

УДК 004.85

DOI: 10.31498/2225-6733.53.1.2026.359776

ПОРІВНЯЛЬНИЙ АНАЛІЗ АЛГОРИТМІВ МАШИННОГО НАВЧАННЯ В ЗАДАЧАХ ПІДТРИМКИ ПРИЙНЯТТЯ РІШЕНЬ ЩОДО АКАДЕМІЧНОЇ УСПІШНОСТІ

Проніна О.І.	канд. техн. наук, доцент, ДВНЗ «Приазовський державний технічний університет», м. Дніпро, ORCID: https://orcid.org/0000-0001-7085-8027 , e-mail: pronina_o_i@pstu.edu ;
Перцев Є.А.	магістр, ДВНЗ «Приазовський державний технічний університет», м. Дніпро, e-mail: pertsev_y_a@students.pstu.edu

У сучасних умовах цифровізації освіти особливої ваги набуває питання об'єктивного оцінювання результатів навчання та своєчасного виявлення студентів, які потребують додаткової підтримки. У даній роботі розглянуто можливості застосування алгоритмів машинного навчання для прогнозування підсумкових результатів семестрового контролю. Актуальність дослідження зумовлена потребою в інструментах об'єктивізації оцінювання та своєчасного виявлення ризиків неуспішності для коригування індивідуальної освітньої траєкторії. Дослідження спрямоване не лише на отримання точного числового бала, а й на визначення рівня успішності студента в узагальненій категоріальній формі (від «Very low» до «Good»), що дозволяє глибше проаналізувати якість освітнього процесу та своєчасно скоригувати індивідуальну траєкторію навчання. У роботі використано набір даних «Students Performance Factors», який охоплює показники відвідуваності занять, інтенсивності навчання, попередньої академічної підготовки та соціально-економічних чинників. Підготовка даних включала очищення від аномалій, нормалізацію ознак і кореляційний аналіз для відбору найбільш інформативних параметрів. Для експериментального дослідження застосовано лінійну регресію, багатошаровий перцептрон (нейронну мережу), дерева рішень та випадковий ліс. Оцінювання моделей здійснювалося за метриками R^2 і MAE для задач регресії та за показником точності для задач класифікації. Результати показали, що для прогнозування точного підсумкового бала найбільш доцільною виявилася лінійна регресія, яка продемонструвала високу точність та простоту інтерпретації. Водночас у задачі розподілу студентів за рівнями успішності кращі результати отримано за допомогою нейронної мережі. Аналіз дерева рішень підтвердив важливу роль відвідуваності як ключового чинника академічного успіху. Отриманні результати показують, що вибір методу має залежати від формату прогнозу: для кількісних оцінок доцільно використовувати регресію, а для якісного розділення студентів за рівнями – нейронні мережі. Дослідження відкриває перспективи впровадження інструментів освітньої аналітики для підвищення ефективності навчання.

Ключові слова: машинне навчання; регресія; дерева рішень; нейронні мережі; оцінка успішності; академічна успішність.

Постановка проблеми

Прогнозування підсумкової оцінки студентів під час семестрового контролю є важливим інструментом об'єктивізації процесу оцінювання та визначення рівня сформованості навчальних компетентностей, ще до проведення фінального іспиту за курс. Використання моделей передбачення дає змогу здійснювати попередній аналіз академічної успішності, своєчасно виявляти потенційні ризики зниження результатів та приймати обґрунтовані управлінські рішення щодо коригування індивідуальної освітньої траєкторії студента.

Застосування прогнозних підходів також створює передумови для підвищення ефективності навчального процесу, оскільки надає студенту можливість цілеспрямовано зосередитися на проблемних аспектах підготовки та оптимізувати розподіл навчального навантаження.

Крім того, у випадках неможливості складання підсумкового контролю з поважних причин, за умови попереднього погодження зі студентом та відповідного нормативного регулювання, результати прогнозної моделі можуть розглядатися як альтернативний механізм визначення очікуваного рівня досягнень. Це дозволяє

забезпечити безперервність освітнього процесу та зберегти принципи академічної доброчесності й об'єктивності оцінювання.

Існує щонайменше два підходи до такого прогнозування.

1. Прогнозування конкретного числового бала. Цей варіант повинен бути максимально точним, оскільки навіть один бал є важливим для оцінювання. Такий підхід доцільний для формування рейтингу, порівняння академічної успішності між студентами та глибокого аналізу якості навчання. У цьому випадку модель має забезпечувати мінімальну абсолютну похибку та високу точність передбачення.

2. Прогнозування категорії успішності. Тут завдання складається з визначення класу, до якого потрапляє студент зі своїми результатами за курс. Важливо правильно віднести результат до певного класу («Відмінний», «Добрий», «Середній», «Задовільний», «Незадовільний» або «Залік» / «Не залік»), в залежності від кінцевого типу оцінювання. Незначні відхилення в числовому значенні допустимі, якщо модель коректно визначає сам клас, до якого потрапляють результати.

