

Наприклад, аналіз даних є ключовим аспектом будь-якої програми соціологічного дослідження, де основні завдання включають визначення типу та обсягу необхідної інформації, а також вибір методів та інструментів для її збору, реєстрації, обробки та перетворення.

Висновки. Методи обробки та аналізу даних є потужними інструментами для отримання цінної інформації з різноманітних джерел. Вони сприяють прийняттю раціональних рішень, розвитку інновацій і дають змогу вирішувати складні завдання в різних галузях. У бізнесі вони допомагають оптимізувати маркетингові кампанії, прогнозувати продажі та сегментувати клієнтів; у медицині – використовуються для діагностики захворювань та визначення ефективності лікування. Однак робота з даними супроводжується викликами, до яких відносять недостовірність даних, великий обсяг інформації, що потребує значних ресурсів, і питання їх безпеки, зокрема захисту персональних даних. Подолання цих викликів вимагає впровадження нових технологій та удосконалення наявних методів.

Сучасний світ вимагає все більшого використання цих методів, що підкреслює їх важливість для науки, бізнесу та суспільства загалом.

Список використаних джерел

1. Методи обробки та інтелектуального аналізу даних з використанням штучних імунних систем / А. С. Бурда, М. А. Прудіус, Я. Г. Стефанюк, О. О. Фомічов. *Системи управління, навігації та зв'язку*. Полтава, 2024. № 3. С. 93–96.
2. Джога М. О., Ніколюк П. К. Методи обробки і аналізу даних. *Комп'ютерні технології обробки даних*. Секція 1. Методи обробки і аналізу даних, 2023. С. 35–37.

УДК 004.422.5

*Курбангалієв Д. М., здобувач вищої освіти,
Поремський Ю. В., канд. техн. наук,
старший викладач кафедри інформаційних
технологій*

АЛГОРИТМИ МАШИННОГО НАВЧАННЯ

Донецький національний університет імені Василя Стуса, м. Вінниця

Алгоритми машинного навчання стали основою сучасного цифрового світу, даючи змогу вирішувати складні завдання, які були неможливі кілька десятиліть тому. Одним із найбільш яскравих прикладів їх впливу є нейронні мережі та їх застосування у комп'ютерному зорі. Нейронні мережі особливо глибоко демонструють неймовірну ефективність у розпізнаванні об'єктів на зображеннях, що допомагає створити системи для медичної діагностики, автономного водіння та навіть мистецької творчості. Оскільки машинне навчання є важливим складником багатьох сучасних систем, знання алгоритмів та методів його реалізації є необхідними не лише для фахівців, а й для широкої аудиторії, зокрема і школярів та студентів [1].

Машинне навчання (Machine Learning) – це підгалузь штучного інтелекту, що розробляє алгоритми, які дають змогу комп'ютерам вивчати закономірності в даних і використовувати ці знання для прийняття рішень. Процес машинного навчання включає кілька етапів:

1. Збір даних: дані збираються з різних джерел (системи, сенсори, інтернет).
2. Попередня обробка даних: очищення та підготовка даних для подальшого аналізу.
3. Навчання моделі: на основі даних модель починає формувати правила, які використовуються для прогнозування або прийняття рішень.
4. Оцінка моделі: перевірка моделі на тестових даних для оцінки її точності та ефективності.
5. Використання та розгортання: готову модель застосовують у реальних задачах [2].

Машинне навчання розподіляється на кілька типів залежно від алгоритмів машинного навчання:

1. Кероване навчання (Supervised Learning) – це підхід у машинному навчанні, коли модель тренується на основі мічених даних.

1.1 Лінійна регресія. Лінійна регресія є одним з простих алгоритмів для прогнозування числових значень. Вона будує лінію (або гіперплощину в багатовимірному просторі), яка мінімізує помилку між передбаченими та фактичними значеннями.

