Tanasiienko N., Vrublevskyi I., Piatnychka V. & Urizchenko V. (2024). Derzhavna investytsiina polityka v Ukraini v umovakh viiny ta povoiennoho periodu [State investment policy in Ukraine during the war and post-war period]. *Visnyk Khmelnytskoho natsionalnoho universytetu. Ekonomichni nauky, 1* (326), 321–326. https://doi.org/10.31891/2307-5740-2024-326-50

United Nations Conference on Trade and Development. (2025). *World Investment Report 2025: International Investment in the Digital Economy*. Country Fact Sheet: Ukraine. https://unctad.org/publication/world-investment-report-2025

Vartsaba V. V. (2024). Priami inozemni investytsii v Ukraini: voiennyi kontekst [Foreign direct investment in Ukraine: military context]. *Naukovyi visnyk Uzhhorodskoho universytetu. Seriia «Ekonomika», 1* (63), 201–209. https://doi.org/10.24144/2409-6857.2024.1(63).201-209

Vasylkivskyi D. & Vyshniuk V. (2025). Formuvannia mekhanizmu zaluchennia inozemnykh investytsii [Formation of the mechanism for attracting foreign investment]. *Visnyk Khmelnytskoho natsionalnoho universytetu. Seriia «Ekonomichni nauky», 2*(3), 378–382. https://doi.org/10.31891/2307-5740-2025-342-3(2)-57

World Bank Group. (2026, April 13). *Donor Financing Mechanisms for Supporting Ukraine*. https://www.worldbank.org/en/brief/2026/04/13/world-bank-group-donor-financing-mechanisms-for-supporting-ukraine

World Bank Group. (2026, June 29). *New World Bank Group Support Will Help Mobilize Investment for Ukraine's Recovery*. https://www.worldbank.org/en/news/press-release/2026/06/29/new-world-bank-group-support-will-help-mobilize-investment-for-ukraine-recovery

World Bank, Government of Ukraine, European Commission & United Nations (2026). *Ukraine – Fifth Rapid Damage and Needs Assessment (RDNA5): February 2022 – December 2025*. Washington, DC: World Bank. https://documents.worldbank.org/en/publication/documents-reports/documentdetail/099022026094036395

Дослідження не отримувало зовнішнього фінансування.

Автор заявляє про відсутність конфлікту інтересів.

Інструменти ШІ в підготовці цієї роботи не використовувалися.

Стаття надійшла до редакції / Received: 13.09.2026
Статтю прийнято до публікації / Accepted: 26.09.2026
Оприлюднено / Published: 18.10.2026

УДК 005.8:330.322.01:004.85
JEL: C53; G17; G31; O22
DOI: https://doi.org/10.32983/2222-4459-2026-9-197-210

ЗАСТОСУВАННЯ МЕТОДІВ МАШИННОГО НАВЧАННЯ ДЛЯ ПРОГНОЗУВАННЯ ВІДХИЛЕНЬ БЮДЖЕТУ ІНВЕСТИЦІЙНОГО ПРОЄКТУ

©2026 **ПРОНЯКІН Є. В., ПРОНОЗА П. В.**

УДК 005.8:330.322.01:004.85
JEL: C53; G17; G31; O22

Пронякін Є. В., Проноза П. В. Застосування методів машинного навчання для прогнозування відхилень бюджету інвестиційного проєкту

Статтю присвячено обґрунтуванню методичного підходу до прогнозування відхилень бюджету інвестиційного проєкту на основі методів машинного навчання та розробленню механізму його інтеграції в систему фінансового контролінгу. Актуальність зумовлена системним характером перевищення кошторисної вартості, стійкість якого не зменшилася попри тривалий розвиток інструментарію кошторисного нормування, та реактивною природою наявних методів фінансового контролю, що генерують сигнал після матеріалізації відхилення у грошових потоках. Систематизовано науковий доробок із застосування машинного навчання до задач прогнозування вартості проєктів і встановлено, що він зосереджений на регресійній постановці, яка не відповідає операційним потребам фінансового контролінгу. Запропоновано переформулювання задачі у класифікаційну: модель оцінює умовну ймовірність виходу накопиченого відхилення за поріг фінансової суттєвості на визначеному горизонті попередження, причому поріг прив'язується до коефіцієнта резерву на непередбачені витрати, що надає йому прямого економічного змісту. Отримано аналітичний вираз оптимального порога ухвалення рішення через співвідношення вартостей помилок класифікації. Доведено, що загальноприйнятий поріг 0,5 економічно виправданий лише за припущення про рівновартісність хибнопозитивних і хибнонегативних помилок, яке в інвестиційному контексті не виконується, а сам поріг є проєктно-специфічним і залежить від структури фінансування, вартості капіталу та розміру резерву. Показано, що застосування синтетичного доповнення міноритарного класу без наступного перекалібрування ймовірностей руйнує економічну інтерпретацію порога. Обґрунтовано процедуру валідації, що усуває дві форми витоку інформації, характерні для панельних даних проєктного обліку, – групову залежність спостережень одного проєкту та доступ моделі до майбутніх макрофінансових режимів – і передбачає стратифіковане оцінювання за стадіями реалізації. Специфіковано простір предикторів із виокремленням макрофінансового блоку, критичного для умов високої інфляційної та курсової волатильності.

Ключові слова: *фінансове управління; інвестиційний проєкт; управління інвестиційними проєктами; машинне навчання.*
Табл.: *5.* ***Формул:*** *8.* ***Бібл.:*** *30.*

Пронякін Єгор Вадимович *– здобувач, кафедра обліку і фінансів, Національний технічний університет «Харківський політехнічний інститут» (вул. Кирпичова, 2, Харків, 61002, Україна)*
E-mail: *Yehor.Proniakin@emmb.khpi.edu.ua*
ORCID: *https://orcid.org/0009-0004-7805-1585*
Проноза Павло Володимирович *– доктор економічних наук, професор, професор кафедри фінансів і кредиту, Харківський національний економічний університет імені Семена Кузнеця (просп. Науки, 9а, Харків, 61165, Україна)*
E-mail: *ppv@hneu.edu.ua*
ORCID: *https://orcid.org/0000-0001-8130-4564*

UDC 005.8:330.322.01:004.85
JEL: C53; G17; G31; O22

Proniakin Ye. V., Pronoza P. V. Machine Learning for Early Warning of Budget Overruns in Investment Projects

The article substantiates a methodological approach to forecasting investment project budget deviations using machine learning and develops a mechanism for embedding it into the project financial control function. The relevance stems from the systemic nature of cost escalation, whose persistence has not diminished despite decades of refinement in cost estimation practice, and from the reactive design of existing financial control instruments, which generate a signal only after the deviation has materialised in cash flows. The body of research applying machine learning to project cost forecasting is systematised, revealing its concentration on the regression formulation, which does not correspond to the operational requirements of financial controlling. A reformulation into a classification task is proposed: the model estimates the conditional probability that cumulative cost deviation will exceed a materiality threshold within a defined warning horizon, with the threshold anchored to the contingency reserve ratio, which endows it with direct economic meaning. An analytical expression for the optimal decision threshold is derived from the ratio of misclassification costs. It is demonstrated that the conventional threshold of 0.5 is economically justified only under the assumption of equal costs of false positive and false negative errors, an assumption that never holds in an investment context, and that the threshold is project-specific, depending on the financing structure, cost of capital and reserve size. It is further shown that synthetic minority oversampling without subsequent probability recalibration invalidates the economic interpretation of the threshold. A validation procedure is substantiated that eliminates two forms of information leakage inherent in project accounting panel data – group dependence among observations of the same project and model access to future macro-financial regimes – and provides for stage-stratified performance assessment. The predictor space is specified with a distinct macro-financial block, critical under conditions of high inflation and exchange rate volatility.

