

---

DOI: <https://doi.org/10.15407/kvt208.02.030>

CC BY-NC

**САВЧЕНКО-СИНЯКОВА Є.А.**, канд. техн. наук, с.н.с.,  
с.н.с. відд. інформаційних технологій індуктивного моделювання,  
<https://orcid.org/0000-0003-4851-9664>, e-mail: [savchenko\\_e@meta.ua](mailto:savchenko_e@meta.ua)  
Міжнародний науково-навчальний центр інформаційних технологій  
та систем НАН України та МОН України,  
пр. Акад. Глушкова, 40, м. Київ, 03187, Україна

## **ПОРІВНЯЛЬНИЙ АНАЛІЗ СТАТИСТИЧНОГО МОДЕЛЮВАННЯ ТА НАПРЯМІВ ШТУЧНОГО ІНТЕЛЕКТУ**

---

***Вступ.** Сьогодні коло застосувань методів та інструментів машинного навчання розширюється дуже швидко, враховуючи загальну автоматизацію та цифровізацію. Використання методів та засобів машинного навчання для моделювання складних процесів дає змогу розв'язувати складні, практично нерозв'язні задачі. Проте інші методи математичного моделювання часто також досить успішно розв'язують задачу побудови моделі на основі вибірки експериментальних даних. Важливим завданням є визначення спільних та відмінних рис різних наукових напрямів штучного інтелекту для підвищення якості побудови математичних моделей.*

***Метою** статті є порівняльний аналіз засобів штучного інтелекту для побудови математичних моделей, таких як статистичне моделювання, машинного навчання, інтелектуальний аналіз даних та глибинне навчання, та визначення умов їхнього застосування.*

***Результати.** Проведено порівняльний аналіз машинного навчання та інших напрямів штучного інтелекту, який показав, що не дивлячись на схожість досліджуваних напрямів, вони мають низку суттєвих відмінностей. Ці напрями можна розрізнити за метою, можливістю перевірити або інтерпретувати отримані результати. На відміну від статистичного моделювання, застосування методів машинного навчання вимагає мінімальних людських зусиль, практично все навантаження лягає на програмне забезпечення.*

*Проведене дослідження показало, що якщо ці два підходи можуть бути використані разом, це може привести до кращого результату, ніж кожен з підходів окремо. Результати порівняльного аналізу може бути використано для побудови системи підтримки прийняття рішень на основі індуктивного підходу та принципів метамоделювання.*

***Висновки.** Виявлені спільні та відмінні риси засобів машинного навчання та статистичного моделювання визначають шляхи їхнього спільного використання для підвищення якості отриманого результату. Водночас треба враховувати, що за використання статистичного моделювання більше уваги приділяється поведінці моделі, а машинне навчання більше зосереджено на інтерпретуванні моделі та точності прогнозування.*

© САВЧЕНКО-СИНЯКОВА Є.А., 2022

**Ключові слова:** машинне навчання, статистичне моделювання, штучний інтелект, інтелектуальний аналіз даних, інтелектуальне моделювання, індуктивне моделювання.

## ВСТУП

Сьогодні, в час загальної комп'ютеризації та цифровізації, все більшу популярність здобувають методи та засоби математичного моделювання, за допомогою яких проводять аналіз великих обсягів даних, шукають в них взаємозв'язки та прогнозують показники на майбутнє. Одним з популярних засобів моделювання є методи машинного навчання, які дають змогу дуже швидко (часто в режимі онлайн) розв'язувати досить складні задачі.

Математичне моделювання залишається класичним інструментарієм отримання математичних моделей складних об'єктів та процесів. Але з появою сучасних продуктивних комп'ютерів з'явилась можливість обробляти величезні масиви даних, що раніше було неможливо. Це стимулювало розроблення нових підходів до аналізу такої інформації для побудови моделей за експериментальними даними. Наприклад, нейромережа навчається за вибіркою вхідних даних таким чином, щоб досягти високої точності обчислення значень вихідної змінної на основі вхідних даних.

Отже, напрям машинного навчання є підходом, що базується на математичній статистиці, але максимально використовує можливості сучасних комп'ютерів.

**Метою** цієї статті є аналіз спільних та відмінних рис різних наукових напрямів штучного інтелекту, таких як машинне навчання, статистичне моделювання та інтелектуальний аналіз даних, та визначення умов їхнього застосування.

## ВИЗНАЧЕННЯ ПОНЯТЬ МАТЕМАТИЧНОГО МОДЕЛЮВАННЯ ТА МАШИННОГО НАВЧАННЯ

Моделювання в широкому сенсі визначають як метод дослідження процесів або явищ шляхом створення їхніх математичних моделей та подальшого дослідження цих моделей. *Метою класичного моделювання* є здобуття, оброблення, подання та використання інформації про об'єкти, які взаємодіють між собою та зовнішнім середовищем, тобто модель тут виступає як засіб пізнання властивостей й закономірностей поведінки об'єкта [1]. Моделювання включає етапи створення, дослідження та використання моделей об'єктів.

