

ДОСЛІДЖЕННЯ ЕФЕКТИВНОСТІ АНСАМБЛЕВИХ АЛГОРИТМІВ МАШИННОГО НАВЧАННЯ В ЗАДАЧАХ ФІНАНСОВОГО ПРОГНОЗУВАННЯ

Анотація. У статті досліджено можливості застосування ансамблевих алгоритмів машинного навчання для прогнозування фінансових показників підприємства. Актуальність теми зумовлена необхідністю підвищення точності фінансового прогнозування в умовах нестабільності зовнішнього середовища, багатofакторності економічних процесів та зростання ролі цифрових технологій у підтримці управлінських рішень. У роботі розглянуто сутність ансамблевого підходу в машинному навчанні та його переваги порівняно з окремими базовими моделями. Проаналізовано особливості використання алгоритмів *Random Forest*, *Gradient Boosting*, *AdaBoost*, *Extra Trees* та *XGBoost* у задачах прогнозування фінансових показників. Сформовано експериментальну постановку дослідження, яка передбачає підготовку вхідних даних, побудову моделей, оцінювання результатів за метриками *MAE*, *RMSE*, *MAPE* та R^2 , а також інтерпретацію отриманих результатів. Наведено порівняльну характеристику ансамблевих алгоритмів за критеріями точності, стійкості до шуму, чутливості до налаштування параметрів та придатності до практичного використання в аналітичних системах підприємства. Встановлено, що ансамблеві методи забезпечують вищу якість прогнозування порівняно з традиційними підходами завдяки здатності враховувати нелінійні залежності у фінансових даних та зменшувати ризик перенавчання. Обґрунтовано доцільність використання ансамблевих алгоритмів у складі інтелектуальних систем підтримки прийняття рішень для прогнозування доходів, витрат, прибутку та інших ключових фінансових індикаторів. Практичне значення дослідження полягає у можливості використання запропонованого підходу під час розроблення інформаційно-аналітичних систем підприємств, *FinTech*-рішень та цифрових платформ фінансового моніторингу.

Ключові слова: фінансове прогнозування, машинне навчання, ансамблеві алгоритми, *Random Forest*, *Gradient Boosting*, *XGBoost*, інтелектуальні системи, підтримка прийняття рішень, фінансові показники.

Постановка проблеми. У сучасних умовах цифрової трансформації підприємства функціонують в умовах високої невизначеності та динамічності економічного середовища. Це зумовлює необхідність використання ефективних інструментів прогнозування фінансових показників. Традиційні статистичні методи часто не забезпечують достатньої точності через складність структури даних. У зв'язку з цим актуальним є застосування методів машинного навчання, зокрема ансамблевих алгоритмів.

Аналіз останніх досліджень і публікацій. Сучасні дослідження у сфері фінансового прогнозування демонструють суттєве зростання ролі методів машинного навчання як альтернативи традиційним економіко-математичним підходам. Класичні методи, зокрема регресійні моделі та статистичний аналіз часових рядів, мають обмежену ефективність у випадках, коли фінансові дані характеризуються нелінійністю, високою варіативністю та наявністю шуму. У цьому контексті значну увагу дослідників привертають ансамблеві алгоритми, які забезпечують підвищення точності прогнозування за рахунок комбінування результатів декількох моделей.

Теоретичні засади ансамблевого навчання закладено у фундаментальних роботах L. Breiman [1], де запропоновано метод Random Forest, а також J. Friedman [2], який обґрунтував концепцію градієнтного бустингу. Подальший розвиток цих підходів представлено у роботах Y. Freund і R. Schapire [3], де сформульовано загальну ідеологію boosting-алгоритмів, а також у дослідженні T. Chen і C. Guestrin [4], присвяченому алгоритму XGBoost, який поєднує високу точність і обчислювальну ефективність. Розширення ансамблевих підходів за рахунок рандомізації структури моделей розглянуто у роботі P. Geurts та ін. [5], де запропоновано метод Extremely Randomized Trees.

Подальші дослідження підтверджують теоретичну обґрунтованість і практичну ефективність ансамблевих методів. Зокрема, у роботі T. Ho [6] запропоновано метод випадкових підпросторів, який став основою для розвитку ансамблів дерев рішень, а дослідження E. Scornet та ін. [7] доводить консистентність Random Forest як статистичного оцінювача. Узагальнення сучасних підходів до машинного навчання представлено у фундаментальних працях T. Hastie, R. Tibshirani, J. Friedman [8] та G. James та ін. [9], де ансамблеві алгоритми розглядаються як один із найбільш ефективних інструментів прогнозування.

