Вінницький національний технічний університет
(повне найменування вищого навчального закладу)

Факультет інтелектуальних інформаційних технологій та автоматизації
(повне найменування інституту, назва факультету (відділення))

Кафедра комп'ютерних наук
(повна назва кафедри (предметної, циклової комісії))

МАГІСТЕРСЬКА КВАЛІФІКАЦІЙНА РОБОТА

на тему:

«Інформаційна технологія прогнозування успішності кінофільму»

Виконав: студент 2-го курсу, групи 2КН-21м
спеціальності 122 «Комп'ютерні науки»
(шифр і назва напряму підготовки, спеціальності)
Борисюк В. М.
(прізвище та ініціали)

Керівник: к.т.н., доцент каф. КН
Барабан С. В.
(прізвище та ініціали)
«15» 12 2022 р.

Опонент: д.т.н. каф. КСУ
Юхимчук М. С
(прізвище та ініціали)
«15» 12 2022 р.

Допущено до захисту
Завідувач кафедри КН
д.т.н., проф. Яровий А.А.
(прізвище та ініціали)
«16» 12 2022 р.

Вінниця ВНТУ - 2022 рік

Вінницький національний технічний університет
Факультет інтелектуальних інформаційних технологій та автоматизації
Кафедра комп'ютерних наук
Рівень вищої освіти ІІ-й (магістерський)
Галузь знань – 12 "Інформаційні технології"
Спеціальність – 122 "Комп'ютерні науки"
Освітньо-професійна програма – "Системи штучного інтелекту"

ЗАТВЕРДЖУЮ
Завідувач кафедри КН
Д.т.н., проф. Яровий А.А.
14.09 **2022 року**

З А В Д А Н Я
НА МАГІСТЕРСЬКУ КВАЛІФІКАЦІЙНУ РОБОТУ СТУДЕНТУ

Борисюку Володимиру Миколайовичу
(прізвище, ім'я, по батькові)

1. Тема роботи: Інформаційна технологія прогнозування успішності кінофільму керівник роботи к.т.н., доцент кафедри КН Барабан С. В.

затверджені наказом вищого навчального закладу від "14" 09 2022 року № 203

2. Строк подання студентом роботи 18 листопада 2022 року

3. Вихідні дані до роботи:
навчальна вибірка – 3000 кінофільмів, тестова вибірка - 600 кінофільмів; мова програмування – об'єктно-орієнтована.

4. Зміст текстової частини:
Вступ, аналіз предметної області прогнозування успішності кінофільму, розробка моделі прогнозування успішності кінофільму, розробка інформаційної технології прогнозування успішності кінофільму, програмна реалізація інформаційної технології прогнозування успішності кінофільму, економічна частина, висновки, перелік використаних джерел, додатки.

5. Перелік ілюстративного матеріалу (з точним зазначенням обов'язкових креслень)

Графік кореляції ознак кінофільму до його доходу, графік розподілу доходу кінофільмів з домашньою і без неї, графіки розподілу доходів для фільмів на різних мовах, загальна структура програмного забезпечення прогнозування успішності фільмів, загальна схема бази-даних кінофільмів, схема алгоритму модуля взаємодії з базою даних, схема алгоритму модуля імпортування фільмів з

tmdb, схема алгоритму модуля прогнозування успішності кінофільму, вікно програми з результатом прогнозування успішності кінофільму.

6. Консультанти розділів роботи

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	виконання прийняв
1-4	Барабан С. В., к.т.н., доц. каф. КН	14.09.2022	14.12.2022
5	Бурєннікова Н. В., д. е. н., проф. каф. ЕПВМ	14.09.2022	12.12.2022

7. Дата видачі завдання 14.09 2022 року

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів магістерської кваліфікаційної роботи	Строк виконання етапів роботи
1	Аналіз сучасних підходів прогнозування успішності кінофільмів.	12.08.2022 - 01.09.2022
2	Побудова моделей прогнозування успішності кінофільму.	01.09.2022 - 15.09.2022
3	Практичне застосування та оцінка ефективності розроблених моделей	15.09.2022 – 20.11.2022
4	Підготовка економічної частини	20.11.2022 - 29.11.2022
5	Апробація та/або впровадження результатів дослідження	29.11.2022 - 10.12.2022
6	Оформлення пояснювальної записки, графічного матеріалу та презентації	10.12.2022 - 16.12.2022

Студент _____ (підпис) Борисюк В. М.

Керівник роботи _____ (підпис) Барабан С. В.

АНОТАЦІЯ

УДК 004.8

Борисюк В. М. Інформаційна технологія прогнозування успішності кінофільму. Магістерська кваліфікаційна робота зі спеціальності 122 – комп’ютерні науки, освітня програма - комп’ютерні науки. Вінниця: ВНТУ, 2022. 103 с.

На укр. мові. Бібліогр.: 28 назв; рис. 40; табл. 26.

У даній магістерській кваліфікаційній роботі на основі проведеного аналізу розроблено програмного забезпечення для прогнозування успішності кінофільму, використовуючи нейромережевий підхід. Запропоновано власну структуру нейронної мережі. Розроблено інформаційну технологія прогнозування успішності кінофільму. Реалізоване програмне забезпечення для прогнозування успішності кінофільму. Серверна частина реалізована на мові програмування Python у програмному середовищі PyCharm, клієнтська частина реалізована у вигляді додатку на мові програмування Dart, з використанням фреймворку Flutter. Розроблений продукт характеризується відсутністю прямих аналогів, наявність графічного інтерфейсу користувача та підвищеною точністю прогнозування успішності кінофільму.

Графічна частина складається з 7 плакатів.

У економічному розділі при оцінюванні за технічними параметрами, згідно узагальненого коефіцієнту якості розробки, визначено, що науково-технічна розробка переважає існуючі аналоги приблизно в 1,56 рази. Також термін окупності становить 0,96 р., що менше 3-х років, що свідчить про комерційну привабливість науково-технічної розробки

Ключові слова: прогнозування, штучний інтелект, нейроні мережі, регресія, *python*.

ABSTRACT

Borysiuk V. M. Information technology for the signature dynamic invariant features formation. Master's thesis in the speciality 122 - Computer Sciences, educational program - Computer Sciences. Vinnytsia: VNTU, 2022. 103 p.

In Ukrainian language. Bibliogr.: 28 titles; fig. 40; table 26.

In this master's thesis, based on the analysis, software was developed for predicting the success of a movie using a neural network approach. A proprietary structure of a neural network is proposed. An information technology for predicting the success of a motion picture has been developed. Implemented software for predicting the success of a motion picture. The server part is implemented in the Python programming language in the PyCharm programming environment, the client part is implemented as an application in the Dart programming language, using the Flutter framework. The developed product has the following advantages: the absence of direct analogues, the presence of a graphical user interface and increased accuracy in predicting the success of a motion picture.

The graphic part consists of 7 posters.

In the economic section, when evaluating according to technical parameters, according to the generalized coefficient of development quality, it was determined that scientific and technical development prevails over existing analogues by approximately 1.86 times. Also, the payback period is 0.96 years, which is less than 3 years, which indicates the commercial attractiveness of scientific and technical development.

Keywords: prediction, artificial intelligence, neural networks, regression, python.

ЗМІСТ

ВСТУП

Актуальність. Кіноіндустрія відіграє дуже важливу роль у нашому житті, вона має вирішальний вплив на суспільство, розширює наші знання про історію та культуру, надихає на розважає. Фільми є високопрофесійною багатомільярдною індустрією з сотнями нових фільми, створені щороку.

Основні складності процесу передбачення успішності кінофільму:

- визначити методи попередньої обробки наборів даних;
- які ознаки є найбільше корисними;
- оцінка показників успішності моделі передбачення.

Для того щоб вирішувати складні і погано формалізовані завдання і виник напрямок, який називається штучні нейронні мережі. Штучні нейронні мережі складаються з нейроноподібних елементів, з'єднаних між собою в мережу. Нейронні мережі знайшли застосування практично.

Якщо можна за допомогою комп'ютера передбачити, наскільки вдалим буде фільм, навіть до його релізу, це був би потужний інструмент для використання. Починаючий голлівудський режисер або кіностудія з деякими технічними навичками могли б передбачити, чи їхня ідея стане успішною і чи це є безпечною інвестицією.

Зв'язок роботи з науковими програмами, планами, темами. Магістерська робота виконана відповідно до напрямку наукових досліджень кафедри комп'ютерних наук Вінницького національного технічного університету 22 К1 «Моделі, методи, технології та пристрої інтелектуальних інформаційних систем управління, економіки, навчання та комунікацій» та плану наукової та навчально-методичної роботи кафедри.

Мета і завдання досліджень. Метою магістерської кваліфікаційної роботи є підвищення точності прогнозування успішності кінофільму програмними засобами за рахунок використання штучних нейронних мереж.

Для досягнення мети розробки необхідно виконати такі задачі:

− провести аналіз предметної області передбачення успішності кінофільмів;

− розглянути існуючі методи передбачення успішності кінофільмів та обрати й обґрунтувати вибір методу, який задовольняє мету даної магістерської кваліфікаційної роботи;

− розробити математичну модель передбачення успішності кінофільмів;

− сформулювати стадії інформаційної технології, розробити структуру та алгоритм роботи програмного засобу;

− виконати програмну реалізацію запропонованої інформаційної технології;

− провести тестування програмного продукту та виконати аналіз отриманих результатів.

Об’єкт дослідження – процес прогнозування успішності кінофільмів комп’ютерними засобами з використанням нейронних мереж.

Предмет дослідження – інформаційна технологія та програмні засоби прогнозування успішності кінофільмів з використанням нейронних мереж та достовірність їх роботи.

Методи дослідження. У роботі використані наступні методи наукових досліджень: системного аналізу, теорії штучних нейронних мереж для реалізації інформаційної технології, методи математичної статистики для розробки процесу розв'язання задачі нейромережевого передбачення успішності кінофільмів та обрахунків результатів експериментів із програмним засобом, об’єктно-орієнтованого програмування.

Наукова новизна одержаних результатів.

Набула подальшого розвитку інформаційна технологія передбачення успішності кінофільмів, в якій на відміно від існуючих використовується нейрона мережа, що дозволяє підвищити точність прогнозування.

Практичне значення одержаних результатів полягає в тому, що на основі проведених досліджень розроблено програмне забезпечення для передбачення успішності кінофільмів.

Запропонована інформаційна технологія сприяє підвищенню точності програмних засобів передбачення успішності кінофільмів, зокрема:

- розроблено алгоритм роботи програмного передбачення успішності кінофільмів;
- розроблено програмні засоби для передбачення успішності кінофільмів.

Достовірність теоретичних положень магістерської кваліфікаційної роботи підтверджується коректністю постановки завдання, коректністю використання математичного апарату методів дослідження, експериментальними дослідженнями, тестування програмної реалізації інформаційної технології розпізнавання музичних патерів. Адекватність розроблених математичних моделей підтверджується результатами експериментальних досліджень.

Особистий внесок здобувача. Усі результати, наведені у магістерській кваліфікаційній роботі, отримані самостійно. У працях, написаних у співавторстві, здобувачу належать: виділення ознак кінофільму, що впливають на його успішність [1].

Апробація результатів роботи. Результати досліджень апробовані на конференції «Молодь в науці: дослідження, проблеми, перспективи-2023», Вінниця, 15 листопада 2022 – 12 травня 2023 року.

Публікації. За результатами досліджень опубліковано одні тези доповіді на науково-технічній конференції [1] та подано заяву на авторське право на програмне забезпечення.

1 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ ПРОГНОЗУВАННЯ УСПІШНОСТІ КІНОФІЛЬМУ

1.1 Постановка задачі прогнозування успішності кінофільму

Прогнозування рейтингу фільму може допомогти людям визначити, який фільм потрібно дивитися. Експерти кіноіндустрії сходяться на думці, що прибуток є ключовим фактором успіху кіно і допомагає кіно виробникам та інвесторам досягти фінансового успіху [2]. Компанії можуть побачити, які фільми, ймовірно, мають хороші рейтинги, і розробити стратегії розробки фільму для збільшення прибутку, наприклад, випускаючи товари для фільмів або створюючи події та рекламні акції, пов’язані з фільмом.

Крім того, використовуючи історичні значення, отримані з раніше випущених фільмів, можна прогнозувати прибуток до того, як фільм буде знятий. Щоб уникнути втрат, компанії-виробники фільмів можуть складати стратегічні плани та приймати рішення щодо виходу фільму в прокат.

