

ВИДІЛЕННЯ ОЗНАК КІНОФІЛЬМІВ, ЩО ВПЛИВАЮТЬ НА УСПІШНІСТЬ КІНОФІЛЬМУ

¹Вінницький національний технічний університет

Анотація

Проаналізовано ознаки кінофільмів, що найбільше впливають на успішність кінофільмів. Отримано показники кореляції характеристик кінофільму до його доходу.

Ключові слова: виділення ознак, математична модель, Python, кореляція, TMDB, кінофільм.

Abstract

The features of motion pictures that have the greatest influence on the success are analyzed. Correlation coefficients between the attributes of a movie and its income have been obtained.

Keywords: prediction, math model, Python, correlation, TMDB, movie.

Останніми роками кіноринок зростає з кожним роком. Ця індустрія щорічно приносить приблизно мільярди доларів доходу [1]. Питання про те, що робить фільм успішним залишається актуальним і по сьогодні, і не тільки для споживачі кіноіндустрії: великі компанії присуджують мільйонні премії тим, хто може покращити їхні рекомендації і алгоритми прогнозування. Метою даної роботи є виділення ознак, що впливають на його кінцевий дохід.

Для знаходження ознак кінофільмів, що впливають на успішність кінофільму використано набір даних TMDB. TMDB – один із найбільших база даних про фільми, що містить інформацію про 568 729 кінофільмів та серіалів [2]. В таблиці 1 перелічено усі характеристики кінофільмів, що надає TMDB.

Таблиця 1 – Характеристики кінофільмів у наборі даних TMDB.

Колонка	Тип даних	Опис
id	integer	Індикатор фільму
belongs_to_collection	json	Приналежність до франшизи
budget	integer	Бюджет
genres	json	Список жанрів фільму
homepage	string	Посилання на домашню сторінку
imdb_id	string	Індифікатор фільму на IMDb
original_language	string	Оригінальна мова фільму
original_title	string	Заголовок фільму
overview	string	Короткий опис фільму
popularity	float	Рейтинг популярності
poster_path	string	Посилання на постер
production_companies	json	Список компаній виробників
production_countries	json	Список країн виробників
release_date	date	Дата релізу
runtime	integer	Хронометраж в хвилинах
status	string	Статус
keywords	json	Ключові слова
cast	json	Список акторів
crew	json	Знімальна команда
revenue	integer	Дохід

Використовуючи набір даних *TMDB* та *Python* пакети *Seaborn*, *Matplotlib* та *Plotly*, побудуємо діаграми кореляції ознак кінофільму до його доходу, та графіки співвідношення бюджету фільму до його доходу [3, 4, 5, 6].

На рисунку 1 зображено кореляцію ознак кінофільмів.

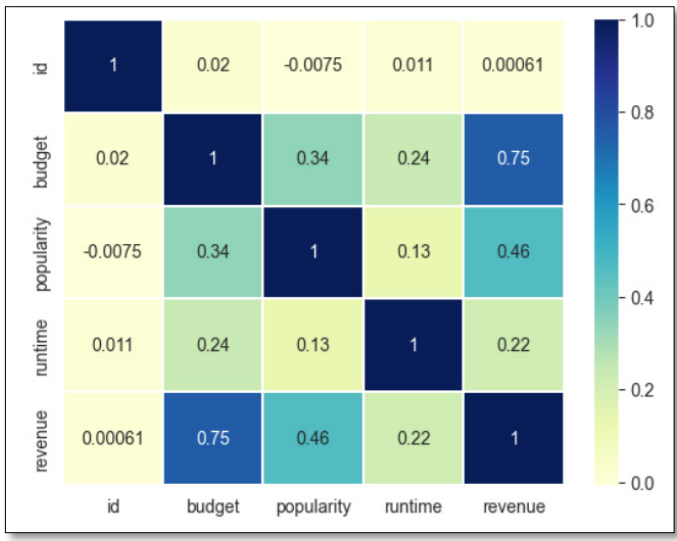


Рисунок 1 – Кореляції ознак кінофільмів.

З рисунка 1 видно що кореляція доходу за бюджетом становить 0.75 одиниць, а кореляція між доходом і хронометражем фільму становить 0.22 одиниць. Це пояснює, що дохід тісно пов’язаний із бюджетом фільму, також з тривалість фільму.

Наступний кроком в дослідженні даних - побудова діаграми розсіювання доходу в порівнянні бюджету кінофільмів. Результат зображено на рисунку 2.

