19. Належне та ефективне державне регулювання трьох колон пенсійного забезпечення громадян України 2020. URL: https://rada.gov.ua/news/Novyny/187818.html.
20. Нагорна А.М., Ніколаєнко Л.А. Диференціація доходів населення та економічне зростання в Україні. Інфраструктура ринку. 2019. Вип. 35. С. 343-350.
21. Про затвердження бюджету Пенсійного фонду України на 2020 р.: Постанова КМУ від 24 січня 2020 р. №22. URL: https://zakon.rada.gov.ua/laws/show/22-2020- %D0 %BF.
22. Борейко В.І. Вплив дефіциту Пенсійного фонду на економічний розвиток України. Вісник економічної науки України. 2019. № 1. С.8-10.
23. Українська міграція в умовах глобальних і національних викликів XXI століття: наукове видання / наук. ред. У.Я. Садова. Львів, 2019. 110 с.
24. Офіційний сайт Державної служби зайнятості. URL: https://www.dcz.gov.ua/analitics/67.
25. Український ринок праці: імперативи та можливості змін: колективна монографія / за наук. ред. д.е.н., проф. І.Л. Петрової, к.е.н. В.В. Близнюк ; НАН України, ДУ «Ін-т екон. та прогнозув. НАН України». Київ. 2018. 356 с.

Статтю подано до редакції 02.10.2020

УДК: 330.4 DOI 10.33111/mise.99.10

Пирогов В.І.
аспірант кафедри економіко-математичного моделювання
ДВНЗ «КНЕУ імені Вадима Гетьмана»

Pyrohov V.I.
postgraduate of economic-mathematical modeling department
SHEI KNEU named after V. Hetman

ВИКОРИСТАННЯ СТРАТИФІКОВАНОГО СЕМПЛІНГУ КОНТРОЛЬНОЇ ВИБІРКИ ДЛЯ ПОКРАЩЕННЯ ПРЕДИКАТИВНОСТІ МОДЕЛЕЙ БУСТІНГОВИХ ДЕРЕВ РІШЕНЬ

USAGE OF STRATIFIED SAMPLING OF CONTROL SUBSET FOR PREDICATIVITY IMPROVEMENT OF BOOSTED DECISION TREE MODELS

Анотація. *У статті проведено дослідження щодо забезпечення стабільності результату класифікації кредитоспроможності позичальника фізичної особи банку за допомогою алгоритму бустінгових дерев рішень з використанням стратифікованого семплінгу.*
Описано загальний принцип роботи платформи для досліджень у сфері науки про дані Kaggle, в рамках якого фахівці зі статистики та добуван-

ня даних конкурують у створенні найкращих моделей для прогнозування та опису даних, запропонованих компаніями або користувачами.
Проаналізовано моделі та програмну реалізацію алгоритму бустінгових дерев рішень для вирішення задачі оцінки кредитоспроможності позичальника банку. Описано найефективніші програмні пакети, що використовуються для програмної реалізації бустінгових дерев рішень — XGBoost та LGBM
Для підтвердження результатів застосовано інструментарій програмного пакету LGBM на даних банку Home Credit доступних у ході Home Credit Competition на платформі з дослідження даних Kaggle.
Наведено деталі змагання Home Credit Competition: проведено опис наданих даних, підхід до створення характеристик для навчання моделі та програмний підхід що був запропонований у ході участі у змаганні.
У ході дослідження запропоновано використання стратифікованого семплінгу контрольної вибірки за цільовою змінною та найбільш значущими характеристиками в ході навчання моделі задля збільшення стабільності результату класифікації і підвищення ефективності валідації модернізації архітектури моделі.
Експериментальним шляхом доведено, що використання стратифікованого семплінгу контрольної вибірки у ході навчання моделей бустінгових дерев рішень дає можливість збільшити стабільність результату моделі, що підвищує ефективність валідації модернізації архітектури моделі.
Ключові слова: *дерева рішень; градієнтний бустінг; стратифікований семплінг; XGBoost; LGBM; Kaggle.*