Прогнозування касових зборів конкретного фільму привернуло багатьох вчених, тому що це передбачення є важкою та складною проблемою. Часто кіноіндустрія залишає у людей враження прибуткової сфери. Цьому враженню сприяють образи знаменитостей та валовий прибуток, який вимірюється сотнями мільйонів доларів. Однак більшість людей лише звертають увагу до найуспішніших фільмів, які, як правило, приносять певний прибуток, але в цілому це враження невірне. За статистикою, з будь-яких десяти великих кінофільмів, створених у середньому, може бути шість-сім кінофільмів характеризується як збитковий [3]. Ці цифри свідчать що кіноіндустрія є одним із найризикованіших ринків індустрії розваг, що виправдовує високу віддачу від успішних фільмів. Це через ці високі ризики у створенні фільмів, які складають адекватний і точний бюджетний план прогнозування доходів стає дуже важливим. Хоча є кілька досліджень щодо прогнозування рейтингу фільмів. Ефективність прогнозування рейтингу перед виходом фільму в

прокат все ще потрібно підвищити. Це дослідження зосереджено на прогнозуванні рейтингу на основі даних, які можна отримати до виходу фільму в прокат. Атрибути метаданих, отримані з IMDb і TMDb, були використані як предикатори. Для створення функцій використовуються такі атрибути, як сюжет, артисти, персонал, режисер, жанр, рейтинг вмісту та рейтинг.

Однією з унікальних рис, які використовуються в цьому дослідженні, є історичні цінності. Ці функції були створені на основі зв'язку між фільмом і раніше випущеними фільмами. Передбачається, що прогнозований рейтинг, заснований на цих історичних значеннях, безумовно, є більш об'єктивним, ніж рейтинг від аудиторії, який з'явився, коли фільм щойно вийшов у прокат. Метод когортного оцінювання шукає фільми, які мають схожість на основі наявних історичних атрибутів і значень, і робить прогноз касового прибутку на основі цих подібностей.

1.2 Огляд відомих методів прогнозування успішності кінофільму

Модель дифузії Басса. Однією з перших моделей поведінки споживачів, які можна застосувати для моделювання касових зборів фільму, є модель Баса [4]. У моделі Басса використовується звичайний диференціал

Рівняння, яке пов'язує, як нові продукти вживаються в популяції:

$$\frac{f(t)}{1-F(t)} = p + qF(t), \tag{1.1}$$

де $f(t)$ – швидкість зміни встановленої базової фракції; $F(t)$ – встановлена базова частка; p – коефіцієнт інноваційності; q – коефіцієнт обмеження.

Розв'язком цієї системи дає відому S-криву (Рисунок 1.1), яка

представляє швидкість прийняття продукту з часом. Ця модель була однією з найвпливовіших і поширених цитованих робіт з історії науки управління.

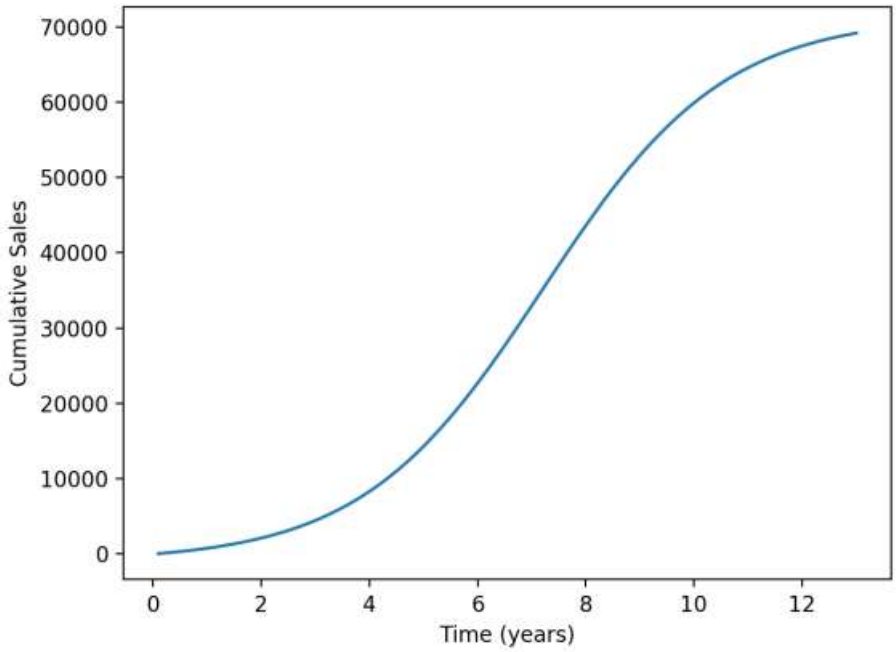


Рисунок 1.1 - S-крива, що показує загальні продажі за період часу

У роботі «*A Parsimonious Model for Forecasting Gross Box-Office Revenues of Motion Pictures*» [5] представлена модель, що використовує експоненційний розподіл з двома параметрами та три параметри гамма-розподілу, яка моделює процес, за допомогою якого людина вирішить подивитися фільм у два етапи. По-перше, який час, необхідний для прийняття рішення про те, чи варто чи ні подивитися фільм. По-друге, змоделювати час, необхідний для виконання цього рішення. У роботі «*The motion picture industry: critical issues in practice, current research, and new research directions*» [6] розробили програму для прогнозування касових зборів для фільму, використовуючи лише дані перед релізом. Зокрема, дана програма здатна передбачити касові збори для фільму на основі моделювання поведінки споживачів за допомогою інтерактивних ланцюгів Маркова [7], де ймовірності переходу залежать від кількості людей, які вже перебувають в інших регіонах.

Модель Едвардса-Бакмайра. Едвардс і Бакмайр розробили систему поєднаних звичайних диференціальних рівняння для моделювання касових зборів фільму після випуску [8]. Модель Едвардса-Бакмайра використовує детермінований підхід, використовуючи керуючі рівняння зі швидкістю зміни доходу в момент t, спочатку заданою:

$$\frac{d\tilde{G}}{dt} = \tilde{S}\tilde{A}, \tag{1.2}$$

$$\frac{d\tilde{A}}{dt} = -a_A\tilde{A}, \tag{1.3}$$

$$\frac{d\tilde{S}}{dt} = -a_S\tilde{S} \tag{1.4}$$

де, $\tilde{G}, \tilde{A}$ та $\tilde{S}$ - прибуток, кількість зароблених грошей за показ за тиждень і кількість показів на якому представлений фільм, відповідно. P - це ціна квитка. Умови включають $\tilde{G}(0) = 0$ та $\tilde{G}(\infty) = \int_0^\infty \tilde{S}\,\tilde{A}dt$.

Модель Едвардса-Бакмайра розвивається далі, намагаючись моделювати негативну реакцію людини ($\tilde{H}$, с відсоток людей, які ненавидять фільм, наданий $H_{\%}$) до фільму, яким керує:

$$\frac{d\tilde{H}}{dt} = \frac{H_{\%}\tilde{S}\tilde{A}}{P} \tag{1.5}$$

Це передбачає попередні знання про загальну кількість людей, які будуть дивитися фільм, і знання кількості людей, яким він не подобається, що недоцільно в реальному світі. Модель Едвардса-Бакмайра додатково розвивається, дозволяючи людям дивитися фільм кілька разів, додаючи параметри для врахування жанру, суми, витраченої на рекламу, та ефективності реклами. Більшість з цих параметрів необхідно оцінювати в

режимі реального часу.

Модель Едвардса-Бакмайра дала багатообіцяючі результати для ряду фільмів, що зображено на рисунках 1.2, 1.3 та 1.4.

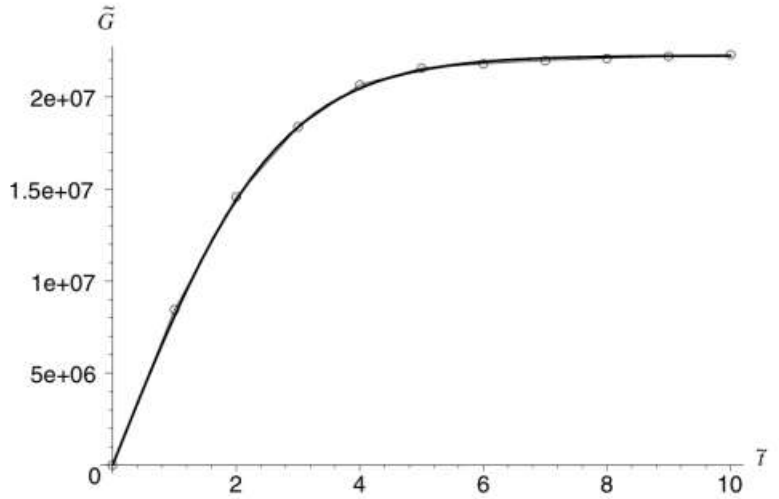


Рисунок 1.2 - Прибуток для фільму *At First Sight*. Крива передбачення

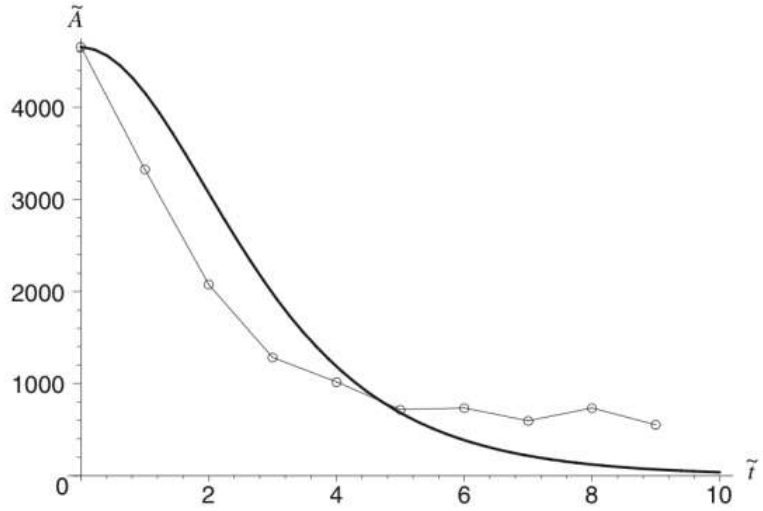


Рисунок 1.3 - Аудиторія для фільму *At First Sight*. Крива передбачення

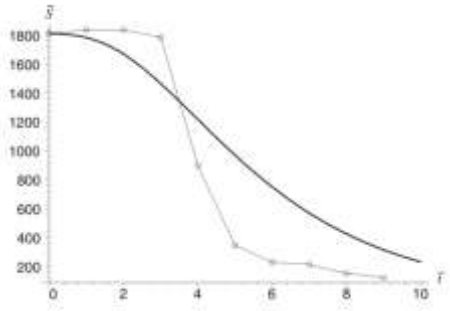


Рисунок 1.4 – Кількісит показів для фільму *At First Sight*. Крива передбачення

Методи стохастичної апроксимації. Касові збори фільмів — це випадковий процес. Прибуток фільму визначається кількість кінотеатрів, які демонструють фільм, і кількість людей, які йдуть дивитися його.

У цьому є випадковість, і стохастичні моделі намагаються змоделювати цю випадковість.

Стохастичні процеси можна моделювати за допомогою ланцюгів Маркова, де наступний крок залежить лише від поточного стану. Марківський процес безпам'ятний – він не знає, як було досягнуто поточного стану або в який час він був досягнутий. Наприклад, за n станів $S = \{s_1, s_2, \ldots, s_n\}$, марковський процес може переходити з одного стану в інший з ймовірністю P_{ij} (у процесі дискретного часу). На рисунку 1.5 зображено ланцюг Маркова.

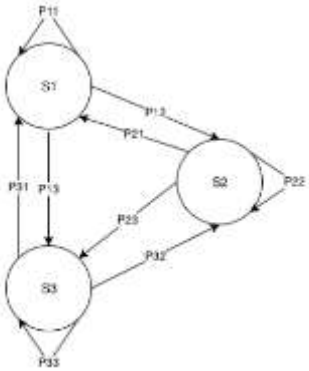


Рисунок 1.5 – Ланцюг Маркова з трьома станами

Ланцюги Маркова може моделювати поведінку доходів окремих

кінотеатрів, але перевести це на макроскопічний погляд на всю систему може бути складно. Моделювання кількох людей за допомогою ланцюгів Маркова призводить до проблеми вибуху простору стану [9], що може зробити цей підхід нежиттєздатним.