Аналіз останніх досліджень та публікацій

В літературі представлено ряд досліджень, які направлені на визначення класу оцінок здобувачів. Так у статті [1] реалізовано різні методи класифікації машинного навчання для академічної успішності студентів з метою прогнозування результатів. Для цього дані студентів магістратури (MS(CS)) були зібрані з державного університету Пакистану через їхні завдання, тести та сесійні оцінки. Для проведення всіх експериментів, а саме попередньої обробки даних, класифікації та візуалізації, було використано інструмент інтелектуального аналізу даних WEKA.

Головною метою роботи [2] є порівняння продуктивності алгоритмів інтелектуального аналізу даних для прогнозування академічної успішності студентів. Набір даних було підготовлено на основі академічних результатів студентів, зібраних з Міжнародного університету науки і технологій армії Бангладеш. Було реалізовано два алгоритми інтелектуального аналізу даних, а саме найвигідніший баєсівський класифікатор та дерево рішень J48, а отриману точність порівняно один з одним. Після попередньої обробки (очищення, дискретизації) як найвигідніший баєсівський класифікатор, так і алгоритм дерева рішень J48 забезпечують точність понад 88,75%.

Дослідження [3] зосереджено на створенні моделі для прогнозування категорій сумарного середнього балу (CGPA) та виявлення студентів, які можуть мати труднощі в навчанні під час навчання в університеті. У дослідженні використовуються дані, пов'язані зі студентами, для класифікації CGPA за певними оцінками (O, A+, A, B+, B, C+, C та F), а також протестовано різні методи машинного навчання, такі як Naive Bayes, JRIP, J48, Random Forest та CatBoost. Оскільки деякі оцінки, такі як C та A+, мали менше студентів, для балансування набору даних було використано метод однокласового схрещування. Моделі були оцінені за допомогою десятикратної перехресної перевірки, і CatBoost показав найкращі результати після балансування даних.

Питання навчання стоїть гостро для різних університетів світу, так наприклад багато студентів болгарських університетів кидають навчання в університеті, не завершивши його. Виявлення студентів, які ризикують кинути навчання, дозволяє своєчасно вживати заходів для їх утримання. У статті [4] представлені результати дослідження, проведеного серед студентів інженерних програм Тракійського університету – Стара Загора. Зібрані дані піддаються обробці, метою якої є пошук найважливіших атрибутів, що визначають ризик кидання навчання в університеті. Обробка проводиться за допомогою програмного забезпечення з відкритим кодом Weka. Застосовуються різні алгоритми вибору атрибутів з різними методами пошуку. Найбільш підходящий алгоритм вибору атрибутів було обрано після застосування класифікатора BayesNet до отриманих результатів. Порівнювалися показники TP

rate, Precision та F-measure. При застосуванні InfoGainAttributeEval отримані найвищі результати щодо точності класифікації.

Точне прогнозування успішності учнів є значним кроком до покращення результатів навчання шляхом прийняття рішень на основі даних. З розвитком інтелектуального аналізу даних про освіту (EDM), алгоритми машинного навчання (ML) стали ефективними інструментами для аналізу даних учнів з метою виявлення прихованих цікавинок. У цій статті представлено комплексне дослідження використання алгоритму випадкового лісу (RF) для прогнозування академічної успішності учнів. Використовуючи різні ознаки, такі як демографічні дані, попередні оцінки, відвідуваність, соціально-економічний статус та позакласна участь, класифікатор випадкового лісу забезпечує високу точність прогнозування, водночас пропонуючи інтерпретованість за допомогою рейтингу важливості ознак. У цьому дослідженні [5] далі обговорюються навчання моделі, оцінка успішності, графічні представлення та наслідки інтеграції таких систем в освітні установи. Крім того, досліджуються обмеження та майбутні напрямки вдосконалення, підкреслюючи ширший потенціал ML у трансформації сектору освіти.

Освітній data mining – це процес застосування інструментів та методів аналізу даних у навчальному закладі. Освітні заклади отримують великі обсяги даних від учнів протягом їхнього навчання. Однак ці дані часто залишаються недостатньо використаними та не покращують академічну успішність учнів. Рівень успішності або невдачі учнів може визначити успіх будь-якого навчального закладу. Запропонована система використовує освітній data mining для аналізу даних, щоб робити прогнози щодо успішності учнів. До даних було застосовано метод класифікації (навчання з учителем) та алгоритм дерева рішень для прогнозування їхньої успішності на випускних іспитах. У цьому проєкті [6] для попередньої обробки даних було використано Weka (Waikato Environment for Knowledge Analysis). Класифікатор дерева рішень з Python sci-kit-learn було застосовано до набору даних та візуалізовано графічно. Мета дослідження полягає в тому, щоб ми могли передавати будь-які нові дані класифікатору, і він зміг би відповідно передбачити правильний клас.