Abstract. *In the article has been conducted a research aiming increase of classification result stability of commercial bank's debtor creditworthiness with usage of boosted decision trees algorithm with application of stratified sampling.*
The general principle of the Kaggle data science research platform is described, in which statistics and data mining specialists compete to create the best models for forecasting and data modelling based on the data offered by companies or users.
Has been conducted an analysis of models and program implementation of boosted decision trees algorithm for estimation of commercial bank's debtor creditworthiness. The most effective program packages are described — XGBoost and LGBM, which are used for program implementation of boosted decision trees.
For confirmation of the results, has been used a program package LGBM on data of Home Credit Bank, available in the scope of Home Credit Competition on data science platform Kaggle.
The details of Home Credit Competition are shared: conducted a description of input data, a description of an approach for creation of characteristics for training a model and technical approach which was proposed during participation in the competition.
During the research proposed to use stratified sampling of control dataset by target variable and the most significant characteristics during training of a model to increase a stability of the result of classification and enhance efficiency during a process of modernization of model's architecture.
Proved experimentally, that the use of stratified sampling of the control sample during the training of boosted decision tree models makes possible to increase the stability of the model result, which increases the efficiency of validation of modernization of the model architecture.
Keywords: *decision trees; gradient boosting; stratified sampling; XGBoost, LGBM; Kaggle.*

Постановка проблеми у загальному вигляді та її зв'язок із важливими науковими або практичними завданнями. Із становленням новітнього інформаційного суспільства на рубежі 20-21 століть перед сучасною економічною наукою постали нові виклики та можливості. Оновлена економічна система генерує колосальні потоки інформації, які можна використовувати для отримання додаткового економічного ефекту, отримуючи додану вартість за допомогою правильного тлумачення даних за допомогою сучасних математичних методів.

Відповідно до останніх досліджень [1], тільки за 2016-2017 роки людство згенерувало об'єм інформації більший ніж за попередні 5000 років розвитку людства.

Незважаючи на великий об'єм генерованої інформації для прийняття операційних рішень використовується лише її невеликий відсоток — 0,5 % [1].

Актуальність дослідження. Виникає необхідність у використанні даних для генерування доданої вартості, а значить формується необхідність і у створенні наукового підґрунтя для використання інформації для досягнення економічних цілей. Сьогодні ми стаємо свідками становлення такої науки, що отримала назву Наука про дані (Data Science).

Аналіз останніх досліджень і публікацій. Серед провідних науковців що працюють у сфері науки про дані, можна виділити праці Джефрі Хінтона [11, 12], Ендрю Иня, Яна ЛеКуна [14], Йошуа Бенджіо, Пітера Норвіга [13], Яна Гудфеллоу [14] та ін.

Kaggle — платформа для досліджень у сфері науки про дані. Молода наука потребує як нового інструментарію, так і нової методології, нових підходів до вирішення специфічного кола задач, що постають перед нею.

Одним із таких сучасних інструментів стають інтернет ресурси, що спеціалізуються на вирішенні задач пов'язаних із наукою про дані. На даний момент найбільшим і найпопулярнішим data science інтернет-хабом є ресурс Kaggle [2].

Kaggle — платформа для змагань з аналітики та передбачувального моделювання, в рамках якого фахівці зі статистики та добування даних конкурують у створенні найкращих моделей для прогнозування та опису даних, запропонованих компаніями або користувачами. Цей краудсорсинговий підхід ґрунтується на тому, що є безліч стратегій, які можуть бути застосовані до будь-якого завдання з передбачувального моделювання, і наперед не відомо, яка методика або аналітичний підхід буде найефективнішим [3].

Метою статті є дослідження теоретичного підґрунтя та практичних підходів вирішення проблеми бінарної класифікації на відкритих даних банку HomeCredit у рамках змагання Home Credit Default Risk competition [4].

Виклад основного матеріалу дослідження. У рамках Home Credit Default Risk competition банком HomeCredit було надано дані по кредитним заявкам на 2 роздрібні кредитні продукти: споживчі кредити та кредитні карти. Специфіка формування вхідної вибірки передбачала вибір популяції unbankable клієнтів, які б отримали відмову у отриманні кредиту за одним із наявних кредитних правил, але були прокредитовані банком задля покращення існуючих моделей прийняття рішень і розширення кола потенційних позичальників.