Підходячи до цього з детермінованої точки зору, моделюючи стохастичний процес за допомогою звичайних диференційних рівнянь. Це означає, що ми втрачаємо дискретні стани моделі ланцюга Маркова, але отримуємо модель, яку набагато легше моделювати, ніж кілька процесів Маркова. Однак ігнорування стохастичних ефектів може означати, що в деяких випадках рівняння створюють значну помилку.

Моделі машинного навчання. Машинне навчання (ML) — це підмножина галузі штучного інтелекту (AI), яка розробляє алгоритми, які можуть навчатися на даних. Контрольований ML включає модель, яка вивчає функцію з позначених навчальних даних.

У разі прогнозування продуктивності фільму історичні дані, що містять таку інформацію, як бюджет, мова тощо, можуть бути використані для навчання моделі. Функція Інженерія даних є важливою частиною розробки ефективної моделі ML. Конструювання ознак даних а є важливою частиною розробка ефективної моделі ML.

Поставлена тут проблема полягає в регресії, де вихідні дані будуть числом з дійсним значенням, наприклад, загальний касовий збір фільму.

Прогнозування касових зборів з певним успіхом розглядалося як проблема класифікації і є потенційно розумний підхід до проблеми, якщо регресія не може дати хороших результатів.

Лінійна регресія. Лінійний регресійний аналіз – це підхід до моделювання зв’язку між незалежними змінними та залежними змінними.

Якщо похідну неможливо обчислити, то ітераційне рішення можна знайти за допомогою градієнтного спуска або при використанні великого набору даних за допомогою стохастичного градієнтного спуску [10].

Нейронні мережі. Нейронна мережа (NN) — це система, натхненна структурою біологічних нейронних мереж. NN можуть бути структурованим так, щоб мати кілька шарів, кожен нейрон з'єднаний з кожним нейроном у наступному шару (крім вхідного шару). На рисунку 1.6 зображено схему нейронної мережі.

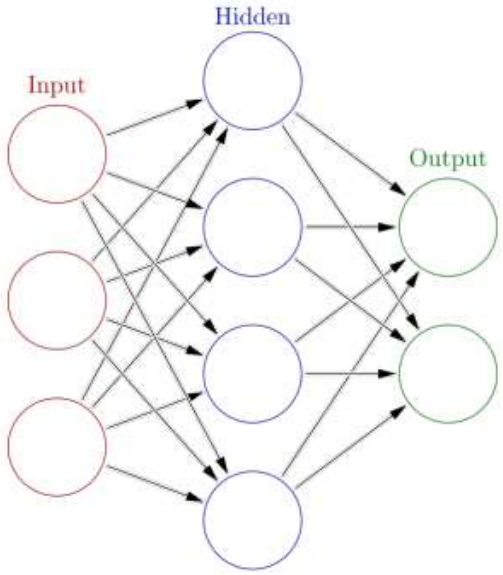


Рисунок 1.6 – Ілюстрація тришарової нейронної мережі

Кожен нейрон пов'язаний із вагою, який нейрон використовує разом зі своїми вхідними параметрами для обчислення вихідного значення. Функцію активації можна застосувати до суми зважених вхідних даних з попереднього шару. Застосована функція активації є гіперпараметром, який необхідно вибрати. Нейронні мережі підлаштовуються до навчального набору даних за допомогою алгоритму зворотного поширення, який може використовувати алгоритми оптимізації, такі як стохастичний градієнтний спуск, які надають додаткові параметри для оптимізації, включаючи швидкість навчання та розмір пакету.

1.3 Аналіз об’єкту проектування

Основною метою розробки інформаційної технології прогнозування успішності кінофільмів є надання кінокомпаніям інформації про успішність кінофільму, що ще знаходиться в розробці. Це дозволить компаніям скорегувати напрямок свої фільмів, щоб отримати максимальну віддачу від аудиторії.

Вимоги інформаційної технології:

- Зручний інтерфейс користувача.
- Наявність персонального кабінету.
- Можливість імпортувати власні дані про фільми.
- Актуальність – система повинна весь час наповнюватись актуальною інформацією, для передбачень.
- Швидкість надання передбачення.
- Інтерпритованість передбачень.

1.4 Висновок до розділу 1

У даному розділі було проведено аналіз предметної області прогнозування успішності кінофільму, а саме – сформульовано постановку задачі, проведено огляд відомих методів розв'язання задачі прогнозування успішності кінофільму.

Виходячи з аналізів предметної області, дало можливість визначити основні вимоги до системи прогнозування успішності кінофільму.

2 РОЗРОБКА МОДЕЛІ ПРОГНОЗУВАННЯ УСПІШНОСТІ КІНОФЛЬМУ

Методологія розробки інформаційної технології прогнозування успішності кінофільму:

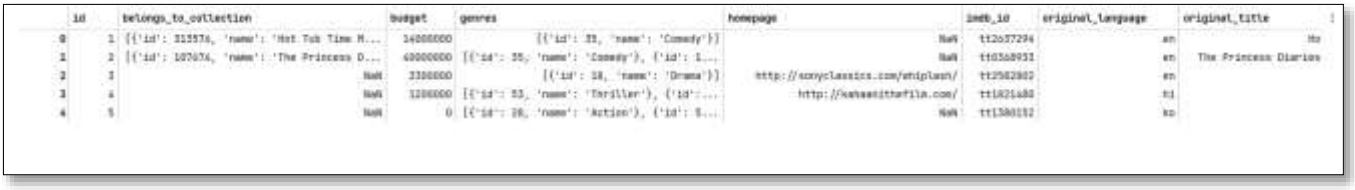
- збір даних з набору даних *TMDB;*
- дослідницький аналіз даних, конструювання ознак, візуалізація даних, дослідження кореляції ознак;
- моделювання експериментів для оцінки продуктивності та вибору методу машинного навчання;
- оцінка моделі на перевірочному наборі даних.

Усі дослідження проводяться з використанням наступних інструментів:

- Python;
- Jupyter;
- Pandas;
- Numpy;
- Scikit-learn, wordcount, elit, TFID.

2.1 Відокремлювання ознак кінофільмів для прогнозування успішності фільму

Для знаходження ознаків кінофільм, що впливають на успішність кінофільму використано набір даних *TMDB* [11]. *TMDB* – один із найбільших база даних про фільми, що містить інформацію про 568 729 кінофільмів та серіалів. На рисунку 2.1 зображено прикладу набору даних *TMDB.*



	id	belongs_to_collection	budget	genres	homepage	imdb_id	original_language	original_title
0	1	[{'id': 313576, 'name': 'Hot Tub Time M...	14000000	[{'id': 35, 'name': 'Comedy'}]	NaN	tt2637294	en	Ho
1	2	[{'id': 107674, 'name': 'The Princess D...	40000000	[{'id': 35, 'name': 'Comedy'}, {'id': 1...	NaN	tt0368933	en	The Princess Diaries
2	3	NaN	3300000	[{'id': 18, 'name': 'Drama'}]	http://sonyclassics.com/whiplash/	tt2582802	en	
3	4	NaN	1200000	[{'id': 53, 'name': 'Thriller'}, {'id':...	http://kahaanithefilm.com/	tt1821480	hi	
4	5	NaN	0	[{'id': 28, 'name': 'Action'}, {'id': 8...	NaN	tt1380152	ko	

Рисунок 2.1 – приклад набору даних *TMDB*

В таблиці 2.1 перелічено усі стовбці дата сету та їх тип даних.

Таблиця 2.1 – Стовбці набору даних

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

Продовження таблиці 2.1

Колонка	Тип даних	Опис
production_countries	json	Список країн виробників
release_date	date	Дата релізу
runtime	integer	Хронометраж в хвилинах
status	stiring	Статус
keywords	json	Ключові слова
cast	json	Список акторів
crew	json	Знімальна команда
revenue	integer	Прибуток

Кореляція ознак. Діапазон значень коефіцієнта кореляції становить від -1.0 до 1.0. Якщо кореляція дорівнює нулю, це означає, що між двома змінними немає лінійного зв’язку. Якщо кореляція наближається до 1, це означає що дві зміні сильно пов’язані між собою. Наприклад, ціна на нафту безпосередньо пов’язана з цінами на авіаквитки. Якщо значення кореляції наближається до -1, це означає що зміні негативно корельовані, тобто якщо одна зміна збільшується, інша зміна зменшується з таким самим співвідношенням і навпаки. На рисунку 2.2 зображено кореляцію ознак кінофільмів.

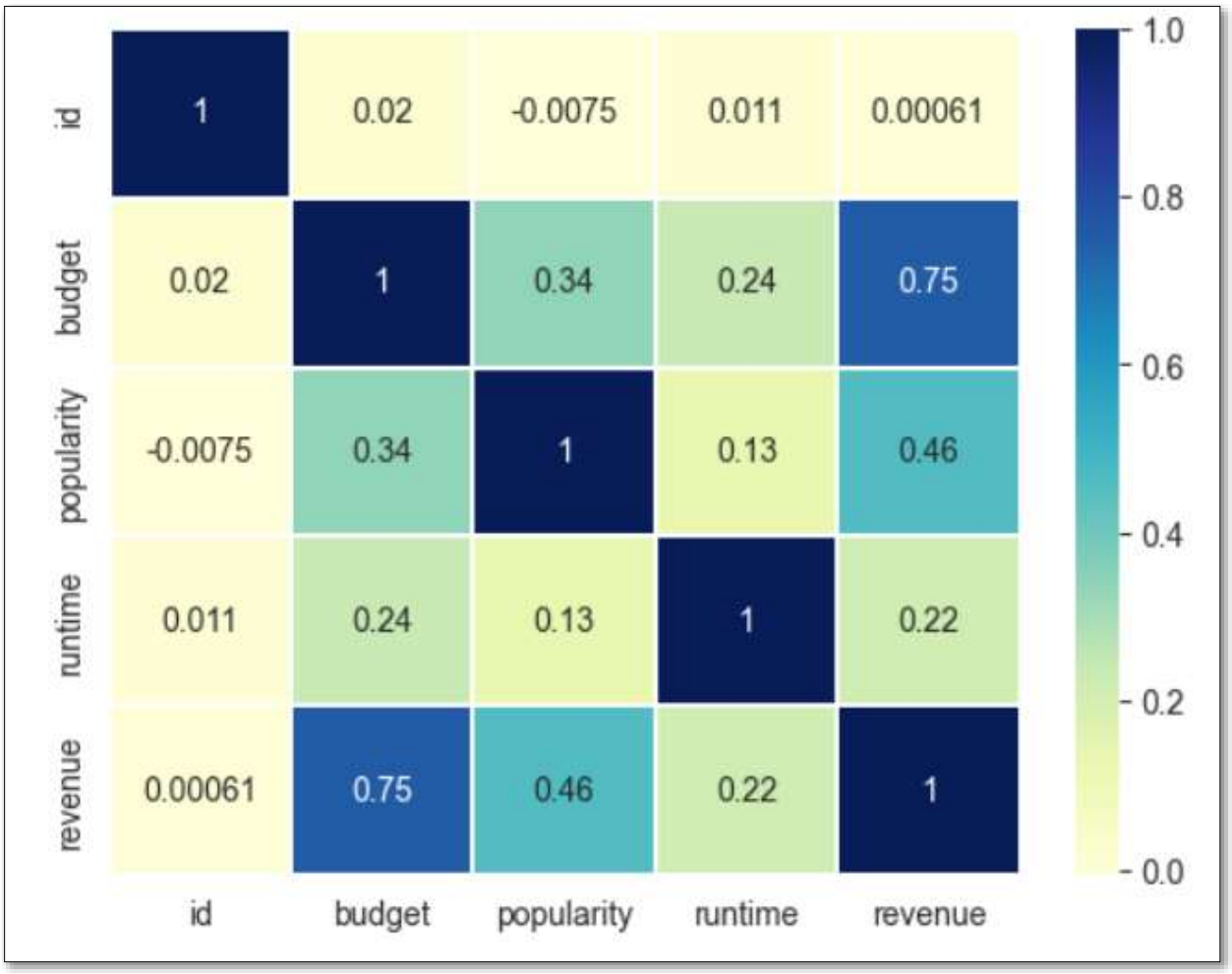


Рисунок 2.2 – Кореляція ознак кінофільмів

З рисунка 2.2 видно що кореляція доходу за бюджетом становить 0.75 одиниць, а кореляція між доходом і хронометражем фільму становить 0.22 одиниць. Це пояснює, що прибуток тісно пов’язаний із бюджетом фільму, також з доходом пов’язаний тривалість фільму.