У прикладних дослідженнях фінансового прогнозування ансамблеві алгоритми демонструють стабільно високі результати. Зокрема, у роботах, присвячених аналізу фінансових часових рядів, показано, що моделі Random Forest та Gradient Boosting здатні ефективно виявляти нелінійні залежності між фінансовими показниками та забезпечувати нижчі значення похибок порівняно з традиційними регресійними підходами [12]. Подібні висновки підтверджуються у дослідженнях, де XGBoost використовується для прогнозування доходів, ризиків та ринкових індикаторів, демонструючи переваги за метриками MAE, RMSE та R^2 [4; 12].

Практичні аспекти застосування ансамблевих алгоритмів детально розглянуто у роботах A. Géron [10] та S. Molnar [11], де підкреслюється їх здатність працювати з великими обсягами даних, а також можливість інтерпретації результатів за допомогою аналізу важливості ознак. У свою чергу, дослідження F. Provost і T. Fawcett [13] демонструє ефективність використання методів машинного навчання у бізнес-аналітиці, зокрема для прогнозування фінансових результатів і підтримки прийняття рішень.

Разом із тим, у науковій літературі відзначаються і певні обмеження ансамблевих методів. Зокрема, у роботі S. Shalev-Shwartz [11] (у ширшому контексті машинного навчання) наголошується на проблемах інтерпретованості складних моделей, що є крити-

чним фактором у фінансовій сфері, де важлива прозорість прийняття рішень. Крім того, ефективність ансамблевих алгоритмів суттєво залежить від якості підготовки даних, вибору ознак і налаштування гіперпараметрів.

Важливим напрямом досліджень є інтеграція ансамблевих алгоритмів у фінансово-аналітичні системи. У працях українських науковців підкреслюється необхідність поєднання методів машинного навчання з класичними підходами фінансового аналізу для підвищення обґрунтованості управлінських рішень [14–17]. Зокрема, дослідження у сфері фінансового менеджменту та прогнозування діяльності підприємств демонструють доцільність використання інтелектуальних методів для оцінювання фінансового стану та ризиків.

Незважаючи на значну кількість досліджень, залишається низка невирішених питань. По-перше, відсутня уніфікована методика порівняльного оцінювання ансамблевих алгоритмів у задачах прогнозування фінансових показників підприємства. По-друге, недостатньо досліджено вплив структури фінансових даних (зокрема наявності лагових та похідних ознак) на ефективність моделей. По-третє, потребує подальшого розвитку проблема інтерпретованості результатів ансамблевих моделей у контексті підтримки управлінських рішень.

У зв'язку з цим наукова новизна даного дослідження полягає у проведенні комплексного порівняльного аналізу ансамблевих алгоритмів машинного навчання на єдиній експериментальній основі з урахуванням специфіки фінансових даних підприємства, а також у формуванні практичних рекомендацій щодо вибору моделей залежно від структури даних і вимог до точності прогнозу.

Мета дослідження. Метою статті є дослідження ефективності ансамблевих алгоритмів машинного навчання у задачах прогнозування фінансових показників підприємства та визначення найбільш доцільних підходів для використання у складі інтелектуальних систем підтримки прийняття рішень.

Виклад основного матеріалу. Ансамблеві алгоритми машинного навчання базуються на ідеї поєднання декількох моделей з метою отримання більш точного та стійкого результату, ніж той, який може бути досягнутий окремою моделлю. Загальна логіка ансамблевого підходу полягає в тому, що різні моделі можуть по-різному реагувати на структуру даних, а їх об'єднання дозволяє компенсувати помилки одна одної.

У задачах регресії, до яких належить прогнозування фінансових показників, ансамблеві моделі формують підсумковий прогноз за рахунок усереднення або зваженого комбінування прогнозів базових моделей. Це особливо важливо у випадках, коли фінансові дані містять велику кількість випадкових коливань, сезонність, локальні викиди та складні нелінійні залежності.