1.2 Логістична регресія. Логістична регресія – це статистичний метод, що використовується для бінарної класифікації. Замість лінії логістична регресія використовує функцію логістичної (сигмоїдної) для прогнозування ймовірностей того, до якого з двох класів належить спостереження.

1.3 Дерево рішень. Дерево рішень – це структура, що розбиває дані на різні класи за допомогою послідовних умов. Кожен вузол дерева – це тест або перевірка, що розділяє дані на підгрупи, а кожен листовий вузол – це передбачення або результат для певної групи.

1.4 Алгоритм К-ближчих сусідів (KNN). Алгоритм К-ближчих сусідів працює за принципом «схожих елементів». Коли потрібно класифікувати новий елемент, алгоритм визначає найближчих К сусідів у навчальних даних і призначає клас на основі більшості голосів.

2. Некероване навчання (Unsupervised Learning). У некерованому навчанні модель працює з даними без міток. Метою є виявлення структур чи патернів в даних, як-от групи або залежності між ознаками.

2.1 Метод головних компонент (PCA). Метод головних компонент (PCA) використовується для зменшення розмірності даних. Це дає змогу зберегти основну інформацію, зменшуючи кількість ознак, які можуть бути використані для подальшого аналізу або моделювання. PCA шукає нові осі (головні компоненти), які максимізують дисперсію в даних.

2.2 Кластеризація К-середніх. Кластеризація К-середніх – це популярний алгоритм для групування схожих елементів у кластери. Користувач визначає кількість кластерів (К), і алгоритм намагається знайти центри кластерів так, щоб елементи всередині кожного кластера були якомога більш схожими.

2.3 *Алгоритм самоорганізуючих карт (SOM)*. Самоорганізуючі карти (SOM) – це тип нейронної мережі, що використовує метод кластеризації для перетворення багатовимірних даних у більш зрозумілу 2D- або 3D-форму.

3. *Напівкеруване навчання (Semi-supervised Learning)*. Напівкеруване навчання є комбінованим підходом, коли навчання проводиться з використанням як мічених, так і немічених даних. Алгоритм намагається використовувати додаткову інформацію з немічених даних, щоб покращити точність моделі.

4. *Машинне навчання з підкріпленням (Reinforcement Learning)*. Машинне навчання з підкріпленням відрізняється від інших типів навчання тим, що агент вчиться через взаємодію з середовищем. Він отримує винагороду або покарання залежно від того, як добре виконана дія, і з часом намагається максимізувати суму отриманих винагород. Класичний приклад – це Q-навчання, де агент поступово вивчає стратегію для досягнення максимальної винагороди [4].

Машинне навчання активно застосовується в багатьох сферах:

Соціальні мережі. Машинне навчання дає змогу платформам, як-от Facebook, Instagram і LinkedIn, аналізувати активність користувачів і на основі цієї інформації пропонувати відповідних друзів, контент або рекламу.

Охорона здоров'я. Алгоритми можуть аналізувати зображення медичних сканувань, як-от рентгенівські знімки або МРТ, для виявлення патологій, наприклад, ракових пухлин або захворювань ока.

Фінанси. У фінансовій сфері машинне навчання застосовується для оцінки кредитного ризику, де системи аналізують фінансову історію клієнта, його платіжну здатність і прогнозують ймовірність неплатежу.

Ритейл (торгівля). Машинне навчання в ритейлі, наприклад, у компаніях Amazon, eBay, використовується для персоналізації рекомендацій товарів.

Транспорт і транспортна логістика. Машинне навчання активно використовується у сфері транспорту для створення систем автоматизованого управління дорожнім рухом. Безпілотні автомобілі – це яскравий приклад використання технологій машинного навчання в транспорті.

Інтернет речей (IoT). Інтернет речей є ще однією сферою, де машинне навчання дає змогу зібрати дані з великої кількості сенсорів і пристроїв, як-от смарт-годинники, домашні термостати, розумні камери спостереження тощо.

