3. Чайка П. Нейронні мережі: пояснення простою і зрозумілою мовою. URL: https://www.poznavayka.org/uk/nauka-i-tehnika-2/neyronni-merezhi-yih-zastosuvannya-robota/

4. Кетеринич Л. О. Прикладне застосування нейронних мереж. URL: http://csc.knu.ua/media/study/master-degree/courses-2020-202x/application-of-neural-networks.pdf

5. Кравченко В. Що таке нейронні мережі та як вони працюють. URL: https://livingfo.com/shcho-take-nejronni-merezhi-ta-iak-vony-pratsiuiut/

6. Добровська Л. М., Добровська І. А. Теорія та практика нейронних мереж. URL: https://ela.kpi.ua/bitstream/123456789/49841/1/Teoriia_praktyka_neironnykh_merezh.pd

7. FutureNow. Що таке нейронна мережа: простими словами. URL: https://futurenow.com.ua/shho-take-nejronna-merezha-prostymy-slovamy/

УДК 004.67

Бойко У. В., здобувач 2 курсу ОС «Магістр» спеціальності 122 Комп'ютерні науки

ВИЯВЛЕННЯ ФАКТОРІВ, ЩО ВПЛИВАЮТЬ НА РЕЙТИНГ І КАСОВІ ЗБОРИ ФІЛЬМУ МЕТОДАМИ DATA SCIENCE

Донецький національний університет імені Василя Стуса, м. Вінниця

Сучасна кіноіндустрія постійно шукає ключі до розуміння успіху фільмів. Ця робота фокусується на використанні методів Data Science для виявлення факторів, що впливають на рейтинг і касові збори фільмів.

У цьому досліджені було проведено аналіз набору даних «*tmdb_movies.csv*». Цей датасет містить широкий спектр інформації про різні фільми, що дає змогу провести глибокий аналіз та визначити показники, які впливають на касовий і глядацький успіх фільмів.

Дослідження базується на застосуванні інструментів програмування R [1], середовища RStudio [2] та пакету Tidyverse для детального аналізу даних про кінофільми. Робота включає вибірку даних, їх підготовку, очищення та аналіз, забезпечуючи об'єктивність і точність результатів.

На рисунку 1 зображено кількісні зведення за кожною змінною в очищеному наборі даних.

Робота включає всебічний статистичний аналіз даних про кінофільми, зокрема описову статистику, кореляційний аналіз та візуалізацію.

На рисунку 2 зображена кореляційна матриця, на якій можна побачити зв'язки між змінними.

```
     budget              revenue            original_title        runtime
 Min.   :     7000   Min.   :     13308   Length:3801        Min.   : 26.0
 1st Qu.: 10000000   1st Qu.:  14460000   Class :character   1st Qu.: 96.0
 Median : 25000000   Median :  46236000   Mode  :character   Median :106.0
 Mean   : 37609526   Mean   : 109155433                      Mean   :109.4
 3rd Qu.: 50000000   3rd Qu.: 126216940                      3rd Qu.:119.0
 Max.   :425000000   Max.   :2781505847                      Max.   :338.0
    genres            vote_count       vote_average     release_year
 Length:3801        Min.   :  10.0   Min.   :2.200   Min.   :1960
 Class :character   1st Qu.:  73.0   1st Qu.:5.700   1st Qu.:1995
 Mode  :character   Median : 207.0   Median :6.200   Median :2004
                    Mean   : 533.8   Mean   :6.174   Mean   :2001
                    3rd Qu.: 584.0   3rd Qu.:6.700   3rd Qu.:2010
                    Max.   :9767.0   Max.   :8.400   Max.   :2015
```