Методи моделювання широко використовують в різних сферах людської діяльності, особливо у проектуванні та керуванні, де основними є процеси ухвалення ефективних рішень на основі інформації про досліджувані об'єкт або процес. Основним призначенням моделі в завданнях керування є прогноз реакції об'єкта на різні керувальні впливи.

Розглянемо детальніше напрям **машинного навчання** (МН) як один із засобів моделювання процесів або об'єктів, та покажемо його місце серед напрямів штучного інтелекту. МН є підмножиною засобів штучного інтелекту, спрямованих на побудову моделей систем навчання за вибіркою експериментальних даних (навчальні дані) без використання явно запрог-

рамованих етапів оброблення [2]. Алгоритми МН забезпечують побудову моделей, наприклад, для прогнозування [3], покладаючись на стандартні методи програмування, такі як об'єктно-орієнтоване програмування [4]. Водночас МН можна вважати спробою описати реальні процеси та системи шляхом апроксимації математичних функцій (рівнянь) [5].

Зазначимо, що основними властивостями методів МН є інтерпретація моделі, зменшення розмірності даних, можливість маніпулювання властивостями, перенесення властивостей на інші об'єкти та отримання нових знань тощо. До задач МН належать такі [5]: побудова регресії, класифікація, кластеризація, багатоваріантний пошук, оцінювання ймовірності, машинний переклад, виявлення аномалій тощо.

Формально завдання МН можна визначити так: комп'ютерна програма має видобути з досвіду  $E$  розв'язання деякого класу задач  $T$  та міри якості (продуктивності)  $P$  за умови, що якість  $P$  в  $T$  задачах покращується зі зростом досвіду  $E$  [6].

Але ж призначення МН набагато ширше, ніж тільки розроблення моделей для прогнозування та визначення коефіцієнтів моделі для опису певної предметної області [7]. МН зорієнтовано на вивчення експериментальних даних за допомогою комп'ютерних програм.

## **ПОРІВНЯННЯ МЕТОДІВ МАШИННОГО НАВЧАННЯ ТА СТАТИСТИЧНОГО МОДЕЛЮВАННЯ**

**Статистичне моделювання** (СМ) є підрозділом штучного інтелекту, який орієнтовано на пошук взаємозв'язків між змінними у вигляді математичних рівнянь, тобто СМ це – формалізація відношень між змінними за невеликого обсягу даних та кількості атрибутів або спостережень, тому неможна нехтувати ймовірністю перенавчання моделей [4].

СМ працює за наявності припущень (обмежень), на відміну від алгоритмів МН, які в цілому позбавлені більшості таких припущень, зокрема для МН, як правило, не потрібно вказувати розподіл залежної або незалежної змінних. Тому методи МН можна застосовувати до різних типів даних, на відміну від методів СМ.

Схожість методів МН та СМ полягає у наявності схожих цілей, але вони не є однаковими напрямками. В [1] зазначено, що спільною рисою цих підходів є етап побудови моделі. Водночас у класичному моделюванні спочатку створюють модель, перевіряють її точність, визначають границі застосування і у разі необхідності здійснюють покращення результатів моделювання. МН дає змогу розв'язувати задачі завдяки застосуванню алгоритму навчання на великій кількості зразків, внаслідок чого можна досягти прийнятних результатів.

Основною відмінністю СМ та МН є те, що СМ заснована на *ймовірнісних просторах*. Статистичну теорію можна вивести з теорії множин [8], де числа можна згрупувати в підмножини, а потім розрахувати чи задати міру для кожної підмножини, щоб гарантувати, що сумарне значення цих мір буде дорівнювати одиниці.

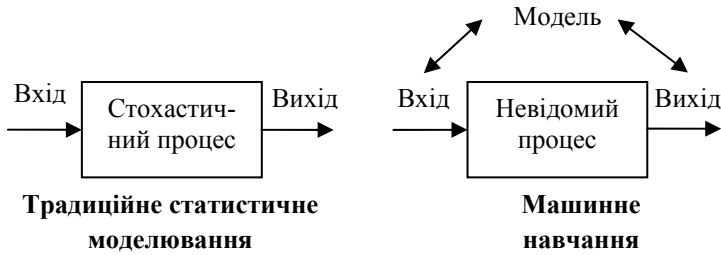


Рис. 1. Підходи до традиційного СМ та МН [9]

Друга значна відмінність СМ та МН полягає в їхніх цілях. Статистичні моделі використовують для пошуку та пояснення зв'язків між змінними, тоді як моделі МН створені для надання точних результатів на тестових даних. Хоча деякими статистичними моделями можна робити прогнози, точність цих моделей зазвичай не є найкращою, оскільки вони не можуть охопити складні зв'язки між даними. З іншого боку, моделі МН можуть надати кращі прогнози, але їх важче пояснити. МН є більш емпіричним підходом [9].

Отже, метою застосування СМ частіше є охарактеризувати взаємозв'язок між даними, а не робити прогнози, тоді як метою МН є отримати повторювані прогнози, і, в першу чергу, важливою є точність моделі, а її інтерпретування [9]. Інакше кажучи, МН дає інформацію про кінцеві результати, СМ — про взаємозв'язок між змінними, припускаючи, що дані є лінійним регресором з додаванням деякого випадкового шуму, який зазвичай має гаусовський характер [8].

