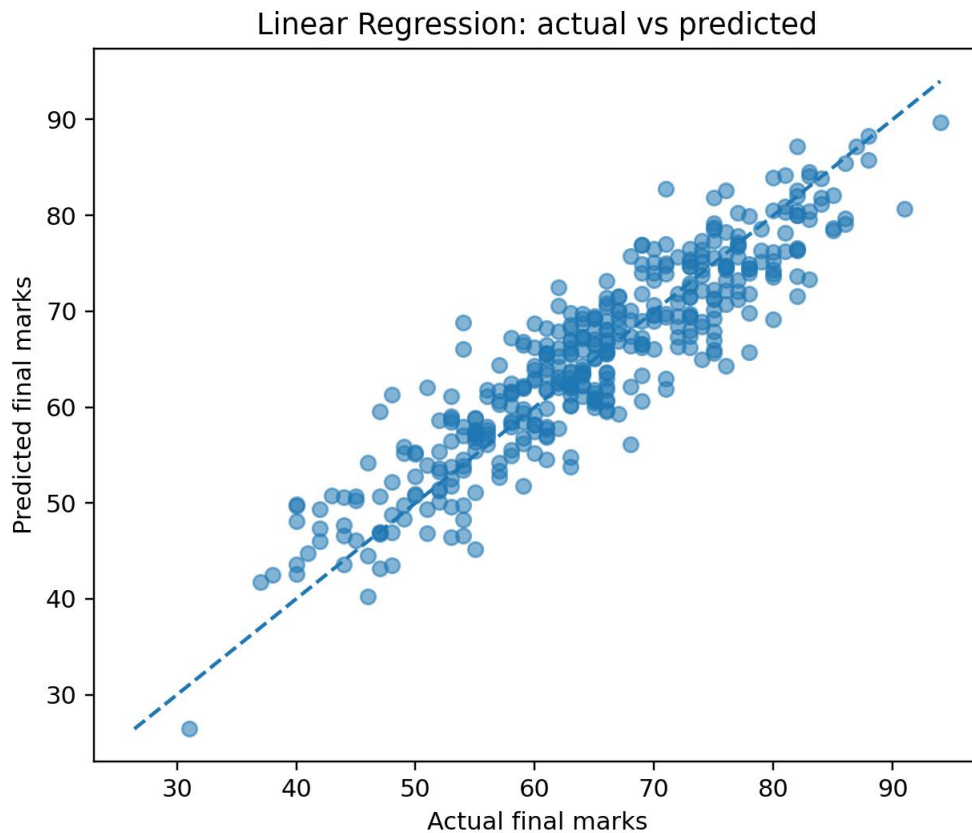
Модель	R ² орташа	SD	MAE	RMSE
Linear Regression	0.836	0.008	3.622	4.587
Gradient Boosting	0.818	0.005	3.823	4.831
Random Forest	0.790	0.007	4.086	5.186
Decision Tree	0.617	0.028	5.598	6.998

Linear Regression моделінің алынған коэффициенттері бойынша эксперименттік теңдеу төмендегідей болды:

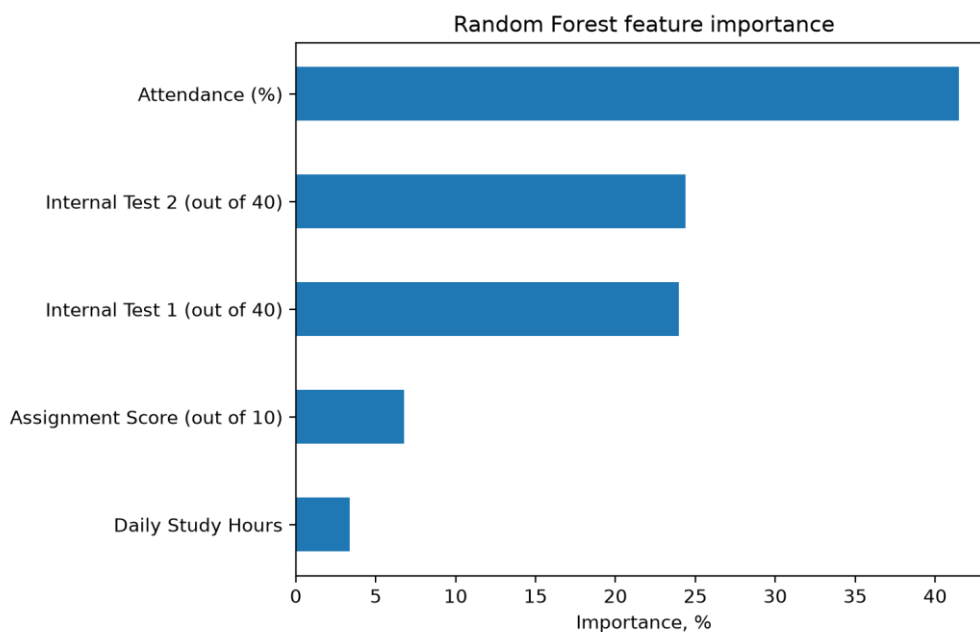
$$\begin{aligned} \text{Final Mark} = & -44.817 + 0.394 \cdot \text{Attendance} + 0.892 \cdot \text{Test1} + \\ & + 0.883 \cdot \text{Test2} + 1.450 \cdot \text{Assignment} + 2.866 \cdot \text{StudyHours}. \end{aligned} \quad (2)$$