У статті [7] виступає проблема прогнозування академічної успішності студентів із застосуванням алгоритмів машинного навчання. Аналіз та апробація відповідних методів стали на етапі у напрямку вдосконалення освітньої аналітики та підвищення якості навчального процесу. В межах дослідження було висвітлено і протестовано кілька моделей машинного навчання з призначенням визначення їх ефективності в завданнях передбачення результатів навчання. Було здійснено оцінку основних показників продуктивності моделей, які дозволили провести якісний аналіз їхньої точності. Результати матриці помилок свідчать про задовільну роботу моделей після здійснення оптимізації їхніх гіперпараметрів. Здебільшого моделі показали високий

рівень точності класифікації. На цьому, було продемонстровано практичне використання побудованих моделей для аналізу нових даних. Загалом реалізовано ефективне рішення задач прогнозного моделювання з використанням алгоритмів RandomForest та XGBoost, яке можна адаптувати для подальшого вдосконалення та реального впровадження на практику.

У статті [8] було зазначено використання різноманітних методів і алгоритмів для обробки даних, зокрема для прогнозування та аналізу поведінки користувачів онлайн-платформи. У процесі дослідження проведено аналіз доступних інструментів для інтеграції моделей машинного навчання та аналітичних інструментів у веб-застосунки. Проведено порівняння кількох алгоритмів машинного навчання, зокрема таких як: Random Forest, KNN-регресії, гребеневої регресії та еластичної мережі, для аналізу залученості користувачів на платформі Prometheus. Результати показують, що використання різних методів машинного навчання дозволяє здійснити точний прогноз щодо залученості користувачів, зокрема з урахуванням змін у соціально-демографічному контексті.

Виходячи з огляду літератури, можна помітити, що актуальним та достовірним методом для оцінки успішності є використання методів машинного навчання. Причому для визначення успішності в літературі застосовуються різні методи машинного навчання.

Можна підсумувати, що для моделей, що прогнозують точний бал, використовують частіше наступне:

- Багатовимірні лінійні регресії – класичний статистичний інструмент, який дозволяє встановити залежність між кількома незалежними змінними та результативним показником [9].

- Багатошаровий перцептрон (MLP) – нейронна мережа прямого поширення з нелінійними функціями активації, здатна виявляти складні приховані закономірності в даних. Для її навчання використовується алгоритм RPROP [10].

Метою досліджень цього напрямку є порівняння ефективності традиційного статистичного підходу та нейромережевої моделі у задачі точного прогнозування результатів.

Коли в задачах стоїть визначення класу успішності, то частіше використовують наступні моделі:

- Дерево рішень (алгоритм SPRINT) – модель, що послідовно розподіляє дані за умовами, формуючи структуру, подібну до схеми прийняття рішень [11].

- Випадковий ліс (Random Forest) – ансамблевий метод, який поєднує результати багатьох дерев рішень для підвищення стабільності та точності класифікації [12].

- Багатошаровий перцептрон, який також застосовується для задач класифікації.

У цьому випадку дослідження показують, що підхід, зоснований на деревоподібній структурі чи нейронній мережі, демонструє кращі результати при визначенні рівня успішності.

Мета статті

Дослідження спрямоване на побудову та порівняння різних моделей прогнозування академічної успішності з метою визначення найбільш ефективного підходу до оцінювання якості освітнього процесу.

В роботі проводяться дослідження можливих варіантів застосування моделей машинного навчання, а саме: дерева рішень, випадковий ліс, нейронні мережі, лінійна регресія, варіантів компенсування дальтонізму.

Виклад основного матеріалу

Було визначено способи оцінки успішності студентів та потенційні моделі нейронних мереж, статистичних алгоритмів для порівняння ефективності при передбаченні успішності студентів. Було розроблено математичну модель для визначення успішності студентів. Для її перевірки використано відкритий набір даних Students Performance Factors [13], що містить показники, які потенційно впливають на підсумкову оцінку. Серед них: кількість навчальних годин на тиждень, відвідуваність занять, участь у додаткових заняттях, попередні результати, якість викладання, рівень доходу та освіти батьків, їхня залученість до освітнього процесу, вплив однолітків тощо. Цільовим параметром виступає екзаменаційна оцінка як інтегральний показник якості навчання.

Перед моделюванням дані було ретельно підготовлено:

- вилучено записи з пропущеними значеннями, щоб уникнути викривлення результатів;

- категоріальні показники, що відображають рівень або ступінь (наприклад, «низький», «середній», «високий»), перетворено на впорядковані числові значення; номінальні ознаки без порядкової природи (наприклад, тип закладу) не враховувалися;

- виконано нормалізацію ознак для приведення їх до спільного масштабу (за винятком цільової змінної для збереження інтерпретованості результату);

- усунуто аномальні спостереження;

- проведено кореляційний аналіз і виключено змінні, що не демонстрували статистично значущого зв'язку з підсумковою оцінкою.

Для класифікації оцінки були розбиті на групи. Високі оцінки (більше 80) мали дуже маленьку кількість для навчання, тому визнані, як аномальні, тому для аналізу оцінки розбиті на рівні групи до 60 («Very low»), від 60 до 67 («Low»), від 67 до 73 («Medium»), від 73 до 80 («Good»).