Також відмінність цих методів вбачають в обсязі оброблюваних даних та у частці участі людини у побудові моделей [2]: МН добре працює з великими масивами даних, з великою кількістю змінних та спостережень; СМ зазвичай застосовують до менших масивів даних з меншою кількістю змінних.

Оскільки алгоритми МН можуть привести до перенавчання, то виникає необхідність поділу вибірки даних на навчальний та тестовий набір даних. Це не притаманно статистичним методам, оскільки у цьому випадку немає намагання мінімізувати емпіричний ризик.

Відмінність МН та СМ полягає в способі перевірки моделей. У традиційній культурі моделювання, побудована модель оцінюється за критерієм точності на тестових даних і це буде свідчити про точність прогнозу на нових даних. На відміну від цього, МН перевіряє модель на основі її попередньої точності, застосовуючи розподіл даних на дві нерівні частини в процесі побудови моделі: на одній частині здійснюється навчання за даними, на другій – перевірка, за якою визначають якість навченої моделі (налаштування на вхідні дані).

В основі методів МН покладено підхід «чорної скриньки», яка є невідомим випадковим процесом, порівняно з раніше передбачуваними простими випадковими процесами, які досліджуються за допомогою традиційного моделювання. За традиційним СМ весь сенс дослідження полягає в тому, щоб відкрити такі «чорні скриньки» для поглиблення знань про змодельовані природні процеси, які лежать в їх основі. На рис. 1 схематично продемонстровано підходи традиційного СМ та МН.

У разі застосування СМ часто розглядають процес моделювання даних як шаблон для статистичного аналізу [10], розробляючи модель, що імітує процес, який намагаються формалізувати. Треба оцінити параметри і зробити висновки, але їх надають стосовно механізму роботи моделі, а не природного процесу, який вона імітує. МН базується безпосередньо на даних спостережень та не прив'язується до природного стохастичного процесу, тому може допомогти уникнути неправильних припущень щодо цього процесу [9].

Ще раз підкреслимо, що статистичні методи зосереджені на висновках про взаємозв'язки за досліджуваними даними, що досягається шляхом створення та налаштування конкретної ймовірнісної моделі, яка дає змогу обчислити кількісну міру впевненості в тому, що виявлений зв'язок описує «справжній» ефект, який навряд чи буде результатом шуму [11]. Крім того, якщо в наявності достатньо даних, ми можемо явно перевірити припущення і за потреби уточнити отриману модель.

Таблиця 1. Порівняння базових понять СМ та МН [4]

Риса	Статистичне моделювання	Машинне навчання
Очікуваний результат	Математична модель	Нейромережі, графи
Параметри для налаштування	Параметри	Ваги
Який механізм використовується	Налагодження	Навчання
Як працюють алгоритми	Продуктивність на тестовій вибірці	Узагальнення
Які завдання розв'язують	Оцінювання/Кластеризація	Навчання без учителя
На яких принципах основано	Математичні методи як основа, з припущеннями та обмеженнями	Вимагає менше припущень.
Спрямованість на використання	рівнянь	алгоритмів
Розмір вибірки даних	Невелика вибірка даних	Може працювати як з малими, так великими обсягами даних
Потреба в зусиллях людини	Зусилля людини потрібні	Мінімальне втручання людини
Властивості отриманої моделі	Найкраща «оцінка»	Гарні прогнозні властивості

Таблиця 2. Порівняння особливих рис СМ та МН [13]

Риса	Статистичне моделювання	Машинне навчання
Мета застосування	Формалізація взаємозв'язків між змінними у вигляді математичних рівнянь	Алгоритм, який може навчитися за даними, не покладаючись на програмування на основі правил.
На що спрямовано	Необхідно визначити функцію, яка найкращим чином опише вхідні дані перед налаштуванням цієї моделі (наприклад, лінійного, полінома тощо)	Не потрібно задавати функцію, оскільки алгоритми машинного навчання можуть автоматично вивчати складні шаблони даних
Точність прогнозування	Прогнозування результатів з точністю 85 %	Прогнозування результатів з точністю 85 %
Чи є тестування даних	Використання різних методів діагностування параметрів	Відсутність жодних статистичних тестів на значущість
Як відбувається тестування	Дані поділено на дві підвибірки: навчальну (70 %) та тестову (30%). Моделі будуються на даних навчальної вибірки, а тестуються як на загальній вибірці, так і для окремих змінних	Дані поділено на дві підвибірки: навчальну (50 %) та тестову (50%). Через відсутність діагностування алгоритмів МН, необхідно здійснювати навчання на двох вибірках даних (навчальній та тестовій), щоб забезпечити дві точки перевірки

Також СМ та МН відрізняються за *обчислювальною здатністю*. Класичне СМ було розроблено для даних із кількома десятками вхідних змінних і розмірами вибірки, які сьогодні вважалися б малими або помірними. Однак із збільшенням кількості вхідних змінних, модель, яка фіксує взаємозв'язки, стає складнішою.