Keywords: *financial management; investment project; investment project management; machine learning.*
Tabl.: *5.* ***Formulae:*** *8.* ***Bibl.:*** *30.*

Proniakin Yehor V. *– Applicant, Department of Accounting and Finance, National Technical University «Kharkiv Polytechnic Institute» (2 Kyrpychova Str., Kharkiv, 61002, Ukraine)*
E-mail: *Yehor.Proniakin@emmb.khpi.edu.ua*
ORCID: *https://orcid.org/0009-0004-7805-1585*
Pronoza Pavlo V. *– D. Sc. (Economics), Professor, Professor of the Department of Finance and Credit, Simon Kuznets Kharkiv National University of Economics (9a Nauky Ave., Kharkiv, 61165, Ukraine)*
E-mail: *ppv@hneu.edu.ua*
ORCID: *https://orcid.org/0000-0001-8130-4564*

Систематичне перевищення кошторисної вартості належить до найстійкіших патологій інвестиційної діяльності, і масштаб цього явища не зменшився за два десятиліття активних досліджень. Перше статистично репрезентативне дослідження ескалації витрат, побудоване на вибірці з 258 транспортних інфраструктурних проєктів сукупною вартістю близько 90 млрд дол. США, показало, що дев'ять із десяти великих проєктів громадського сектора завершуються з перевищенням початкового бюджету, причому середнє перевищення становить близько 45% для залізничних проєктів, 34% – для мостів і тунелів та 20% – для автомобільних доріг [1]. Пізніші узагальнення сформулювали цю закономірність як «залізний закон мегапроєктів» і оцінили світові витрати на мегапроєкти в 6–9 трлн дол. США щорічно, або близько 8% глобального валового внутрішнього продукту [2]. Принципово важливим є не сам факт наявності відхилень, а їхня однобічна спрямованість: якби йшлося про випадкову похибку прогнозування, розподіл був би приблизно симетричним відносно нуля, чого емпірично не спостерігається. Розподіл величини перевищення є асиметричним із вираженим правостороннім «важким хвостом», унаслідок чого інвестор наражається не лише на ризик перевищення бюджету, а й на ризик того, що величина цього перевищення виявиться значно більшою за очікувану. Стандартні підходи до резервування, побудовані на середніх або медіанних значеннях історичних відхилень, за таких умов систематично недооцінюють потребу в резерві.

Для економіки України проблема набуває особливої гостроти, оскільки прогнозована потреба в капітальних вкладеннях на відбудову супроводжується жорсткими обмеженнями внутрішніх фінансових ресурсів і підвищеними вимогами донорів щодо ефективності використання коштів, а в умовах капітального раціонування перевищення бюджету одного проєкту означає пряме витіснення іншого проєкту з інвестиційної програми.

Ключова методологічна проблема полягає в тому, що домінуючий інструментарій фінансового контролю інвестиційних проєктів має переважно реактивний характер. Класична методологія освоєного обсягу, аналіз відхилень від кошторису та ковзний бюджетний контроль фіксують розбіжність між плановими та фактичними показниками після того, як вона матеріалізувалася у грошових потоках. Прогнозні індикатори освоєного обсягу формально є зверненими в майбутнє, однак їхня прогнозна здатність спирається на екстраполяційну гіпотезу про збереження досягнутої результативності, що робить їх малочутливими до нелінійних і порогових ефектів; спроби підвищити передбачувану здатність планового обсягу через лінійне моделювання підтвердили обмеженість самої індексної конструкції [3].

Показовим є й те, що побудова статистично обґрунтованих меж толерантності для сигналів контролю потребує виходу за межі детермінованої логіки освоєного обсягу та звернення до імітаційного моделювання [4]. Емпіричні порівняння свідчать, що індексні формули дають найменш надійні оцінки саме на ранніх стадіях реалізації, коли кількість звітних періодів є малою, тобто тоді, коли управлінська цінність прогнозу є максимальною [5]. Виникає центральний парадокс наявної системи контролю: надійність сигналу та його управлінська цінність змінюються у протилежних напрямках, оскільки простір фінансового маневру звужується нелінійно протягом життєвого циклу – рішення, доступні на стадії проєктування, стають недоступними після укладення контрактів і авансування підрядника.

ОГЛЯД ЛІТЕРАТУРИ ТА АНАЛІЗ ПОПЕРЕДНІХ ДОСЛІДЖЕНЬ

Дослідницьке поле, у межах якого формується розв'язання цієї проблеми, склалося на перетині трьох традицій. Перша з них – *поведінково-економічна та інституційна діагностика причин перевитрат* – довела, що перевищення бюджету є системним явищем, зумовленим комбінацією оптимістичного зміщення та стратегічного спотворення оцінок з боку ініціаторів проєктів [1; 6]. Запропонований у межах цієї традиції метод прогнозування за референтним класом став першим кроком до дата-орієнтованого прогнозування, оскільки замінює внутрішній погляд на проєкт зовнішнім, побудованим на емпіричному розподілі результатів зіставних проєктів [7; 8]. Емпірична перевірка методу на реальних портфелях підтвердила його працездатність, водночас виявивши залежність результату від коректності формування референтного класу [9]. Принципове обмеження підходу полягає в тому, що він оперує агрегованими розподілами та не забезпечує індивідуалізації прогнозу для конкретного проєкту. Паралельно розвивалася критична лінія, яка ставить під сумнів монокаузальні пояснення перевитрат і обстоює перехід до ймовірнісної теорії причинності, що враховує взаємозалежність чинників [10]. Ця теза є концептуальним містком до машинного навчання: якщо причинність перевитрат є множинною та контекстно зумовленою, класичний регресійний апарат, який вимагає апріорної специфікації функціональної форми та переліку взаємодій, виявляється структурно непристосованим до задачі.

Друга традиція охоплює *кількісні моделі оцінювання та контролю вартості*. Ранні порівняльні дослідження зіставили множинну регресію, штучні нейронні мережі та міркування за прецедентами на масиві з 530 житлових проєктів і продемонстрували перевагу нейромережевих моделей у задачах параметричної оцінки на ранніх стадіях життєвого циклу [11]. Практично одночасно було показано, що для прогнозування саме відхилення вартості, а не її абсолютного рівня, перевага нейронних мереж над регресією не є безумовною та суттєво залежить від обсягу й якості навчальної вибірки [12]. Подальший розвиток пов'язаний із застосуванням еволюційних алгоритмів для налаштування моделей опорних векторів у задачі прогнозування кінцевої вартості проєкту [13] та з використанням ансамблевого моделювання й непараметричного бутстрепу на масиві з 1600 завершених проєктів, що дозволило отримати прогнози в межах ±10% фактичної вартості для 92% валідаційних спостережень [14]. Окремою гілкою є нечітке моделювання, у межах якого ймовірність зростання витрат визначається не для проєкту в цілому, а для окремих елементів кошторису [15; 16].

Третя традиція – *сучасні ансамблеві алгоритми в поєднанні з методологією інтерпретованого машинного навчання*. Застосування градієнтного бустингу до архівних даних будівельних проєктів разом із аналізом адитивних пояснень на основі значень Шеплі дозволило не лише отримати прийнятну точність, а й ранжувати чинники перевитрат, серед яких домінували початкова сума контракту, зміни обсягу робіт і поверховість об'єкта [17]. Розширення емпіричної бази до 836 публічних проєктів підтвердило перевагу алгоритмів типу CatBoost і стекінгових ансамблів з коефіцієнтом детермінації 0,883, а найвагомішим предиктором виявилися додаткові угоди щодо зміни обсягів із часткою впливу понад 41% [18]. Аналогічні результати отримано для енергетичних об'єктів із використанням випадкового лісу та інтерпретації

за Шеплі [19] і для змішаних вибірок будівельних проєктів на основі нейромережевих моделей з метаевристичним налаштуванням [20].

Методологічну основу цієї гілки становлять роботи, що визначили конструкцію ансамблів дерев рішень [21], масштабовану реалізацію градієнтного бустингу [22] та єдиний підхід до пояснення прогнозів складних моделей [23]. Особливе значення для фінансового управління має дослідження, у якому модель градієнтного бустингу побудовано безпосередньо на метриках освоєного обсягу за даними 110 реальних проєктів, що містили 1268 точок спостереження, і продемонстровано стійку перевагу над традиційними індексними формулами та моделями зростання на ранній, середній і пізній стадіях реалізації [5].

Критичний аналіз розробленості проблеми дає підстави виокремити п'ять невирішених питань. Перше стосується *перенесення моделей:* абсолютна більшість емпіричних досліджень виконана на вибірках проєктів країн зі стабільним макрофінансовим середовищем, тоді як перенесення отриманих специфікацій на вітчизняний ґрунт є некоректним через принципово інші режими інфляційної та валютної волатильності, специфіку публічних закупівель і наявність безпекових ризиків, які у зарубіжних моделях відсутні як клас предикторів. Друге питання – *домінування регресійної постановки задачі*, тобто прогнозування кінцевої вартості; значно менше опрацьовано постановку, релевантну саме для фінансового контролінгу, а саме, прогнозування ймовірності виходу відхилення за встановлений менеджментом поріг толерантності в межах заданого горизонту попередження. Винятками є лише поодинокі роботи, що працюють із класами перевищення [24; 25], причому жодна з них не доводить прогноз до регламентованого управлінського рішення. Третє питання – *економічна асиметрія помилок:* вартість хибнонегативного прогнозу, тобто пропущеного перевищення, і хибнопозитивного, що спричиняє безпідставну мобілізацію резерву або зупинку робіт, для інвестора принципово різна, однак поріг класифікації в наявних роботах фіксується за замовчуванням на рівні 0,5 без економічного обґрунтування. Четверте питання має суто технічний характер, але наслідки має змістовні: *у роботах, що застосовують синтетичне доповнення міноритарного класу, як правило, відсутній етап перекалібрування ймовірностей*, унаслідок чого будь-які твердження про оптимальний поріг втрачають чинність. П'яте питання – *організаційно-управлінський контур:* поза увагою залишається механізм трансформації прогнозу моделі в коригування резерву на непередбачені витрати, перегляд графіка вибірки кредитних коштів або ініціювання перепроєктування.