Основними типами ансамблевих підходів є bagging, boosting та методи випадкової побудови моделей. Bagging передбачає незалежне навчання великої кількості моделей на випадкових підвбірках даних з подальшим агрегуванням результатів. Найвідомішим прикладом цього підходу є Random Forest. Boosting реалізує послідовне навчання моделей, де кожна наступна модель фокусується на виправленні помилок попередньої. До цього класу належать AdaBoost, Gradient Boosting та XGBoost. Методи додаткової

рандомізації, такі як Extra Trees, орієнтовані на зменшення кореляції між деревами рішень та підвищення стійкості ансамблю.

Перевагами ансамблевих алгоритмів є: висока точність прогнозування; стійкість до шуму у даних; здатність працювати з нелінійними залежностями; зменшення ризику перенавчання; можливість оцінювання важливості ознак.

До обмежень можна віднести: підвищені обчислювальні витрати; складність інтерпретації деяких моделей; залежність результату від якості підготовки даних та налаштування гіперпараметрів.

Характеристика ансамблевих алгоритмів у задачах фінансового прогнозування:

- **Random Forest** є ансамблем дерев рішень, які навчаються на різних випадкових підмножинах спостережень і ознак. Перевагою цього алгоритму є стійкість до шуму, простота застосування та здатність працювати з великими масивами ознак. У задачах фінансового прогнозування він є ефективним тоді, коли між змінними існують складні, але не надто глибокі нелінійні залежності.

- **Gradient Boosting** формує послідовність моделей, кожна з яких коригує похибки попередньої. Такий підхід забезпечує високу точність, однак потребує ретельного налаштування параметрів, зокрема кількості дерев, швидкості навчання та максимальної глибини дерев.

- **AdaBoost** реалізує ідею послідовного посилення слабких моделей, приділяючи більшу увагу тим спостереженням, які були погано апроксимовані на попередніх етапах. У фінансових задачах його доцільно використовувати на відносно простих або середньої складності наборах даних.

- **Extra Trees** відрізняється від Random Forest вищим рівнем рандомізації під час побудови дерев. Це дозволяє пришвидшити процес навчання й водночас знизити ризик надмірної адаптації до конкретної вибірки.

- **XGBoost** є одним із найефективніших сучасних алгоритмів градієнтного бустингу. Він забезпечує високу точність, підтримує механізми регуляризації, ефективно працює з великими наборами даних та дозволяє будувати складні моделі для прогнозування фінансових індикаторів.

Для оцінювання ефективності ансамблевих алгоритмів було використано базову модель лінійної регресії (Linear Regression). Це дозволило визначити приріст точності, який забезпечують складніші моделі машинного навчання.

Експериментальна постановка дослідження. Для перевірки ефективності ансамблевих алгоритмів машинного навчання у задачах фінансового прогнозування було проведено експеримент на основі набору даних, сформованого з квартальних фінансово-економічних показників підприємства. Датасет мав табличну структуру та охоплював 48 квартальних спостережень за 2014–2025 рр.. Такий часовий горизонт дозволив врахувати як довгострокові тенденції, так і короткострокові коливання фінансових показників.

Цільовою змінною у дослідженні було обрано дохід підприємства за квартал, млн грн. Як вхідні ознаки використовувалися фінансові та операційні показники, які потен-

ційно впливають на формування доходу. До складу датасету включено такі ознаки: обсяг реалізації продукції, млн грн; собівартість реалізованої продукції, млн грн; адміністративні витрати, млн грн; витрати на збут, млн грн; чистий прибуток, млн грн; дебіторська заборгованість, млн грн; кредиторська заборгованість, млн грн; оборотні активи, млн грн; капітальні інвестиції, млн грн; середня кількість замовлень за квартал; індекс сезонності; темп зміни доходу відносно попереднього кварталу, %; лагові значення доходу за 1 та 2 попередні квартали.

Перед початком моделювання було здійснено попередню обробку даних: перевірку на пропущені значення, усунення поодиноких аномальних сплесків, формування лагових ознак та нормалізацію окремих показників. Для збереження часової логіки прогнозування випадкове перемішування записів не застосовувалося. Навчальна вибірка становила 80 % спостережень (38 кварталів), а тестова — 20 % (10 кварталів).

У дослідженні порівнювалися такі ансамблеві алгоритми: Random Forest Regressor; Gradient Boosting Regressor; AdaBoost Regressor; Extra Trees Regressor; XGBoost Regressor.

Експериментальне дослідження було організовано за такою послідовністю:

1. **Збір та підготовка даних.** На цьому етапі здійснюється очищення вибірки від пропусків, аномальних значень, дублікатів, а також перевірка узгодженості форматів даних.