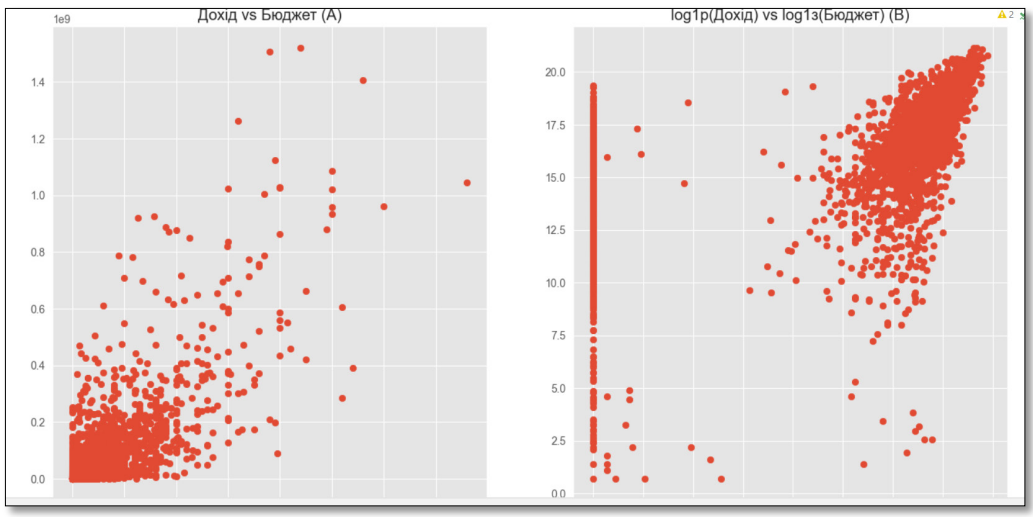


Рисунок 2 – Графік співвідношення доходу кінофільму до його бюджету.

На графіку 1 (A) видно, що бюджет та дохід має певну кореляція. Також можна побачити, що в наборі даних є 815 фільмів з нульовим бюджетом.

Важливою ознакою успішності кінофільму є наявність домашньої сторінки фільму, де користувачі можуть дізнатись усю необхідно інформацію про кінофільм та слідкувати за оновленнями. На рисунку 3 зображено статистику доходу фільмів з домашньою сторінкою та без неї.

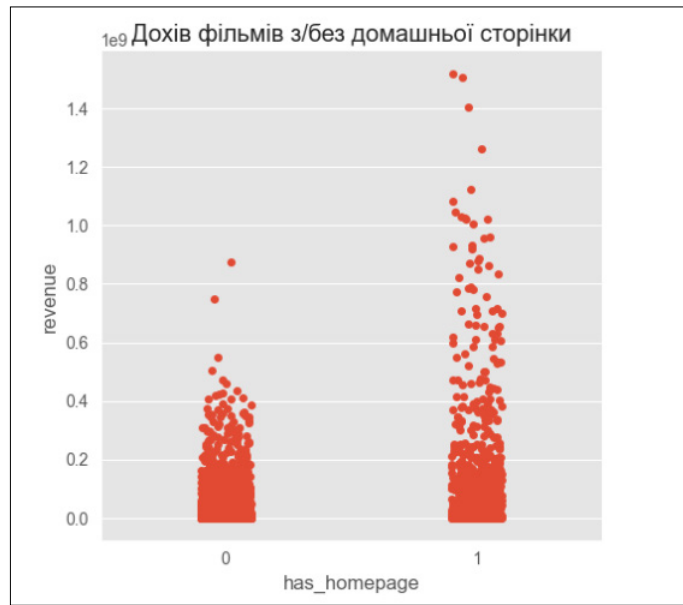


Рисунок 3 – Графік доходу фільмів з домашньою сторінкою і без неї.

З рисунка 3 видно, що фільми, які мають домашню сторінку, мають більший дохід порівняно з фільмами, що не мають домашньої сторінки.

На рисунку 2.8 зображено графік розподілу доходів між різними мовами фільмів.