Огляди процесів цифрової трансформації управління проєктами підтверджують, що інтеграція предиктивних інструментів у контур прийняття рішень залишається найслабшою ланкою впровадження [26], а у вітчизняному дискурсі ця проблематика представлена переважно на рівні інформаційного забезпечення управління проєктами [27] та загальних можливостей аналітики великих даних у фінансових оцінках [28].

МЕТА ТА МЕТОДИ ДОСЛІДЖЕННЯ

Метою статті є обґрунтування методичного підходу до прогнозування відхилень бюджету інвестиційного проєкту на основі методів машинного навчання та розроблення механізму його інтеграції в систему фінансового контролінгу проєктної діяльності. Досягнення мети передбачає формалізацію цільової змінної в класифікаційній постановці з визначеним горизонтом попередження та економічно обґрунтованим порогом суттєвості; специфікацію простору предикторів із урахуванням вітчизняного макрофінансового та інституційного контексту; виведення оптимального порога ухвалення рішення з функції очікуваних грошових втрат замість статистичних критеріїв; розроблення процедури валідації, що враховує панельну та часову структуру даних; а також обґрунтування регламенту трансформації прогнозного сигналу в управлінське рішення.

Методологічну основу дослідження становлять такі методи. *Порівняльний аналіз* використано при зіставленні наявних моделей за єдиною системою критеріїв. *Формалізацію та абстрагування* – при побудові конструкції цільової змінної та простору предикторів. *Методи теорії статистичних рішень* – при виведенні оптимального порога класифікації з функції очікуваних грошових втрат. *Методи фінансового аналізу* – при вартісній параметризації наслідків помилок прогнозування. *Системний підхід* – при обґрунтуванні механізму вбудовування предиктивного модуля в регламент фінансового управління проєктом.

РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ

Дослідження побудовано за логікою послідовного звуження: від систематизації наявних підходів – до виявлення структурного розриву між постановкою прогнозної задачі в наукових роботах і операційними потребами фінансового контролінгу; від формалізації цільової змінної – до виведення критерію ухвалення рішення та процедури його емпіричної перевірки. Першим кроком цієї послідовності є *систематизація емпіричного матеріалу,*

що дозволяє зіставити наявні роботи не за декларованою тематикою, а за фактичною постановкою прогнозної задачі та за придатністю отриманого результату для ухвалення фінансового рішення.

Відбір робіт для систематизації здійснено за чотирма критеріями. Робота мала спиратися на емпіричну базу реальних інвестиційних проєктів, а не на змодельовані дані. Вона мала містити кількісно виміряний результат прогнозування, придатний для зіставлення з іншими роботами. Її цільовою змінною мав бути вартісний, а не календарний параметр проєкту. Нарешті, вона мала належати до однієї з трьох окреслених вище дослідницьких традицій, що дозволяє простежити еволюцію інструментарію від параметричних моделей до ансамблевих алгоритмів з інтерпретацією. Отримана вибірка охоплює період з 2003 по 2026 рік. Вона не є результатом систематичного огляду за формалізованим протоколом і не претендує на повноту відображення публікаційного потоку: відбір є цілеспрямованим і покликаний показати спектр методичних рішень, між якими фактично обирає дослідник, а також межі, у яких цей спектр досі залишався.

Зіставлення відібраних робіт здійснено за п'ятьма критеріями: емпірична база дослідження, застосований метод, постановка прогнозної задачі, ключовий кількісний результат і *обмеження, що перешкоджає використанню моделі в контурі фінансового управління*. Останній критерій є визначальним для мети статті, оскільки саме він переводить зіставлення з площини опису методів у площину придатності результату для ухвалення фінансового рішення. Результати систематизації наведено в *табл. 1*.

Таблиця 1

Порівняльний аналіз емпіричних досліджень із прогнозування перевитрат за інвестиційними проєктами

Дослідження	Емпірична база	Метод	Постановка задачі	Ключовий результат	Обмеження для фінансового управління
1	2	3	4	5	6
Attalla & Hegazy [12]	Проєкти реконструкції	ANN vs множинна регресія	Регресія (відхилення вартості)	Перевага ANN не є безумовною	Нестійка генералізація на малих вибірках
Kim et al. [11]	530 житлових проєктів, Сеул, 1997–2000	MRA / ANN / CBR	Регресія (вартість на ранній стадії)	ANN точніші за регресію та прецеденти	Прогноз одноразовий, не оновлюється в ході реалізації
Cheng et al. [13]	Будівельні проєкти	Еволюційний SVM	Регресія (EAC)	Підвищення точності EAC	Відсутній зв'язок із порогом суттєвості
Ahiaga-Dagbui & Smith [14]	1600 завершених проєктів	ANN, бутстреп, ансамблі	Регресія (кінцева вартість)	92% прогнозів у межах ±10% факту	Точковий прогноз без горизонту попередження
Asiedu et al. [24]	321 освітній проєкт, Гана	Множинна лінійна регресія	Ймовірність перевищення до підписання контракту	5 значущих предикторів; MAPE 30,2%	Низька пояснювальна сила; лише переддоговірна стадія
Coffie & Cudjoe [17]	Проєкти Гани, 2016–2018	XGBoost + SHAP	Регресія (коефіцієнт перевищення)	RMSE 0,202; провідні чинники – сума контракту, зміни обсягу	Прогноз не трансформовано в управлінське рішення
Turkyilmaz & Polat [25]	Кейс будівельної компанії	6 алгоритмів класифікації	Класи коефіцієнта перевищення за ризик-скорингом	Оновлення прогнозу на будь-якій стадії виконання	Не враховано асиметрію вартості помилок
Narbaev et al. [5]	110 проєктів, 1268 точок, Бельгія	XGBoost на метриках EVM	Регресія (EAC)	Перевага над EVM і growth-моделями на всіх стадіях	Прогнозується вартість, а не кероване відхилення

Закінчення табл. 1

1	2	3	4	5	6
Alazawy et al. [19]	Енергетичні об'єкти	Random Forest, SVR, KNN + SHAP	Регресія (вартість спорудження)	R^2 = 0,956; RMSE = 29,27	Галузева вузькість вибірки
Hamdan et al. [18]	836 публічних проєктів, Йорданія	CatBoost, стекінг, XGBoost	Регресія (коефіцієнт перевищення)	R^2 = 0,883; зміни обсягів – 41,2% важливості	Макрофінансові предиктори відсутні

Джерело: авторська розробка на основі [5; 11–14; 17–19; 24–25].

Порівняння за критерієм постановки задачі дає результат, який визначає подальшу логіку дослідження. Вісім із десяти розглянутих робіт формулюють задачу як регресійну, тобто прогнозують абсолютну вартість або коефіцієнт перевищення, і лише дві оперують імовірністю чи класами перевищення [24; 25]. Жодна з робіт не вводить явного горизонту попередження, який визначав би простір управлінського маневру, і жодна не пов'язує поріг ухвалення рішення з економічними наслідками помилки прогнозування. Другим суттєвим результатом зіставлення є структура використаних даних: лише дві роботи явно спираються на повторні спостереження в ході реалізації проєкту [5; 25], тоді як решта оперує одним спостереженням на проєкт, зафіксованим на момент завершення. Це означає, що проблематика залежності спостережень усередині проєкту та зміни макрофінансових режимів у часі, критична для коректного оцінювання прогнозної якості, у наявних роботах фактично не поставлена. Третім результатом є склад простору предикторів: за стабільно високої важливості контрактно-виконавчого блоку [17; 18] макрофінансові чинники в жодній із розглянутих специфікацій не виокремлено як самостійний блок ознак.

