УДК 004.8:378.14

ПОРІВНЯЛЬНЕ ОЦІНЮВАННЯ МОДЕЛЕЙ МАШИННОГО НАВЧАННЯ ТА БАЗОВОГО ПРОГНОЗУ ДЛЯ ВИЗНАЧЕННЯ НАВЧАЛЬНОЇ УСПІШНОСТІ

Кириллова С.К.
Національний університет «Одеська юридична академія»
Україна, Одеса
sonia.kirillova2014@gmail.com

***Анотація**. Досліджено якість прогнозування навчальної успішності за допомогою моделей машинного навчання та класичного базового прогнозу. Проведено порівняння логістичної регресії, дерева рішень, випадкового лісу й багатошарового персептрона за трьома прогнозованими показниками: ризиком неуспішності, рівнем успішності та підсумковим балом. Експериментально встановлено залежність результативності моделей від доступності поточних оцінок і визначено випадки, у яких застосування машинного навчання забезпечує перевагу над базовим підходом.*
***Ключові слова**: машинне навчання; навчальна успішність; порівняльне оцінювання; базовий прогноз; освітня аналітика.*

Вступ. Прогнозування навчальної успішності є одним із практичних завдань освітньої аналітики, оскільки дає змогу не лише оцінювати результати навчання, а й завчасно виявляти студентів, яким може знадобитися додаткова підтримка. Для розв'язання цього завдання застосовують методи машинного навчання, що дають змогу виявляти закономірності у даних про оцінки, відвідуваність та інші характеристики освітнього процесу. Водночас висока точність складнішої моделі сама по собі не свідчить про її практичну перевагу: результат необхідно зіставляти з простішими методами, які можуть забезпечувати порівнянну якість прогнозу за менших обчислювальних витрат і вищої зрозумілості.

У сучасних дослідженнях розглядаються різні алгоритми прогнозування успішності, а їх результативність залежить від характеристик навчальної вибірки, складу ознак і способу оцінювання. Зокрема, огляд методів освітнього інтелектуального аналізу даних засвідчує різноманітність підходів до прогнозування та відсутність універсально найкращого алгоритму [1]. У дослідженні [2] розглянуто оцінювання ефективності моделей машинного навчання для прогнозування успішності. Новіші роботи також демонструють актуальність порівняння алгоритмів та ансамблевих підходів для освітніх даних [3, 4].

Однак для практичного використання прогнозних моделей важливо оцінювати не тільки їхні абсолютні показники якості, а й додатковий результат порівняно з базовим прогнозом, що відтворює просте правило «буде як зараз». Такий підхід дає змогу встановити, чи виправдане ускладнення системи прогнозування та на якому етапі навчання воно має найбільшу цінність.

Метою роботи є порівняльне оцінювання моделей машинного навчання та класичного базового прогнозу для визначення навчальної успішності за різних обсягів доступної інформації про освітній процес, а також встановлення умов, за яких застосування машинного навчання забезпечує покращення якості прогнозування.

Основна частина. Для проведення експерименту використано відкритий набір даних UCI Student Performance, який містить відомості про 649 учнів двох португальських шкіл, зокрема проміжні оцінки, кількість пропусків занять, попередні незадовільні результати та

інші характеристики навчання [5]. Підсумкова оцінка за курс португальської мови позначена як *G3* і використана для формування трьох цільових показників: бінарного класу ризику неуспішності (*G3*<10), рівня успішності та числового значення підсумкового бала. Частка записів, віднесених до групи ризику, становить 15,4%, що зумовлює необхідність урахування дисбалансу класів під час оцінювання класифікаційних моделей.

Для порівняння вибрано чотири алгоритми машинного навчання: логістичну регресію, дерево рішень, випадковий ліс (Random Forest) і багатошаровий персептрон (MLP). У задачі прогнозування числового балу замість логістичної регресії застосовано лінійну регресію. Як точку відліку використано базовий прогноз без машинного навчання. За наявності поточних оцінок він передбачає, що підсумковий результат буде відповідати останній відомій оцінці; якщо оцінки ще відсутні, використовується найчастіший клас або середнє значення навчальної вибірки. Таке порівняння дає змогу визначити не лише найкращий алгоритм серед моделей машинного навчання, а й доцільність його застосування порівняно з простим правилом.

Експеримент використовував три набори вхідних ознак, які відповідають різним моментам прогнозування: без поточних оцінок; із додаванням оцінки першого періоду (*G1*); із додаванням оцінок першого та другого періодів (*G1*, *G2*). Така організація дослідження дала змогу простежити, як зміна доступності інформації впливає на співвідношення результативності моделей і базового прогнозу.

