

СИСТЕМА АНАЛІЗУ ПОПУЛЯРНОСТІ КОМП'ЮТЕРНИХ ІГОР

Плешивцев Є.О., науковий керівник Голуб Б.Л.

Актуальність. Ефективний аналіз популярності комп'ютерних ігор відіграє важливу роль для видавців, інвесторів та розробників, оскільки він допомагає у стратегічному плануванні: від вибору оптимального часу релізу та жанрової спрямованості гри до прийняття обґрунтованих інвестиційних і маркетингових рішень. Аналіз великих масивів даних про перегляди стримів, обсяги продажів, відгуки гравців та активність користувачів дозволяє виявляти ключові тенденції у спільноті геймерів і на ринку відеоігор. Це, своєю чергою, сприяє ефективнішому просуванню продуктів та максимізації прибутку.

З огляду на те, що індустрія відеоігор генерує значні обсяги інформації, виникає потреба у створенні сховищ даних (СД) для ефективної організації й аналізу накопичених даних. У цьому контексті особливо актуальними є технології ETL (Extract, Transform, Load) та OLAP (Online Analytical Processing), які дозволяють автоматизувати процеси перенесення, обробки й багатовимірного аналізу даних.

Об'єктом дослідження є процес аналізу популярності комп'ютерних ігор, що охоплює дані про стримінгову активність, продажі, рецензії та активність гравців.

Предметом дослідження є методи аналізу даних, технології побудови сховищ даних та підходи до класифікації відеоігор за рівнем їх популярності.

Мета дослідження: розробити та реалізувати систему, яка дозволяє аналізувати популярність комп'ютерних ігор і прогнозувати її на основі накопичених даних (метрик стримінгової активності, обсягів продажів, рецензій та онлайн-активності гравців).

Для досягнення мети було вирішено наступні завдання:

- Спроектовано структуру оперативної бази даних, яка забезпечує зберігання поточних даних про відеоігри, показники стримінгу, продажі, рецензії та активність гравців. Структура представлена на рисунку 1.
- Реалізовано механізми ETL (Extract, Transform, Load) для завантаження даних у сховище (деталі поза межами тез).
- Розроблено та програмно реалізовано на C#[1,2] та Python[3,4] алгоритми для вирішення задачі класифікації (наприклад, визначення класу популярності для комбінацій Гра-Регіон).

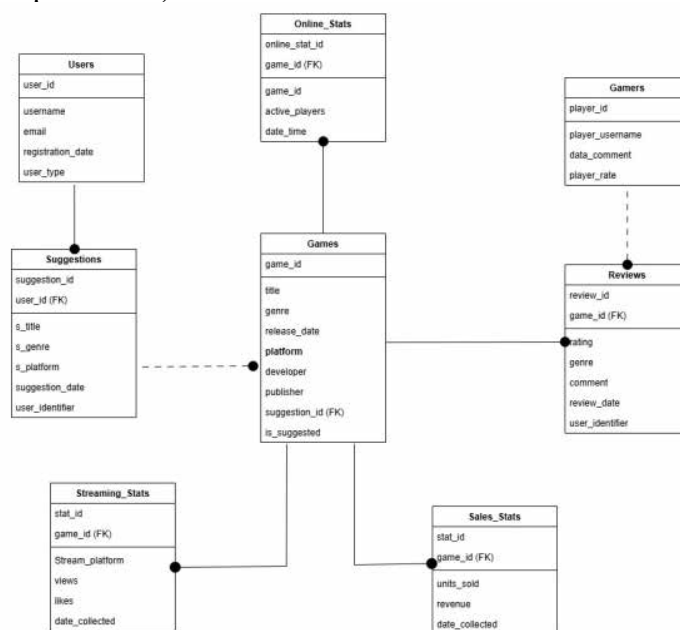


Рис. 1 Структура оперативної бази даних

Оперативна база даних зберігає актуальну інформацію про комп'ютерні ігри, включаючи стримінгову статистику, продажі, онлайн-активність гравців і відгуки. Центральною є таблиця Games, пов'язана з іншими таблицями через зовнішні ключі. Структура є нормалізованою, що забезпечує швидке оновлення даних та підготовку їх до аналітики.