Враховуючи, що Home Credit банком було обрано вибірку клієнтів із низьким кредитним рейтингом для забезпечення достатньої прогностичної сили аналітичних моделей банком було надано додаткові джерела даних, такі як:

— детальна поведінкова інформація по балансу наявних і попередніх кредитів клієнта та його платежам за даними кредитного бюро та внутрішніми даними банку;

— інформація із реєстрів нерухомості про стан і середні значення факторів що характеризують нерухомість, що перебуває у власності клієнта;

— оцінки регіону проживання клієнта;

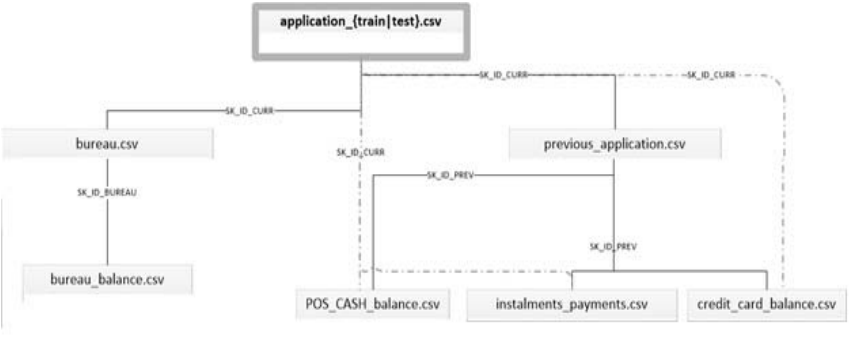

Рис. 1. Структура вхідних даних Home Credit Competition (побудовано на основі публічних даних [5])

Навчальна вибірка, надана банком для побудови прогностичної моделі включала 307 511 спостережень (табл. 1).

Таблиця 1

СТРУКТУРА ВХІДНОЇ НАВЧАЛЬНОЇ ВИБІРКИ HOMECREDIT COMPETITON

Тип кредиту	Кількість спостережень	% спостережень
Кредит виплачено	282 686	91.93%
Дефолт	24 825	8.07%
Grand Total	**307 511**	**100.00%**

Тестова вибірка, надана банком для валідації побудованої моделі включала 48 744 спостережень.

Виклад основного матеріалу дослідження.

Метод бустінгових дерев рішень. Одним із найпопулярніших та найефективніших алгоритмів, що використовуються у ході data science змагань є метод gradient boosted trees (бустингові дерева рішень).

Класичною працею, що заклала теоретичний фундамент для створення бустингових дерев рішень є робота Ж. Фрідмана «Жадібна апроксимація функцій: машина градієнтного бустінгу»[6]

В основу праці Фрідмана закладена ідея, що сама по собі базова предикативна модель є «слабкою», і може бути посиленою завдяки побудові ансамблів моделей, характеристики яких будуть перевизначитися за допомогою оптимізаційних алгоритмів (наприклад алгоритм градієнтного спуску). Результат отриманого ансамблю моделей агрегується і вихідна модель рахується «сильною» за рахунок зменшення дисперсії вихідного результату і оптимізації параметрів. Загальне представлення вихідної моделі матиме вигляд:

$$F(x;\{b_m,a_m\}_1^M) = \sum_{m=1}^{M} b_m h(x;a_m) \qquad (1)$$

де $h(x;a_m)$ – параметрична функція із вхідними змінними x та параметрами $a = \{a_1, a_2 \dots\}$

Розглянемо випадок коли кожна базова модель є деревом рішень. В такому разі кожне дерево рішень має адитивну форму:

$$h\left(x;\{b_j,R_j\}_1^J\right) = \sum_{j=1}^{J} b_j 1(x \in R_j) \qquad (2)$$

У даному виразі $\{R_j\}_1^J$ це простір кінцевих вузлів дерева рішень, що повністю покриває діапазон значень незалежної змінної **x.**

Функція-індикатор 1(*) має значення 1 якщо її аргумент справджується, і 0 у протилежному випадку.

Параметрами даної базової моделі є коефіцієнти $\{b_j\}_1^J$ які визначають границі просторів $\{R_j\}_1^J$., що в свою чергу представляють розподіли некінцевих вузлів дерева.

