

СИСТЕМНИЙ АНАЛІЗУ ПОВЕДІНКИ КОРИСТУВАЧІВ НА РІЗНИХ ІГРОВИХ ПЛАТФОРМАХ ЗА ДОПОМОГОЮ МЕТОДІВ МАШИННОГО НАВЧАННЯ

Вінницький національний технічний університет

Анотація

Робота присвячена розвідувальному аналізу даних та кластеризації поведінки користувачів на ігрових платформах (Steam, PlayStation, Xbox) із застосуванням методів машинного навчання. Було проведено аналіз набору даних Kaggle «Gaming Profiles 2025» та визначено основні поведінкові ознаки користувачів.

Ключові слова: системний аналіз, поведінка користувачів, машинне навчання, кластеризація, аналіз даних, Kaggle, ігрові платформи.

Abstract

The paper is devoted to exploratory data analysis and clustering of user behavior on gaming platforms (Steam, PlayStation, Xbox) using machine learning methods. The Kaggle dataset "Gaming Profiles 2025" was analyzed and the main behavioral features of users were identified.

Keywords: system analysis, user behavior, machine learning, clustering, data analysis, Kaggle, gaming platforms.

Вступ

Поведінка користувачів ігрових платформ є комплексним феноменом, що має кількісні та якісні характеристики. Вона може значно відрізнятися залежно від типу гравця, ігрових жанрів, рівня залученості та платформи. Аналіз такої поведінки дозволяє зрозуміти вподобання користувачів, спрогнозувати їх активність та створити більш ефективні маркетингові стратегії.

Особливо важливою ця задача є для провідних ігрових платформ, таких як Steam, PlayStation та Xbox. В умовах високої конкуренції та швидкого розвитку ігрової індустрії сегментація гравців та визначення ключових поведінкових груп дозволяють утримувати наявних користувачів та залучати нових. Знання про стилі гри, рівень активності, типові вподобання гравців можуть бути використані для розробки індивідуальних рекомендацій та спеціалізованих пропозицій.

У зв'язку з цим виникає необхідність у розробці інформаційних технологій, здатних аналізувати та класифікувати поведінку гравців на основі великих обсягів даних. Машинне навчання дозволяє виявляти приховані закономірності, які неможливо визначити простими статистичними методами. Це дає можливість компаніям перейти від загального до персоналізованого підходу в роботі з користувачами, тим самим значно підвищуючи рівень їх залученості та задоволеності ігровим досвідом.

Розвідувальний аналіз

Для проведення аналізу було обрано набір даних з платформи Kaggle під назвою «Gaming Profiles 2025», автором якого є користувач «Artyom Kruglov». Даний датасет містить відкриту інформацію про користувачів трьох ігрових платформ (Steam, PlayStation, Xbox). У наборі даних міститься інформація про кількість ігор у бібліотеці кожного гравця, кількість досягнень, дні активності, середню кількість ігор та досягнень за день та інші поведінкові характеристики гравців. Датасет містить більше мільйона записів та понад десять поведінкових ознак для кожного гравця.

У наборі даних наявна наступна інформація:

- Ідентифікатори користувачів (playerid) — унікальні ідентифікатори гравців.
- Бібліотеки ігор (library) — список ігор, що є у бібліотеці користувача.

- Кількість ігор (num_games) — загальна кількість ігор у бібліотеці.
- Досягнення (achievements) — інформація про кількість та типи досягнень користувачів.
- Кількість досягнень (num_achievements) — загальна кількість отриманих досягнень.
- Дні активності (active_days) — кількість днів, коли гравець був активним на платформі.
- Кількість ігор на день (games_per_day) — середня кількість ігор, які користувач запускав щоденно.
- Кількість досягнень на день (achievements_per_day) — середня кількість досягнень, що отримує гравець за день.
- Кількість досягнень на гру (achievements_per_game) — середня кількість досягнень на одну гру.
- Інформація про ігри (жанри, дата виходу, рейтинг, назва гри).
- Країни користувачів (country) — географічні дані про гравців.
- Історія придбань ігор (покупки, ціни, дати покупок).
- Рецензії та відгуки користувачів (оцінки, текстові відгуки).