Виявлені обмеження визначають конструкцію запропонованого підходу. Його вихідною тезою є перехід від реактивної логіки контролю, за якої сигнал формується після відображення витрат в обліку, до предиктивної, за якої на кожну звітну дату оцінюється ймовірність настання фінансово суттєвої події на визначеному горизонті попередження. Принципові відмінності двох контурів за десятьма критеріями узагальнено в *табл. 2*.

Вихідною умовою коректної постановки задачі є відмова від інтуїтивного визначення перевищення бюджету на користь формалізованої конструкції, яка одночасно задає одиницю спостереження, подію, що прогнозується, та момент, на який прогноз є актуальним. Позначимо через $i = 1, \ldots, N$ окремий інвестиційний проєкт, а через $t = 1, \ldots, T_i$ – послідовність звітних періодів у межах його реалізації. Для кожної пари (i, t) з облікової системи доступні бюджет при завершенні BAC_i, фактичні витрати наростаючим підсумком $AC_{i,t}$, освоєний обсяг $EV_{i,t}$ і плановий обсяг $PV_{i,t}$, а похідними є частка виконання $c_{i,t} = EV_{i,t} / BAC_i$, індекс виконання вартості $CPI_{i,t} = EV_{i,t} / AC_{i,t}$ та індекс виконання розкладу $SPI_{i,t} = EV_{i,t} / PV_{i,t}$. Накопичене відносне відхилення витрат на звітну дату визначається як:

$$d_{i,t} = (AC_{i,t} - EV_{i,t}) / BAC_i. \quad \textbf{(1)}$$

Ця величина є нормованим на бюджет проєкту від'ємним значенням відхилення за вартістю та має принципову перевагу перед простим зіставленням фактичних витрат із плановим обсягом: вона співвідносить витрати з фактично виконаним обсягом робіт, а не з календарним планом, і тому не змішує вартісне відхилення з відставанням у графіку. Нормування на *BAC* робить показник порівнянним між проєктами різного масштабу, що є обов'язковою умовою навчання моделі на змішаній вибірці. Підсумкове відхилення проєкту визначається як $\delta_i = (AC_{i,T} - BAC_i) / BAC_i$, де T – момент завершення проєкту.

Ключовим елементом постановки є поріг фінансової суттєвості θ, який не слід задавати довільно або за міркуваннями зручності. Економічно обґрунтованим є прив'язування порога до коефіцієнта резерву на непередбачені витрати r_i, закладеного у фінансову модель проєкту: доки $d_{i,t} \le r_i$, відхилення поглинається резервом і не потребує зовнішніх дій, щойно $d_{i,t} > r_i$, виникає потреба в додатковому фінансуванні або в перегляді обсягу проєкту, а отже $\theta_i = r_i$. Альтернативною прив'язкою для проєктів із борговим фінансуванням є рівень, за якого порушуються фінансові ковенанти кредитної угоди. Цільова змінна формулюється у двох варіантах. Ковзна постановка з горизонтом попередження h задається як:

$$y_{i,t}^{(h)} = 1, \text{ якщо } d_{i,t+h} > \theta_i, \text{ та } 0 - \text{в іншому разі.} \quad \textbf{(2)}$$

Ця постановка полягає в тому, що накопичене відхилення перевищить поріг суттєвості через h звітних періодів після дати прогнозу. Терміналь-

Таблиця 2

Порівняльна характеристика реактивного та предиктивного контурів фінансового контролю інвестиційного проєкту

Критерій зіставлення	Реактивний контур на базі методології освоєного обсягу	Предиктивний контур на базі класифікаційної ML-моделі
Об'єкт вимірювання	Відхилення, що вже відбулося	Ймовірність відхилення, що очікується
Момент формування сигналу	Після відображення витрат в обліку	На звітну дату з випередженням на визначений горизонт
Форма результату	Точкова оцінка (CV, CPI, EAC)	Калібрована ймовірність перевищення порогу суттєвості
Гіпотеза, покладена в основу	Збереження досягнутої результативності	Емпірично виявлені нелінійні залежності та взаємодії
Чутливість до порогових ефектів	Низька [3]	Висока за побудовою [25]
Вимоги до даних	Дані поточного проєкту	Ретроспективний масив зіставних проєктів
Критерій оптимізації	Статистичні межі толерантності [4]	Мінімум очікуваних грошових втрат від помилок класифікації
Інтерпретованість результату	Прозора за конструкцією	Забезпечується інструментами XAI [23]
Управлінський тригер	Пояснення відхилення постфактум	Коригування резерву, перегляд графіка вибірки коштів
Ключове обмеження	Запізнення сигналу	Залежність від обсягу та якості історичних даних

Джерело: авторська розробка.

на постановка (3) стосується підсумкового результату проєкту, а прогноз формується на кожну звітну дату:

$$y_i^{(term)} = 1, \text{ якщо } \delta_i > \theta_i, \text{ та } 0 - \text{в іншому разі.} \quad \textbf{(3)}$$

Ковзна постановка генерує максимальну кількість спостережень, явно вводить горизонт попередження та безпосередньо відповідає операційній логіці бюджетного контролю, тоді як термінальна ближча до кінцевого інтересу інвестора, але позбавлена явного горизонту та дає лише один клас події на проєкт. Методично доцільним є використання ковзної постановки як основної, а термінальної – як перевірки на робастність, причому розбіжність результатів між ними є діагностично цінною, оскільки вказує на проєкти, у яких раннє відхилення було успішно компенсоване. Розширенням є порядкова постановка з кількома порогами $\theta^{(1)} < \theta^{(2)} < \theta^{(3)}$, що відповідає градації сигналів контролю за рівнями суттєвості й емпірично реалізована для класів коефіцієнта перевищення на основі динамічного ризик-скорингу [25], однак вимагає більшої вибірки через фрагментацію міноритарних класів.

Одиницею спостереження є пара «проєкт – звітна дата», що формує панельну структуру даних і породжує дві методичні вимоги, порушення яких є найпоширенішою помилкою в прикладних дослідженнях цього класу. Перша вимога – *відсутність витоку інформації з майбутнього:* кожна ознака, розрахована на момент *t*, має бути фактично відомою на цю дату, що стосується насамперед макрофінансових показників, які публікуються із затримкою і тому мають бути лаговані на реальний строк публікації, а не на нуль. Друга вимога – *нормування вартісних ознак на масштаб проєкту*, оскільки абсолютні показники непорівнянні між проєктами. Специфікацію простору предикторів наведено в *табл. 3*.

Частка виконання $c_{i,t}$ виконує подвійну функцію: вона є предиктором і водночас змінною стратифікації при оцінюванні якості моделі, оскільки прогнозна здатність істотно різниться між ранньою, середньою та пізньою стадіями реалізації.

Модель повертає оцінку умовної ймовірності $\hat{p}_{i,t} = P(y_{i,t}^{(h)} = 1 \mid x_{i,t})$, а управлінське рішення формується правилом генерування сигналу тривоги за умови $\hat{p}_{i,t} \geq \tau$. Вибір порога τ є не технічною деталлю, а самостійним фінансовим рішенням, оскільки саме він визначає баланс між двома типами помилок, вартість яких для інвестора принципово різна. Економічні наслідки чотирьох можливих поєднань рішення та фактичного стану систематизовано в *табл. 4*.

Сукупні очікувані втрати за заданого порога становлять:

$$L(\tau) = \Sigma_i \Sigma_t \left[(1 - y_{i,t}) \cdot 1\{\hat{p}_{i,t} \geq \tau\} \cdot C^{FP}_i + \right.$$
$$\left. + y_{i,t} \cdot 1\{\hat{p}_{i,t} < \tau\} \cdot C^{FN}_i + y_{i,t} \cdot 1\{\hat{p}_{i,t} \geq \tau\} \cdot C^{TP}_i\right]. \quad \textbf{(4)}$$

Таблиця 3

Специфікація простору предикторів

Блок	Змінні	Доступність	Обґрунтування включення
Проєктно-технічний (статичний)	Логарифм BAC, планова тривалість, тип об'єкта, спосіб закупівлі, кількість учасників торгів, відносне відхилення ціни переможця від очікуваної	З дати укладення контракту	Значущість підтверджено на переддоговірних моделях [11; 24]
Контрактно-виконавчий (кумулятивний)	Кількість і сумарна вартість додаткових угод (у частках BAC), кількість змін обсягу, індикатор зміни підрядника, фінансова класифікація підрядника	На кожну звітну дату	Зміни обсягів стабільно посідають перше місце за важливістю [17; 18]
Динамічний (EVM)	$CPI_{i,t}$, $SPI_{i,t}$, $c_{i,t}$, $d_{i,t}$, перші різниці ΔCPI, ΔSPI за 1–3 періоди, дисперсія CPI за ковзним вікном	На кожну звітну дату	Ефективність конфігурації на EVM-метриках підтверджено емпірично [5]
Макрофінансовий	Індекс цін виробників у будівництві (лагований), місячна зміна курсу, ставка за новими кредитами нефінансовим корпораціям, індекс вартості ключових ресурсів	З урахуванням лагу публікації	Блок, слабко представлений у зарубіжних специфікаціях; критичний для умов високої волатильності
Контекстний	Регіон реалізації, індикатор безпекового ризику, джерело фінансування (бюджет / донор / приватний)	З дати ухвалення рішення	Специфічний клас чинників без аналогів у зарубіжних моделях

Джерело: авторська розробка.