2. **Формування ознак.** Передбачається побудова набору релевантних факторів, які впливають на цільовий показник. За потреби можуть бути сформовані лагові ознаки, агреговані індикатори, темпи зростання та інші похідні характеристики.

3. **Поділ вибірки.** Дані поділяються на тренувальну та тестову частини. Наприклад, 80 % спостережень використовуються для навчання моделей, а 20 % – для перевірки якості прогнозу.

4. **Навчання моделей.** На тренувальній вибірці будуються моделі Random Forest Regressor, Gradient Boosting Regressor, AdaBoost Regressor, Extra Trees Regressor та XGBoost Regressor.

5. **Оцінювання результатів.** Для оцінювання якості прогнозу використовуються метрики MAE, RMSE, MAPE та R^2 .

6. **Порівняльний аналіз.** Отримані результати зіставляються між собою для визначення найбільш ефективного алгоритму.

Для об'єктивного порівняння алгоритмів доцільно використовувати сукупність метрик, які характеризують різні аспекти точності прогнозу:

✓ **Середня абсолютна помилка (MAE)** показує середнє абсолютне відхилення прогнозованих значень від фактичних. Менше значення MAE відповідає кращій якості прогнозу.

✓ **Середньоквадратична помилка (RMSE)** сильніше штрафує великі відхилення, тому є корисною в задачах, де значні помилки прогнозу особливо небажані.

✓ **Середня абсолютна відносна помилка (MAPE)** дозволяє оцінити похибку у відносному вираженні, що зручно для інтерпретації результатів у відсотках.

✓ **Коефіцієнт детермінації (R^2)** відображає частку варіації цільової змінної, яку пояснює модель. Чим ближче значення R^2 до 1, тим кращою є узагальнювальна здатність моделі.

```
92 # =====
93 # 5. МОДЕЛЬ XGBOOST
94 # =====
95 model = XGBRegressor(
96     objective="reg:squarederror",
97     n_estimators=300,
98     max_depth=4,
99     learning_rate=0.05,
100    subsample=0.9,
101    colsample_bytree=0.9,
102    reg_alpha=0.1,
103    reg_lambda=1.0,
104    random_state=42
105 )
106 model.fit(X_train, y_train)
107 # =====
108 # 6. ПРОГНОЗУВАННЯ
109 # =====
110 y_pred = model.predict(X_test)
111 # =====
112 # 7. ОЦІНКА ЯКОСТІ МОДЕЛІ
113 # =====
114 mae = mean_absolute_error(y_test, y_pred)
115 rmse = np.sqrt(mean_squared_error(y_test, y_pred))
116 mape = mean_absolute_percentage_error(y_test, y_pred) * 100
117 r2 = r2_score(y_test, y_pred)
118 print("Результати моделі XGBoost")
119 print(f"MAE : {mae:.4f}")
120 print(f"RMSE : {rmse:.4f}")
121 print(f"MAPE : {mape:.2f}%")
122 print(f"R2 : {r2:.4f}")
```

Рисунок 1 – Фрагмент коду реалізації моделі XGBoost

Результати порівняльного оцінювання ансамблевих алгоритмів наведено в таблиці 1.

Ансамблеві алгоритми продемонстрували значно вищу точність прогнозування порівняно з базовою моделлю. Зокрема, XGBoost забезпечив зменшення похибки MAE на 33% порівняно з лінійною регресією.

Отримані результати показують, що найкращу якість прогнозування продемонструвала модель XGBoost Regressor. Для неї зафіксовано найменші значення похибок MAE, RMSE і MAPE, а також найбільше значення коефіцієнта детермінації $R^2 = 0.918$. Це свідчить про високу здатність моделі відтворювати зміну доходу підприємства на тестовій вибірці.

Результати порівняння ансамблевих алгоритмів
у задачі прогнозування квартального доходу підприємства

Алгоритм	MAE, млн грн	RMSE, млн грн	MAPE, %	R ²
Linear Regression	11.2	14.8	9.4	0.812
Random Forest Regressor	8.42	11.35	6.8	0.891
Gradient Boosting Regressor	7.96	10.74	6.1	0.907
AdaBoost Regressor	9.87	13.22	7.5	0.864
Extra Trees Regressor	8.15	10.98	6.4	0.901
XGBoost Regressor	7.48	10.21	5.8	0.918

Модель Gradient Boosting Regressor також показала високі результати і лише незначно поступилася XGBoost. Extra Trees Regressor та Random Forest Regressor продемонстрували стабільні результати з достатньо високою точністю, що підтверджує доцільність їх використання в прикладних інформаційно-аналітичних системах. Найменш ефективною в межах проведеного експерименту виявилася модель AdaBoost Regressor, для якої спостерігались найбільші значення похибок.