На рисунку 2.3 зображено графік розподіл доходу.

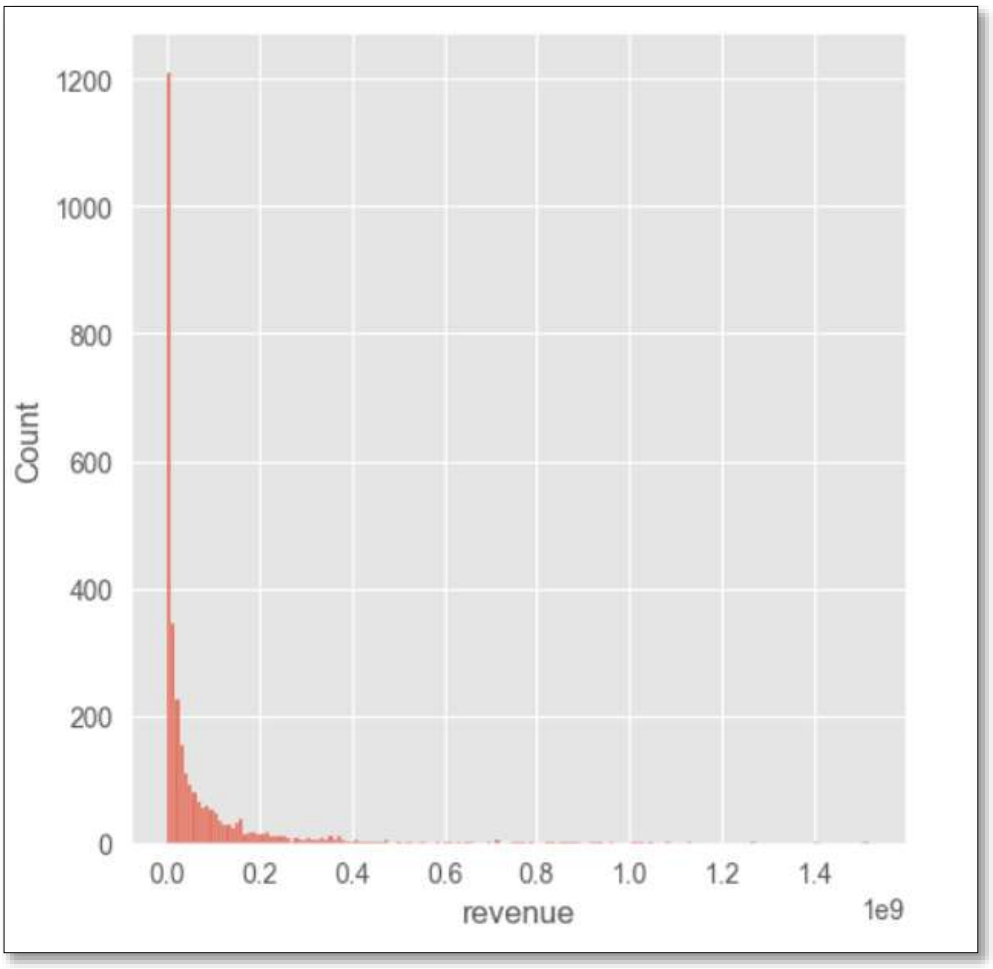


Рисунок 2.3 – Графік розподіл доходу

З рисунку 2.3 видно, що дані дуже спотворені, тому важко зробити висновки з графіку, тому потрібно нормалізувати дані. Для цього використовується логарифмічне трансформування, формула (1).

$$log1p = \log(1 + x) \tag{2.1}$$

2.1.1 Відношення бюджету кінофільму до доходу

На рисунку 2.4 зображено результат логарифмічного трансформування. Після трансформування дані стали нормально розподілені, що має меншу асиметрію та ексцес.

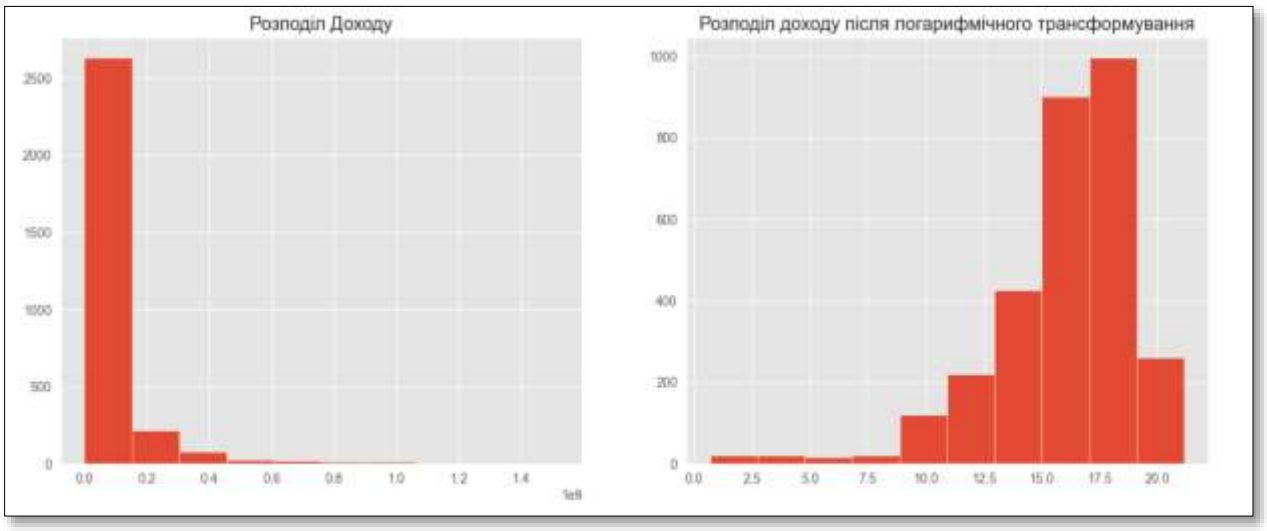


Рисунок 2.4 – Логарифмічне трансформування графіку доходу

На рисунку 2.5 зображено результат логарифмічного трансформування для бюджету.

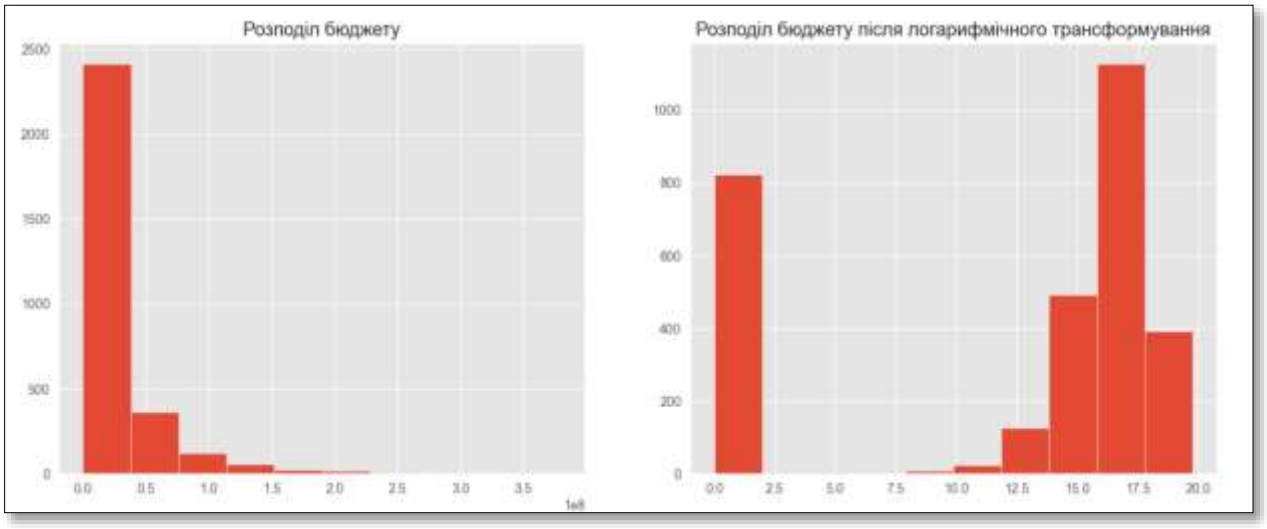


Рисунок 2.5 – розподіл бюджету

Наступний кроком в досліджені даних буде побудова діаграми розсіювання доходу.

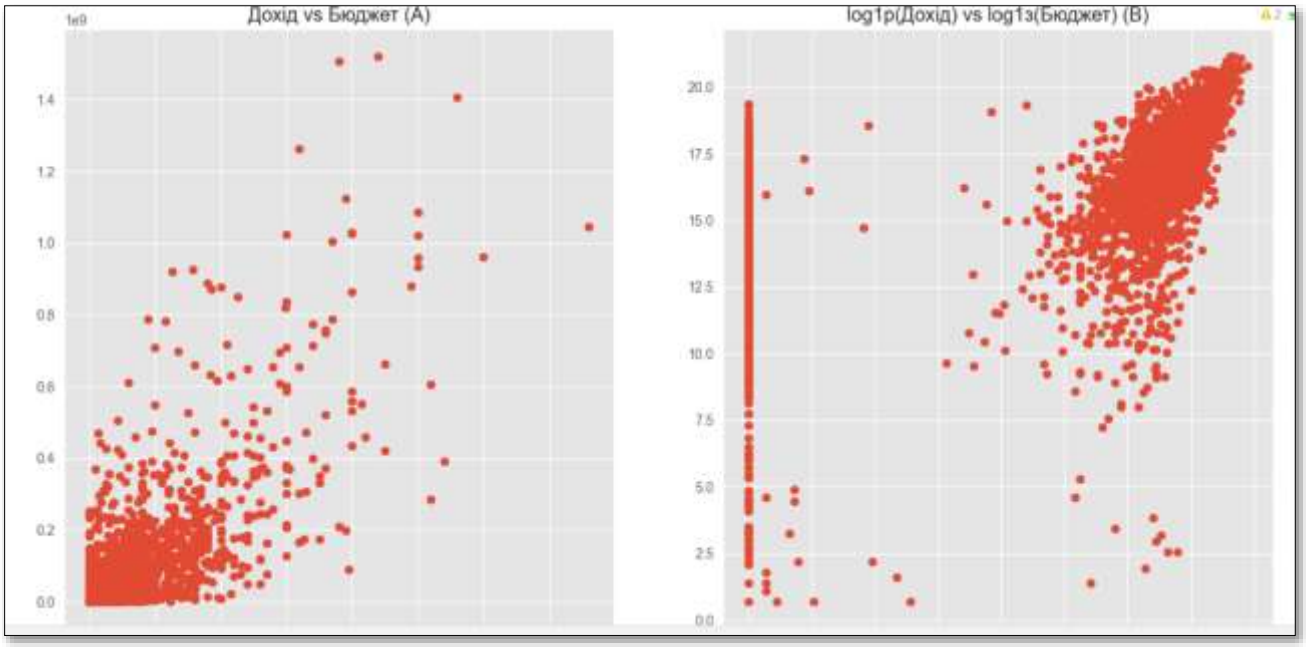


Рисунок 2.6 – Графік діаграми розсіювання бюджету проти доходу

На графіку 2.6 (А) видно, що бюджет та прибуток має певну міру кореляцій. Також можна побачити, що в наборі даних є 815 фільмів з нульовим бюджетом.

2.1.2 Зв’язок між домашньою сторінкою та доходом.

Також важливою ознакою успішності кінофільму є наявність домашньої сторінки фільму, де користувачі можуть дізнатись усю необхідно інформацію про кінофільм та слідкувати за оновленнями. На рисунку 2.7 зображено статистику доходу фільмів з домашньою сторінкою та без неї.