Для дерева рішень визначення бустінгового алгоритму набуває вигляду:

$$F_m(x) = F_{m-1}(x) + p_m \sum_{j=1}^{J} b_{jm} 1(x \in R_j) \quad (3)$$

де $\{R_{jm}\}_1^J$ — простори визначені кінцевими вузлами дерева рішень у ході ітерації m.

Призначення даних просторів полягає у прогнозуванні псевдовідповідей $\{\bar{y}_i\}_1^N$.

p_m — фактор масштабування для алгоритму лінійного пошуку.

Запис (3) може бути зведений до:

$$F_m(x) = F_{m-1}(x) + \sum_{j=1}^{J} \gamma_{jm} 1(x \in R_j) \quad (4)$$

де $\gamma_{jm} = p_m b_{jm}$

В цілому алгоритм може бути описаний наступним циклом:

$$F_0(x) = median\{y_i\}_1^N$$

For m = 1 to M do:

$$\bar{y}_i = sign(y_i - F_{m-1}(x_i)), i = 1, N$$

$$\{R_{jm}\}_1^J = \text{дерево рішень } \{\bar{y}_i, x_i\}_1^N$$

$$\gamma_{jm} = median_{x_i \in R_{jm}}\{y_i - F_{m-1}(x_i)\}, j = 1, J$$

$$F_m(x) = F_{m-1}(x) + \sum_{j=1}^{J} \gamma_{jm} 1(x \in R_j)$$

end For

end Algorythm [6]

Градієнтне посилення (бустінг) дерев рішень створює конкурентоспроможні, надійні, інтерпретовані моделі для вирішення задач класифікації, причому хороші результати досягаються навіть в умовах низької якості вхідних даних.

Програмні пакети XGBoost та LGBM. XGBoost — це бібліотека програмного забезпечення з відкритим кодом, що є фреймворком з підтримкою алгоритму градієнтного бустінгу для мов програмування C ++, Java, Python, R, і Julia, створена в 2014 році. Бібліотека працює на Linux, Windows, та macOS.

Крім роботи на одному комп'ютері, XGBoost також підтримує розподілені структури обробки даних, такі як Apache Hadoop, Apache Spark і Apache Flink. Вона отримала велику популярність і увагу нещодавно, оскільки цей алгоритм використовувався багатьма командами-переможцями на змаганнях з машинного навчання.

XGBoost було засновано як дослідницький проект у рамках групи Distributed Machine Learning Communities (DMLC) [9]. Спочатку бібліотека представляла собою додаток що настроювався за допомогою конфігураційного файла. Після перемоги в програмі Higgs Machine Learning Challenge бібліотека стала відомою у колах змагань ML. Незабаром до XGBoost було додано пакети для Python і R, зараз існують пакети для багатьох інших мов, таких як Julia, Scala, Java та ін. Можливість використання різних мов програмування розширила коло розробників і принесла XGBoost популярність серед спільноти Kaggle. Робота над XGBoost була опублікована авторами бібліотеки Тіагі Ченом (Tianqi Chen) та Карлосом Гуестріном (Carlos Guestrin) на науковому інтернет-ресурсі arxiv.org та знаходиться у вільному доступі.[7]

LightGBM (LGBM) — фреймворк для градієнтного бустінгу що використовує навчальні алгоритми дерев рішень. LightGBM є сучаснішою оптимізованою програмною реалізацією алгоритму жадібної апроксимації функцій з використанням дерев рішень. Основні сильні сторони LightGBM:

1) більша швидкість та ефективність навчання моделей;

2) вища точність отриманих моделей;

3) більш ефективне використання оперативної пам'яті у ході побудови моделей;

4) підтримка паралельного навчання та навчання за допомогою графічних процесорів [8].

Порівняльний аналіз результатів на відкритих наборах даних показав, що LightGBM перевершує існуючі бустінгові фреймворки як за ефективністю, так і за точністю.

Опис початкових предикативних моделей побудованих у ході змагання Home Credit. Першим етапом у підготовці предикативної моделі була первинна обробка даних і розробка характеристик на основі незалежних змінних моделі.