Таблиця 4

Специфікація простору предикторів

Рішення / фактичний стан	Перевищення відбудеться (y = 1)	Перевищення не відбудеться (y = 0)
Сигнал згенеровано ($\hat{p} \geq \tau$)	Істинно позитивний. Чиста вартість C^{TP}: витрати на превентивне втручання за вирахуванням відвернутої частини перевищення; може бути від'ємною, тобто становити чисту вигоду	Хибно позитивний. Вартість C^{FP}: альтернативна вартість надлишково іммобілізованого резерву, адміністративні витрати на перегляд, можливі втрати від призупинення робіт
Сигнал не згенеровано ($\hat{p} < \tau$)	Хибно негативний. Вартість C^{FN}: непокрите перевищення, премія за екстрене залучення фінансування, штрафні санкції, репутаційні втрати	Істинно негативний. Вартість дорівнює нулю

Джерело: авторська розробка.

Оптимальний поріг визначається як $\tau^* = \arg\min L(\tau)$. За умови сталості вартостей у межах проєкту задача має аналітичний розв'язок, що випливає безпосередньо з порівняння очікуваних наслідків двох рішень: сигнал доцільно генерувати тоді, коли $p \cdot C^{TP} + (1 - p) \cdot C^{FP} < p \cdot C^{FN}$, звідки:

$$\tau^*_i = C^{FP}{}_i / (C^{FP}{}_i + C^{FN}{}_i - C^{TP}{}_i). \quad (5)$$

Ця формула має пряме управлінське прочитання: що дорожчою є пропущена подія відносно хибної тривоги, то нижчим є оптимальний поріг і то чутливішою має бути система раннього попередження. Симетричний випадок $C^{FP} = C^{FN}$ і $C^{TP} = 0$ дає $\tau^* = 0{,}5$ – стандартний поріг, який у прикладних роботах застосовується за замовчуванням і який,

як видно з виведення, є економічно виправданим лише за припущення про рівновартісність помилок, що в інвестиційному контексті не виконується ніколи. Параметризація складових здійснюється як:

$$C^{FN}_i = E[\max(\delta_i - r_i, 0)] \cdot BAC_i \cdot (1 + k_i) + P_i, \quad \textbf{(6)}$$

де: k_i – премія за екстрене залучення фінансування відносно вартості планового, а P_i – очікувані штрафні санкції та втрати від затримки введення в експлуатацію. Очікувана величина непокритого перевищення $E[\max(\delta_i - r_i, 0)]$ розраховується за емпіричним розподілом відхилень відповідного референтного класу проєктів, що є безпосередньою точкою поєднання запропонованого підходу з методом прогнозування за референтним класом [7; 9] та з підходами до інтеграції останнього з методологією освоєного обсягу [29]. Вартість хибної тривоги параметризується як:

$$C^{FP}_i = r_i \cdot BAC_i \cdot (WACC_i - r_f) \cdot (h/12) + A_i, \quad \textbf{(7)}$$

де: перший доданок відображає альтернативну вартість іммобілізації резерву на горизонті попередження; A_i – адміністративні витрати на процедуру перегляду.

Принциповим є висновок, що τ^* є проєктно-специфічним, а не єдиним для всієї моделі: проєкт із дорогим борговим фінансуванням і жорсткими ковенантами вимагає нижчого порога, ніж проєкт, що фінансується з власних коштів із широким резервом. Це відрізняє запропонований підхід від наявної практики, у якій поріг класифікації фіксується глобально.

Застосування формули (5) коректне лише за умови, що $\hat{p}$ є каліброваною ймовірністю, тобто серед спостережень із прогнозом близько 0,3 частка позитивних випадків справді наближається до 0,3. Алгоритми градієнтного бустингу за замовчуванням цієї властивості не мають, тому обов'язковим етапом є пост-калібрування методом ізотонічної регресії або логістичного перетворення Платта на окремій валідаційній підвибірці, що не використовувалася для навчання. Якість калібрування контролюється за оцінкою Брієра:

$$B = (1/n) \cdot \Sigma (\hat{p} - y)^2 \quad \textbf{(8)}$$

та за діаграмою надійності. Система метрик відповідно перебудовується: частка правильних відповідей відхиляється як нерелевантна за дисбалансу класів, основною метрикою якості ранжування є площа під кривою «точність – повнота», основною метрикою якості ймовірностей – оцінка Брієра, а основною метрикою управлінської цінності – величина $L(\tau^*)$ у грошовому вимірі, зіставлена з аналогічним показником для базового сценарію контролю на основі індексних формул освоєного обсягу.

Подія перевищення порога суттєвості є відносно рідкою, що породжує дисбаланс навчальної вибірки. Стандартним рішенням є синтетичне доповнення міноритарного класу [30], однак у розглядуваній задачі його застосування потребує застереження, яке в прикладних роботах зазвичай ігнорується. Ресемплінг змінює апріорний розподіл класів, а отже, систематично зміщує вихідні ймовірності моделі вгору; оскільки вся конструкція оптимального порога спирається на каліброваність $\hat{p}$, некритичне застосування синтетичного доповнення руйнує економічну інтерпретацію τ^*. Методично коректними є два шляхи: або застосування вагових коефіцієнтів класів безпосередньо у функції втрат алгоритму без зміни складу вибірки, або застосування ресемплінгу з обов'язковим подальшим перекалібруванням на невибірковій частині даних, що зберегла початковий розподіл класів. Перший шлях є переважним через меншу кількість етапів, на яких може виникнути похибка.

Панельна структура даних породжує дві форми залежності спостережень, кожна з яких робить стандартну k-кратну кросвалідацію недійсною. Групова залежність полягає в тому, що спостереження, які належать одному проєкту, описують ту саму сутність у різні моменти часу; випадкове розбиття помістить частину звітних дат проєкту до навчальної вибірки, а частину – до тестової, унаслідок чого модель фактично прогнозуватиме поведінку проєкту, який вона вже спостерігала, і оцінка якості виявиться систематично завищеною. Часова залежність полягає у зміні макрофінансових режимів: навчання на даних пізніших періодів із тестуванням на ранніх дає моделі доступ до інформації про майбутні умови, недоступної в реальному застосуванні. Запропонована схема усуває обидві проблеми через дворівневу вкладену конструкцію, порівняльні характеристики якої наведено в *табл. 5*.

Зовнішній контур реалізує розбиття з рухомим початком: обирається послідовність часових меж $T_1 < T_2 < \ldots < T_m$, і на кроці m модель навчається на всіх проєктах, завершених до T_m, і тестується на проєктах, розпочатих після T_m; використання розширюваного вікна є переважним перед ковзним за обмеженого обсягу даних, оскільки зберігає максимум навчальної інформації. Буферна зона виключення передбачає вилучення з навчальної вибірки проєктів, періоди реалізації яких перекриваються з тестовим вікном, навіть за формальної відповідності критерію дати завершення, оскільки без цього кроку модель отримує інформацію про макроекономічні умови тестового періоду через проєкти-«сусіди», що є прихованою формою ви-

Таблиця 5

Порівняння схем валідації

Характеристика	Стандартна k-кратна крос-валідація	Запропонована групова валідація з рухомим початком
Принцип розбиття	Випадковий за спостереженнями	Хронологічний за датою завершення проєкту, групування за проєктом
Групова залежність	Не враховується; витік у межах проєкту	Усувається: усі спостереження проєкту в одному фолді
Часова залежність	Не враховується; доступ до майбутніх режимів	Усувається: навчання лише на проєктах, завершених до межі T
Перекриття періодів реалізації	Не контролюється	Контролюється буферною зоною виключення
Налаштування гіперпараметрів	На тих самих фолдах	На вкладених групових фолдах усередині навчального блоку
Характер оцінки якості	Систематично завищена	Наближена до умов реального застосування

Джерело: авторська розробка.