Датасет розділено на навчальну групу та тестувальну. Після навчання було здійснено передбачення на основі тестувальної вибірки.

На рисунку 1 зображено графік лінійної регресії, де показано відповідність передбачених значень оригінальним. Зеленим кольором позначено найвищу точність моделі, а градація від зеленого кольору показує

ступінь похибки. Як видно, найбільша від’ємна похибка складає -3 (синій колір), а найбільша позитивна похибка сягає 5 (блідий зелено-жовтий колір). По графіку можна оцінити кількість відхилень, і як видно, є багато занижень оцінок.

Тобто модель схильна до невеликих похибок у вузькому діапазоні, що є прийнятним для формування рейтингів та глибокого аналізу якості навчання.

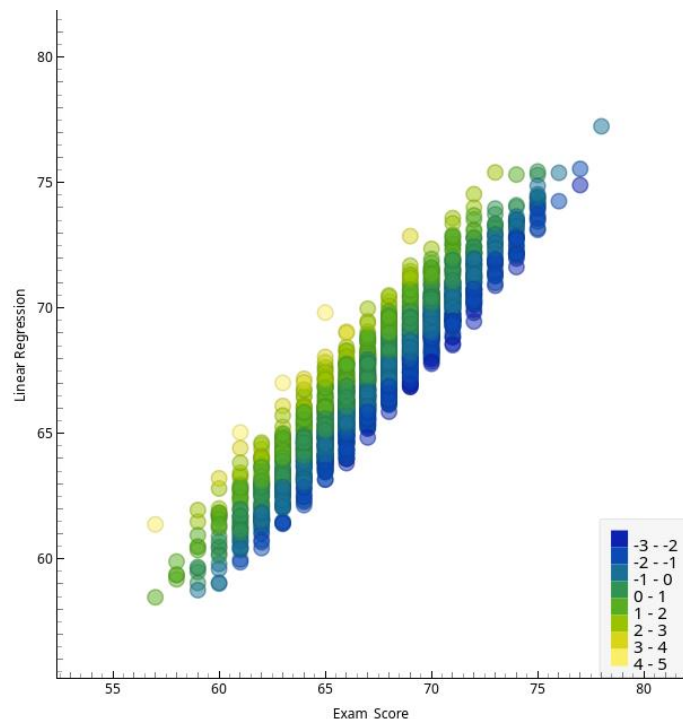


Рис. 1 – Графік лінійної регресії

Результати якості моделей для регресійного аналізу наведено у таблиці 1.

Таблиця 1

Порівняння моделей для регресійного аналізу

Параметри / Моделі	Лінійна регресія	Багато шаровий перцептрон
R2 – Коефіцієнт детермінації	0.974	0.949
MAPE – Середнє абсолютне відсоткове відхилення	0.009	0.006
MAE – Середня абсолютна похибка	0.428	0.57
RMSE – корінь від MSE	0.531	0.738
MSE – Середньоквадратична помилка	0.282	0.545

Як видно, обидві моделі ефективні для передбачення, але точність має велике значення тут, тому лінійна регресія є ефективнішою. Завдяки високій

точності та простішій конфігурації порівняно зі складними алгоритмами, регресійна модель є оптимальною для впровадження в системи освітньої аналітики, де важливо отримати точний числовий результат

В порівнянні у таблиці 2 наведені результати якості моделей для класифікації.

Таблиця 2

Порівняння моделей класифікації

Параметри / Моделі	Дерево рішень	Випадковий ліс	Багатошаровий перцептрон
PREC – Точність серед позитивних випадків	0.802	0.881	0.919
CA – Точність моделі	0.832	0.874	0.919
RECALL – чутливість	0.572	0.575	0.919
F1 – ідеальна влучність та повнота	0.668	0.696	0.918

У задачах класифікації за рівнями успішності, згідно з результатами, наведеними у таблиці 2, більш високі результати показує багатошаровий перцептрон. Дана модель продемонструвала точність та чутливість (RECALL) на рівні 0,919, що суттєво випереджає дерево рішень (0,832) та випадковий ліс (0,874).

На рисунку 2 наведено дерево рішень. Побудоване дерево рішень демонструє чітку ієрархію впливу ознак на прогнозований результат. Найбільш інформативним атрибутом виступає Attendance, який формує кореневий поділ вибірки, що свідчить про його

визначальний вплив на цільову змінну. Подальші розгалуження відбуваються переважно за ознаками Hours_Studied та Previous_Scores, що підтверджує логічну залежність результату від інтенсивності навчання та попередньої академічної підготовки. У правій гілці дерева (Attendance > 82) спостерігається стабільно вища частка позитивного класу (до 87,5%), що вказує на сильний позитивний зв'язок між високою відвідуваністю та досягненням результату 67–73. У лівій частині дерева (Attendance ≤ 82) переважає клас 61–67, що додатково підтверджує вагомість цієї ознаки.