Розваги та медіа. Машинне навчання також застосовується в індустрії розваг для персоналізації контенту. Алгоритми рекомендують фільми, музику, книги або відео на основі попередніх вподобань користувача [3].

Отже, машинне навчання є не тільки важливим компонентом сучасних технологій, але й потужним інструментом для трансформації численних індустрій. Його здатність автоматизувати складні процеси, обробляти великі обсяги даних та робити прогнози має величезний потенціал. Вивчення основ машинного навчання дає змогу студентам і школярам розвивати критичне мислення і готуватися до роботи в інноваційних сферах. Просте освоєння алгоритмів, як-от KNN або дерево рішень, створює хорошу основу для розуміння складніших концепцій і сприяє підготовці до кар'єри в технологічних компаніях [1].

Список використаних джерел

1. Alpaydin E. Introduction to Machine Learning. Book. The MIT Press. 2014. 600 p.
2. Bishop C. M. Pattern Recognition and Machine Learning. Book. Springer. 2006. 738 p.
3. Goodfellow I., Bengio Y., Courville A. Deep Learning. Book. MIT Press. 2016. 775 p.
4. Hastie T., Tibshirani R., Friedman J. The Elements of Statistical Learning. Book. Springer 2009. 745 p.
5. Mitchell T. M. Machine Learning. Book. McGraw-Hill. 1997. 414 p.
6. Russell S. J., Norvig P. Artificial Intelligence: A Modern Approach. Book. Prentice Hall. 2010. 1152 p.

УДК 004.9

*Лавренюк Б. В., здобувач вищої освіти,
Потапова Н. А., канд. екон. наук, доцент,
доцент кафедри інформаційних технологій*

ВИКОРИСТАННЯ АНАЛІЗУ ДАНИХ ДЛЯ ПРОГНОЗУВАННЯ ТА ВИЯВЛЕННЯ ТРЕНДІВ У СВІТІ КІНЕМАТОГРАФІЇ

Донецький національний університет імені Василя Стуса, м. Вінниця

Сучасний кінематограф все більше інтегрується з технологіями, що дає змогу глибше розуміти вподобання аудиторії та передбачати майбутні тенденції. Одним із найпотужніших інструментів для цього є аналіз даних. Використання великих обсягів інформації дає змогу не лише прогнозувати успіх кінопроектів, але й оперативно реагувати на зміни у запитах глядачів. У цьому контексті розкриття можливостей аналізу даних у кіноіндустрії набуває стратегічної значущості.

З огляду на те, що стримінгові сервіси, соціальні мережі та онлайн-рецензії формують глобальний кіноринок, кінокомпанії зіштовхуються з викликом не лише створення якісного продукту, а й конкурентної боротьби за увагу глядача. Аналіз даних допомагає не лише передбачити успішність майбутніх проєктів, але й знижує фінансові ризики за допомогою обґрунтованого прийняття рішень. У світі, де смаки аудиторії змінюються надзвичайно швидко, здатність виявляти тренди стає визначальним чинником успіху.

Аналіз даних став одним із ключових інструментів у світі кінематографії для виявлення трендів і прогнозування успіху проєктів. Провідні гравці індустрії активно використовують сучасні технології аналізу, як-от машинне навчання, обробка природної мови (NLP), кластеризація даних і алгоритми глибокого навчання, щоб зрозуміти вподобання аудиторії та адаптуватися до змін на ринку.

Netflix, один із лідерів стримінгового ринку, активно застосовує алгоритми машинного навчання для аналізу даних про глядачів. Система аналізує мільярди годин переглядів, включно з даними про час перегляду, пропуски, паузи та повторні перегляди. Наприклад, під час розробки серіалу «Дивні дива» (Stranger Things) компанія враховувала популярність жанру наукової фантастики серед молодшої аудиторії, що була підтверджена аналізом історичних