Датасет містить більше мільйона записів, що забезпечує достатню репрезентативність і дозволяє виявляти як загальні тренди, так і окремі сегменти гравців із специфічною поведінкою. Це надає значний потенціал для подальшого аналізу та кластеризації за допомогою методів машинного навчання.

У датасеті присутні такі основні таблиці:

- players.csv – унікальні користувачі з країною, нікнеймом, кількістю ігор, активних днів тощо;
- purchased_games.csv – інформація про придбані ігри кожного гравця;
- achievements.csv – досягнення по кожному користувачу та грі;
- games.csv – додаткова інформація про кожен гру (назва, жанр, дата виходу);
- history.csv – історія активності (Steam);
- reviews.csv – рецензії користувачів на ігри;
- prices.csv – ціни ігор.

Для кожної платформи (Steam, PlayStation, Xbox) дані структуровані окремо, що дає змогу проводити як платформенний, так і загальний аналіз.

Було виконано попередню статистичну обробку (describe), що дозволило виявити такі закономірності:

- Діапазон кількості ігор у бібліотеці гравця — від 1 до кількох тисяч.
- Середня кількість ігор для більшості гравців: Steam — близько 80–150, PlayStation та Xbox — менше 50.
- Кількість досягнень, як правило, пропорційна кількості ігор, але існують гравці-колекціонери з сотнями/тисячами досягнень.
- Значна частина гравців активна протягом менше ніж 100 днів, хоча є унікальні користувачі з багаторічною історією активності.
- У таблиці досягнень спостерігаються як масові (типові) досягнення, так і рідкісні/унікальні.

Було побудовано гістограми розподілу кількості ігор у бібліотеках гравців, розподілу досягнень, активних днів та країнової структури користувачів (рис. 1-4). Виявлено:

- Три основні групи гравців: казуальні (маленька бібліотека), середньоактивні, “колекціонери” (багато ігор).
- Найбільша частка користувачів — з США, Бразилії, Китаю, Росії.
- В кожній платформі ТОП-10 ігор відрізняється, що свідчить про різну аудиторію та переваги.

