

**ГНАТЮК ОЛЕКСАНДР**

Національна академія Державної прикордонної служби України  
імені Богдана Хмельницького  
<https://orcid.org/0009-0005-4569-0302>  
e-mail: [alexgnatyuk71@gmail.com](mailto:alexgnatyuk71@gmail.com)

**СМОЛІНСЬКИЙ ЮРІЙ**

Національна академія Державної прикордонної служби України  
імені Богдана Хмельницького  
<https://orcid.org/0009-0003-9141-5710>  
e-mail: [smol12021974@gmail.com](mailto:smol12021974@gmail.com)

**ВИНТУ АРТУР**

Національна академія Державної прикордонної служби України  
імені Богдана Хмельницького  
<https://orcid.org/0009-0007-6154-1993>  
e-mail: [arturomake@gmail.com](mailto:arturomake@gmail.com)

## ВИКОРИСТАННЯ МАШИННОГО НАВЧАННЯ ДЛЯ ОБРОБКИ ДАНИХ НА МІКРОКОМП'ЮТЕРАХ

*Машинне навчання - це метод аналізу даних, який автоматизує побудову аналітичної моделі. Це галузь штучного інтелекту, яка ґрунтується на ідеї, що машини повинні вміти вчитися та адаптуватися через досвід. Воно тісно пов'язане з обчислювальною статистикою, яка робить прогнози на основі статистичних даних, зібраних комп'ютером. Його іноді плутають з інтелектуальним аналізом даних, але це більше орієнтоване на аналіз розвідувальних даних, тоді як машинне навчання містить складні алгоритми, які використовуються в основному для прогнозування, коли машинне навчання концентрується на прогнозуванні на основі вже відомих атрибутів, отриманих через навчальні дані, тоді пошук даних більше фокусується на пошуку невідомих атрибутів у будь-яких даних. атрибутів у будь-яких даних.*

*Ключові слова: машинне навчання, аналіз, дані, алгоритм.*

HNATIUK OLEKSANDR, SMOLINSKYI YURI, VINTU ARTUR

Bohdan Khmelnytskyi National Academy of the State Border Guard Service of Ukraine, Khmelnytskyi, Ukraine

## USING MACHINE LEARNING FOR DATA PROCESSING ON MICROCOMPUTERS

*Machine learning is a data analysis method that automates the construction of an analytical model. It is a branch of artificial intelligence based on the idea that machines should be able to learn and adapt through experience. It is closely related to computational statistics, which makes predictions based on statistical data collected by a computer. It is sometimes confused with data mining, but that is more focused on analyzing intelligence data, while machine learning contains complex algorithms that are used mainly for prediction, where machine learning concentrates on predicting based on already known attributes obtained through training data, then data mining focuses more on finding unknown attributes in any data. Machine learning applications may be associated with spam filtering, optical character recognition, and search engines. Machine learning uses data to determine which algorithm is best for generating results based on the quantity, quality, and nature of the data. This data is then used for intelligent analysis in various ways, such as recommendation systems such as similar products on eBay, personalized content on google plus pages, video ads on sites such as YouTube, and last but not least, suggestions from friends on Facebook. Machine learning and artificial intelligence can be used to solve business problems, namely: Automating repetitive tasks such as data entry, document processing and customer service. This frees up employees' time for more complex work. Optimization of business processes. ML/AI can analyze data to identify inefficiencies and optimize business processes, increasing company productivity. Service personalization. ML/AI helps analyze customer data to provide them with personalized offers, recommendations and support. Automation of answers to questions and anticipation of customer needs. Detection of fraudulent transactions and cyber security. Risk management.*

*Keywords: machine learning, analysis, data, algorithm.*

### Вступ

Машинне навчання концентрується на розробці таких комп'ютерних програм і алгоритмів, які самі вчаться рости й адаптуватися під час подачі нових даних. Цей процес не схожий на процес інтелектуального аналізу даних. Обидві системи проходять через надані їм дані або збираються в пошуках шаблонів. Однак у додатках для інтелектуального аналізу даних, дані витягуються для розуміння людиною, в той час як алгоритми машинного навчання використовують ці дані для пошуку шаблонів у даних і відповідно зміни дій програми.

### Історія

У 1959 році компанія ІВМ опублікувала статтю в журналі IBM Journal of Research and Development з незрозумілою та цікавою на той час назвою [1]. Автором статті був Артуром Семюелом з ІВМ, у статті йшлося про використання машинного навчання у шашках, «щоб перевірити той факт, що комп'ютер можна запрограмувати так, що він навчиться грати в шашки краще, ніж може грати людина, яка написала програму».