Коэффициенттердің өлшем бірліктері әртүрлі болғандықтан, олардың абсолюттік шамаларын тікелей фактор маңыздылығы ретінде салыстыруға болмайды. Осы себепті

қосымша Random Forest feature importance талдауы жүргізілді. Нәтижесінде сабаққа қатысу 41.49% маңыздылықпен бірінші орын алды; Internal Test 2 – 24.39%, Internal Test 1 – 23.96%, Assignment Score – 6.79%, Daily Study Hours – 3.37% (4-сұрет).



Сурет 3. Linear Regression моделі үшін нақты және болжанған қорытынды бағалардың салыстырылуы.



Сурет 4. Random Forest моделі бойынша факторлардың салыстырмалы маңыздылығы.

Корреляциялық талдау мен Random Forest маңыздылығының бір бағытты нәтиже беруі сабаққа қатысудың осы деректер жиынында маңызды болжамдық белгі екенін көрсетеді.

Дегенмен бұл байланыс себеп-салдарлық әсер ретінде түсіндірілмеуі тиіс: қолданылған деректер бақылаулық сипатта, ал модельдер ассоциациялар мен болжамдық заңдылықтарды анықтайды.

Біздің экспериментте Linear Regression-ның күрделірек ағаштық ансамбльдерден жоғары нәтиже көрсетуі деректердегі негізгі байланыстардың елеулі бөлігі сызықтық компонентпен сипатталатынын болжауға мүмкіндік береді. Бұл қорытынды күрделі модель әрдайым тиімді деген түсінікті шектейді және алгоритм таңдауда эмпирикалық салыстыру қажеттігін көрсетеді. Әдебиетте де әртүрлі деректер жиындары үшін оңтайлы алгоритмдердің өзгеретіні байқалады: кейбір жұмыстар ансамбльдік және нейрондық модельдердің артықшылығын көрсетсе (Tang et al., 2024; Xue & Niu, 2023; Yu & Liu, 2022), басқа зерттеулер интерпретацияланатын белгілер мен классикалық ML әдістерінің практикалық құндылығына назар аударады (Guanin-Fajardo et al., 2024; Jin, 2026).

Практикалық тұрғыдан модельді оқу процесінің аралық кезеңінде академиялық мониторинг құралы ретінде қолдануға болады. Сабаққа қатысу мен екі аралық тест нәтижелері қорытынды бағамен жоғары байланыс көрсеткендіктен, осы белгілер төмендеген кезде білім алушыға қосымша қолдау ұсыну сценарийін қарастыруға болады. Алайда автоматтандырылған болжам студентке қорытынды баға қоюдың жалғыз негізі болмауы керек; ол педагогтің шешімін қолдайтын қосымша аналитикалық құрал ретінде пайдаланылуы тиіс.

Зерттеудің шектеулері де бар. Біріншіден, ашық деректер жиынының контексті нақты университеттің оқу процесін толық бейнелемеуі мүмкін. Екіншіден, деректерде пән күрделілігі, оқытушы ерекшелігі, алдыңғы GPA, әлеуметтік және мотивациялық факторлар жоқ. Үшіншіден, модель сыртқы тәуелсіз университет деректерінде тексерілген жоқ. Болашақ зерттеуде нақты LMS деректерімен сыртқы валидация жүргізіп, түсіндірілетін AI әдістерін және объективті бағалау моделіне интеграциялау мүмкіндігін зерттеу орынды.