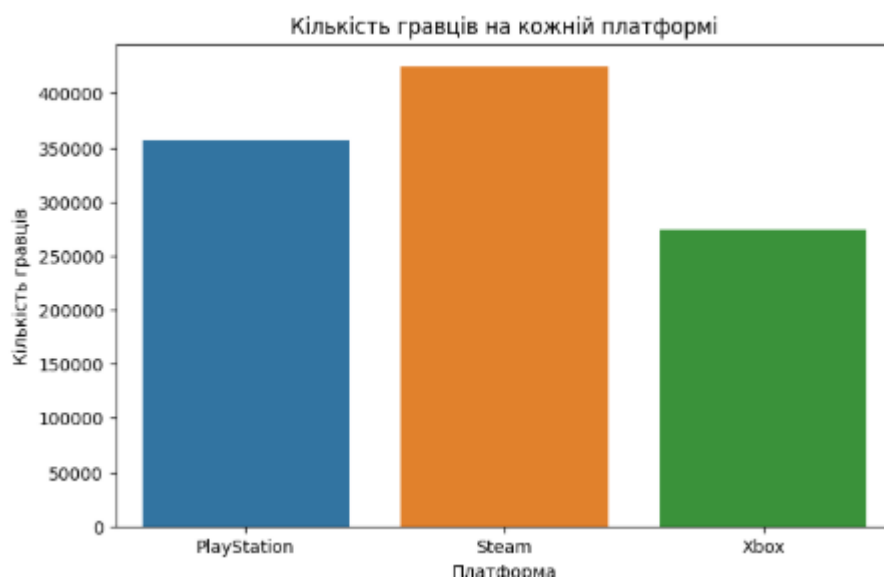


Рисунок 1 – Розподіл гравців по платформах

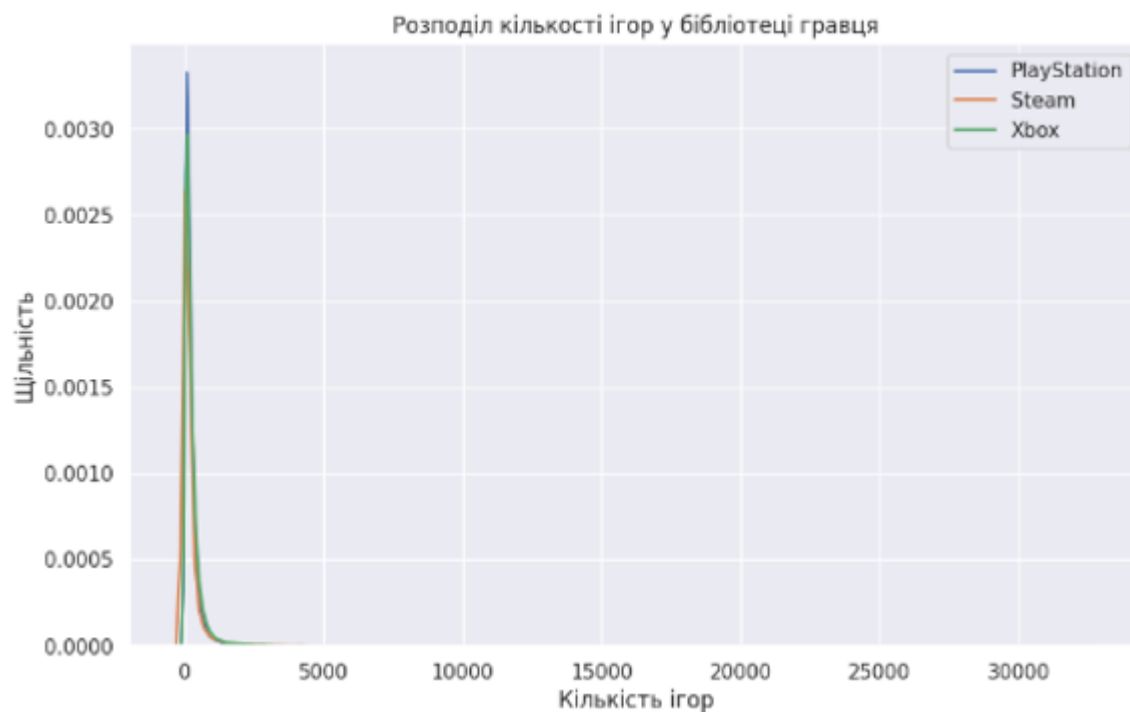


Рисунок 2 – Розподіл кількості ігор в бібліотеці гравця



Рисунок 3 – Визначення гравців по ігровій активності

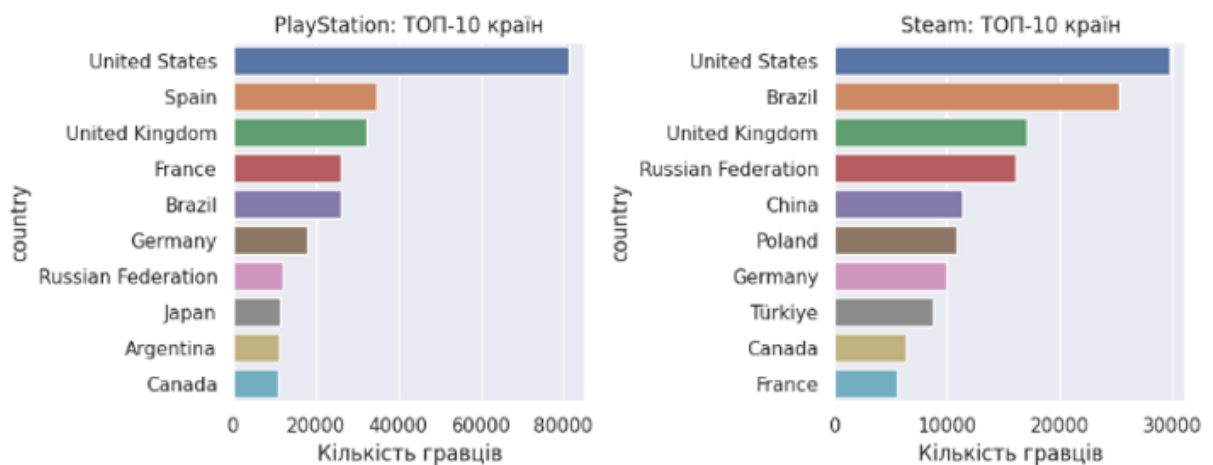


Рисунок 4 - Країнова структури користувачів

На основі первинних даних сформовано нові ознаки:

- Кількість ігор на день;
- Кількість досягнень на день;
- Кількість досягнень на одну гру;
- Активність за жанрами ігор (для Steam).

Ці ознаки використано для побудови моделей кластеризації та подальшого аналізу типів гравців.

Висновки

У процесі системного аналізу даних ігрових платформ було здійснено детальний огляд наявних датасетів, їх основних ознак і базових статистик. Встановлено, що дані охоплюють ключові параметри активності гравців — кількість ігор у бібліотеці, досягнення, кількість днів активності, а також низку додаткових ознак, які можуть характеризувати стиль і інтенсивність взаємодії користувачів із платформою.

Під час первинного аналізу даних було виявлено наявність значних варіацій у кількості ігор і досягнень серед різних користувачів, а також певні крайні (аномальні) значення, які свідчать про наявність як «казуальних», так і «активних» гравців. Аналіз розподілу числових ознак показав, що більшість гравців мають відносно невелику бібліотеку ігор, тоді як невелика частина користувачів вирізняється особливо високою ігровою активністю.