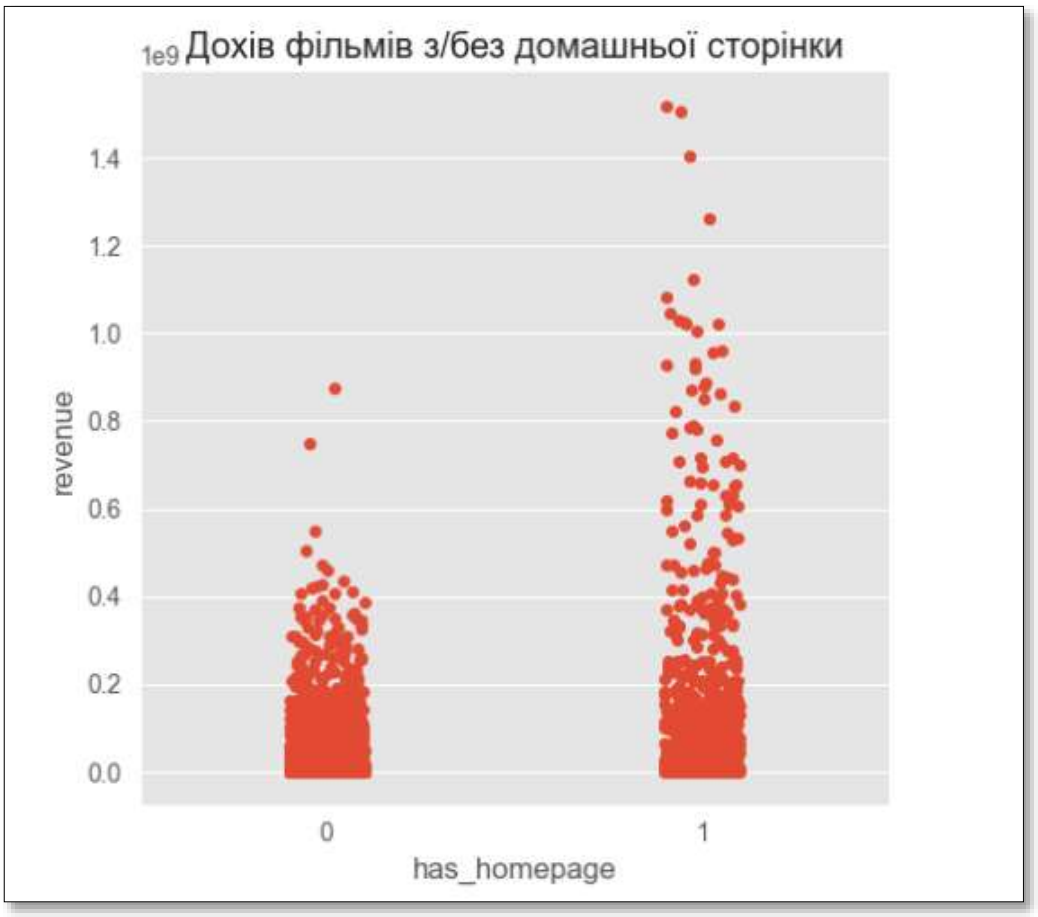


Рисунок 2.7 – Графік доходу фільмів з домашньою сторінкою і без неї

З рисунка 2.7 видно, що фільми, які мають домашню сторінку, мають більший прибуток порівняно з фільмами, що не мають домашньої сторінки. Отже, звідси можна зробити висновок, що наявність домашньої сторінки та прибуток корелюють між собою.

2.1.3 Зв’язок між оригінальною мовою фільму (original_language) і середнім доходом.

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

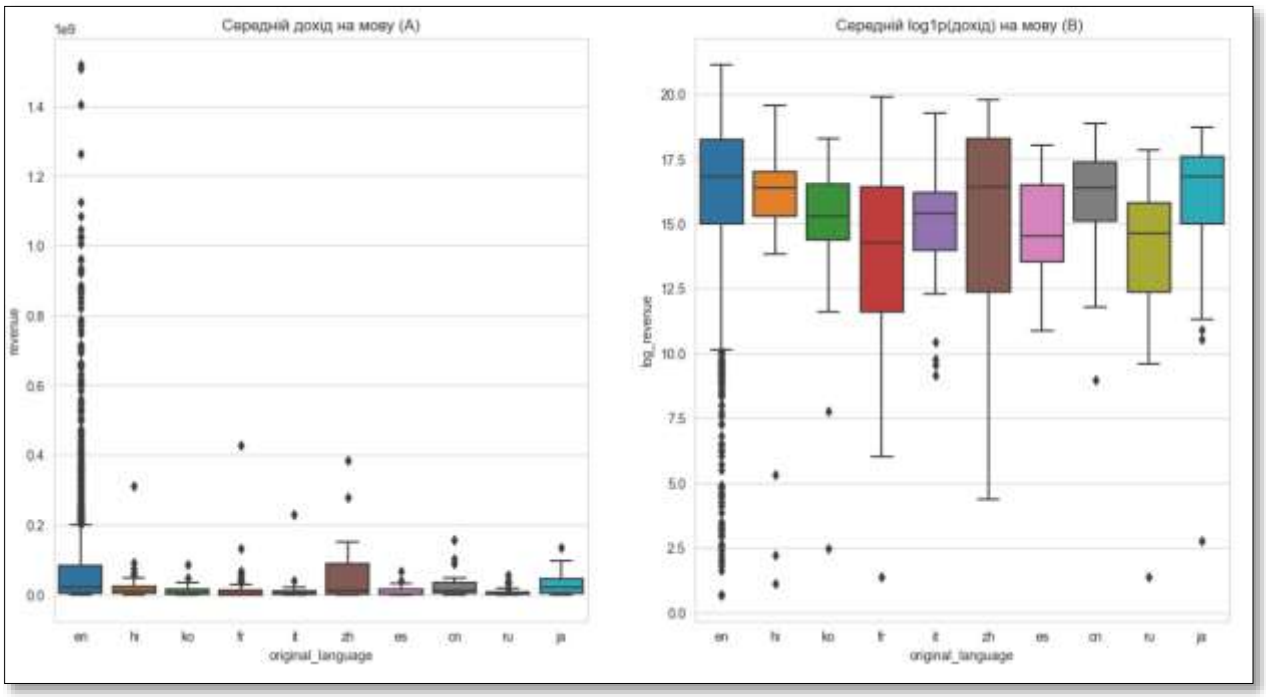


Рисунок 2.8 - Графік розподілу доходів для фільмів на різних мовах

З графіку 2.8 (А), що англійська мова має значно більший прибуток порівняно з іншими мовами. Проте після логарифмічно трансформування, рисунок 2.8, видно, що інші мови також створюють прибуток порівнянний з англійською мовою. Однак саме англомовні фільми лідирують на діаграмі доходів.

2.1.4 Найбільш вживані слова в фільмах.

Далі отримаємо ілюстрацію найбільш вживаних слів в назвах фільмів, за допомогою технології *WordCloud. WordCloud* – це метод візуалізації даних, який використовується для представлення текстових даних, де розмір слова вказує на його частоту чи важливість [12]. Для даного експерименту використано *Python* бібліотеку *wordcloud* [13]. Результат роботи *WordCloud* зображено на рисунку 2.9, з нього видно, що найбільш вживаними словами в назвах фільмів є – *Man, Last, Love, La, Life, Death.*

.

Рисунок 2.9 – Найбільш вживанні слова в назвах фільмів

Таким же чином, як отримали статистику часто вживаних слів в назвах фільмах, отримаємо статистику слів, що використовуються для опису фільмів. Результат зображено на рисунку 2.10, з нього видно, що найбільш вживаними словами в описах до фільмів є – *Find*, *Love*, *One*, *Life*, *World*, *Family*.

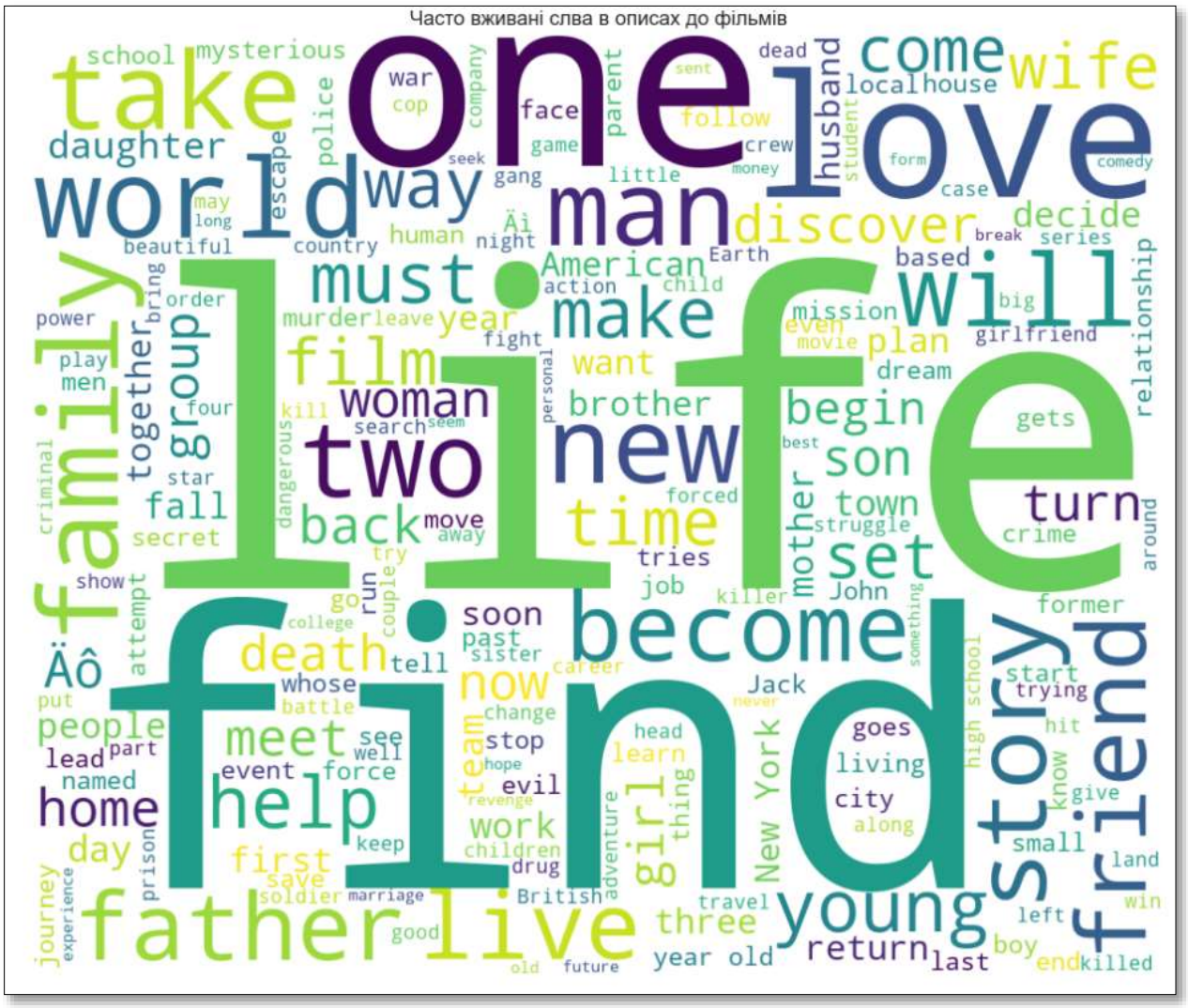


Рисунок 2.10 - Найбільш вживанні слова в описах до фільмів

2.1.5 Вплив опису фільму на прибуток

Для цього підходу застосуємо лінійну регресію, для цього потрібно дізнатись частоту окремих слів між усіма фільмами, за допомогою TFID векторизації [14]:

- TFID допомагає перетворити текс на вектори ознак, які можна пізніше використовувати для оцінювання.
- TFID – числове представлення частоти слів серед документів (описів до фільмів).

Щоб спрогнозувати прибуток, потрібно підібрати лінійну регресію до набору даних. Для цього буде використовуватись *Python* пакет *eli5* [15]. *Eli5*

допоможе налагодити класифікатор машинного навчання, а також допомагає пояснити прогноз. Це допоможе виявити слова, які найбільше впливають до прибуток фільму. На рисунку 2.11 зображено вагу впливу деяких слів на прибуток фільму.

Рисунок 2.11 – Ваги впливу сливі з описів до фільму на прибуток

З рисунка 2.11 видно, що слова можу впливати як позитивно так і негативно на прибуток фільму. Такі слова як *to, blombin, complication* мають позитивний влив, а такі слова як *politicians, 18, violence* мають негативний вплив на прибуток.

Аналогічним способом, отримаємо ваги слів з назв фільмів. Результат зображено на рисунку 3.12.