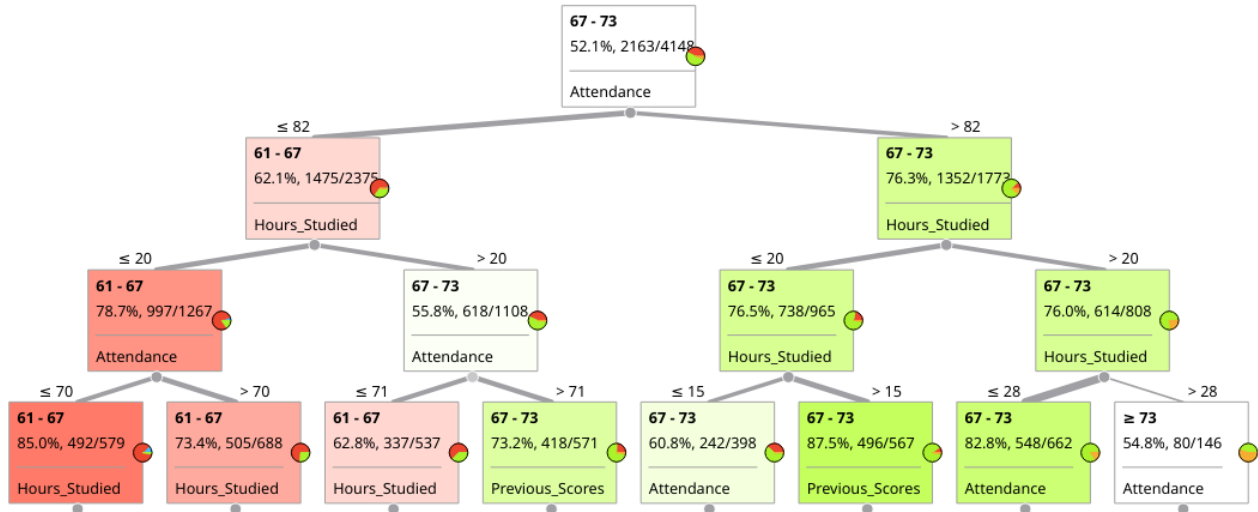


Рис. 2 – Дерево рішень для передбачення класу оцінки

Отримані метрики якості моделі свідчать про її помірно високу ефективність. Загальна точність (CA = 0,832) вказує на достатньо стабільну класифікацію в межах вибірки, а значення PREC = 0,802 підтверджує високу точність прогнозування позитивного класу. Водночас чутливість (RECALL = 0,572) є нижчою, що означає часткову втрату позитивних випадків під час класифікації. Значення F1 = 0,668 відображає компроміс між точністю та повнотою і свідчить про збалансованість моделі без суттєвого переобучення. Таким чином, модель є інтерпретованою та придатною для пояснювального аналізу факторів впливу, однак для підвищення виявлення позитивних випадків доцільним є подальше налаштування параметрів або балансування класів.

Як видно, нейронна мережа справляється краще за класичні дерева рішень.

На рисунку 3 зображено відповідність початкової та прогнозованої оцінки. Чим більше відхилення, тим

світліше колір групи. Як видно, реальні оцінки дуже часто відповідають прогнозованим.

Аналізуючи отримані результати можна помітити, що скупчення точок розташовано в зонах, де фактичний бал відповідає прогнозованому інтервалу (61–67, 67–73, ≥73). Оскільки основна маса студентів у вибірці зосереджена в інтервалах (61–67, 67–73) це вказує на те, що у виборці більшість студентів середнього рівня успішності, саме тому прогноз стабільний саме для цих категорій. В класифікації наявні точки поза головною діагоналлю (наприклад, фактичний бал 67–73 при прогнозі 61–67), що вказує на помилки класифікації. Невелика кількість таких помилок спостерігається переважно на межах інтервалів, що є типовим для багатокласової класифікації. Проте їхня кількість є відносно невеликою, що свідчить про високу узгодженість моделі з реальними даними та підтверджує класифікаційну здатність. Тобто можна зробити висновок що нейронна мережа справляється з задачею класифікацією студентів за рівнями успішності.

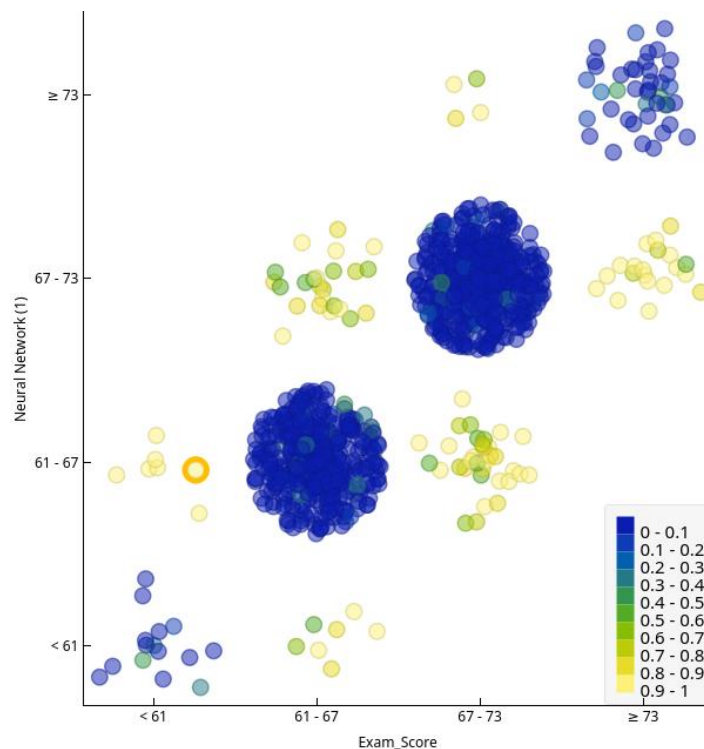


Рис. 3 – Графік відповідності прогнозованих нейронною мережею значень оригінальним