току. Внутрішній контур призначено для підбору гіперпараметрів і здійснюється методом групової k-кратної валідації всередині навчального блоку, де групувальною змінною є ідентифікатор проєкту. Підсумкові метрики розраховуються окремо для ранньої (частка виконання 1–29%), середньої (30–69%) та пізньої (70–95%) стадій реалізації – розмежування, що застосовується в емпіричних дослідженнях прогнозування вартості на EVM-даних [5]. Стратифікація за стадіями є принциповою, оскільки саме на ранній стадії прогноз має найвищу управлінську цінність і водночас найнижчу очікувану точність, а агрегована метрика по всій вибірці приховує цей розрив і створює хибне враження про придатність моделі.

Пояснення прогнозів здійснюється методом адитивних пояснень на основі значень Шеплі [23] на двох рівнях. *Глобальний рівень* дає ранжування чинників за середнім внеском у прогноз і слугує для змістовної валідації моделі: якщо ранжування суперечить економічній логіці або галузевому досвіду, це є сигналом про артефакт даних, а не про наукове відкриття. *Локальний рівень* дає розклад конкретного прогнозу за внесками окремих ознак і є безпосереднім входом для управлінського рішення, оскільки відповідає на питання, які саме чинники спричинили сигнал і на що має бути спрямоване втручання. Контроль стабільності здійснюється через зіставлення ранжування важливості ознак між фолдами зовнішнього контуру, а суттєва розбіжність ранжувань свідчить про перенавчання або нестабільність залежностей у часі та є підставою для спрощення моделі. Додатково запроваджується моніторинг зміщення розподілів ознак за індексом стабільності популяції з визначенням регламентного строку перенавчання.

Практична цінність підходу реалізується лише за умови його вбудовування в регламент фінансового управління. Прогнозна ймовірність, що перевищує калібрований поріг τ^*, має запускати визначену послідовність дій: перерахунок необхідного обсягу резерву на непередбачені витрати, перегляд графіка вибірки кредитних коштів або траншів донорського фінансування, ініціювання процедури аналізу вартості проєктних рішень. Локальний розклад прогнозу за внесками ознак визначає, який саме з цих важелів є релевантним у конкретному випадку: домінування контрактно-виконавчого блоку вказує на потребу в перегляді умов договору, домінування макрофінансового блоку – на потребу в хеджуванні або перегляді структури фінансування. Забезпечення інтерпретованості в умовах донорського фінансування є не побажанням, а обмеженням, оскільки модель, результат якої не може бути пояснений наглядовій раді, аудитору чи представнику донора, не буде впроваджена незалежно від її прогнозної якості.

ВИСНОВКИ

Проведене дослідження дозволяє сформулювати такі узагальнення. Систематичне перевищення бюджету інвестиційних проєктів має структурний, а не випадковий характер, що підтверджується однобічною спрямованістю відхилень, їхньою стійкістю протягом десятиліть попри методичні вдосконалення та асиметричністю розподілу з важким правостороннім хвостом; це означає, що подальше вдосконалення кошторисного нормування саме по

собі не здатне усунути проблему та має бути доповнене інструментами раннього виявлення. Множинна та взаємозалежна природа причинності перевитрат є методологічним аргументом на користь методів машинного навчання, оскільки чинники перевищення діють у взаємодії, а їхня функціональна форма не може бути специфікована апріорі. Наявні наукові напрацювання із застосування машинного навчання до задач прогнозування вартості є значними, однак зосереджені переважно на регресійній постановці, яка не відповідає операційним потребам фінансового контролінгу; переформулювання задачі з регресійної у класифікаційну з визначеним горизонтом попередження та порогом суттєвості, прив'язаним до коефіцієнта резерву на непередбачені витрати, є *ключовим елементом наукової новизни запропонованого підходу*.

Отримано аналітичний вираз оптимального порога ухвалення рішення через співвідношення вартостей помилок класифікації, з якого випливає, що загальноприйнятий поріг 0,5 економічно виправданий лише за припущення про рівновартісність хибнопозитивних і хибнонегативних помилок, яке в інвестиційному контексті не виконується, а сам поріг є проєктно-специфічним і залежить від структури фінансування, вартості капіталу та розміру резерву конкретного проєкту. Показано, що застосування синтетичного доповнення міноритарного класу без подальшого перекалібрування ймовірностей руйнує економічну інтерпретацію порога, унаслідок чого методично коректним є або використання вагових коефіцієнтів класів у функції втрат алгоритму, або обов'язкове перекалібрування на підвибірці з непорушеним розподілом класів.

Обґрунтовано процедуру валідації, що усуває дві форми витоку інформації, характерні для панельних даних проєктного обліку, – групову залежність спостережень одного проєкту та доступ моделі до майбутніх макрофінансових режимів, – і передбачає стратифіковане оцінювання за стадіями реалізації, без якого агрегована метрика приховує розрив між управлінською цінністю та точністю прогнозу на ранній стадії.

Пряме перенесення зарубіжних специфікацій на вітчизняний ґрунт є методологічно некоректним, оскільки макрофінансова волатильність, інституційна специфіка публічних закупівель і наявність безпекових ризиків формують класи предикторів, які у зарубіжних моделях або відсутні, або мають незначну вагу. Прогнозна модель набуває управлінського сенсу лише за умови вбудовування в регламент фінансового управління, оскільки прогноз, що не запускає визначеної послідовності дій, залишається аналітичною вправою.

Напрямом подальших досліджень є емпірична апробація запропонованого підходу на масиві вітчизняних інвестиційних проєктів із перевіркою гіпотези про суттєвість макрофінансового блоку предикторів, а також розроблення методики вартісного оцінювання управлінського ефекту від упровадження предиктивного модуля, яка дозволила б порівнювати альтернативні моделі за критерієм фінансового результату, а не статистичної точності. ■

БІБЛІОГРАФІЯ

1. Flyvbjerg B., Holm M. K. S., Buhl S. L. Underestimating costs in public works projects: Error or lie? *Journal of the American Planning Association*. 2002. Vol. 68. Iss. 3. P. 279–295. DOI: https://doi.org/10.1080/01944360208976273
2. Flyvbjerg B. What you should know about megaprojects and why: An overview. *Project Management Journal*. 2014. Vol. 45. Iss. 2. P. 6–19. DOI: https://doi.org/10.1002/pmj.21409
3. Chen H. L., Chen W. T., Lin Y. L. Earned value project management: Improving the predictive power of planned value. *International Journal of Project Management*. 2016. Vol. 34. Iss. 1. P. 22–29. DOI: https://doi.org/10.1016/j.ijproman.2015.09.008
4. Colin J., Vanhoucke M. Setting tolerance limits for statistical project control using earned value management. *Omega*. 2014. Vol. 49. P. 107–122. DOI: https://doi.org/10.1016/j.omega.2014.06.001
5. Narbaev T., Hazır Ö., Khamitova B., Talgat S. A machine learning study to improve the reliability of project cost estimates. *International Journal of Production Research*. 2024. Vol. 62. Iss. 12. P. 4372–4388. DOI: https://doi.org/10.1080/00207543.2023.2262051
6. Flyvbjerg B., Holm M. K. S., Buhl S. L. What causes cost overrun in transport infrastructure projects? *Transport Reviews*. 2004. Vol. 24. Iss. 1. P. 3–18. DOI: https://doi.org/10.1080/0144164032000080494a
7. Flyvbjerg B. From Nobel Prize to project management: Getting risks right. *Project Management Journal*. 2006. Vol. 37. Iss. 3. P. 5–15. DOI: https://doi.org/10.1177/875697280603700302
8. Flyvbjerg B. Quality control and due diligence in project management: Getting decisions right by taking the outside view. *International Journal of Project Management*. 2013. Vol. 31. Iss. 5. P. 760–774. DOI: https://doi.org/10.1016/j.ijproman.2012.10.007
9. Batselier J., Vanhoucke M. Practical application and empirical evaluation of reference class forecasting for project management. *Project Management Journal*. 2016. Vol. 47. Iss. 5. P. 36–51. DOI: https://doi.org/10.1177/875697281604700504
10. Love P. E. D., Ahiaga-Dagbui D. D., Irani Z. Cost overruns in transportation infrastructure projects: Sowing the seeds for a probabilistic theory of causation. *Transportation Research Part A: Policy and Practice*. 2016. Vol. 92. P. 184–194. DOI: https://doi.org/10.1016/j.tra.2016.08.007