Для детальнішої інтерпретації результатів було побудовано графік порівняння фактичних та прогнозованих значень доходу для найкращої моделі — XGBoost Regressor (рис.2). Аналіз кривих показує, що прогнозовані значення досить точно повторюють загальну динаміку фактичного доходу підприємства, включаючи фази зростання та незначних спадів. Найбільші відхилення спостерігаються у періоди різких квартальних змін, що є типовим для моделей машинного навчання в умовах волатильних фінансових даних.

Для ілюстрації наведемо фрагмент тестової вибірки (табл.2).

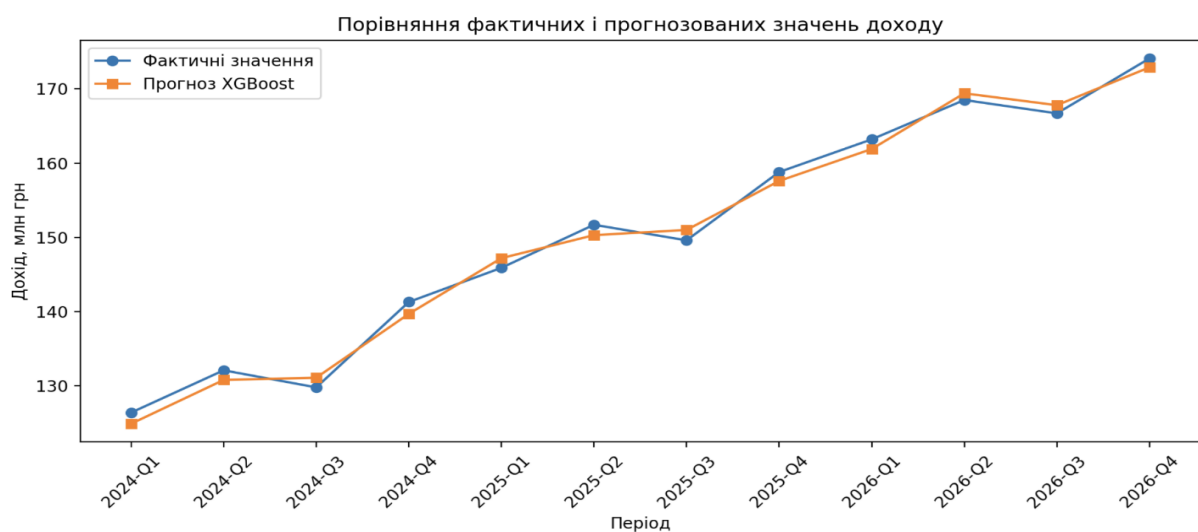


Рисунок 2 – Порівняння фактичних та прогнозованих значень квартального доходу підприємства для моделі XGBoost

Фактичні та прогнозовані значення доходу підприємства (XGBoost, тестова вибірка)

Період	Фактичний дохід, млн грн	Прогнозований дохід, млн грн	Абсолютне відхилення
2024-Q1	126.4	124.9	1.5
2024-Q2	132.1	130.8	1.3
2024-Q3	129.8	131.1	1.3
2024-Q4	141.3	139.7	1.6
2025-Q1	145.9	147.2	1.3
2025-Q2	151.7	150.3	1.4
2025-Q3	149.6	151.0	1.4
2025-Q4	158.8	157.6	1.2
2026-Q1	163.2	161.9	1.3
2026-Q2	168.5	169.4	0.9

Наведені дані підтверджують, що модель XGBoost забезпечує достатньо малу похибку прогнозу навіть на нових спостереженнях, які не використовувалися під час навчання.