Contribution	Feature
+12.762	<BIAS>
+1.302	the chaos
+0.917	to
+0.874	fred
+0.760	chaos and
+0.633	s home
+0.555	return to
+0.504	home
+0.462	her job
+0.456	the
+0.390	creates
+0.355	escape from
+0.354	her
+0.321	childhood
+0.307	husband
+0.278	mother s
+0.221	from
+0.196	after
+0.196	her mother
+0.179	to win
+0.135	elizabeth
+0.129	s
+0.127	marriage
+0.108	up
+0.093	to her
+0.089	and her
+0.088	husband and
+0.074	between
+0.071	that
+0.068	of
+0.060	returns to
+0.057	and
+0.050	when
+0.047	win
+0.024	losing her
+0.003	breaks
-0.042	she
-0.057	between the
-0.062	in
-0.086	her husband
-0.100	job
-0.113	losing
-0.130	attempts
-0.145	after her
-0.232	friend
-0.255	returns
-0.261	escape
-0.284	attempts to
-0.290	mother
-0.327	to escape
-0.419	back
-0.478	job in
-0.481	from the
-0.504	return
-0.695	mayhem
-0.913	and return
-0.927	chaos

Рисунок 2.12 – Ваги впливу слів з назв фільмів на прибуток

2.1.6 Вплив дати виходу на прибуток кінофільму

Побудуємо діаграми з розбиттям доходу по місяцям, кварталам року, днів тижня, щоб з’ясувати чи вливає дата виходу фільму на його прибуток. На

рисунка 2.13, 2.14, 2.15 та 2.16 зображено розбиття доходу за місяцем, кварталом, днем тижня відповідно.

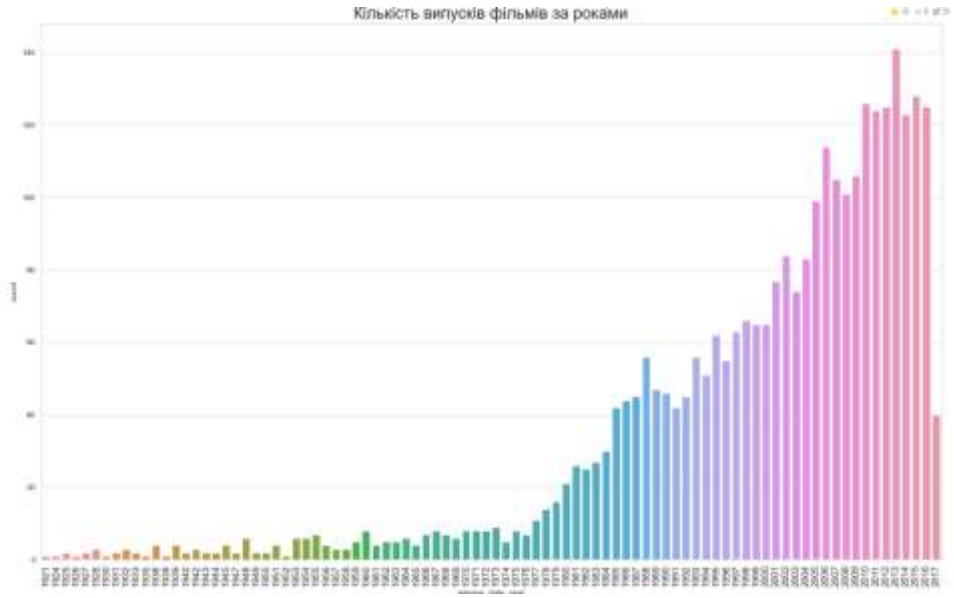


Рисунок 2.13 – Кількість випусків фільмів за роками

На рисунку 2.13 видно, що з 2000-х років спостерігається значний стрибок в кількості випущених фільмів. З діаграми випливає, що в 2013 році було випущено найбільше фільмів, тобто 140 фільмів на рік.

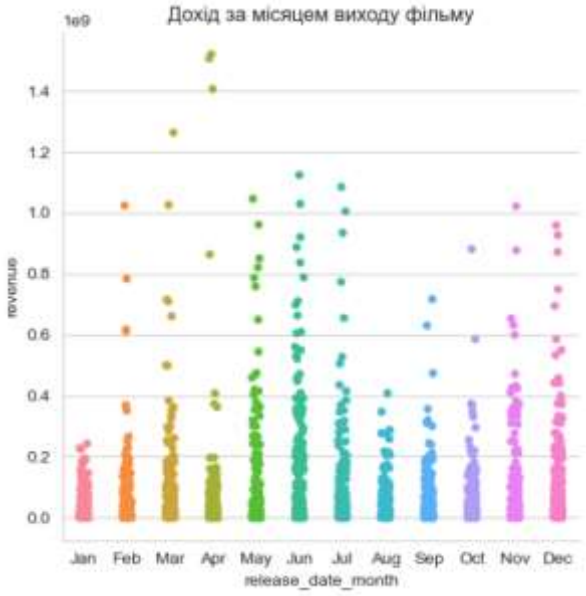


Рисунок 2.14 – Графік доходу за місяцем виходу фільму

З рисунка 2.14 випливає, що фільм, випущений у квітні, має максимальний прибуток, тоді як фільм, випущений у січні, має менший прибуток порівняно з іншими місяцями.

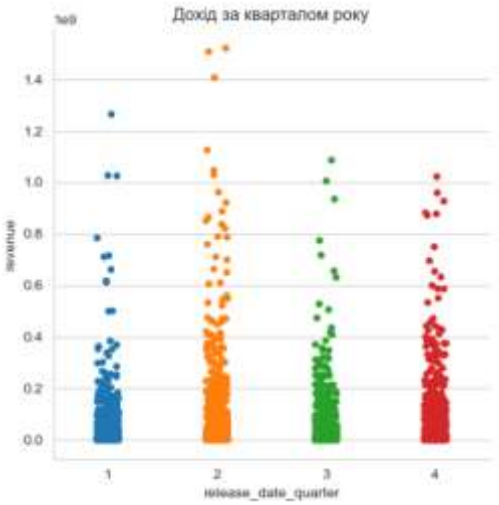


Рисунок 2.15 – Графік доходу за кварталом року

З рисунка 2.15 випливає, що фільм, випущений у другому кварталі (квітень-червень), має більший прибуток порівняно з фільмом, випущеним у іншому кварталі.

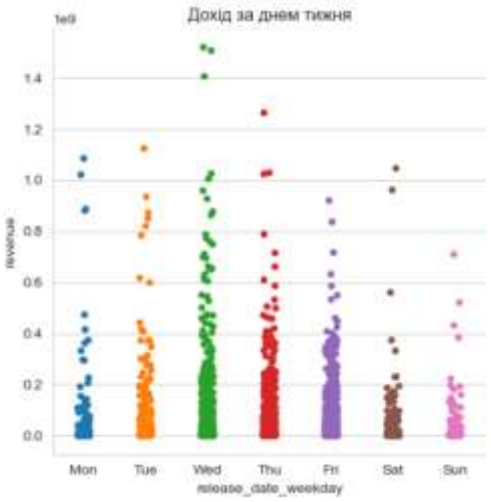


Рисунок 2.16 – Графік доходу за днем тижня

З рисунка 2.16, випливає що фільми, випущений в середу та четверг має більший прибуток. Але це може корелювати з тим, що найбільш очікувані прем’єри зазвичай виходять в четверг, а попередній показ в середу.

2.1.7 Співвідношення між хронометражем кінофільму та доходом.

На рисунку 2.17 зображено розподіл доходу між фільмами з різним хронометражем.

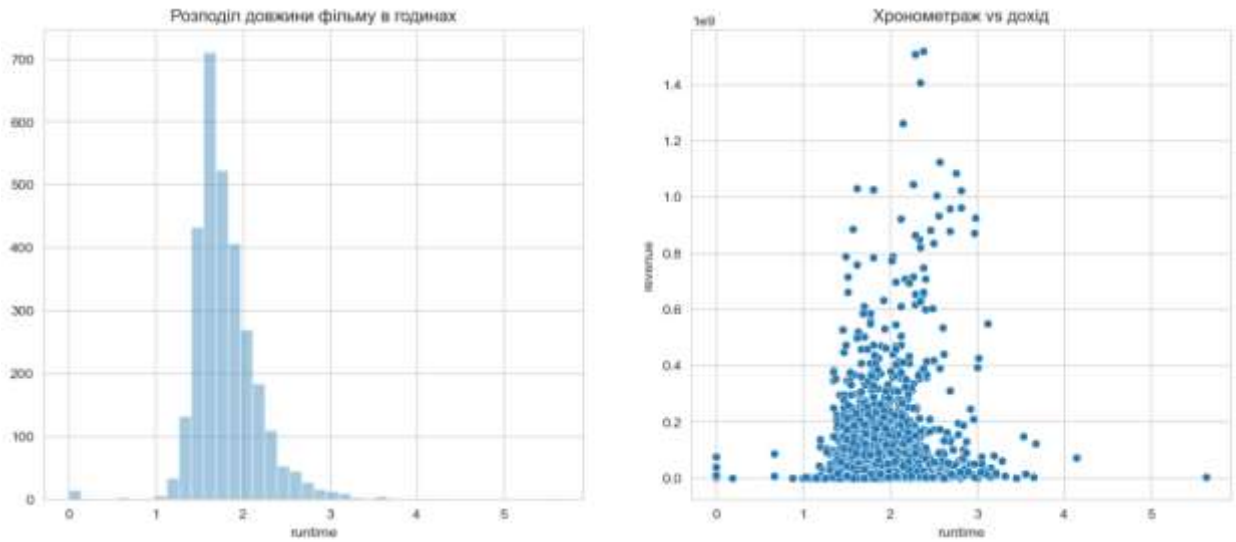


Рисунок 2.17 – Співвідношення хронометражу фільму з його доходом

З рисунка 2.17 випливає, що більшість фільмів триває від 1 до 3 годин. І фільм, який припадає на цей проміжок часу, має найвищий прибуток.

2.2 Розробка моделі прогнозування успішності кінофільму

Для тренування та валідації моделі прогнозування успішності кінофільму вхідний набір даних розбитий на дві секції – дані для тренування моделі та дані для валідації. Ознаки кінофільмів, що впливають на його прибуток та зміні передбачення. Ознаки фільмів, що будуть брати участь в розробці моделі прогнозування успішності кінофільму, перелічені у таблиці 2.2.

Таблиця 2.2 – ознаки для побудови моделі прогнозування успішності фільмів

Назва ознаки	Тип даних	Опис
budget	integer	Бюджет фільму
popularity	float	Коефіцієнт популярності
runtime	integet	Хронометраж
revenue	integer	Прибуток
log1p(revenue)	flaot	log1p(Прибуток)
log1p(budget)	flaot	log1p(Бюджет)
has_homepage	boolean	Наявність домашньої сторінки
release_date_year	integer	Рік виходу
release_date_weekday	integer	День тижня виходу
release_date_month	integer	Місяць виходу
release_date_day	integer	День місяць виходу
release_date_quarter	integer	Квартал виходу

2.2.1 Метод лінійної регресії

Побудуємо модель використовуючи лінійну регресі. Для цього використаємо *Python* пакет *scikit-learn* та клас з цього пакету *LinearRegression*. Для перевірки точності використано коефіцієнт детермінації (*r2_core*) [16]. На рисунку 2.18 зображено застосування лінійної регресії на ознаках виділених в таблиці 2.2.

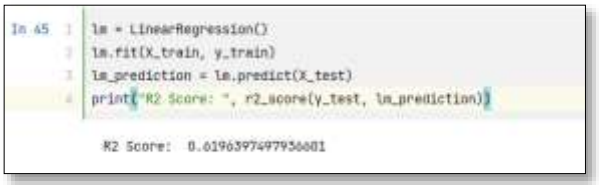


Рисунок 2.18 – Застосування лінійної регресії

З рисунка 2.18 видно, що точність лінійної регресії складає **61.96%.**

2.2.2 Метод Random Forest.