Аналізуючи результати класифікації методом дерева рішень (Random Forest), можна побачити відносно помірні результати, що помітно поступаються нейронним мережам. У задачі регресії нейронна мережа теж показує високу точність, достатню для успішних передбачень, проте лінійна регресія при використанні простішої конфігурації має точність вищу, через що вона виграє у більшості критеріїв, таких як швидкість, простота і сама точність.

Висновки

Процес вибору оптимальних методів для передбачення результатів семестрового контролю студентів безпосередньо залежить від конкретної мети дослідження та формату, у якому потрібно отримати прогноз.

У ситуаціях, коли необхідно встановити точну кількісну оцінку, раціональним вибором є використання моделей лінійної регресії. Порівняно зі складними нейронними мережами, регресійні моделі є значно швидшими у розрахунках та простішими в реалізації. Окрім того, у певних ситуаціях регресійні моделі здатні демонструвати навіть вищу ефективність, забезпечуючи меншу похибку при визначенні конкретних балів.

В свою чергу нейронні мережі показують високу точність при розподілі студентів по рівням, модель демонструє при цьому високу точність та збалансованість та здатність виявити складні закономірності в даних.

Хоча нейронні мережі, а саме багатошаровий перцептрон зазвичай демонструють вищу точність прогнозування (0,919 проти 0,832 у дерева рішень), дерева

рішень все одно залишаються конкурентоспроможними для поставленої задачі оцінки успішності студентів. Перевагою дерев рішень є подальша можливість побудови більш складних ансамблевих моделей, таких як випадковий ліс. Це дозволить ще більше підвищити стабільність та точність результатів.

Результати проведеного дослідження показують, що поєднання регресійних моделей та нейронних мереж є потужним фундаментом для створення всебічної системи сучасної освітньої аналітики. Оскільки впровадження таких прогнозних моделей в освітній процес дозволяє своєчасно виявляти студентів групи ризику, коригувати їхні індивідуальні траєкторії та об'єктивізувати систему оцінювання, дотримуючись принципів академічної доброчесності.

Перелік використаних джерел

- [1] Students' class performance prediction using machine learning classifiers / A. Ahmed et al. *Quaid-e-Awam University Research Journal of Engineering Science & Technology*. 2021. Vol. 19(1). Pp. 112–121. DOI: <https://doi.org/10.52584/qj.1901.16>.
- [2] Exploring Weka and Python for educational data mining: Naïve Bayes vs. J48 / Sheikh T., Saha R., Jahan N., Al Mamun A. *Proceedings of the International Conference on Machine Learning and Computational Systems (ICMCSI)*, Lalitpur, Nepal, 18-19 January 2024. Pp. 470–475. DOI: <https://doi.org/10.1109/ICMCSI61536.2024.00074>.