Здійснений розвідувальний аналіз дозволив виділити основні фактори для подальшої кластеризації та моделювання поведінки гравців. Одержані висновки стали підґрунтям для вибору ознак і постановки задачі сегментування користувачів на основі їхньої активності та досягнень, що дозволяє ефективніше аналізувати відмінності між різними групами геймерів на платформах Steam, PlayStation і Xbox.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Kruglov, Artyom. Gaming Profiles 2025: Steam, PlayStation, Xbox. Kaggle Dataset. URL: <https://www.kaggle.com/datasets/artiomkruglov/gaming-profiles-2025-steam-playstation-xbox>
2. Pedregosa, F., Varoquaux, G., Gramfort, A. et al. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, 12, 2825–2830, 2011. URL: <https://scikit-learn.org/stable/>
3. Facebook Prophet Documentation. URL: <https://facebook.github.io/prophet/>
4. Matplotlib Pyplot Documentation. URL: https://matplotlib.org/stable/api/pyplot_api.html

Хащеватський Олександр Вікторович – студент групи СА-21б, факультет інтелектуальних інформаційних технологій та автоматизації, Вінницький національний технічний університет, Вінниця, e-mail: alexsandr0982716823@gmail.com

Козачко Олексій Миколайович – к.т.н., доцент кафедри системного аналізу та інформаційних технологій, Вінницький національний технічний університет, e-mail: lekoz80@gmail.com

Khaschevatsky Oleksandr Viktorovich – student of group SA-21b, Faculty of Intelligent Information Technologies and Automation, Vinnytsia National Technical University, Vinnytsia, e-mail: alexsandr0982716823@gmail.com

Kozachko Oleksii M. - Ph.D., Assistant Professor of the Department of System Analysis and Information Technologies, Vinnytsia National Technical University, Vinnytsia, e-mail: lekoz80@gmail.com