У ході розробки характеристик використовувались:

1. математико-статистичний підхід — застосування набору математичних функцій для агрегації наявних даних по обслуговуванню кредитів клієнта;

- середнє значення;
- мінімум;
- максимум;
- сума;
- стандартне відхилення;
- кількість унікальних записів;

2. експертний аналіз даних на основі економічного змісту базових незалежних змінних:

- **Income per Person** — дохід позичальника в розрахунку на 1 члена його сім'ї.
- **Children ratio** — відношення кількості дітей у сім'ї позичальника до загальної кількості членів його сім'ї.
- **Credit to Goods ratio —** відношення суми кредиту до вартості товарів що були придбані в кредит.

3. специфічні показники що використовуються у банківському секторі у сфері роздрібного кредитування:

- **DPD** — *days past due* — просрочення платежу за кредитом.
- **DBD** — *days before due* — виконання платежу за кредитом раніше графіку.
- **Loan to Income ratio** — відношення суми кредиту до доходу позичальника.

Загальна кількість характеристик, що була включена до остаточної моделі — 591.

Використання стратифікованого семплінгу для створення контрольної вибірки у ході навчання моделі бустінгових дерев рішень. У ході побудови бустінгових предикативних моделей важливим завданням дослідника є формування контрольної вибірки.

Використання контрольної вибірки у ході підготовки моделі дає можливість запобігти перенавчанню (overfitting) предикативної моделі — запобігти випадків коли модель здатна успішно розпізнавати лише конкретний набір навчальних даних.

Процес формування контрольної вибірки включає відбір спостережень із загального набору таким чином, щоб контрольна

вибірка була якомога наближенішою до навчальної за своїми властивостями.

Для збереження властивостей навчальної вибірки у контрольній використовується метод стратифікованого семплінгу.

Стратифікований семплінг — метод випадкового відбору, що передбачає поділ загальної популяції на менші підгрупи (страти) і проведення випадкового відбору із страт. Страти формуються базуючись на гомогенних характеристиках популяції, що дає можливість відтворити гетерогенність загальної популяції у вибірці.

Класичною працею що описує використання стратифікованого семпінгу у статистиці є робота Дж. Неймана «Про два різні аспекти репрезентативного методу: метод стратифікованого семплінгу і метод цільового вибору» [10].

Базова ідея стратифікованого семплінгу:

1. розбиття гетерогенної вибірки на менші групи, або страти (підпопуляції), такі що групи відбору є:

- гомогенними відносно цільових характеристик страти;
- гетерогенними відносно цільових характеристик між стратами;

2. проведення випадкового відбору спостережень із кожної страти у відповідності із розподілом цільових характеристик страт у початкових даних;

Загальний підхід до проведення стратифікованого відбору наведено на рис. 2.

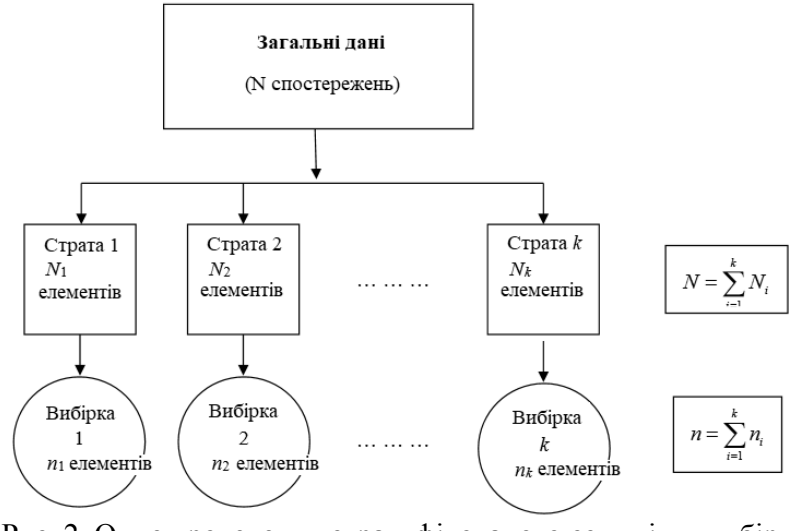

Рис. 2. Опис проведення стратифікованого семплінгу вибірки (побудовано автором)

Результати XGBoost та LGBM із використанням стратифікованого семплінгу вибірки.

У ході дослідження проведено експеримент з використання стратифікованого семплінгу для формування контрольної вибірки для навчання економіко-математичної моделі з використанням моделі бустінгових дерев рішень.