Алынған эксперименттік нәтижелерді талдау барысында модель күрделілігінің жоғары болуы әрқашан болжам сапасының артуына алып келмейтіні анықталды. Зерттеуде Linear Regression моделінің ансамбльдік және ағаш негізіндегі алгоритмдермен салыстырғанда жоғары нәтиже көрсетуі қарастырылған деректер жиынындағы кіріс белгілері мен қорытынды емтихан бағасы арасында айқын сызықтық тәуелділіктің бар екенін көрсетеді. Атап айтқанда, Linear Regression моделінің R^2 көрсеткіші 0.830 деңгейіне жетіп, Gradient Boosting (0.817), Random Forest (0.780) және Decision Tree (0.582) нәтижелерінен жоғары болды. Сонымен қатар MAE=3.675 болуы модель болжаған қорытынды бағаның нақты бағадан орта есеппен шамамен 3.7 баллға ғана ауытқитынын көрсетеді. Білім алушылардың нәтижелері 100 балдық шкалада бағаланатынын ескерсек, мұндай ауытқу зерттелген деректер үшін модельдің жеткілікті болжамдық қабілетін сипаттайды.

Эксперимент барысында анықталған тағы бір маңызды ерекшелік – академиялық көрсеткіштердің оқу нәтижесіне әсер ету деңгейінің біркелкі болмауы. Корреляциялық талдау бойынша сабаққа қатысу көрсеткіші қорытынды емтихан нәтижесімен ең жоғары оң байланысты көрсетті ($r=0.726$). Бірінші және екінші аралық тест нәтижелері де қорытынды бағамен жоғары байланысқа ие болды. Бұл білім алушының оқу нәтижесін болжау кезінде тек бір бақылау көрсеткішіне сүйенудің орнына бірнеше академиялық факторды бірге қарастыру қажеттігін көрсетеді. Сонымен қатар күнделікті оқу уақытының корреляция коэффициенті басқа көрсеткіштермен салыстырғанда төмен болғанымен ($r=0.413$), оның оң мәні жүйелі өзіндік дайындықтың қорытынды нәтижемен байланысты екенін көрсетеді.

Зерттеу нәтижелерінің практикалық маңыздылығы машиналық оқыту моделін білім беру үдерісінде ерте академиялық мониторинг құралы ретінде пайдалану мүмкіндігімен байланысты. Семестр барысында жинақталатын сабаққа қатысу, аралық бақылау және тапсырма нәтижелері негізінде білім алушының ықтимал қорытынды нәтижесін алдын ала бағалауға болады. Мұндай тәсіл төмен нәтиже көрсету қаупі бар білім алушыларды қорытынды емтиханға дейін анықтап, оларға қосымша оқу материалдарын, консультацияларды немесе жеке қолдау шараларын ұсынуға мүмкіндік береді.

Сонымен бірге ұсынылған тәсілді білім алушыға қорытынды бағаны автоматты түрде тағайындайтын жүйе ретінде қарастыру дұрыс емес. Машиналық оқыту нәтижесі оқытушының кәсіби бағалауын алмастырмай, шешім қабылдауды қолдайтын қосымша аналитикалық құрал қызметін атқаруы тиіс. Болашақ зерттеулерде модельге білім алушының оқу платформасындағы белсенділігі, тапсырмаларды орындау тұрақтылығы, білім деңгейінің уақыт бойынша өзгеруі және басқа да академиялық көрсеткіштерді енгізу арқылы болжамның сенімділігі мен объективтілігін одан әрі зерттеу жоспарланады.

ҚОРЫТЫНДЫ

Зерттеуде 2000 білім алушының деректері негізінде қорытынды емтихан нәтижесін болжауға арналған төрт машиналық оқыту алгоритмі салыстырылды. Кіріс белгілері ретінде сабаққа қатысу, екі аралық тест, тапсырма бағасы және күнделікті оқу уақыты қолданылды. Эксперимент бірыңғай 80/20 train-test бөлінісі және 5-fold cross-validation арқылы жүргізілді.

Linear Regression тест жиынында $R^2=0.830$, $MAE=3.675$ және $RMSE=4.618$ көрсетіп, Gradient Boosting, Random Forest және Decision Tree модельдерінен жоғары нәтиже берді. Кросс-валидацияда оның орташа R^2 көрсеткіші 0.836 ± 0.008 болды, бұл модель нәтижесінің тұрақтылығын растады.

Факторлық талдау бойынша сабаққа қатысу қорытынды емтихан бағасымен ең жоғары корреляцияны көрсетті ($r=0.726$) және Random Forest feature importance бойынша 41.49% үлеспен бірінші орын алды. Осылайша, зерттеудің практикалық нәтижесі – салыстырмалы түрде қарапайым, түсіндірілетін регрессиялық модель арқылы оқу нәтижесін жеткілікті дәл болжауға және негізгі академиялық факторларды айқындауға болатындығы.