Побудуємо модель використовуючи random forest. Для цього використаємо *Python* пакет *scikit-learn* та клас з цього пакету *RandomForestRegressor*. Для перевірки точності використано коефіцієнт детермінації. На рисунку 2.19 зображено застосування random forest на ознаках виділених в таблиці 2.2.

```
In 46  from sklearn.ensemble import RandomForestRegressor

       RF_model = RandomForestRegressor(random_state =0, n_estimators=500, max_depth=10)
       RF_model.fit(X_train, y_train)

       y_hat = RF_model.predict(X_test)
       print ("R2 score:", r2_score(y_hat, y_test))

       R2 score: 0.5396510534381079
```

Рисунок 2.19 - Застосування random forest

З рисунка 2.18 видно, що коефіцієнт детермінації складає **53.96%.** Однієї з переваг RandomForestRegressor в *scikit-learn*, що можна отримати коефіцієнти випливу ознак на результат передбачення. На рисунку 2.20 зображено диграму впливу ознак на результат передбачення.

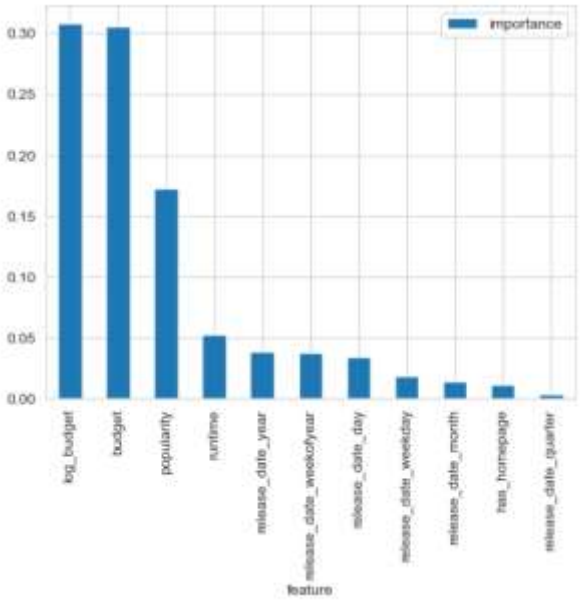


Рисунок 2.20 – Діаграма впливу ознак на результат передбачення

2.2.3 Метод Gradient Boosting.

Побудуємо модель використовуючи gradien boosting. Для цього використаємо *Python* пакет *scikit-learn* та клас з цього пакету *GradientBoostingRegressor*. Для перевірки точності використано коефіцієнт детермінації та середньоквадратичну похибку (mse) [17]. На рисунку 2.21 зображено застосування *GradientBoostingRegressor* на ознаках виділених в таблиці 2.2.

Рисунок 2.21 - Застосування *GradientBoostingRegressor*

З рисунку 2.21 видно, що коефіцієнт детермінації досягає **82.52%**, а точність моделі на тестовому наборі даних дорівнює **85.37%.**

2.2.4 Метод регресії на основі нейронної мережі.

Оскільки матеріалів, щодо прогнозування успішності кінофільмів, виявлено не так багато, запропоновану власну модель машинного навчання. Дана модель базується на регресійній моделі машинного навчання для прогнозування цін на будинки [18]. Модель змінено під предметну область кінофільмів, тож було добавлено нові жари та оновлено параметри для існуючих шарів. На рисунку 2.22 зображено фінальний вид моделі машинного навчання.

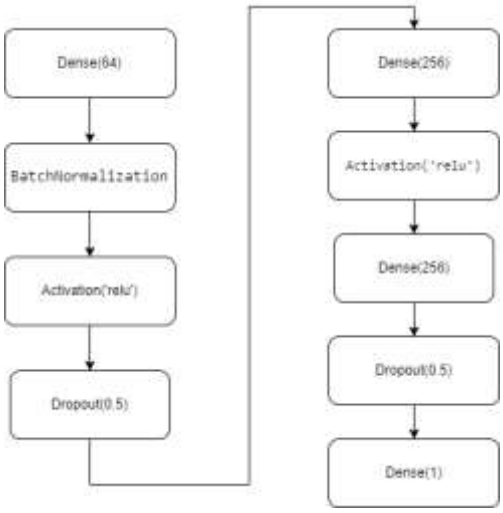


Рисунок 2.22 – структура регресійної моделі прогнозування успішності кінофільму

Розглянемо шари, присутні в нейронній мережі:

- *Dense (256)* – звичайний повно зв'язний шар, з 256 нейронами [19].
- *BatchNormalization* – застосовує перетворення, щоб підтримувати середнє значення вихідних значень шару до проміжку від 0 до 1 [20].
- *Activation('relu')* - функція активації *relu*, графік функції зображений на рисунку 2.23 [21].
- *Dropout* – випадково встановлює деякі елементів вхідної матриці рівним нулю, з вказаною частотою [22].

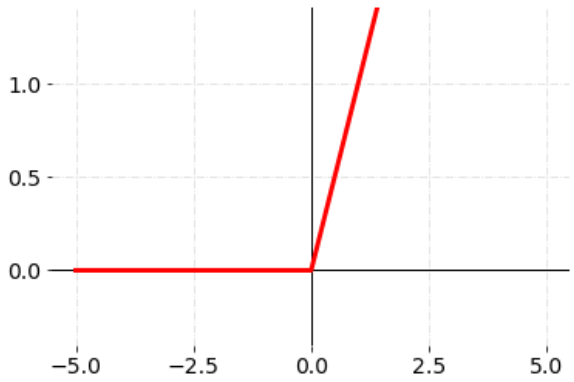


Рисунок 2.23 – Графік функції активації *relu*

Важливо зазначити, що шар *BatchNormalization* працює по різному під час навчання і в режими прогнозування мережі. Під час навчання даний шар нормалізує дані використовуючи середнє значення та стандартне відхилення вхідних даних. Під час прогнозування шар нормалізує значення використовуючи тієї самі параметри, що і для тренування.

Нейронна мережа закінчується шаром *Dense*, що має лише один нейрон, без функції активації. Це типова конфігурація для скалярної регресії (метою якої є прогнозування одного значення на неперервній числовій прямій).

Застосування функції активації могло б обмежувати діапазон вихідних значень: наприклад, якщо в останньому шарі застосувати сигмоїдну функцію активації, мережа навчалась би прогнозувати тільки значення в проміжку між 0 та 1.

Модель компілюється з функцією втрат – середня квадратична похибка. Ця функція широко використовується в задачах регресії.

2.3 Висновок до розділу 2

Проаналізовано та досліджено набір даних TMDB. Побудовано графіки кореляцій між ознаками кінофільму та його доходу. Відокремлено основні ознаки, що впливають на прибуток кінофільму.

У даному розділі було проаналізовано різні підходи для побудови моделі передбачення успішності кінофільму. Розглянуто наступні алгоритми: лінійна регресія, random forest, gradient boosting.

Запропоновано власну модель прогнозування успішності кінофільму на основі нейронної мережі.

3 РОЗРОБКА ІНФОРМАЦІЙНОЇ ТЕХНОЛОГІЇ ПРОГНОЗУВАННЯ УСПІШНОСТІ КІНОФІЛЬМУ

3.1 Розробка структури програмного забезпечення прогнозування успішності кінофільму

Для того, щоб визначитись з структурою програмного забезпечення для прогнозування успішності кінофільмам потрібно виділити вимоги до системи:

- наївність інтерфейсу користувача;
- база даних кінофільмів;
- можливість вносити дані про фільмі та редагувати їх;
- можливість імпортувати фільми з існуючих джерел;
- прогнозування успішності кінофільму повинно бути відносно швидким;
- результат прогнозування успішності кінофільму повинен бути зрозумілим користувачу.

Також система повинна виділяти бізнес логіку від інтерфейсу користувача. У такий спосіб інтерфейс може бути реалізованим різними способами, в залежності від вимог системи та бізнес плану. Наприклад інтерфейс може бути реалізованим у наступні способи:

- веб-сайт;
- мобільний додаток;
- у вигляді командного рядка;
- декстопний додаток;
- бот в месенджері;
- інтеграції з іншими сервісами.

З вище наведених вимог можна згенерувати загальну структуру системи

прогнозування успішності кінофільмів, що зображена на рисунку 3.1

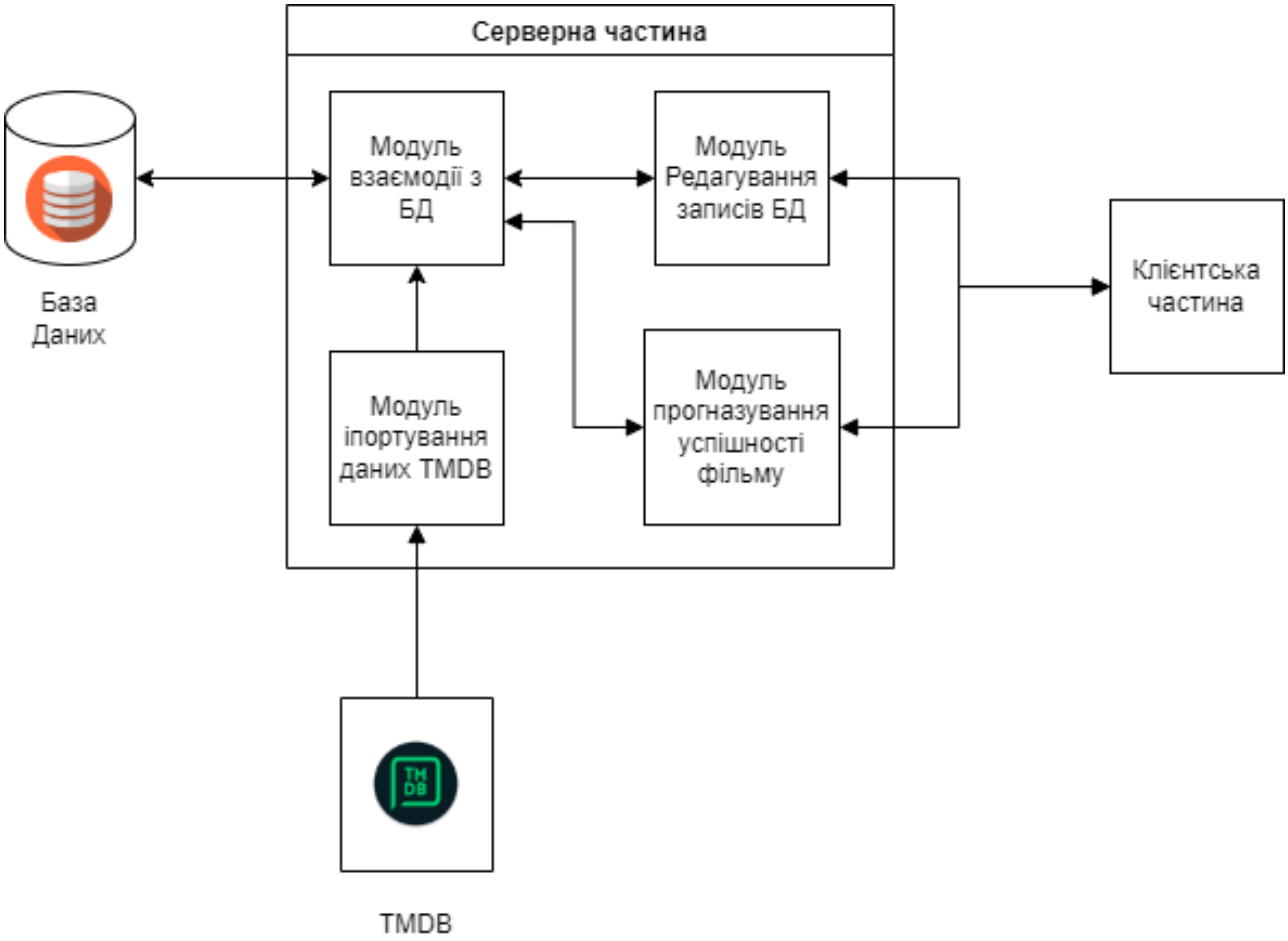


Рисунок 3.1 – Загальна структура програмного забезпечення прогнозування успішності фільмів

Джерелом імпортування даних кінофільмів, обрано найбільшу базу даних фільмів TMD. Обраний саме цей сервіс, оскільки він надає доступ до найширшої бібліотеки кінофільмів, містить різноманітні мета дані про фільмів, має інтерфейси інтеграцій.

Система прогнозування кінофільмів буде виконана у вигляді клієнт-серверного програми. Таким чином, логіка зберігання фільмів та прогнозування буде інкапсульована в серверній частині, а логіка інтерфейсу користувача в клієнтській частині системи.

Для реалізації серверної частини потрібно обрати архітектуру. Для налаштування спілкування між двома програма розробники будують мости –

програмний інтерфейс програма (API), щоб дозволити одній системі отримати доступ до даних або функціональності іншої.