Хоча це була не перша публікація, в якій використовувався термін «машинне навчання» як такий, Артур Семюел вважається першою людиною, яка ввела його в обіг і визначив машинне навчання в тому вигляді, в якому ми знаємо його сьогодні. Знакова робота Семюела стаття Семюела «Деякі дослідження в

галузі машинного навчання на прикладі гри в шашки» також є одним з перших прикладів «Деякі дослідження машинного навчання з використанням гри в шашки» також є раннім свідченням прагнення homo sapiens передати нашу власну систему навчання створеним людиною машинам.

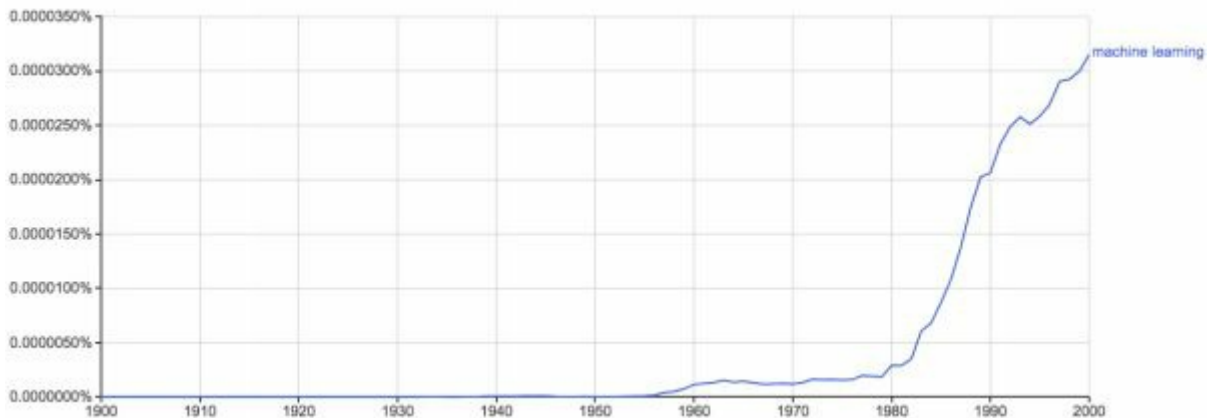


Рис.1. Історичні згадки про «машинне навчання» в опублікованих книгах.  
Джерело: Google Ngram Viewer, 2017

Артур Семюел у своїй статті вводить машинне навчання як підгалузь комп'ютерних наук, яка дає комп'ютерам можливість навчатися без явного програмування. Майже шість десятиліть по тому це визначення залишається загальноприйнятим.

Хоча у визначенні Артура Семюела це прямо не згадується, ключовою особливістю машинного навчання є концепція самонавчання. Це стосується застосування статистичного моделювання для виявлення закономірностей і підвищення продуктивності на основі даних та емпіричної інформації; і все це без прямих команд програмування. Це те, що Артур Семюел описав як здатність вчитися без явного програмування. Але він не робить висновку, що машини формують рішення без попереднього програмування. Навпаки, машинне навчання значною мірою залежить від комп'ютерного програмування. Натомість, Семюел зауважив, що машинам не потрібна пряма команда на вході, щоб для виконання поставленого завдання, а лише вхідні дані.

### Використання в інформатиці

Машинне навчання виникло через прагнення до штучного інтелекту. У перші дні штучного інтелекту вже як інтелектуального поля дослідники були дуже зацікавлені в тому, щоб машини навчалися на даних [2]. Тож вони намагалися підійти до проблеми за допомогою різноманітних символічних методів, а також з методів, які на той час називалися нейронними мережами, зазвичай це були лише моделі, які згодом були виявлені для перепакування загальних лінійних моделей ймовірності та статистики.

З приділенням особливої уваги логічному та заснованому на знаннях підходам було спричинено розрив між штучним інтелектом і машинним навчанням. Ймовірнісні системи були заражені як теоретичними, так і практичними питаннями збирання та подання даних. К 1980 року виникли експертні системи, щоб домінувати над ШІ, але статистика була невдалою.

Тривали роботи над символічними і заснованими на знаннях системами, які призводять до індуктивного логічного програмування, але статистична сфера досліджень наразі перебуває поза галуззю ШІ, в лінії розпізнавання образів і пошуку даних.

Машинне навчання стало окремим полем і почало розширюватися в 1990-х роках. Лінія змінила свою мету - досягти ШІ, намагаючись розв'язати розв'язні проблеми більш практичного характеру. Потім поле відсувало його увагу від символічних методологій, які воно успадкувало від штучного інтелекту, і натомість перейшло до методів і моделей, узятих із ймовірності та статистики [3].