Алынған модель білім алушылардың оқу динамикасын ерте бақылауға арналған интеллектуалды жүйелердің аналитикалық компоненті ретінде қолданылуы мүмкін. Кейінгі зерттеулерде нақты жоғары оқу орнының LMS және бағалау деректерімен модельді тексеру, қосымша мінез-құлық белгілерін енгізу және болжам нәтижесін объективті бағалау моделіне біріктіру жоспарланады.

ҚАРЖЫЛАНДЫРУ: Зерттеу арнайы гранттық қаржыландырусыз орындалды.

ЖАСАНДЫ ИНТЕЛЛЕКТ ТЕХНОЛОГИЯЛАРЫН ПАЙДАЛАНУ ТУРАЛЫ

ХАБАРЛАМА: Қолжазбаны дайындау барысында генеративті жасанды интеллект ChatGPT (OpenAI, 2026) мәтінді құрылымдау, тілдік редакциялау және эксперимент нәтижелерін түсіндіруге көмекші құрал ретінде қолданылды. Деректерді өңдеу мен машиналық оқыту экспериментінің сандық нәтижелері бағдарламалық есептеулер арқылы тексерілді. Автор(лар) қолжазбаның мазмұны мен ғылыми тұжырымдары үшін толық жауапкершілікті сақтайды.

ПАЙДАЛАНЫЛҒАН ӘДЕБИЕТТЕР ТІЗІМІ

Abougalala, R. A., Alharbi, N., Amasha, M. A., Areed, M. F., Alkhalaf, S., & Khairy, D. (2025). Predicting student performance academic using Automated Machine Learning (AutoML): In medical academic institutions. *Journal of New Approaches in Educational Research*, 14(1), Article 19. <https://doi.org/10.1007/s44322-025-00038-9>

Baashar, Y., Alkaws, G., Mustafa, A., Alkahtani, A. A., Alsariera, Y. A., Ali, A. Q., Hashim, W., & Tiong, S. K. (2022). Toward predicting student's academic performance using artificial neural networks (ANNs). *Applied Sciences*, 12(3), Article 1289. <https://doi.org/10.3390/app12031289>

Borna, M.-R., Saadat, H., Hojjati, A. T., & Akbari, E. (2024). Analyzing click data with AI: Implications for student performance prediction and learning assessment. *Frontiers in Education*, 9, Article 1421479. <https://doi.org/10.3389/educ.2024.1421479>

- Chen, S., & Ding, Y. (2023). A machine learning approach to predicting academic performance in Pennsylvania's schools. *Social Sciences*, 12(3), Article 118. <https://doi.org/10.3390/socsci12030118>
- Guanin-Fajardo, J. H., Guaña-Moya, J., & Casillas, J. (2024). Predicting academic success of college students using machine learning techniques. *Data*, 9(4), Article 60. <https://doi.org/10.3390/data9040060>
- Hegazi, M. O., Almaslukh, B., & Siddig, K. (2023). A fuzzy model for reasoning and predicting student's academic performance. *Applied Sciences*, 13(8), Article 5140. <https://doi.org/10.3390/app13085140>
- Holicza, B., & Kiss, A. (2023). Predicting and comparing students' online and offline academic performance using machine learning algorithms. *Behavioral Sciences*, 13(4), Article 289. <https://doi.org/10.3390/bs13040289>
- Huynh-Cam, T.-T., Chen, L.-S., & Huynh, K.-V. (2022). Learning performance of international students and students with disabilities: Early prediction and feature selection through educational data mining. *Big Data and Cognitive Computing*, 6(3), Article 94. <https://doi.org/10.3390/bdcc6030094>
- Ji, M., Le, J., Chen, B., & Li, Z. (2024). A predictive model for classifying college students' academic performance based on visual-spatial skills. *Frontiers in Psychology*, 15, Article 1434015. <https://doi.org/10.3389/fpsyg.2024.1434015>
- Jin, X. (2026). Predicting academic success: Machine learning analysis of student, parental, and school efforts. *Asia Pacific Education Review*, 27(1), 1–22. <https://doi.org/10.1007/s12564-023-09915-4>
- Liu, C., Wang, H., & Yuan, Z. (2022). A method for predicting the academic performances of college students based on education system data. *Mathematics*, 10(20), Article 3737. <https://doi.org/10.3390/math10203737>
- Liu, Y., Fan, S., Xu, S., Sajjanhar, A., Yeom, S., & Wei, Y. (2023). Predicting student performance using clickstream data and machine learning. *Education Sciences*, 13(1), Article 17. <https://doi.org/10.3390/educsci13010017>
- Nnamoko, N., Barrowclough, J., Onikoyi, B., & Liptrott, M. (2025). Predicting student academic outcome using online behavioural statistics and neighbourhood influences: A machine learning approach. *SN Computer Science*, 6(6), Article 661. <https://doi.org/10.1007/s42979-025-04189-6>
- Olabanjo, O. A., Wusu, A. S., & Manuel, M. (2022). A machine learning prediction of academic performance of secondary school students using radial basis function neural network. *Trends in Neuroscience and Education*, 29, Article 100190. <https://doi.org/10.1016/j.tine.2022.100190>
- OpenAI. (2026, August 12). *Assistance with manuscript structuring, language editing, and interpretation of machine-learning experiment results* [Large language model]. ChatGPT. <https://chatgpt.com/>
- Pallathadka, H., Sonia, B., Sanchez, D. T., De Vera, J. V., Godinez, J. A. T., & Pepito, M. T. (2022). Investigating the impact of artificial intelligence in education sector by predicting student performance. *Materials Today: Proceedings*, 51, 2264–2267. <https://doi.org/10.1016/j.matpr.2021.11.395>
- Poudyal, S., Mohammadi-Aragh, M. J., & Ball, J. E. (2022). Prediction of student academic performance using a hybrid 2D CNN model. *Electronics*, 11(7), Article 1005. <https://doi.org/10.3390/electronics11071005>
- Romero, C., & Ventura, S. (2024). *Educational data mining and learning analytics: An updated survey* (arXiv:2402.07956). arXiv. <https://doi.org/10.48550/arXiv.2402.07956>
- Tang, B., Li, S., & Zhao, C. (2024). Predicting the performance of students using deep ensemble learning. *Journal of Intelligence*, 12(12), Article 124. <https://doi.org/10.3390/jintelligence12120124>