Створення аналітичної системи аналізу популярності ігор на базі сховища даних із використанням сучасних методів аналізу даних (таких як алгоритм Наївного Байєса) є актуальним і перспективним напрямком. Такий підхід дозволяє перейти від інтуїтивних припущень до науково обґрунтованих рішень у сфері розробки та маркетингу ігор. Отримані класифікації, як видно на рисунку 2, надають корисні прогностні дані для інвесторів та розробників ігор.

	GameName	RegionName	AvgViewers	PredictedClass	ProbHigh	ProbLow
	20 Minutes Till ...	Oceania	1267	L	14.95	85.05
	2XKO	Global	2876	L	13.13	86.87
	60 Seconds! Re...	Oceania	2455	L	6.57	93.43
	60 Seconds! Re...	North America	1004	L	6.23	93.77
	60 Seconds! Re...	Africa	1961	L	6.64	93.36
	60 Seconds! Re...	Asia	1352	L	6.87	93.13
	7 Days to Die	Europe	2011.5	L	3.64	96.36
	7 Days to Die	Oceania	1913.62	L	4.21	95.79
	7 Days to Die	Global	4001.25	L	3.64	96.36
	7 Days to Die	South America	2653.67	L	4.33	95.67
	7 Days to Die	North America	2593	L	3.98	96.02
	7 Days to Die	Africa	1709.75	L	4.26	95.74

Рис. 2. Приклад результатів класифікації методом Naive Bayes.

На рисунку 2 продемонстровано результати застосування алгоритму Наївного Байєса, де для кожної комбінації «Гра - Регіон» розраховано ймовірність належності до класу високої популярності ('H') та визначено фінальний клас популярності ('H' або 'L'). Клас H відповідає тим випадкам, коли середня кількість глядачів на місяць для гри в певному регіоні дорівнює або перевищує загальне середнє значення по всіх іграх (у вибірці, наприклад, ~11 550), а клас L – коли цей показник нижчий за усереднений по галузі рівень.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Документація мови C#: [Електронний ресурс]. – Режим доступу: <https://learn.microsoft.com/en-us/dotnet/csharp/> (дата звернення: 19.04.2025).
2. ML.NET – офіційна бібліотека машинного навчання для .NET: [Електронний ресурс]. – Режим доступу: <https://dotnet.microsoft.com/en-us/apps/machinelearning-ai/ml-dotnet> (дата звернення: 19.04.2025).
3. Документація мови Python: [Електронний ресурс] - Режим доступу: <https://docs.python.org/3/> (дата звернення 19.05.2025).
4. Документація Scikit-Learn: [Електронний ресурс] - Режим доступу: <https://scikit-learn.org/stable/> (дата звернення 19.05.2025).

**НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ БІОРЕСУРСІВ І
ПРИРОДОКОРИСТУВАННЯ УКРАЇНИ
ФАКУЛЬТЕТ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ**



**ТЕОРЕТИЧНІ ТА ПРИКЛАДНІ
АСПЕКТИ РОЗРОБКИ
КОМП'ЮТЕРНИХ СИСТЕМ**

**ЗБІРНИК НАУКОВИХ ПРАЦЬ ЗА МАТЕРІАЛАМИ
VII ВСЕУКРАЇНСЬКОЇ НАУКОВО-ПРАКТИЧНОЇ КОНФЕРЕНЦІЇ
СТУДЕНТІВ І АСПІРАНТІВ
*24 квітня 2025 року***

Київ 2025

УДК 004

Відповідальна за випуск: М.І. Лендел

Збірник наукових праць за матеріалами VII Всеукраїнської науково-практичної конференції студентів і аспірантів «ТЕОРЕТИЧНІ ТА ПРИКЛАДНІ АСПЕКТИ РОЗРОБКИ КОМП'ЮТЕРНИХ СИСТЕМ '2025», 24 квітня 2025 року, НУБіП України, Київ. – 452 с. (електронне видання)

Відповідальність за зміст публікацій несуть автори.

Передрук матеріалів, а також використання їх будь-якій формі допускається лише з дозволу авторів