Для проведення експерименту використано 3 методики відбору для контрольної вибірки:

1. випадковий відбір;
2. стратифікований відбір за залежною змінною;
3. стратифікований відбір за залежною змінною та найбільш значущими змінними моделі із групи EXT_SRC;

Для усіх методик використовувалась спільна архітектура моделі та вхідні характеристики.

У результаті проведених 5 досліджень для кожного із типів відбору отримано наступні результати LGBM моделей на контрольній вибірці (табл..і 2-4):

Таблиця 2

ВИПАДКОВИЙ ВІДБІР

№	1	2	3	4	5
AUC	0.7963041	0.7902307	0.7859878	0.7872575	0.7750518

Дисперсія $\partial^2 = 6.02121 * 10^{-5}$

Таблиця 3

СТРАТИФІКОВАНИЙ ВІДБІР ЗА ЗАЛЕЖНОЮ ЗМІННОЮ

№	1	2	3	4	5
AUC	0.7951976	0.7903783	0.7896288	0.7850145	0.7907957

Дисперсія $\partial^2 = 1.314457 * 10^{-5}$

Таблиця 4

СТРАТИФІКОВАНИЙ ВІДБІР ЗА ЗАЛЕЖНОЮ ЗМІННОЮ ТА ЗМІННИМИ EXT_SRC

№	1	2	3	4	5
AUC	0.7934122	0.7911496	0.7871199	0.7895246	0.7876644

Дисперсія $\partial^2 = 6.671408 * 10^{-6}$

Базуючись на отриманих результатах можна зробити висновок що використання стратифікованого відбору за показниками, які

мають найбільший вплив на модель, призводить до зменшення дисперсії результатів ітерацій моделі на контрольній вибірці.

Даний результат дає можливість збільшити стабільність результату моделі, що стає у нагоді у ході валідації ефективності модернізації архітектури моделі.

Наступним етапом експерименту є порівняння предикативної сили моделей побудованих із використанням різних методів відбору валідаційної вибірки. Для оцінки предикативної здатності використовувалась тестова вибірка, яка, за правилами конкурсу, була недоступна для дослідників. Результат оцінки предикативної здатності для тестової вибірки відбувався на стороні системи Kaggle базуючись на оцінці цільової змінної.

Результат побудованих моделей на тестовій вибірці наведено у табл. 5.

Таблиця 5

РЕЗУЛЬТАТИ ПОБУДОВАНИХ МОДЕЛЕЙ НА ТЕСТОВІЙ ВИБІРЦІ

Тип валідаційної множини	Результат (AUC)	
	Валідаційна вибірка	Тестова вибірка
Випадковий відбір	0.78697	0.7907
Стратифікований відбір за залежною змінною	0.7902	0.79156
Стратифікований відбір за залежною змінною та змінними EXT_SRC	0.78977	0.79354

Як видно із отриманих результатів, навіть враховуючи неоднозначний результат отриманий безпосередньо на контрольній вибірці (стратифікований відбір тільки за залежної змінною показав кращий результат, ніж відбір за залежною змінною і найбільш значущими змінними моделі), на тестовій вибірці предикативна сила отриманої моделі прямо залежить від глибини стратифікації при відборі спостережень до контрольної вибірки.

Висновок. Моделі бустінгових дерев рішень впевнено тримають лідерство серед алгоритмів, що використовуються у змаганнях з класифікації даних.

Найпопулярнішими програмними пакетами, які використовуються для створення моделей бустінгових дерев рішень є класичний XGBoost та більш оптимізоване рішення з використанням аналогічного алгоритму — LGBM.

У ході змагань з класифікації даних виникає необхідність забезпечення стабільного результату за інших рівних умов, відпо-

відно виникає необхідність елімінувати флуктуації розподілу характеристик у контрольній вибірці у порівнянні із загальним набором даних.

Запропоновано використання стратифікованого семплінгу за цільовою змінною та найбільш значущими характеристиками моделі для вирішення описаної проблеми.

За результатами проведеного дослідження можна зробити висновок, що:

1. додаткова стратифікація у ході відбору контрольної вибірки позитивно впливає на предикативну силу моделі за рахунок збереження гетерогенності загального набору даних у контрольній вибірці;

2. окрім позитивного впливу на предикативну силу, використання стратифікованого відбору за найбільш значущими показниками моделі, призвело до зменшення дисперсії результатів ітерацій моделі на контрольній вибірці.