Навпаки, МН зосереджується на прогнозуванні за допомогою алгоритмів навчання для пошуку закономірностей часто у великих та громіздких даних. Методи МН будуть найкорисніші, коли кількість об'єктів більша, ніж кількість вхідних змінних. МН робить мінімальні припущення щодо систем генерування даних; вони можуть бути ефективними, навіть якщо дані збираються без ретельно контрольованого плану експерименту та за наявності складних нелінійних взаємозв'язків у даних. Однак, незважаючи на переконливі результати прогнозування, є ймовірність ускладненої інтерпретації отриманих результатів.

Навпаки, МН зосереджується на прогнозуванні за допомогою алгоритмів навчання для пошуку закономірностей часто у великих та громіздких даних. Методи МН будуть найкорисніші, коли кількість об'єктів більша, ніж кількість вхідних змінних. МН робить мінімальні припущення щодо систем генерування даних; вони можуть бути ефективними, навіть якщо

дані збираються без ретельно контролюваного плану експерименту та за наявності складних нелінійних взаємозв'язків у даних. Однак, незважаючи на переконливі результати прогнозування, є ймовірність ускладненої інтерпретацію отриманих результатів.

Треба враховувати відмінність у термінології у разі застосування СМ та МН [4].

Хоча є схожість між методичними підходами СМ та МН, але вона не завжди очевидна [12]. Обидва підходи вимагають мінімізації помилок, але використовують різні стратегії оптимізації для вдосконалення своїх алгоритмів. Кожний підхід має свої сильні сторони, і, як наслідок, можуть розглядатися як додаткова стратегія. Наприклад, МН може забезпечити певне розуміння процесу, який описують часові ряди, виділивши певний їхній клас, якому будуть відповідати різні методи побудови моделей. У табл. 2 показано риси, що притаманні і СМ, і МН, та їхні відмінності [13].

Отже, якщо потрібно вибрати алгоритм, який з високою точністю надаватиме прогнози, то слід вибрати МН. Якщо потрібно довести зв'язок між змінними або робити висновки за наявними даними, СМ буде найкращим [8].

## **ПОРІВНЯННЯ МАШИННОГО НАВЧАННЯ ТА ІНШИХ НАПРЯМІВ ШТУЧНОГО ІНТЕЛЕКТУ**

До засобів штучного інтелекту також належить **інтелектуальний аналіз даних** (ІАД), який випередив МН майже на два десятиліття, спершу за назвою «виявлення знань у базах даних» (KDD) [14], а потім — Data Mining. ІАД — це сукупність методів виявлення за первинними даними раніше невідомих, нетривіальних, корисних на практиці та доступних для інтерпретації знань, які є необхідними для прийняття рішень в різних сферах людської діяльності [15]. ІАД використовують з 1930-х років, МН з'явився в 1950-х роках, вперше використаний в програмі гри в шашки.

ІАД є методом дослідження зібраних вхідних даних для визначення конкретного результату. Суть та мету технології ІАД можна охарактеризувати як технологію, яка призначена для пошуку в великих обсягах даних прихованих закономірностей. З іншого боку, у разі застосування МН здійснюється навчання для виконання складних завдань, з подальшим використанням зібраних даних для розв'язання все складніших завдань.

Отже, як МН, так і ІАД дають змогу відкривати закономірності в даних, тобто нові знання, але з різною метою: МН — здійснювати комп'ютерне навчання для розв'язання завдань; ІАД — для надання людині можливості досліджувати дані та інтерпретувати отримані результати. Водночас ІАД, в основному, зосереджується на розвідному аналізі за допомогою неконтрольованого навчання, тобто коли структура вхідних даних невідома. За [16], ІАД оперує змістовними завданнями, а МН — застосуванням математичної теорії, працює з алгоритмами [17], тому виникли такі терміни як, наприклад, «алгоритми машинного навчання в аналізі даних».

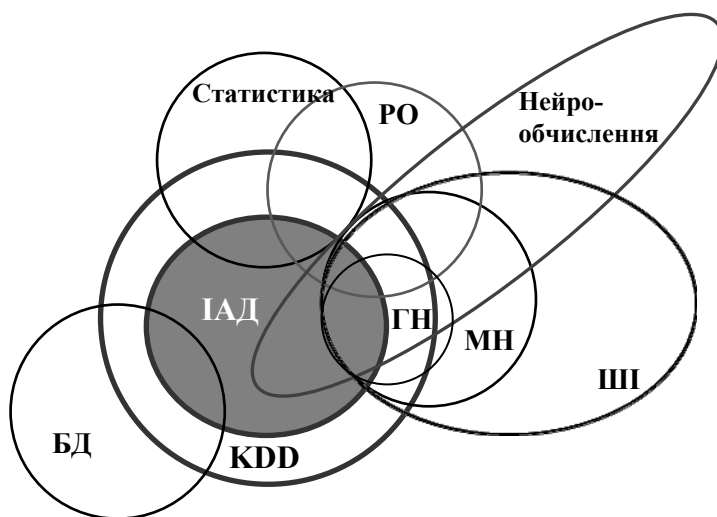


Рис. 2. Зіставлення машинного навчання та інших підходів до розв'язання завдань штучного інтелекту [9]

Отже, можна підсумувати особливості описаних напрямів.