- [3] Interpreting the machine learning approaches to predict CGPA of the university students / R. Paul et al. *TEM Journal*. 2025. Vol. 14(4). Pp. 3438–3447. DOI: <https://doi.org/10.18421/TEM144-50>.
- [4] Pehlivanova T., Nedeva V. Attributes selection using machine learning for analysing students' dropping out of university: a case study. *IOP Conference Series: Materials Science and Engineering*. 2021. Vol. 1031. Article 012055. DOI: <https://doi.org/10.1088/1757-899X/1031/1/012055>.
- [5] Predicting Student Performance Using Random Forest Algorithm / R. Wania et al. *International Conference on Innovations and Trends in Engineering Technologies*, Biel, Switzerland, 25–26 November 2025. Vol. 01, no. 01. Pp. 01-02. DOI: <https://doi.org/10.5281/zenodo.15704741>.
- [6] Atokolo N. C. Design and implementation of students academic performance prediction model using decision tree classifiers : Project work submitted in partial fulfilment of the requirements for the award of a Bachelor of Science (B.Sc.) degree in Computer Science. Benin City : Benson Idahosa University, 2021. 78 p. DOI: <https://doi.org/10.13140/RG.2.2.20977.04962/1>.
- [7] Нікіфоров Р. О., Ткаченко Л. А. Моделювання навчальних результатів студентів із використанням алгоритмів машинного навчання. *Наукові записки*. 2025. № 162. С. 21–35. DOI: <https://doi.org/10.31392/NZ-udu-162.2025.03>.
- [8] Прогнозування залученості користувачів освітніх вебплатформ за допомогою алгоритмів машинного навчання / Лях І., Дудник В., Ціпіню Ю., Ціпіню А. *Науковий вісник Ужгородського університету*. Серія: Математика і інформатика. 2025. Вип. 46. С. 218–225. DOI: [https://doi.org/10.24144/2616-7700.2025.46\(1\).218-225](https://doi.org/10.24144/2616-7700.2025.46(1).218-225).
- [9] Regression analysis of multivariate recurrent event data allowing time-varying dependence with application to stroke registry data / W. Li et al. *Statistical Methods in Medical Research*. 2024. Vol. 33, no. 2. Pp. 309–320. DOI: <https://doi.org/10.1177/09622802231226330>.
- [10] Поркуян О. В., Самойлова Ж. Г. Моделювання лінійної нейронної мережі з зворотним поширенням помилки для основних каналів керування реактором синтезу оцтової кислоти. *Вісник Східноукраїнського національного університету імені Володимира Даля*. 2023. № 3(279). С. 31–36. DOI: <https://doi.org/10.33216/1998-7927-2023-279-3-31-36>.
- [11] Scalable random forest with data-parallel computing / Vázquez-Novoa F., Conejero J., Tatu C., Badia R. M. *Communications in Computer and Information Science*. 2023. Vol. 1858. Pp. 396–411. DOI: https://doi.org/10.1007/978-3-031-39698-4_27.
- [12] Scornet E., Biau G., Vert J.-P. A Random Forest Guided Tour: Non-expert infers the forest from interconnected trees. *Annals of Statistics*. 2022. Vol. 50(6). Pp. 3861–3885. DOI: <https://doi.org/10.1214/22-AOS2231>.
- [13] Dataset «Student Performance Factors». URL: <https://www.kaggle.com/datasets/lainguyn123/student-performance-factors> (дата звернення: 15.12.2025).

A COMPARATIVE ANALYSIS OF MACHINE LEARNING ALGORITHMS FOR DECISION SUPPORT IN PREDICTING ACADEMIC PERFORMANCE

Pronina O.I.	<i>PhD (Engineering), associate professor, SHEI «Priazovskyi state technical university», Dnipro, ORCID: https://orcid.org/0000-0001-7085-8027, e-mail: pronina_o_i@pstu.edu;</i>
Pertsev Ye.A.	<i>M.Sc., SHEI «Priazovskyi state technical university», Dnipro, ORCID: e-mail: pertsev_y_a@students.pstu.edu</i>

In the modern conditions of digitalization of education, the issue of objective assessment of learning outcomes and timely identification of students who need additional support is of particular importance. This work considers the possibilities of using machine learning algorithms to predict the final results of semester control. The relevance of the study is due to the need for tools for objectifying assessment and timely identification of risks of failure to adjust the individual educational trajectory. The study is aimed not only at obtaining an accurate numerical score, but also at determining the level of student success in a generalized categorical form (from «Very low» to «Good»), which allows for a deeper analysis of the quality of the educational process and timely adjustment of the individual learning trajectory. The work uses the «Students Performance Factors» data set, which includes indicators of class attendance, intensity of study, previous academic training, and socio-economic factors. Data preparation included anomaly removal, feature normalization, and correlation analysis to select the most informative parameters. For the experimental study, linear regression, multilayer perceptron (neural network), decision trees, and random forest were used. The models were evaluated using R^2 and MAE metrics for regression tasks and accuracy for classification tasks. The results showed that linear regression was the most appropriate for predicting the exact final score, demonstrating high accuracy and ease of interpretation. At the same

time, in the task of distributing students by achievement levels, the best results were obtained using a neural network. Decision tree analysis confirmed the important role of attendance as a key factor in academic success. The results obtained show that the choice of method should depend on the forecast format: for quantitative assessments, it is advisable to use regression, and for qualitative separation of students by levels, neural networks. The study opens up prospects for implementing educational analytics tools to improve learning efficiency.

Keywords: machine learning; regression; decision trees; neural networks; performance evaluation; academic performance.