Додатки машинного навчання можуть бути пов'язані з фільтрацією спаму, оптичним розпізнаванням символів і пошуковими системами. Машинне навчання використовує дані для визначення того, який алгоритм є найкращим для створення результатів, заснованих на кількості даних, якості та характері даних. Ці дані потім використовуються для інтелектуального аналізу різними способами, наприклад, такими системами рекомендацій, як схожі продукти на eBay, персоналізованим контентом на сторінках google plus, відеорекламами на сайтах, наприклад, YouTube, і останніми, але не найменшими пропозиціями друзів на Facebook. Також використовується для інтелектуального пошуку в пошукових системах Google і Bing.

### Методи машинного навчання

Машинне навчання в основному використовує діапазон або спектр на основі методу оптимізації великої кількості параметрів [4]. Для людей недоцільно знаходити таке оптимальне налаштування вручну. Наприклад, розпізнавання динаміка з тону, тону й амплітуди. Немає гарантії, що машинне навчання працюватиме в кожному випадку. Іноді машинне навчання зазнає невдачі, вимагаючи розуміння проблеми, яка має бути вирішена, щоб застосувати правильний алгоритм.

- Контрольоване навчання. Машині задаються вхідні дані та їхні бажані виходи, об'єкти, звані «вчителем», і мета полягає в тому, щоб вивчити загальне правило, яке відображає вхідні дані для виходів. Ці

алгоритми застосовують усе, що вони дізналися раніше, до будь-яких нових даних.

- Неконтрольоване навчання. Мітки / теги або пояснення не даються алгоритму навчання щодо введення, і він залишається сам по собі, щоб знайти в ньому структуру. Використовується для виявлення прихованих шаблонів у даних. Ці алгоритми можуть отримувати свої власні висновки або висновки з даних наборів даних.

- Навчання в дії. Програмне забезпечення взаємодіє з мінливим середовищем, у якому воно повинно виконувати певну задачу (наприклад, керування транспортним засобом), не повідомляючи, чи наближається воно до місця призначення, чи дізнається, як грати в гру, граючи проти когось.

- Напівкероване машинне навчання. Суб'єкт «учитель» дає машині дані з деякими недоліками, виходи відсутній.

Програмне забезпечення для машинного навчання широко доступне, а організації, які бажають розвивати можливості в цій галузі, мають безліч варіантів. При оцінці машинного навчання слід враховувати такі вимоги:

- Швидкість
- Час для оцінювання
- Точність моделі
- Легка інтеграція
- Гнучке розгортання
- Зручність використання
- Візуалізація

У наші дні дані занадто великі для людей, щоб обробляти й аналізувати їх самим.

Дуже великі вимоги до даних. Ці алгоритми навчання вимагають великої кількості даних навчання.

Було б дуже складно працювати з такими великими обсягами даних або збирати такі дані. Але такі речі, як збільшення кількості та варіації доступних даних, різноманітність обробки, що є дешевшою та потужнішою, та більш доступне сховище даних, у наші дні ми можемо швидко та автоматично створювати моделі та алгоритми, що можуть аналізувати більші та складніші дані, що забезпечують більш швидкі та точніші результати у великих масштабах. Тому машинне навчання швидко стає дуже важливою і широко впроваджуваною частиною нашого повсякденного життя.

#### **Навчальні та тестові дані**

У машинному навчанні дані поділяються на навчальні та тестові. Перша частина даних, тобто початковий запас даних, який ви використовуєте для розробки моделі, є навчальними даними. У випадку з виявленням спаму в електронній пошті на основі навчальних даних можуть бути виявлені помилкові спрацювання, подібні до автовідповіді PayPal або Facebook.

Після того, як ви успішно розробили модель на основі навчальних даних і задоволені її точністю, ви можете протестувати модель на решті даних, відомих як тестові дані. Якщо ви задоволені результатами як навчальних, так і тестових даних, модель машинного навчання готова до фільтрації вхідних електронних листів і генерування рішень про те, як класифікувати ці вхідні повідомлення.

Різниця між машинним навчанням і традиційним програмуванням на перший погляд може здатися тривіальною, але вона стане зрозумілою, коли ви розглянете подальші приклади і побачите особливу силу самонавчання в більш складних ситуаціях.

#### **Категорії машинного навчання (ML)**

Машинне навчання включає в себе кілька сотень алгоритмів, заснованих на статистиці, і вибір правильного алгоритму або комбінації алгоритмів для роботи є постійним викликом для всіх, хто працює в цій галузі. Але перш ніж ми розглянемо конкретні алгоритми, важливо зрозуміти три основні категорії машинного навчання. Ці три категорії - контрольоване, неконтрольоване та підкріплення.

#### **Навчання під контролем**

Як перша гілка машинного навчання, контрольоване навчання зосереджується на вивченні закономірностей шляхом встановлення взаємозв'язку між змінними та відомими результатами і роботи з маркованими наборами даних [3, 4].