*Статистичне моделювання:*

- більше, ніж ІАД, базується на статистичній теорії;
- більше зосереджується на перевірці гіпотез.

• *Машинне навчання:*

- має значно більшою мірою евристичний характер;
- концентрується на поліпшенні процесу навчання.

• *Інтелектуальний аналіз даних*

- характеризується інтеграцією теорії і евристик;
- дає змогу сконцентруватися на єдиному процесі аналізу даних, охоплює очищення даних, навчання, інтеграцію та візуалізацію результатів.

**Data Sciences (DS)** — наука про дані. Це обчислювальні та статистичні методи, які застосовуються до даних — як невеликих, так і великих наборів даних [8]. DS включає попереднє оброблення та аналіз даних, кодування, встановлення конвертора, з'єднання між БД, веб-сервісами тощо. Дослідницький або розвідувальний аналіз даних включає вивчення та візуалізацію даних, щоб допомогти фахівцю краще розуміти ці дані та робити з них висновки [18].

Ще однією групою методів МН є методи **глибинного навчання** або Deep Learning. Термін «глибинне навчання» (ГН) виник в середині 2000-х років, хоча окремі методи ГН відомі з 80-х років минулого сторіччя [20, 21]. Побудова нейронних мереж є найбільш широко використовуваним алгоритмом МН для прогнозування. Застосування нейронної мережі до часових рядів має різні варіанти залежно від структури та класу часових рядів і дає можливість обробляти складні структури даних [13].



Основою ГН є нейронна мережа з кількома рівнями для повторного навчання за вхідними даними [22]. Нейронна мережа імітує роботу людського мозку, що надає можливість вирішувати погано визначені проблеми. Прикладами засобів розв'язання таких завдань є програми розпізнавання зображень, мовлення та комп'ютерного зору.

Останнім часом все ширше використовують термін **інтелектуальне моделювання** (ІМ). Інтелектуальне моделювання визначено як процес побудови моделей об'єктів із застосуванням знань та інструментальних засобів, які забезпечують досягнення якості моделей на рівні кваліфікованого конструктора моделей (користувача, модельєра) [19]. Інтелектуальне моделювання включає набагато більше, ніж інструменти і методи розкриття шаблонів (залежностей) в даних. Модель має бути налаштована таким чином, щоб людина змогла зрозуміти та кількісно оцінити точність прогнозування для майбутніх (ще не отриманих) даних [7].

Для аналізу співвідношення МН, СМ та інших напрямів штучного інтелекту наведемо діаграму Венна [2], яка схематично демонструє відношення цих напрямів (рис. 2): ІАД — інтелектуальний аналіз даних, РО — розпізнавання образів, МН — машинне навчання, KDD (Knowledge Discovery in Database) — відкриття знань у базі даних, ГН — глибинне навчання. Ця діаграма візуально демонструє певні перетини напрямків штучного інтелекту між собою.

Kelleher, J. D. підкреслив, що першим розробником нейромережі ГН є О.Г. Івахненко, **метод групового урахування аргументів** (МГУА) якого вважається першою нейронною мережею, яка навчається за вхідними даними [21]. Перший загальний робочий алгоритм керованого навчання багаторівневої мережі перцептронів був опублікований у [23]. У [24] було описано нейронну мережу з восьмима шарами, навченими методом групового урахування аргументів, який широко використовують і досі [25].

МГУА також може бути названо одним з алгоритмів інтелектуального моделювання, як метод синтезу моделей з автоматичним вибором структури та параметрів лінійних, нелінійних, різницевих та інших моделей на основі короткої вибірки даних в умовах невизначеності та неповноти вихідної інформації з метою виявлення невідомої закономірності функціонування досліджуваного об'єкта або процесу, інформація про яку неявно міститься у вибірці даних [19].

В алгоритмах МГУА поділ вибірки неявно (автоматично) забезпечує дотримання відомого принципу компромісу між складністю моделі та її точністю для побудови моделі оптимальної складності. Це дає можливість віднести МГУА до найефективніших методів обчислювального інтелекту, тому цей метод може бути основою для розроблення інструментальних засобів інтелектуального моделювання.

## **ПОСІДНАННЯ РІЗНИХ МЕТОДІВ ШТУЧНОГО ІНТЕЛЕКТУ**

Як вказувалось, метою застосування як МН так і СМ є навчання на основі даних. Але між цими підходами немає жорсткого розмежування. Деякі фахівці зі статистичного моделювання запозичили методи МН, рухаючись до об'єднаної області, яку вони називають «статистичним навчанням»

(англ. statistical learning) [26]. Термін «статистичне навчання» є еквівалентом прогнозного моделювання [7].