Використання стратифікованого семплінгу контрольної вибірки у ході навчання моделей бустінгових дерев рішень дає можливість збільшити стабільність результату моделі, що підвищує ефективність валідації модернізації архітектури моделі.

Бібліографічні посилання

1. Harris R. More data will be created in 2017 than the previous 5,000 years of humanity. *App Developer Magazine*. 2016. URL: https://appdevelopermagazine.com/more-data-will-be-created-in-2017-than-the-previous-5,000-years-of-humanity-/ (дата звернення 01.05.2020)

2. Платформа для змагань з аналітики та передбачувального моделювання Kaggle: веб-сайт. URL: https://www.kaggle.com/ (дата звернення 01.05.2020)

3. Kaggle. *Вікіпедія* : веб-сайт. URL: https://uk.wikipedia.org/wiki/Kaggle (дата звернення 01.05.2020)

4. Home Credit Default Risk. *Kaggle*: веб-сайт. URL: https://www.kaggle.com/c/home-credit-default-risk (дата звернення 01.05.2020)

5. Home Credit Default Risk Competition Data Description. *Kaggle*: веб-сайт. URL: https://www.kaggle.com/c/home-credit-default-risk/data (дата звернення 01.05.2020)

6. Friedman J.H. Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, Vol. 29, No. 5. P. 1189-1232. URL: https://projecteuclid.org/download/pdf_1/euclid.aos/1013203451 (дата звернення 01.05.2020)

7. Chen T., Guestrin C. XGBoost: A Scalable Tree Boosting System. *arXiv*:1603.0275. URL: https://arxiv.org/abs/1603.02754 (дата звернення 01.05.2020)

8. LightGBM source code. *Github*. URL: https://github.com/Microsoft/LightGBM (дата звернення 01.05.2020)
9. Chen T. Story and lessons behind the evolution of XGBoost. URL: https://homes.cs.washington.edu/~tqchen/2016/03/10/story-and-lessons-behind-the-evolution-of-xgboost.html (дата звернення 01.05.2020)
10. Neyman J. On the two different aspects of the representative method: the method of stratified sampling and the method of purposive selection. Journal of the Royal Statistical Society, 97(4). 1934. P. 558-625. URL: http://www.stat. cmu.edu/~brian/905-2008/papers/neyman-1934-jrss.pdf (дата звернення 01.05.2020)
11. Krizhevsky A., Sutskever I., Hinton G. E. ImageNet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information Processing*. №25, MIT Press, Cambridge. URL: http://www.cs.toronto.edu/~hinton/absps/imagenet.pdf (дата звернення 01.05.2020)
12. Salakhutdinov R.R., Mnih A., Hinton, G.E. Restricted Boltzmann Machines for Collaborative Filtering. *International Conference on Machine Learning*. Corvallis, Oregon. 2007. URL: http://www.cs.toronto.edu/~hinton/absps/netflix.pdf (дата звернення 01.05.2020)
13. Russell S.J., Norvig P. Artificial Intelligence: A Modern Approach. 2nd edition. New Jersey: Prentice Hall, 2003.
14. Goodfellow I., Bengio Y., Courville A. Deep Learning (Adaptive Computation and Machine Learning series). Cambridge: The MIT Press, 2016

Статтю подано до редакції 12.09.2020

УДК: 519.71: 336.71 DOI 10.33111/mise.99.11

Піскунова О. В.,
доктор економічних наук, професор кафедри
економіко-математичного моделювання
Водзянова Н. К.,
старший викладач кафедри
економіко-математичного моделювання
Панченко К. С.,
здобувач кафедри
економіко-математичного моделювання,
ДВНЗ «КНЕУ імені Вадима Гетьмана»

Piskunova O.V.,
Doctor of Economics,
Professor of the Department of Economic and Mathematical Modeling
Vodzyanova N.K.,
Senior Lecturer of the Department of Economic and Mathematical Modeling
Panchenko K.S.,
Graduate Student of the Department of Economic and Mathematical Modeling,
SHEI KNEU named after V. Hetman