11. Kim G.-H., An S.-H., Kang K.-I. Comparison of construction cost estimating models based on regression analysis, neural networks, and case-based reasoning. *Building and Environment*. 2004. Vol. 39. Iss. 10. P. 1235–1242. DOI: https://doi.org/10.1016/j.buildenv.2004.02.013
12. Attalla M., Hegazy T. Predicting cost deviation in reconstruction projects: Artificial neural networks versus regression. *Journal of Construction Engineering and Management*. 2003. Vol. 129. Iss. 4. P. 405–411. DOI: https://doi.org/10.1061/(ASCE)0733-9364(2003)129:4(405)
13. Cheng M.-Y., Peng H.-S., Wu Y.-W., Chen T.-L. Estimate at completion for construction projects using evolutionary support vector machine inference model. *Automation in Construction*. 2010. Vol. 19. Iss. 5. P. 619–629. DOI: https://doi.org/10.1016/j.autcon.2010.02.008
14. Ahiaga-Dagbui D. D., Smith S. D. Dealing with construction cost overruns using data mining. *Construction Management and Economics*. 2014. Vol. 32. Iss. 7–8. P. 682–694. DOI: https://doi.org/10.1080/01446193.2014.933854
15. Plebankiewicz E. Model of predicting cost overrun in construction projects. *Sustainability*. 2018. Vol. 10. Iss. 12. Art. 4387. DOI: https://doi.org/10.3390/su10124387
16. Plebankiewicz E., Wieczorek D. Prediction of cost overrun risk in construction projects. *Sustainability*. 2020. Vol. 12. Iss. 22. Art. 9341. DOI: https://doi.org/10.3390/su12229341
17. Coffie G. H., Cudjoe S. K. F. Using extreme gradient boosting (XGBoost) machine learning to predict construction cost overruns. *International Journal of Construction Management*. 2024. Vol. 24. Iss. 16. P. 1742–1750. DOI: https://doi.org/10.1080/15623599.2023.2289754
18. Hamdan M. M., Thneibat M., Hyari K. Predicting cost overrun in construction projects using machine learning algorithms: The case of Jordan. *Engineering, Construction and Architectural Management*. 2026. Vol. 33. Iss. 6. P. 4732–4763. DOI: https://doi.org/10.1108/ECAM-09-2024-1209
19. Alazawy S. F. M., Ahmed M. A., Raheem S. H., et al. Explainable machine learning to predict the construction cost of power plant based on random forest and Shapley method. *CivilEng*. 2025. Vol. 6. Iss. 2. Art. 21. DOI: https://doi.org/10.3390/civileng6020021
20. Al-Mnaseer R., Al-Smadi S., Al-Bdour H. Machine learning-aided time and cost overrun prediction in construction projects: Application of artificial neural network. *Asian Journal of Civil Engineering*. 2023. Vol. 24. P. 2583–2593. DOI: https://doi.org/10.1007/s42107-023-00665-7
21. Breiman L. Random forests. *Machine Learning*. 2001. Vol. 45. Iss. 1. P. 5–32. DOI: https://doi.org/10.1023/A:1010933404324
22. Chen T., Guestrin C. XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York : ACM, 2016. P. 785–794. DOI: https://doi.org/10.1145/2939672.2939785
23. Lundberg S. M., Lee S.-I. A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*. 2017. Vol. 30. URL: https://proceedings.neurips.cc/paper_files/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf
24. Asiedu R. O., Frempong N. K., Alfen H. W. Predicting likelihood of cost overrun in educational projects. *Engineering, Construction and Architectural Management*. 2017. Vol. 24. Iss. 1. P. 21–39. DOI: https://doi.org/10.1108/ECAM-06-2015-0103
25. Turkyilmaz A. H., Polat G. Risk-based completion cost overrun ratio estimation in construction projects using machine learning classification algorithms: A case study. *Buildings*. 2024. Vol. 14. Iss. 11. Art. 3541. DOI: https://doi.org/10.3390/buildings14113541
26. Nenni M. E., De Felice F., De Luca C., Forcina A. How artificial intelligence will transform project management in the age of digitisation: A systematic literature review. *Management Review Quarterly*. 2025. Vol. 75. P. 1669–1716. DOI: https://doi.org/10.1007/s11301-024-00418-z
27. Панас Я. В. Інформаційне забезпечення управління інноваційними проектами в умовах цифрової економіки. *Економіка та суспільство*. 2024. Вип. 62. DOI: https://doi.org/10.32782/2524-0072/2024-62-119
28. Остапюк Н. А., Нікіфоренко В. В. Використання сучасних аналітичних інструментів та Big Data в оцінці вартості підприємств. *Економіка та суспільство*. 2025. Вип. 75. DOI: https://doi.org/10.32782/2524-0072/2025-75-42
29. Batselier J., Vanhoucke M. Improving project forecast accuracy by integrating earned value management with exponential smoothing and reference class forecasting. *International Journal of Project Management*. 2017. Vol. 35. Iss. 1. P. 28–43. DOI: https://doi.org/10.1016/j.ijproman.2016.10.003
30. Chawla N. V., Bowyer K. W., Hall L. O., Kegelmeyer W. P. SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*. 2002. Vol. 16. P. 321–357. DOI: https://doi.org/10.1613/jair.953

REFERENCES

Ahiaga-Dagbui D. D. & Smith S. D. (2014). Dealing with construction cost overruns using data mining. *Construction Management and Economics, 32*(7–8), 682–694. https://doi.org/10.1080/01446193.2014.933854

Al-Mnaseer R., Al-Smadi S. & Al-Bdour H. (2023). Machine learning-aided time and cost overrun prediction in construction projects: Application of artificial neural network. *Asian Journal of Civil Engineering, 24*, 2583–2593. https://doi.org/10.1007/s42107-023-00665-7

Alazawy S. F. M., Ahmed M. A. & Raheem S. H. (2025). Explainable machine learning to predict the construction cost of power plant based on random forest and Shapley method. *CivilEng, 6*(2). https://doi.org/10.3390/civileng6020021

Asiedu R. O., Frempong N. K. & Alfen H. W. (2017). Predicting likelihood of cost overrun in educational projects. *Engineering, Construction and Architectural Management, 24*(1), 21–39. https://doi.org/10.1108/ECAM-06-2015-0103

Attalla M. & Hegazy T. (2003). Predicting cost deviation in reconstruction projects: Artificial neural networks versus regression. *Journal of Construction Engineering and Management, 129*(4), 405–411. https://doi.org/10.1061/(ASCE)0733-9364(2003)129:4(405)

Batselier J. & Vanhoucke M. (2016). Practical application and empirical evaluation of reference class forecasting for project management. *Project Management Journal, 47*(5), 36–51. https://doi.org/10.1177/875697281604700504

Batselier J. & Vanhoucke M. (2017). Improving project forecast accuracy by integrating earned value management with exponential smoothing and reference class forecasting. *International Journal of Project Management, 35*(1), 28–43. https://doi.org/10.1016/j.ijproman.2016.10.003

Breiman L. (2001). Random forests. *Machine Learning, 45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Chawla N. V., Bowyer K. W., Hall L. O. & Kegelmeyer W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research, 16*, 321–357. https://doi.org/10.1613/jair.953

Chen T. & Guestrin C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (p. 785–794). New York: ACM. https://doi.org/10.1145/2939672.2939785

Chen H. L., Chen W. T. & Lin Y. L. (2016). Earned value project management: Improving the predictive power of planned value. *International Journal of Project Management, 34*(1), 22–29. https://doi.org/10.1016/j.ijproman.2015.09.008

Cheng M.-Y., Peng H.-S., Wu Y.-W. & Chen T.-L. (2010). Estimate at completion for construction projects using evolutionary support vector machine inference model. *Automation in Construction, 19*(5), 619–629. https://doi.org/10.1016/j.autcon.2010.02.008

Coffie G. H. & Cudjoe S. K. F. (2024). Using extreme gradient boosting (XGBoost) machine learning to predict construction cost overruns. *International Journal of Construction Management, 24*(16), 1742–1750. https://doi.org/10.1080/15623599.2023.2289754

Colin J. & Vanhoucke M. (2014). Setting tolerance limits for statistical project control using earned value management. *Omega, 49*, 107–122. https://doi.org/10.1016/j.omega.2014.06.001

Flyvbjerg B., Holm M. K. S. & Buhl S. L. (2002). Underestimating costs in public works projects: Error or lie? *Journal of the American Planning Association, 68*(3), 279–295. https://doi.org/10.1080/01944360208976273

Flyvbjerg B., Holm M. K. S. & Buhl S. L. (2004). What causes cost overrun in transport infrastructure projects? *Transport Reviews, 24*(1), 3–18. https://doi.org/10.1080/0144164032000080494a