Теорія статистичного навчання є основою для МН, яке базується на статистичній теорії та функціональному аналізі [27]. Теорія статистичного навчання має справу з проблемою статистичного висновку знаходження прогнозної функції на основі даних та успішно застосовується в таких галузях, як комп'ютерний зір, розпізнавання мови та біоінформатика.

Статистичне моделювання вимагає знати, як і чому була вибрана конкретна модель, а не як і чому робляться конкретні прогнози. Це не менш важливо, ніж інтерпретованість моделі.

Але висловлено думку [10], що область використання статистичного аналізу може зменшитись як через втрату актуальності, так і «крихкість» методів через ігнорування алгоритмічного підходу [7]. Тому спільне застосування таких підходів може давати кращі моделі, ніж кожен з підходів окремо [1].

Сьогодні застосування МН стало важливим елементом у прийнятті рішень. Це революціонізувало процес прийняття рішень завдяки мінімально можливому часу, необхідному для їх прийняття. Кожен рух людей, матеріалів, готової продукції тощо фіксується та зберігається як дані та використовується для прийняття рішень за допомогою засобів штучного інтелекту.

Поява та розвиток інструментів МН спростили статистичне прогнозування завдяки доступу до більшого обсягу даних порівняно з минулим.

Найкращим варіантом буде підхід, за яким враховуватиметься найкраще, що є в МН та СМ. Проведені дослідження демонструють, що значно корисніше використовувати ці підходи разом, ніж окремо, якщо є така можливість.

Отже, справжній фахівець з аналізу даних повинен мати в своєму арсеналі обидва підходи — МН та СМ. Сьогодні засоби МН не можна реалізувати без надійної статистичної бази [28]. Програмний код, написаний для полегшення роботи спеціаліста з аналізу даних, не скасовує потреби в глибокому розумінні теорії об'єктів чи процесів, що досліджуються.

## **ВИСНОВКИ**

Проведений порівняльний аналіз різних напрямів штучного інтелекту, таких як машинне навчання, статистичне моделювання, інтелектуальний аналіз даних, глибинне навчання та інтелектуальне моделювання, показав, що не дивлячись на те, що всі ці напрями схожі між собою, вони мають значні відмінності та обмеження для опрацювання даних.

Показано, що головна відмінність цих підходів — їхня мета. Так методи СМ вимагають у розробника моделі розуміння взаємозв'язку змінних у рівнянні, щоб якнайкраще оцінити функцію виходу з прийнятною помилкою. Методи МН здійснюють навчання для розв'язання складних завдань, методи ІАД — надають людині можливість за допомогою комп'ютера розв'язувати складні задачі. Крім того, застосування методів МН вимагає мінімальних людських зусиль, практично все навантаження лягає на програмне забезпечення.

Проведений аналіз показав, що ті галузі досліджень, де застосовують спільно класичне (статистичне) моделювання і машинне навчання дають успішніші моделі ніж ті, які є результатом застосування цих підходів окре-

мо. Результати проведеного дослідження може бути використано для побудови системи підтримки прийняття рішень на основі спільного використання статистичного моделювання і машинного навчання.

#### ЛІТЕРАТУРА

1. Поняття моделі та моделювання. Властивості та класифікація моделей. URL: [https://web.posibnyky.vntu.edu.ua/fksa/2kvetnyj\\_komp'yuterne\\_modelyuvannya\\_system\\_procesiv/t1/11.htm](https://web.posibnyky.vntu.edu.ua/fksa/2kvetnyj_komp'yuterne_modelyuvannya_system_procesiv/t1/11.htm) (Дата звернення: 2 травня 2022).
2. Srivastava, T. Difference between Machine Learning & Statistical Modeling. URL: <https://www.analyticsvidhya.com/blog/2015/07/difference-machine-learning-statistical-modeling> (Дата звернення: 25 квітня 2022).
3. Sharma, D., & Kumar, N. A review on machine learning algorithms, tasks and applications. *International Journal of Advanced Research in Computer Engineering & Technology* (IJARCET). 2017. 6(10). P. 2278–1323.
4. Machine Learning vs Statistical Modeling. URL: <https://www.youtube.com/watch?v=ja15aheBOi0> (Дата звернення: 2 травня 2022).
5. Most Common Machine Learning Tasks. URL: <https://vitalflux.com/7-common-machine-learning-tasks-related-methods> (Дата звернення: 25 травня 2022).
6. Mitchell, T. (1997). *Machine Learning*. New York: McGraw Hill. ISBN 0-07-042807-7. OCLC 36417892.
7. Difference Between Algorithm and Model in Machine Learning. URL: <https://machinelearningmastery.com/difference-between-algorithm-and-model-in-machine-learning> (Дата звернення: 12 квітня 2022).
8. The Actual Difference Between Statistics and Machine Learning. URL: <https://towardsdatascience.com/the-actual-difference-between-statistics-and-machine-learning-64b49f07ea3> (Дата звернення: 5 травня 2022).
9. Traditional Modeling vs. Machine Learning. URL: <https://medium.com/thal%C4%93s/thales-traditional-modeling-vs-machine-learning-b7996342ea93> (Дата звернення: 12 квітня 2022).
10. Breiman, L. Statistical Modeling: The Two Cultures (with comments and a rejoinder by the author). *Statistical Science*, 2001. 16(3). P. 199–231. doi: 10.1214/ss/1009213726.
11. Ij, H. Statistics versus machine learning. *Nat Methods*, 2018. 15(4). 233 p. <https://www.nature.com/articles/nmeth.4642>.
12. Dangeti, P. *Statistics for machine learning*. Packt Publishing Ltd., 2017. Packt Publishing. ISBN: 9781788295758.
13. Statistical Modeling vs Machine Learning. URL: <https://blog.arkieva.com/statistical-modeling-vs-machine-learning/> (Дата звернення: 2 грудня 2021).
14. Mannila, H. Data mining: machine learning, statistics, and databases. *Proceedings of 8th International Conference on Scientific and Statistical Data Base Management*, 1996, pp. 2–9, doi: 10.1109/SSDM.1996.505910.
15. Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI magazine*. 17(3). P. 37–37.
16. Van Calster, B., Verbakel, J. Y., Christodoulou, E., Steyerberg, E. W., & Collins, G. S. Statistics versus machine learning: definitions are interesting (but understanding, methodology, and reporting are more important). *Journal of clinical epidemiology*, 2019. 116.
17. Mitchell, Tom M. Machine learning and data mining. *Communications of the ACM*. 42.11, 1999. pp. 30–36.
18. Hripcsak, G., Duke, J. D., Shah, N. H., et al. Observational Health Data Sciences and Informatics (OHDSI): opportunities for observational researchers. *Studies in health technology and informatics*. 2015. 216. 574 p.
19. Goodfellow, I., Bengio, Y., & Courville, A. *Deep learning*. MIT press. 2016. 800 p. ISBN: 0262035618. DOI: <https://doi.org/10.1007/s10710-017-9314-z>.
20. Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural networks*. 61. pp. 85–117.