Wen, X., & Juan, H. (2023). Early prediction of students' performance using a deep neural network based on online learning activity sequence. *Applied Sciences*, 13(15), Article 8933. <https://doi.org/10.3390/app13158933>

Xue, H., & Niu, Y. (2023). Multi-output based hybrid integrated models for student performance prediction. *Applied Sciences*, 13(9), Article 5384. <https://doi.org/10.3390/app13095384>

Yu, F., & Liu, X. (2022). Research on student performance prediction based on stacking fusion model. *Electronics*, 11(19), Article 3166. <https://doi.org/10.3390/electronics11193166>

Zhao, L., Ren, J., Zhang, L., & Zhao, H. (2023). Quantitative analysis and prediction of academic performance of students using machine learning. *Sustainability*, 15(16), Article 12531. <https://doi.org/10.3390/su151612531>

Авторлар туралы мәліметтер

Информация об авторах

Information about authors

№	Информация об авторах (Ф.И.О. полностью, ученая степень, место работы или учебы, город, страна, E-mail, сотовый номер автора корреспондента:) на казахском, русском, английском языках
1	Сұлтанәлі Айгерім Ердалықызы – магистр, оқытушы, докторант, Қожа Ахмет Ясауи атындағы Халықаралық қазақ-түрік университеті, Түркістан қ., Қазақстан Sultanali Aigerim Yerdalykyzy – магистр, преподаватель, докторант, Международный казахско-турецкий университет имени Ходжи Ахмеда Ясави, г. Туркестан, Казахстан Sultanali Aigerim Yerdalykyzy – Master, Lecturer, Doctoral Student, Khoja Akhmet Yassawi International Kazakh-Turkish University, Turkestan, Kazakhstan E-mail: aigerim.sultanali@ayu.edu.kz ORCID: 0009-0008-6948-7271
2	Казбекова Гулнур Нагиметовна (ғылыми жетекші) – техника ғылымдарының кандидаты, қауымдастырылған профессор, Қожа Ахмет Ясауи атындағы Халықаралық қазақ-түрік университеті, Түркістан қ., Қазақстан Казбекова Гулнур Нагиметовна (научный руководитель) – кандидат технических наук, ассоциированный профессор, Международный казахско-турецкий университет имени Ходжи Ахмеда Ясави, г. Туркестан, Казахстан Kazbekova Gulnur Nagimetovna (scientific supervisor) – Candidate of Technical Sciences, Associate Professor, Khoja Akhmet Yassawi International Kazakh-Turkish University, Turkestan, Kazakhstan E-mail: gulnur.kazbekova@ayu.edu.kz ORCID: 0000-0002-2756-7926