Вибірку стратифіковано поділено на навчальну частину (486 записів) і тестову (163 записи), яка містила 25 представників групи ризику. На навчальній частині додатково виконано п’ятикратну кросвалідацію. Для задачі класифікації ризику обчислювали точність (Accuracy), влучність (Precision), повноту (Recall), F1-міру та площу під ROC-кривою (ROC-AUC). Для прогнозування рівня успішності використовували точність і макроусереднені класифікаційні метрики, а для прогнозування підсумкового бала – середню абсолютну похибку (MAE), корінь із середньоквадратичної похибки (RMSE) та коефіцієнт детермінації (R^2). Результати порівняння моделей у задачі прогнозування ризику неуспішності наведено в табл. 1.

Таблиця 1 – Порівняння якості прогнозу ризику неуспішності на тестовій вибірці

Набір вхідних ознак	Модель	Recall	F1	ROC-AUC
Без проміжних оцінок	Random Forest	0,48	0,387	0,719
Без проміжних оцінок	Базовий прогноз	0,00	0,000	0,500
З оцінкою першого періоду (*G1*)	Дерево рішень	0,80	0,645	0,834
З оцінкою першого періоду (*G1*)	Random Forest	0,84	0,609	0,859
З оцінкою першого періоду (*G1*)	Базовий прогноз	0,84	0,618	0,931
З оцінками першого та другого періодів (*G1*, *G2*)	Дерево рішень	0,88	0,698	0,908
З оцінками першого та другого періодів (*G1*, *G2*)	Базовий прогноз	0,96	0,750	0,975

На початковому етапі, коли поточні оцінки відсутні, базовий прогноз не виявляє жодного студента групи ризику, оскільки для всіх записів визначає найчастіший клас. Його точність становить 0,847, проте F1-міра дорівнює нулю, а ROC-AUC – 0,500. Натомість найкращий результат серед моделей машинного навчання отримано для Random Forest: ROC-AUC становить 0,719 на тестовій вибірці, а за п’ятикратною кросвалідацією – 0,877 ± 0,049. Отже, за відсутності поточних оцінок машинне навчання забезпечує додаткову прогностичну можливість, якої не має базовий підхід.

Після появи оцінки першого періоду відмінності між підходами зменшуються. Дерево рішень демонструє найвищу серед моделей машинного навчання F1-міру (0,645), а Random Forest – найбільшу повноту (0,84). Водночас базовий прогноз досягає ROC-AUC 0,931, перевищуючи відповідні значення моделей машинного навчання на тестовій вибірці. За наявності оцінок двох періодів базовий прогноз також показує найвищі тестові значення F1-міри (0,750) і ROC-AUC (0,975). Проте за кросвалідацією Random Forest має дещо вищий ROC-AUC (0,969 ± 0,009 проти 0,957 ± 0,020 у базового прогнозу). Це свідчить про необхідність оцінювати моделі не за одним показником або одним поділом даних, а з урахуванням стабільності результатів.

Порівняння прогнозування рівня успішності підтверджує аналогічну тенденцію. Без поточних оцінок точність моделей становить лише 0,264–0,337, тобто вони не забезпечують достатньо надійного визначення рівня. Після додавання оцінки першого періоду найкращі результати демонструють дерево рішень і базовий прогноз – по 0,632. За наявності оцінок двох періодів точність дерева рішень зростає до 0,730, тоді як базовий прогноз досягає 0,748. Частка прогнозів, що відрізняються від фактичного рівня не більше ніж на один рівень, для обох підходів становить 1,000. Отже, у цій задачі складніші алгоритми не забезпечують істотної переваги над простим підходом, коли доступні проміжні оцінки.

У задачі прогнозування підсумкового бала найкращий результат без поточних оцінок отримано для Random Forest: RMSE становить 3,008 бала порівняно з 3,357 у базового прогнозу. Після додавання *G1* похибки обох підходів практично збігаються: RMSE дорівнює 1,853 для Random Forest і 1,845 для базового прогнозу. За наявності *G1* і *G2* випадковий ліс демонструє найменшу тестову похибку – RMSE 1,158 та R^2 0,881, тоді як базовий прогноз має RMSE 1,226 та R^2 0,867. Таким чином, у регресійній задачі машинне навчання забезпечує помітнішу перевагу на початковому етапі та незначне покращення за наявності оцінок двох періодів.

Окремого врахування потребує стабільність моделей. За результатами кросвалідації логістична регресія характеризується найменшим розкидом ROC-AUC, тоді як дерево рішень є чутливішим до складу навчальних підвибірок. Багатошаровий персептрон у більшості експериментальних комбінацій поступається іншим алгоритмам, зокрема за прогнозування ризику з оцінкою першого періоду його повнота становить лише 0,40. Це вказує на обмежену доцільність застосування нейронної мережі для набору такого обсягу.