21. Степашко В.С. Концептуальные основы интеллектуального моделирования. *Управляющие системы и машины*. 2016. № 4. С. 3–15.
22. Kelleher, J. D. Deep learning. Cambridge, Massachusetts: The MIT Press; 2019. 280 p. DOI: <https://doi.org/10.7551/mitpress/11171.001.0001>.
23. Ивахненко, А.Г.; Лапа, В.Г. (1965). Кибернетические предсказывающие устройства. Киев: Наукова думка.
24. Ivakhnenko, A. (1971). Polynomial theory of complex systems. *IEEE Transactions on Systems, Man and Cybernetics* (4): 364–37.
25. Schmidhuber, J. My First Deep Learning System of 1991. Deep Learning Timeline 1960–2013.
26. Bzdok, D., Altman, N. & Krzywinski, M. Statistics versus machine learning. *Nature Methods*. 2018. 15. P. 233–234. <https://doi.org/10.1038/nmeth.4642>.
27. James, G., Witten, D., Hastie, T., & Tibshirani, R. An introduction to statistical learning. 2013. Vol. 112, New York: Springer. p. 18.
28. Statistical Modeling vs Machine Learning. URL: <https://www.kdnuggets.com/2019/08/statistical-modeling-vs-machine-learning.html> (Дата звернення: 2 квітня 2022).

Отримано 12.06.2022

## REFERENCES

1. Ponyattya modeli ta modelyuvannya. Vlastyovosti ta klasyfikatsiya modeley. URL: [https://web.posibnyky.vntu.edu.ua/fksa/2kvetnyj\\_komp'yuterne\\_modelyuvannya\\_system\\_procesiv/t1/11.htm](https://web.posibnyky.vntu.edu.ua/fksa/2kvetnyj_komp'yuterne_modelyuvannya_system_procesiv/t1/11.htm) [Last accessed 2 May 2022] (In Ukrainian).
2. Srivastava, T. Difference between Machine Learning & Statistical Modeling. URL: <https://www.analyticsvidhya.com/blog/2015/07/difference-machine-learning-statistical-modeling/> [Last accessed 25 Apr. 2022].
3. Sharma, D., & Kumar, N. (2017). A review on machine learning algorithms, tasks and applications. *International Journal of Advanced Research in Computer Engineering & Technology* (IJARCET), 6(10), pp. 2278–1323.
4. Machine Learning vs Statistical Modeling. URL: <https://www.youtube.com/watch?v=jaI5aheBOi0> [Last accessed 2 Apr. 2022].
5. Most Common Machine Learning Tasks. URL: <https://vitalflux.com/7-common-machine-learning-tasks-related-methods/> [Last accessed 25 May 2022].
6. Mitchell, T. (1997). *Machine Learning*. New York: McGraw Hill. ISBN 0-07-042807-7. OCLC 36417892.
7. Difference Between Algorithm and Model in Machine Learning. URL: <https://machinelearningmastery.com/difference-between-algorithm-and-model-in-machine-learning/> [Last accessed 2 Apr. 2022].
8. The Actual Difference Between Statistics and Machine Learning. URL: <https://towardsdatascience.com/the-actual-difference-between-statistics-and-machine-learning-64b49f07ea3> [Last accessed 5 May 2022].
9. Traditional Modeling vs. Machine Learning. URL: <https://medium.com/thal%C4%93s/thales-traditional-modeling-vs-machine-learning-b7996342ea93> [Last accessed 12 Apr. 2022].
10. Breiman, L. (2001). Statistical Modeling: The Two Cultures (with comments and a rejoinder by the author). *Statistical Science*, 16(3), 199–231. doi: 10.1214/ss/1009213726.
11. Ij, H. (2018). Statistics versus machine learning. *Nat Methods*, 15(4), 233. <https://www.nature.com/articles/nmeth.4642>.
12. Dangeti, P. (2017). *Statistics for machine learning*. Packt Publishing Ltd.
13. Statistical Modeling vs Machine Learning. URL: <https://blog.arkieva.com/statistical-modeling-vs-machine-learning/> [Last accessed 2 Dec. 2021].
14. Mannila, H. (1996). Data mining: machine learning, statistics, and databases," Proceedings of 8th International Conference on Scientific and Statistical Data Base Management, pp. 2–9, doi: 10.1109/SSDM.1996.505910.

15. Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI magazine*, 17 (3), pp. 37–37.
16. Van Calster, B., Verbakel, J. Y., Christodoulou, E., Steyerberg E. W. & Collins G. S. (2019). Statistics versus machine learning: definitions are interesting (but understanding, methodology, and reporting are more important). *Journal of clinical epidemiology*, 116.
17. Mitchell, Tom M. (1999). Machine learning and data mining. *Communications of the ACM* 42.11, pp. 30–36.
18. Hripcsak, G., Duke, J. D., Shah, N. H., et all. (2015). Observational Health Data Sciences and Informatics (OHDSI): opportunities for observational researchers. *Studies in health technology and informatics*, 216, 574 p.
19. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT press. 800 p. ISBN: 0262035618. DOI:10.1007/s10710-017-9314-z.
20. Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural networks*, 61, pp. 85–117.
21. Stepashko V.S. (2016). Conceptual Fundamentals of Intelligent Modeling. *Upravlausie sistemy i masiny*. №4. pp. 3–15 (In Russian).
22. Kelleher, J. D. (2019). Deep learning. MIT press. DOI: <https://doi.org/10.7551/mitpress/11171.001.0001>.
23. Ivakhnenko, O.H.; Lapa, V.H. (1965). Kibernetichni prorochi prystroyi. Kyiv: Naukova dumka (In Russian).
24. Ivakhnenko, A. (1971). Polynomial theory of complex systems. *IEEE Transactions on Systems, Man and Cybernetics* (4): 364–37.
25. Schmidhuber, J. My First Deep Learning System of 1991. Deep Learning Timeline 1960–2013.
26. Bzdok, D., Altman, N. & Krzywinski, M., (2018). Statistics versus machine learning. *Nature Methods*. 15, pp. 233–234. <https://doi.org/10.1038/nmeth.4642>.
27. James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning (Vol. 112, p. 18). New York: Springer.
28. Statistical Modeling vs Machine Learning. URL: <https://www.kdnuggets.com/2019/08/statistical-modeling-vs-machine-learning.html> [Last accessed 2 Apr. 2022].

Received 12.06.2022

Savchenko-Syniakova Ye.A., PhD (Engineering),  
Senior Researcher of the Department for Information  
Technologies of Inductive Modeling  
<https://orcid.org/0000-0003-4851-9664>, e-mail: savchenko\_e@meta.ua  
International Research and Training Center  
for Information Technologies and Systems  
of the National Academy of Sciences of Ukraine  
and Ministry of Education and Science of Ukraine,  
40, Acad. Glushkov av., Kyiv, 03187, Ukraine

#### COMPARATIVE ANALYSIS STATISTICAL MODELING AND APPROACHES OF ARTIFICIAL INTELLIGENCE

**Introduction.** Nowadays, the application of machine learning methods and tools is developing very rapidly, given the overall automation and digitalization. The use of machine learning methods and tools for modeling complex processes makes it possible to solve problems that were previously difficult or impossible to solve.

However other methods of mathematical modeling also make it possible to solve the problem of constructing a model based on a sample of experimental data. The task was to compare various scientific areas of artificial intelligence, such as machine learning,

mathematical modeling, statistics, data mining and inductive modeling in terms of building mathematical models, to find out what common and distinctive features they have.

**The purpose** of the research is a comparative analysis of the areas of mathematical modeling, statistics and machine learning.

**Results.** A comparative analysis of machine learning and other approaches to solving artificial intelligence problems was carried out.

**Conclusion.** The conducted analysis shows that machine learning and mathematical (statistical) modeling are similar concepts, but not the same, and it does not depend on the purpose of applying algorithms based on these approaches. Where it is possible to use these two approaches together, this seems appropriate as it can increase the amount of data processed and increase the "understandability" of the resulting models.

**Keywords:** *machine learning, mathematical modeling, artificial intelligence, data mining, statistics.*