Контрольоване навчання працює, подаючи машині вибірку даних з різними ознаками (позначеними як «X») і правильними вихідними значеннями даних (позначеними як «y»). Той факт, що значення вихідних даних і ознак відомі, кваліфікує набір даних як «маркований». Потім алгоритм розшифровує закономірності, що існують у даних, і створює модель, яка може відтворити ті ж самі основні правила з новими даними.

Наприклад, щоб спрогнозувати ринкову ціну на купівлю вживаного автомобіля, керований алгоритм може формулювати прогнози, аналізуючи зв'язок між атрибутами автомобіля (включаючи рік випуску, марку автомобіля, пробіг тощо) та ціною продажу інших автомобілів, проданих на основі історичних даних [5]. Враховуючи, що алгоритм під наглядом знає кінцеву ціну інших проданих карток, він може працювати в зворотному напрямку, щоб визначити зв'язок між характеристиками автомобіля та його вартістю.

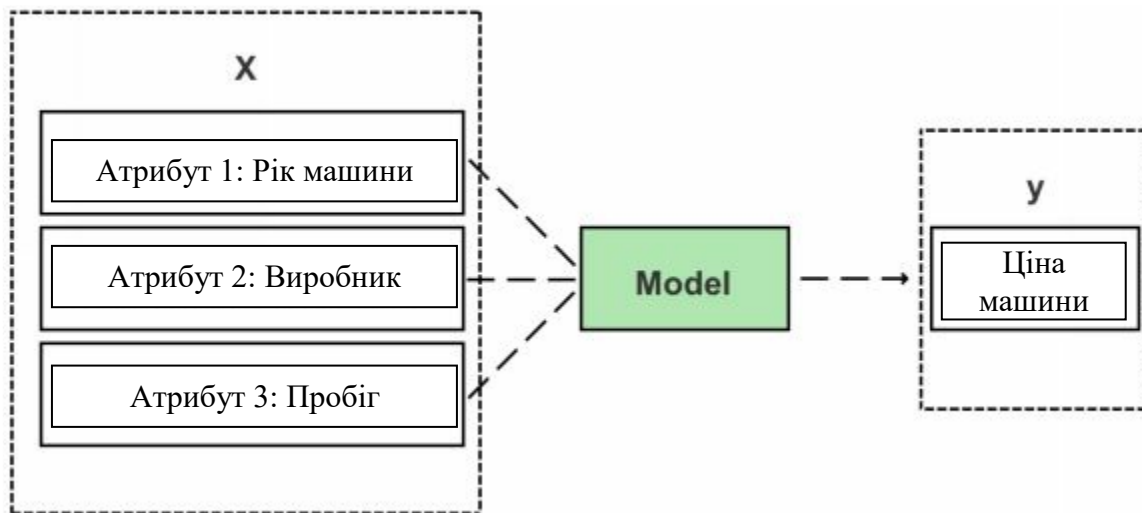


Рис.2. Модель прогнозування вартості автомобіля

### Висновки

Машинне навчання (ML) і штучний інтелект (ШІ, AI) - це потужні інструменти, які можуть допомогти бізнесу в багатьох аспектах. Наприклад ML/AI дають змогу автоматизувати багато рутинних завдань, підвищують ефективність бізнес-процесів, покращують обслуговування клієнтів, сприяють розробці нових продуктів і послуг, можуть видавати цінні рекомендації на основі аналізу величезних масивів даних.

Machine Learning і штучний інтелект можуть бути використані для вирішення завдань бізнесу, а саме:

- Автоматизація завдань, що повторюються, таких як введення даних, обробка документів і обслуговування клієнтів. Це дає змогу вивільнити час співробітників для складнішої роботи.
- Оптимізація бізнес-процесів. ML/AI може аналізувати дані для виявлення неефективності та оптимізації бізнес-процесів, підвищуючи продуктивність компанії.
- Персоналізація обслуговування. ML/AI допомагає аналізувати дані про клієнтів, щоб надавати їм персоналізовані пропозиції, рекомендації та підтримку.
- Автоматизація відповідей на запитання і передбачення потреб клієнтів.
- Виявлення шахрайських транзакцій і кібербезпека.
- Управління ризиками.

### References

1. Arthur Samuel, Some Studies in Machine Learning Using the Game of Checkers, IBM Journal of Research and Development, Vol. 3, Issue. 3, 1959.
2. Shalev-Shwartz, Shai, and Shai Ben-David. Understanding machine learning: From theory to algorithms. Cambridge university press, 2014.
3. Alpaydin, Ethem. Machine learning. MIT press, 2021.
4. Sammut, Claude, and Geoffrey I. Webb, eds. Encyclopedia of machine learning. Springer Science & Business Media, 2011.
5. Kevin Kelly, The Inevitable: Understanding the 12 Technological Forces That Will Shape Our Future, Penguin Books, 2016.