Flyvbjerg B. (2006). From Nobel Prize to project management: Getting risks right. *Project Management Journal, 37*(3), 5–15. https://doi.org/10.1177/875697280603700302

Flyvbjerg B. (2013). Quality control and due diligence in project management: Getting decisions right by taking the outside view. *International Journal of Project Management, 31*(5), 760–774. https://doi.org/10.1016/j.ijproman.2012.10.007

Flyvbjerg B. (2014). What you should know about megaprojects and why: An overview. *Project Management Journal, 45*(2), 6–19. https://doi.org/10.1002/pmj.21409

Hamdan M. M., Thneibat M. & Hyari K. (2026). Predicting cost overrun in construction projects using machine learning algorithms: The case of Jordan. *Engineering, Construction and Architectural Management, 33*(6), 4732–4763. https://doi.org/10.1108/ECAM-09-2024-1209

Kim G.-H., An S.-H. & Kang K.-I. (2004). Comparison of construction cost estimating models based on regression analysis, neural networks, and case-based reasoning. *Building and Environment, 39*(10), 1235–1242. https://doi.org/10.1016/j.buildenv.2004.02.013

Love P. E. D., Ahiaga-Dagbui D. D. & Irani Z. (2016). Cost overruns in transportation infrastructure projects: Sowing the seeds for a probabilistic theory of causation. *Transportation Research Part A: Policy and Practice, 92*, 184–194. https://doi.org/10.1016/j.tra.2016.08.007

Lundberg S. M. & Lee S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, *30*. https://proceedings.neurips.cc/paper_files/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf

Narbaev T., Hazır Ö., Khamitova B. & Talgat S. (2024). A machine learning study to improve the reliability of project cost estimates. *International Journal of Production Research, 62*(12), 4372–4388. https://doi.org/10.1080/00207543.2023.2262051

Nenni M. E., De Felice F., De Luca C. & Forcina A. (2025). How artificial intelligence will transform project management in the age of digitisation: A systematic literature review. *Management Review Quarterly, 75*, 1669–1716. https://doi.org/10.1007/s11301-024-00418-z

Ostapiuk N. A. & Nikiforenko V. V. (2025). Vykorystannia suchasnykh analitychnykh instrumentiv ta Big Data v otsintsi vartosti pidpryiemstv [Use of modern analytical tools and Big Data in enterprise valuation]. *Ekonomika ta suspilstvo, 75*. https://doi.org/10.32782/2524-0072/2025-75-42

Panas Ya. V. (2024). Informatsiine zabezpechennia upravlinnia innovatsiinymy proektamy v umovakh tsyfrovoi ekonomiky [Information support for managing innovative projects in the conditions of the digital economy]. *Ekonomika ta suspilstvo, 62*. https://doi.org/10.32782/2524-0072/2024-62-119

Plebankiewicz E. (2018). Model of predicting cost overrun in construction projects. *Sustainability, 10*(12). https://doi.org/10.3390/su10124387

Plebankiewicz E. & Wieczorek D. (2020). Prediction of cost overrun risk in construction projects. *Sustainability, 12*(22). https://doi.org/10.3390/su12229341

Turkyilmaz A. H. & Polat G. (2024). Risk-based completion cost overrun ratio estimation in construction projects using machine learning classification algorithms: A case study. *Buildings, 14*(11). https://doi.org/10.3390/buildings14113541

Дослідження не отримувало зовнішнього фінансування.
Автори заявляють про відсутність конфлікту інтересів.
Інструменти ШІ в підготовці цієї роботи не використовувалися.

Стаття надійшла до редакції / Received: 28.08.2026
Статтю прийнято до публікації / Accepted: 10.09.2026
Оприлюднено / Published: 18.10.2026

УДК 330.322.011:005.336.4
JEL: 330.322.011:005.336.4
DOI: https://doi.org/10.32983/2222-4459-2026-9-210-220

МЕЖА ЗАСТОСОВНОСТІ КРИТЕРІЮ ОЧІКУВАНОЇ ВІДДАЧІ В ІНВЕСТУВАННІ В ЛЮДСЬКИЙ КАПІТАЛ БІЗНЕС-СТРУКТУР

©2026 **ПЕТРУШКА Т. О., ГУРСЬКИЙ С. І.**

УДК 330.322.011:005.336.4
JEL: 330.322.011:005.336.4

Петрушка Т. О., Гурський С. І. Межа застосовності критерію очікуваної віддачі в інвестуванні в людський капітал бізнес-структур

Дослідження спрямоване на закриття прогалини, що залишається на перетині двох теоретичних ліній, які в економічній науці формувалися відносно незалежно одна від одної: теорії людського капіталу, у межах якої інвестиційне рішення оцінюється за критерієм очікуваної віддачі, та теорії інвестування за умов невизначеності, у якій розмежовано вимірювану та невимірювану форми незнання майбутнього. У класичній теорії людського капіталу, започаткованій працями Т. Шульца, Г. Беккера та Дж. Мінсера, а також у вітчизняних методиках оцінювання інвестицій у розвиток персоналу імпліцитно приймається, що ймовірнісна структура майбутніх станів середовища є заданою або принаймні надійно оцінюваною за історичними даними. Зазначене припущення втрачає чинність для бізнес-структур, які ухвалюють інвестиційні рішення щодо людського капіталу в середовищі, здатному до стрибкоподібних, непередбачуваних зрушень. Метою статті є обґрунтування на основі розмежування ризику та невизначеності, за Ф. Найтом, причин втрати визначеності критерієм максимізації очікуваної віддачі за умов необчислюваної невизначеності, а також встановлення того, який критерій ефективності інвестиційного забезпечення розвитку людського капіталу є коректним у такому середовищі. Методичну основу дослідження становлять концептуально-порівняльний аналіз двох теоретичних ліній, критичний аналіз вітчизняного та закордонного наукового доробку, а також ілюстративний кількісний фрагмент, побудований на панельних даних 41 машинобудівного підприємства України за 2021–2025 рр., що охоплюють 205 спостережень підприємств протягом року та стратифіковані за локусом ухвалення інвестиційного рішення. Показано, що подія лютого 2022 р. належить не до ризикової, а до найтівської категорії, що випливає з двох самостійних підстав: неповторюваності самої події та відсутності бази для оцінювання ймовірностей подальших переходів середовища. Обґрунтовано, що стандартний стохастичний аналіз чутливості, попри зовнішню подібність, залишається в парадигмі ризику, оскільки передбачає варіювання параметрів у межах заданого розподілу, тоді як аналіз невизначеності потребує варіювання самого розподілу. Сформульовано положення про те, що за умов необчислювальної невизначеності критерій очікуваної віддачі не набуває визначеного значення, оскільки розподіл, за яким очікуване значення могло би бути коректно обчислене, не існує. За результатами ілюстративного емпіричного фрагмента встановлено, що зв'язок між динамікою чисельності персоналу та динамікою витрат на одного працівника змінив знак після лютого 2022 р. неоднаково для підприємств із різним локусом ухвалення рішення, що узгоджується зі структурним, а не суто рівневим характером зрушення. Наукову новизну становить критерій, за яким інвестиційне забезпечення розвитку людського капіталу оцінюється за здатністю траєкторії результату витримувати кожен із правдоподібних станів середовища, а не за приростом усередненого за цими станами результату. Новизну становить також поширення найтівського розмежування ризику та невизначеності на клас інвестиційних рішень щодо персоналу бізнес-структур. Практична цінність одержаних результатів полягає в наданні менеджменту методологічної підстави для відмови від оцінювання інвестиційних альтернатив за єдиним найбільш імовірним сценарієм на користь перевірки їх стійкості до множини правдоподібних сценаріїв. Перспективу подальших досліджень становить операціоналізація критерію стійкості через множину правдоподібних матриць імовірностей переходу між режимами середовища.

Ключові слова: *невизначеність за Найтом; людський капітал; критерій стійкості траєкторії; ризик; режимне зрушення; бізнес-структура; інвестиційне рішення.*

Рис.: *2.* ***Табл.:*** *1.* ***Формул:*** *2.* ***Бібл.:*** *25.*

Петрушка Тетяна Олексіївна *– кандидат економічних наук, доцент, доцент кафедри економіки підприємства та інвестицій, Національний університет «Львівська політехніка» (вул. Степана Бандери, 12, Львів, 79013, Україна)*
E-mail: *Tetiana.O.Petrushka@lpnu.ua*
ORCID: *https://orcid.org/0000-0002-2005-5573*