References

- [1] A. Ahmed, K. Kumar, M. Khuhro, A. Wagan, I. Halepoto, and R. Hussain, "Students' class performance prediction using machine learning classifiers," *Quaid-e-Awam University Research Journal of Engineering Science & Technology*, vol. 19, no. 1, pp. 112–121, 2021. doi: [10.52584/qrj.1901.16](https://doi.org/10.52584/qrj.1901.16).
- [2] T. Sheikh, R. Saha, N. Jahan, and A. Al Mamun, "Exploring Weka and Python for educational data mining: Naïve Bayes vs. J48," in *Proc. of the Int. Conf. on Machine Learning and Computational Systems (ICMCSI)*, Lalitpur, Nepal, January 18–19, 2024, pp. 470–475. doi: [10.1109/ICMCSI61536.2024.00074](https://doi.org/10.1109/ICMCSI61536.2024.00074).
- [3] R. Paul, S. Hussain, K. Kannan, B. Mondal, A. Baruah, and S. Gaftandzhieva, "Interpreting the machine learning approaches to predict CGPA of the university students," *TEM Journal*, vol. 14, no. 4, pp. 3438–3447, 2025. doi: [10.18421/TEM144-50](https://doi.org/10.18421/TEM144-50).
- [4] T. Pehlivanova, and V. Nedeva, "Attributes selection using machine learning for analysing students' dropping out of university: A case study," *IOP Conference Series: Materials Science and Engineering*, vol. 1031, Art. no. 012055, 2021. doi: [10.1088/1757-899X/1031/1/012055](https://doi.org/10.1088/1757-899X/1031/1/012055).
- [5] R. Wania, S. Mukherjee, S. Mondal, S. Mukherjee, and S. Hazra, "Predicting student performance using random forest algorithm," in *Proc. of the Int. Conf. on Innovations and Trends in Engineering Technologies*, Biel, Switzerland, November 25–26, 2025, vol. 01, no. 01, pp. 01–02. doi: [10.5281/zenodo.15704741](https://doi.org/10.5281/zenodo.15704741).
- [6] N. C. Atokolo, "Design and implementation of students academic performance prediction model using decision tree classifiers," B.Sc. thesis, Dept. Computer Science, Benson Idahosa University, Benin City, Nigeria, 2021. doi: [10.13140/RG.2.2.20977.04962/1](https://doi.org/10.13140/RG.2.2.20977.04962/1).
- [7] R. O. Nikiforov, and L. A. Tkachenko, "Modeliuvannia navchalnykh rezultativ studentiv iz vykorystanniam alhorytmiv mashynnoho navchannia" ["Modeling students' learning results using machine learning algorithms"], *Naukovi zapysky – Scientific Notes*, № 162, pp. 21–35, 2025. doi: [10.31392/NZ-udu-162.2025.03](https://doi.org/10.31392/NZ-udu-162.2025.03). (Ukr.)
- [8] I. Liakh, V. Dudnyk, Yu. Tsipino, and A. Tsipino, "Prohnozuvannia zaluchenosti korystuvachiv osvitnikh vebplatform za dopomohoiu alhorytmiv mashynnoho navchannia" ["Forecasting user engagement in educational web platforms using machine learning algorithms"], *Naukovyi visnyk Uzhhorodskoho universytetu. Seriya: Matematika i informatyka – Scientific Bulletin of Uzhhorod University. Series of Mathematics and Informatics*, no. 46, pp. 218–225, 2025. doi: [10.24144/2616-7700.2025.46\(1\).218-225](https://doi.org/10.24144/2616-7700.2025.46(1).218-225). (Ukr.)
- [9] W. Li et al., "Regression analysis of multivariate recurrent event data allowing time-varying dependence with application to stroke registry data," *Statistical Methods in Medical Research*, vol. 33, no. 2, pp. 309–320, 2024. doi: [10.1177/09622802231226330](https://doi.org/10.1177/09622802231226330).
- [10] O. V. Porkuyan, and Zh. G. Samojlova, "Modeliuvannia liniinoi neuronnoi merezhi z zvorotnym poshyrenniam pomylky dlia osnovnykh kanaliv keruvannia reaktorom syntezu otstovoi kysloty" ["Modeling of a linear neural network with inverse error propagation for the main channels of acetic acid synthesis reactor control"], *Visnyk Skhidnoukrajinskoho natsionalnoho universytetu imeni Volodymyra Dalia – Visnik of the Volodymyr Dahl East Ukrainian National University*, № 3(279), pp. 31–36, 2023. doi: [10.33216/1998-7927-2023-279-3-31-36](https://doi.org/10.33216/1998-7927-2023-279-3-31-36). (Ukr.)
- [11] F. Vázquez-Novoa, J. Conejero, C. Tatu, and R. M. Badia, "Scalable random forest with data-parallel computing," *Communications in Computer and Information Science*, vol. 1858, pp. 396–411, 2023. doi: [10.1007/978-3-031-39698-4_27](https://doi.org/10.1007/978-3-031-39698-4_27).
- [12] E. Scornet, G. Biau, and J.-P. Vert, "A random forest guided tour: Non-expert infers the forest from interconnected trees," *Annals of Statistics*, vol. 50, no. 6, pp. 3861–3885, 2022. doi: [10.1214/22-AOS2231](https://doi.org/10.1214/22-AOS2231).
- [13] "Student Performance Factors," Kaggle, [Online]. Available: <https://www.kaggle.com/datasets/lainguyun123/student-performance-factors>.

Стаття надійшла 03.02.2026
Стаття прийнята 01.03.2026
Стаття опублікована 26.03.2026

Цитуйте цю статтю як: Проніна О. І., Перцев Є. А. Порівняльний аналіз алгоритмів машинного навчання в задачах підтримки прийняття рішень щодо академічної успішності. *Вісник Приазовського державного технічного університету. Серія: Технічні науки*. 2026. Вип. 53, том 1. С. 61–68. DOI: <https://doi.org/10.31498/2225-6733.53.1.2026.359776>.