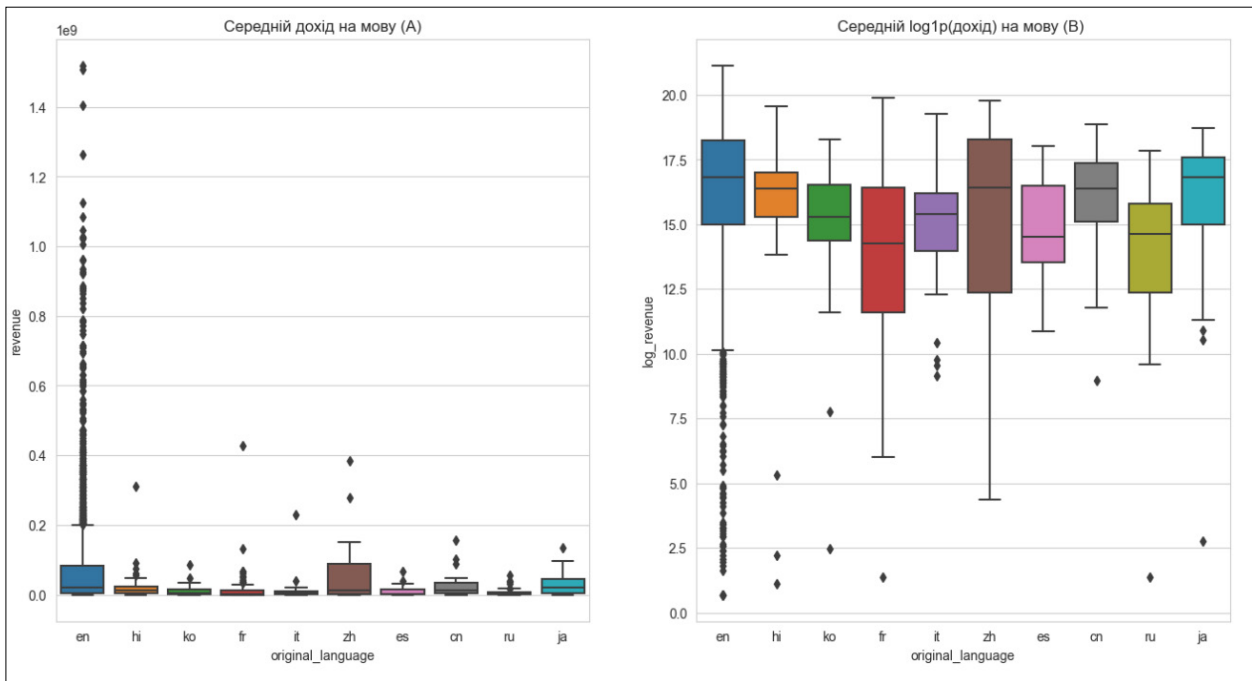


Рисунок 4 - Графік розподілу доходів для фільмів на різних мовах.

З рисунку 4 (A), що англійська мова має значно більший дохід порівняно з іншими мовами. Проте після логарифмічно трансформування, рисунок 4(B), видно, що інші мови також створюють дохід порівняний з англійською мовою. Однак саме англійські фільми лідирують на діаграмі доходів.

Отже, проаналізовано ознаки, що найбільше впливають на майбутній дохід кінофільму. Серед ознак можна виділити, ті що найбільше впливають на дохід кінофільму:

- **бюджет**, сильно впливає на кінцевий дохід фільму, проте якщо брати пропорційне співвідношення бюджету до доходу, то бюджет має менший вплив на дохід, хоча залишається все ще однією з найважливіших ознак;
- кінофільм з **головною сторінкою** в інтернеті чи соціальних мережах мають значно більші доходи;
- кінофільми на англійській мові збирають більше доходів в прокаті, порівняно з іншими мовами.

В подальшому дані дослідження та отримані показників кореляції ознак кінофільмів до його доходу можна використовувати для створення математичних моделей чи моделей нейронних мереж для прогнозування успішності кінофільму.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Статистика доходів кінофільмів [Електронний ресурс]. – Режим доступу: <https://www.boxofficemojo.com/year/>
2. Офіційна сторінка набору даних *TMDB* [Електронний ресурс]. – Режим доступу: <https://www.kaggle.com/deepmatrix/imdb-5000-movie-dataset>
3. Офіційна сторінка *Python* [Електронний ресурс]. – Режим доступу: <https://www.python.org/>
4. Офіційна сторінка *Seaborn* [Електронний ресурс]. – Режим доступу: <https://seaborn.pydata.org/>
5. Офіційна сторінка *Matplotlib* [Електронний ресурс]. – Режим доступу: <https://matplotlib.org/>
6. Офіційна сторінка *Plotly* [Електронний ресурс]. – Режим доступу: <https://plotly.com/>

Борисюк Володимир Миколайович – студент групи 2КН-21М, факультет інформаційних технологій та комп'ютерної інженерії, Вінницький національний технічний університет, Вінниця, e-mail: volodymyr.borysiuk0@gmail.com

Барабан Сергій Володимирович – к. т. н., доцент, доцент кафедри комп'ютерних наук, Вінницький національний технічний університет, м. Вінниця.

Volodymyr Borysiuk – student of the Faculty of Information Technologies and Computer Engineering, group 2CS-21M, Vinnytsia National Technical University, Vinnytsia, e-mail: volodymyr.borysiuk0@gmail.com

Sergiy Baraban – Cand. Sc., Assistant Professor of the Chair of Computer Science, Vinnytsia National Technical University, Vinnytsia.