Рисунок 1. Кількісні зведення за кожною змінною

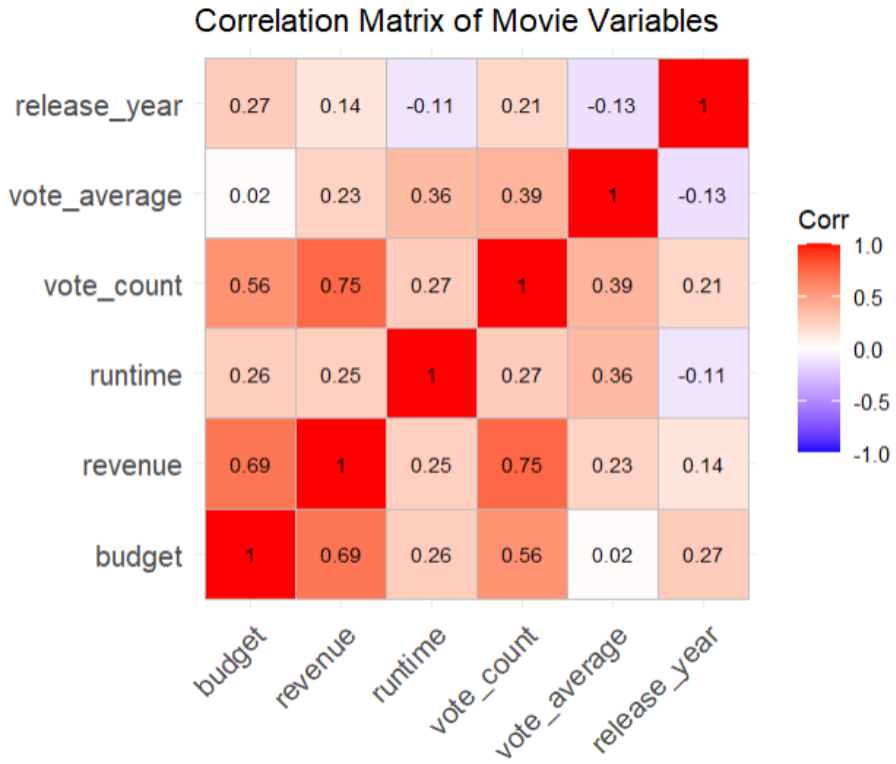

Рисунок 2. Кореляційна матриця

Використано багатофакторну лінійну регресію для дослідження впливу різних змінних, як-от бюджет, рік випуску, кількість голосів оцінки, тривалість і жанр, на рейтинг і касові збори фільмів. Для зручності дослідження впливу жанрів їх було розбито на окремі бінарні змінні. Виявлено значимі фактори, які впливають на успіх фільму.

Регресійна модель для передбачення касових зборів показала, що ключові фактори, як-от бюджет, тривалість фільму і кількість голосів, мають міцний позитивний зв’язок з касовими зборами. Певні жанри, як-от анімація та пригоди, мають статистично значимий позитивний вплив на касовий успіх.

Регресійна модель для передбачення рейтингу показала, що фільми з більшою кількістю голосів мають тенденцію до вищих оцінок, що може відображати схильність популярних фільмів мати вищий рейтинг. Тривалість фільму також має позитивний вплив на рейтинг. Це може бути пов'язано з тим, що довші фільми часто є більш глибокими або детальними в розкритті сюжету і персонажів. Серед жанрів анімація, документальні фільми, драми мають позитивний вплив, тоді як жанри, як-от наукова фантастика, пригоди, комедія, трилери, фентезі і жахи вказують на негативний вплив на середню оцінку.

У майбутніх дослідженнях можна додатково дослідити вплив акторського складу, репутації режисера, сезонності випуску фільму та ін. на його рейтинг та комерційний успіх. На основі проведеного аналізу можна зробити певні рекомендації:

1. Продюсерам та інвесторам треба ретельно планувати бюджет, зважаючи на його вплив на потенційні збори та рейтинг.

2. Важливе інвестування в маркетинг для збільшення кількості голосів та популярності фільму, оскільки це має великий вплив на досліджувані показники.

3. Треба забезпечити оптимальну тривалість фільму, яка забезпечуватиме глибоку розповідь, але не буде втомлювати аудиторію.

Розробники фільмів могли б зосередитися на жанрах, які історично показують високі рейтинги та касові збори.

Список використаних джерел

1. R Programming Language – Introduction. *GeeksforGeeks*. 2021. URL: https://www.geeksforgeeks.org/r-programming-language-introduction/

2. What Is RStudio? A Beginner's Guide. *CareerFoundry*. 2023. URL: https://careerfoundry.com/en/blog/data-analytics/what-is-rstudio/

3. Wickham H. Ggplot2: Elegant Graphics for Data Analysis. *New York: Springer International Publishing*, 2016. 260 с.

4. TMDb. *Вікіпедія*. 2023. URL: https://uk.wikipedia.org/wiki/TMDb

5. Hiller W. What Is Data Cleaning and Why Does It Matter? *CareerFoundry*. 2023. URL: https://careerfoundry.com/en/blog/data-analytics/what-is-data-cleaning/