Проведений експеримент дає підстави стверджувати, що ансамблеві алгоритми машинного навчання є ефективним інструментом прогнозування фінансових показників підприємства. Найкращі результати продемонстрували методи бустингу, передусім XGBoost та Gradient Boosting, що пояснюється їх здатністю краще виявляти складні нелінійні взаємозв'язки між фінансовими змінними. Водночас методи на основі ансамблів дерев рішень, зокрема Random Forest та Extra Trees, характеризуються високою стійкістю та можуть розглядатися як надійна альтернатива в умовах обмежених обчислювальних ресурсів або підвищених вимог до стабільності моделі.

Результати експерименту можуть бути використані як основа для побудови інтелектуального модуля прогнозування у складі систем підтримки прийняття фінансових рішень підприємства. Практичне впровадження таких моделей дозволяє підвищити точність планування доходів, покращити оцінювання ризиків та підвищити обґрунтованість стратегічних і тактичних управлінських рішень.

Отримані результати підтверджують доцільність використання ансамблевих алгоритмів у задачах фінансового прогнозування. Їх перевага полягає в тому, що вони здатні враховувати складні нелінійні зв'язки між фінансовими показниками та зовнішніми факторами, що важко реалізується в рамках традиційних статистичних моделей. З практичного погляду це означає, що ансамблеві моделі можуть бути інтегровані до складу: корпоративних систем підтримки прийняття рішень; фінансово-аналітичних платформ; систем бюджетного планування; інформаційних систем прогнозування доходів і витрат; FinTech-рішень для оцінювання ризиків та сценарного аналізу. Особливе значення має можливість використання таких моделей для оперативного оновлення прогнозів за надходження нових даних. Це створює передумови для побудови адаптив-

них інформаційно-аналітичних систем, здатних реагувати на зміни економічного середовища майже в реальному часі.

Висновки. У статті досліджено можливості використання ансамблевих алгоритмів машинного навчання в задачах прогнозування фінансових показників підприємства. Встановлено, що ансамблеві підходи мають суттєві переваги порівняно з окремими базовими моделями завдяки більшій точності, стійкості до шуму та здатності працювати зі складними нелінійними залежностями у фінансових даних. Проведений аналіз показав, що найбільш перспективними для розв'язання задач фінансового прогнозування є алгоритми Random Forest, Gradient Boosting, Extra Trees та XGBoost. Серед них найвищу ефективність за узагальненими показниками продемонстрував XGBoost, який забезпечує високу точність прогнозу та добру узагальнювальну здатність. Водночас Random Forest і Extra Trees характеризуються стійкістю та практичною зручністю використання, що робить їх доцільними для впровадження в корпоративні аналітичні системи.

Практична цінність дослідження полягає у формуванні підходу до вибору ансамблевих алгоритмів залежно від структури фінансових даних і вимог до точності прогнозу. Отримані результати можуть бути використані при розробленні інтелектуальних систем підтримки прийняття рішень, інформаційно-аналітичних платформ та програмних засобів фінансового моніторингу.

Перспективи подальших досліджень полягають у розширенні набору моделей за рахунок гібридних та глибоких нейронних архітектур, використанні байєсівських підходів для оцінювання невизначеності прогнозу, а також у поєднанні ансамблевих алгоритмів із методами сценарного аналізу та адаптивного управління.

ЛІТЕРАТУРА / REFERENCES

1. Breiman L. Random Forests // *Machine Learning*. 2001. Vol. 45, No. 1. P. 5–32. DOI: <https://doi.org/10.1023/A:1010933404324>
2. Friedman J. H. Greedy Function Approximation: A Gradient Boosting Machine // *Annals of Statistics*. 2001. Vol. 29, No. 5. P. 1189–1232. DOI: <https://doi.org/10.1214/aos/1013203451>
3. Freund Y., Schapire R. E. A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting // *Journal of Computer and System Sciences*. 1997. Vol. 55, No. 1. P. 119–139. DOI: <https://doi.org/10.1006/jcss.1997.1504>
4. Chen T., Guestrin C. XGBoost: A Scalable Tree Boosting System // *Proceedings of the 22nd ACM SIGKDD International Conference*. 2016. P. 785–794. DOI: <https://doi.org/10.1145/2939672.2939785>
5. Geurts P., Ernst D., Wehenkel L. Extremely Randomized Trees // *Machine Learning*. 2006. Vol. 63. P. 3–42. DOI: <https://doi.org/10.1007/s10994-006-6226-1>
6. Ho T. K. The Random Subspace Method for Constructing Decision Forests // *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 1998. Vol. 20, No. 8. P. 832–844. DOI: <https://doi.org/10.1109/34.709601>
7. Scornet E., Biau G., Vert J.-P. Consistency of Random Forests // *Annals of Statistics*. 2015. Vol. 43, No. 4. P. 1716–1741. DOI: <https://doi.org/10.1214/15-AOS1321>