Отримані результати дають підстави стверджувати, що вибір моделі має визначатися не її складністю, а типом прогнозованого показника, моментом формування прогнозу та перевагою над базовим підходом. Для оцінювання ризику на ранньому етапі доцільно розглядати Random Forest та інші моделі машинного навчання, оскільки вони дають змогу виявляти частину студентів групи ризику до появи поточних оцінок. Якщо ж оцінки вже доступні, базовий прогноз стає сильним конкурентом складніших алгоритмів. У таких умовах застосування машинного навчання слід обґрунтовувати додатковими перевагами, зокрема стабільністю, можливістю врахування інших ознак та поясненням прогнозу.

Водночас результати мають обмеження: експеримент проведено на відкритому наборі даних учнів середніх шкіл, а не на локальних даних закладу фахової передвищої освіти. Крім того, тестова вибірка містить лише 25 записів групи ризику, тому окремі значення метрик можуть змінюватися внаслідок помилки класифікації одного або кількох записів.

Висновки. У результаті порівняльного експерименту встановлено, що перевага моделей машинного навчання над базовим прогнозом залежить від доступності поточних оцінок і типу прогнозованого показника. За відсутності оцінок Random Forest забезпечив найкращий

результат у прогнозуванні ризику неуспішності серед досліджених моделей (ROC-AUC 0,719 на тестовій вибірці та 0,877 ± 0,049 за кросвалідацією), тоді як базовий прогноз не виявив жодного представника групи ризику. Після появи проміжних оцінок якість базового прогнозу суттєво зросла, і в низці експериментів він досягав результатів, порівнянних із моделями машинного навчання або кращих за них. У прогнозуванні підсумкового бала Random Forest забезпечив найменшу тестову похибку за наявності оцінок двох періодів (RMSE 1,158), однак його перевага над базовим прогнозом була невеликою.

Отже, експериментально підтверджено доцільність обов'язкового порівняння моделей машинного навчання з базовим прогнозом перед їх практичним застосуванням. Для невеликих освітніх вибірок модель слід вибирати з урахуванням не лише точності, а й повноти виявлення ризику, F1-міри, стабільності за кросвалідацією та зрозумілості результатів. Практичне використання таких моделей у системах електронного журналу доцільно розглядати як засіб підтримки рішень викладача, а не як механізм автоматичного визначення навчального статусу студента.

Використання інструментів ШІ. Під час підготовки тексту тез використано ChatGPT (GPT-6) для структурування та мовного редагування матеріалу, формулювання анотації.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

[1] Zhang, Y., Yun, Y., An, R., Cui, J., Dai, H., & Shang, X. (2021). Educational data mining techniques for student performance prediction: Method review and comparison analysis. *Frontiers in Psychology, 12*, Article 698490. https://doi.org/10.3389/fpsyg.2021.698490

[2] Стаценко, В.В. & Пилипенко, В.І. (2024). Оцінювання ефективності моделі прогнозування успішності методами машинного навчання. *Вісник Хмельницького національного університету. Технічні науки, 1*(331), 271–276. https://doi.org/10.31891/2307-5732-2024-331-41

[3] Oreški, D. (2025). Student success prediction based on machine learning and learning styles. *International Journal of Education and Information Technologies, 19,* 96–99. https://doi.org/10.46300/9109.2025.19.10

[4] Савенко, В.Д., Ярема, О.Р., & Бабічев, С.А. (2026). Стратегії ансамблевого навчання для прогнозування успішності студентів: порівняльна та інтерактивна структура. *Прикладні питання математичного моделювання, 9*(1), 327–339. https://doi.org/10.32782/mathematical-modelling/2026-9-1-30

[5] Cortez, P., & Silva, A. (2008). Using data mining to predict secondary school student performance. In *Proceedings of the 5th Annual Future Business Technology Conference*, 5–12. https://www.researchgate.net/publication/228780408_Using_data_mining_to_predict_secondary_school_student_performance

Kyryllova S.K.

Comparative evaluation of machine learning models and a baseline prediction for determining academic performance

Abstract. *The study compares machine learning models and a classical baseline approach for predicting students' academic performance. Logistic regression, decision tree, Random Forest, and multilayer perceptron models are evaluated across three prediction tasks: identifying academic failure risk, classifying performance levels, and predicting final grades. The experiments reveal how prediction quality depends on the availability of interim grades and identify the conditions under which machine learning provides an advantage over the baseline.*

Keywords*: machine learning; academic performance; comparative evaluation; baseline prediction; educational analytics.*