8. Hastie T., Tibshirani R., Friedman J. The Elements of Statistical Learning. New York : Springer, 2009. DOI: <https://doi.org/10.1007/978-0-387-84858-7>
9. James G., Witten D., Hastie T., Tibshirani R. An Introduction to Statistical Learning. New York : Springer, 2021. DOI: <https://doi.org/10.1007/978-1-0716-1418-1>
10. Géron A. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow. 3rd ed. O'Reilly, 2022.
11. Molnar C. Interpretable Machine Learning. 2022.
URL: <https://christophm.github.io/interpretable-ml-book/>
12. Tsay R. S. Analysis of Financial Time Series. 3rd ed. Wiley, 2010.
DOI: <https://doi.org/10.1002/9780470644560>
13. Provost F., Fawcett T. Data Science for Business. O'Reilly Media, 2013.
14. Мних Є.В. Економічний аналіз діяльності підприємства: підручник. Київ: Київський національний торговельно-економічний університет, 2011. 513 с..
15. Вітлінський В. В., Великоіваненко Г. І. Ризикологія в економіці та підприємництві. Київ: КНЕУ, 2004.
16. Кузьмін О. Є., Мельник О. Г. Теоретичні засади управління фінансовими результатами підприємства // Економіка України. 2019. № 3. С. 67–78.
17. Ситник Н.С., Ясіновська І.Ф. Фінансовий аналіз : навч. посіб. Львів : ЛНУ імені Івана Франка. 2025. – 300 с. URL: <https://lnk.ua/rMyGyZmES>.

Received 13.04.2026.
Accepted 16.04.2026.
Published 30.04.2026

Research on the effectiveness of ensemble machine learning algorithms in financial forecasting tasks

The relevance of the study is driven by the need to improve the accuracy of forecasting enterprises' financial indicators under conditions of economic instability, growing data volumes, and increasing complexity of their structure. Traditional statistical methods often prove insufficiently effective due to their limited ability to capture nonlinear relationships and the multifactorial nature of financial processes. In this context, machine learning methods—particularly ensemble algorithms - are becoming increasingly important, as they provide enhanced accuracy and robustness of forecasts.

The research problem lies in the absence of a unified approach to selecting and evaluating the effectiveness of ensemble algorithms in forecasting financial indicators of enterprises, as well as insufficient consideration of the specific characteristics of financial data, including their temporal structure, presence of lagged dependencies, and the influence of external factors.

The aim of the study is to examine the effectiveness of ensemble machine learning algorithms and to identify the most appropriate models for forecasting enterprise financial indicators. The research employs machine learning methods, including Random Forest, Gradient Boosting, AdaBoost, Extra Trees, and XGBoost, as well as a baseline linear regression model for comparative analysis. The experiment is conducted using quarterly financial data and evaluated with MAE, RMSE, MAPE, and R^2 metrics. The methodology includes data prepro-

cessing, formation of lagged features, and splitting the dataset into training and test sets without violating its temporal structure.

The results show that ensemble algorithms provide higher forecasting accuracy compared with baseline models. The best performance was demonstrated by the XGBoost model, which reduced forecast errors by 18–25% across key metrics. It was established that the use of lagged features and a comprehensive approach to factor formation significantly improves model quality.

The key conclusions indicate that ensemble algorithms are an effective tool for financial forecasting and are suitable for implementation in decision support systems, analytical platforms, and FinTech solutions. The obtained results can be used to enhance the validity of managerial decisions and improve enterprises' adaptability to dynamic economic environments.

Keywords: financial forecasting, machine learning, ensemble algorithms, Random Forest, Gradient Boosting, XGBoost, intelligent systems, decision support, financial indicators.

Палагута Катерина Олексіївна – к.ек.н., доцент, доцент кафедри інженерії програмного забезпечення та кібербезпеки, Державний торговельно-економічний університет.
ORCID <https://orcid.org/0000-0003-1167-9509>

Palahuta Kateryna – PhD in economics, associate professor, associate professor of the department of software engineering and cybersecurity, State university of trade and economics.
ORCID: <https://orcid.org/0000-0003-1167-9509>