Для швидкої та масштабної інтеграції програм *API* реалізується з використанням протоколів та/або специфікації для визначення семантики та синтаксису повідомлень, що передаються по зв'язку. Ці специфікації складають архітектуру *API*. З часом були створенні різні архітектури стилі *API*. Кожен з них має свої стандарти обміну даними. У таблиці 3.1 наведено порівняльну характеристику існуючих архітектур серверних додатків [23].

Таблиця 3.1 порівняльна характеристика архітектури серверних додатків

	RPC	SOAP	REST	GraphQL
Опис	Виклик локальних процедур	Обгорнута структура повідомлень	Дотримання 6 архітектурних обмежень	Схема та система типів
Формат	JSON, XML, Protobuf, Thrift, FlatBuffres	XML	XML, JSON, HTML, TEXT	JSON
Складність	Проста	Складна	Проста	Середня

Оскільки *REST* архітектура відносно проста в реалізації там підтримує багато форматів спілкування, було обрано саме цю архітектуру клієнт-серверної частини.

3.2 Проектування бази-даних інформаційної технології прогнозування успішності фільмів

Для зберігання інформації, що буде використовуватись для прогнозування успішності кінофільму, потрібні наступні сутності:

- фільм – movie;

- жанр – genre;
- студія – production_company;
- країна виробник – production_country;
- ключове слово – key_word;
- актор – actor;
- знімальний персонал – credit.

Знімальний персонал – сутність, що відображає людей задіяних в зйомці фільму, таки-як: режисер, сценарист монтажер і тд. В таблиці 3.1 відображено поля сутності – знімальний персонал.

Таблиця 3.2 – Поля сутності знімального персоналу (credit)

Назва поля	Тип даних	Опис
credit_id	integer, primary key	Індифікатор
department	string	Тип департаменту: Directing, Production …
gender	enum(male, female)	Стать
job	string	Опис посади
name	string	Ім’я

Жанр – сутність, що відображає жанр фільму. В таблиці 3.2 відображено поля сутності – жанр.

Таблиця 3.3 – Поля сутності жанр (genre)

Назва поля	Тип даних	Опис
genre_id	integer, primary key	Індифікатор
name	string	Ім’я жанра

Студія – сутність, що студія розробки фільму. В таблиці 3.3 відображено

поля сутності – **студія**.

Таблиця 3.4 – Поля сутності студія (production_company)

Назва поля	Тип даних	Опис
id	integer, primary key	Індифікатор
name	string	Ім’я студії

Країна виробник– сутність, що студія розробки фільму. В таблиці 3.4 відображено поля сутності – країна виробник.

Таблиця 3.5 – Поля сутності студія (production_country)

Назва поля	Тип даних	Опис
Id	integer, primary key	Індифікатор
Name	string	Назва країни

Ключове слово – ключове слово, що використовується для характеристики кінофільму. В таблиці 3.5 відображено поля сутності – країна виробник.

Таблиця 3.6 – Поля сутності ключове слово (key_word)

Назва поля	Тип даних	Опис
Id	integer, primary key	Індифікатор
Name	string	Назва слова

Актор – актор, що приймає участь у фільмі. В таблиці 3.6 відображено поля сутності – актор.

Таблиця 3.7 – Поля сутності актор (cast)

Назва поля	Тип даних	Опис
id	integer, primary key	Індифікатор
Назва поля	Тип даних	Опис
charachter	string	Ім’я персонажа
gender	Enum(male ,female)	Стать
name	Enum(male ,female)	Ім’я актора

Фільм. В таблиці 3.7 відображено поля сутності – фільм.

Таблиця 3.8 – Поля сутності фільм (movie)

Назва поля	Тип даних	Опис
id	integer, primary key	Індифікатор
budget	integer	Бюджет
genre_ids	List[integer], зовнішній ключ на genre	Список жанрів
homepage	string	Посилання на головну сторінку
Imdb_id	string	Індифікатор IMDB
original_language	string	Мова фільму
Назва поля	Тип даних	Опис
original_title	string	Заголовок фільму
popularity	float	Оцінка популярності
production_companies	List[integer], зовнішній ключ на product_company	Компанії виробники
production_countries	List[integer], зовнішній ключ на product_country	Країни виробники

Продовження таблиці 3.9.

release_date	Datetime	Дата релізу
runtime	Integer	Хронометраж фільму
status	String	Статус фільму
keywords	List[integer], зовнішній ключ key_wrod	Список ключових слів
Cast	List[integer], зовнішній ключ actor	Список акторів
Crew	List[integer], зовнішній ключ credit	Персонал
revenue	Integer	Прибуток

На основі вище наведених сутностей, можна згенерувати загальну схему з’язків бази даних, що зображено на рисунку 3.2.

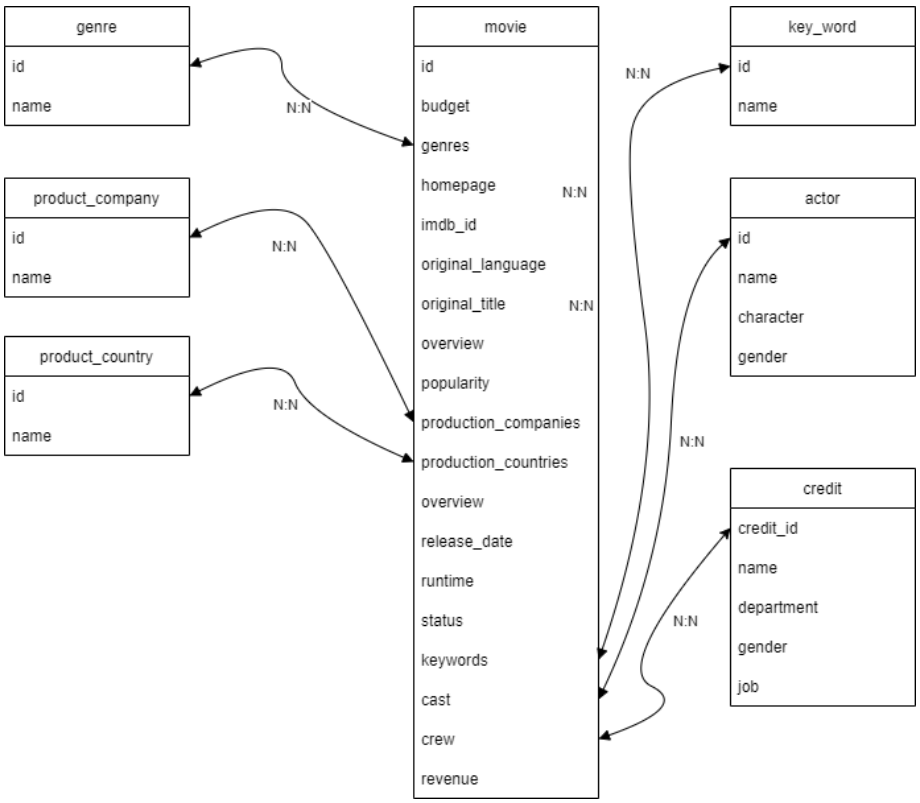


Рисунок 3.2 – Загальна схема бази-даних кінофільмів

3.3 Розробка алгоритму модуля взаємодії з базою даних кінофільмів

Модуль взаємодії з базою даних кінофільмів відповідає за програмне з’єднання з базою даних. Модуль повинен виконувати наступний функціонал:

- додавання нового фільму
- редагування існуючого фільму
- видалення фільму
- отримання фільмів

На рисунку 3.3 зображено схему алгоритму роботи модуля взаємодії з базою даних.

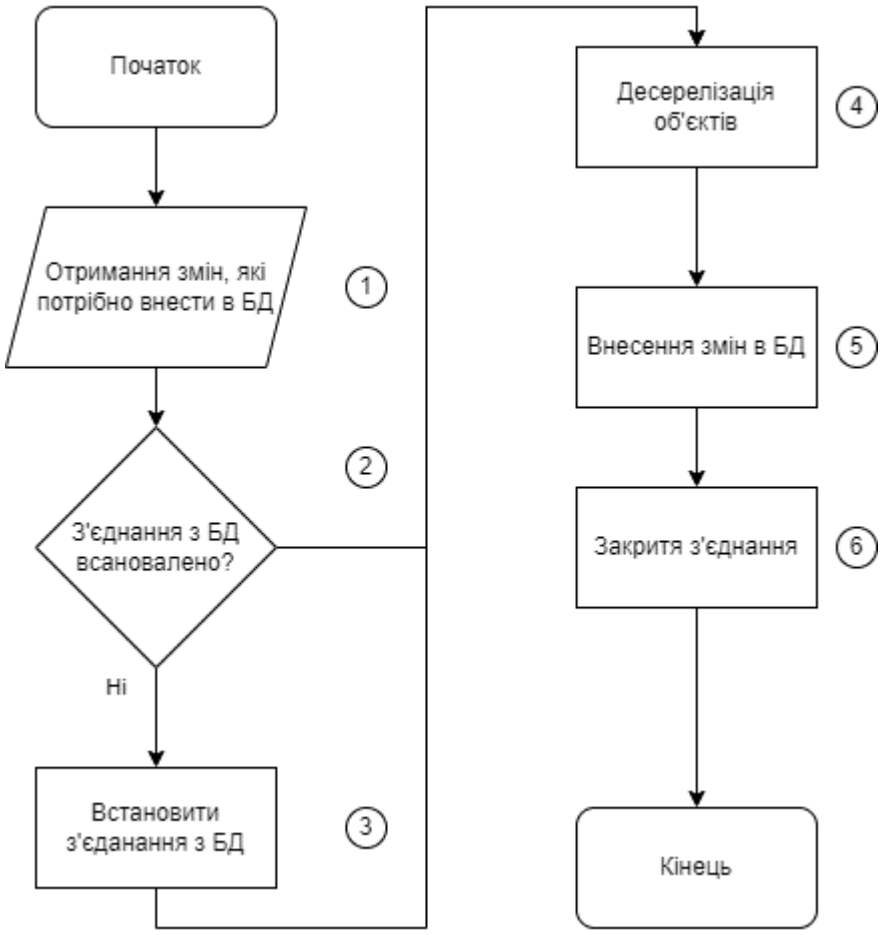


Рисунок 3.3 – Схема алгоритму модуля взаємодії з базою даних

Опишемо даний алгоритм:

1. Отримання змін, які потрібно ввести в базу даних.
2. Перевірка, чи встановлене з’єднання з базою даних.
3. Встановлення з’єднання з базою даних.
4. Десерілізація об’єктів – перетворення типів даних об’єктів мови програмування в типи даних бази даних.
5. Закриття з’єднання з базою даних.

3.4 Розробка алгоритму модуля імпортування кінофільмів з TMDB

Основною задачею модуля імпортування фільмів – є імпортування фільмів з найбільшої бази даних фільмів TMDB та в подальшому зберігання отриманих даних в локальній базі. При цьому модуль повинен вміти об’єднувати існуючі дані про фільми з даними отриманими з TMDB. На рисунку 3.4 зображено схему алгоритму модуля.

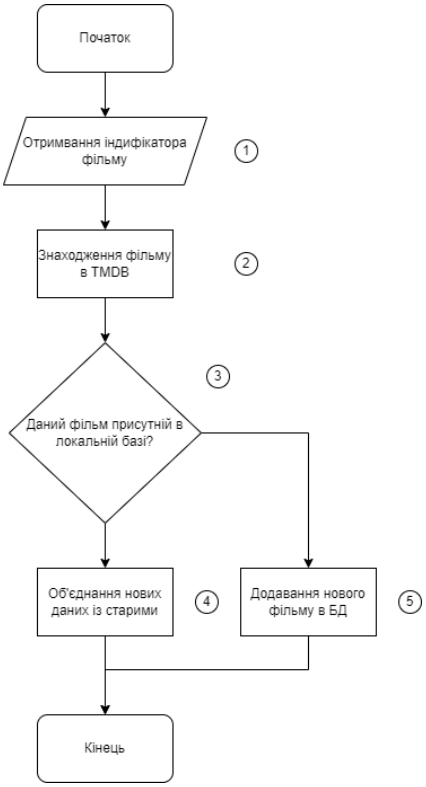


Рисунок 3.4 – Схема алгоритму модуля імпортування фільмів з TMDB

Опишемо даний алгоритм:

1. Отримання індикатора фільму.
2. Знаходження фільму в базі даних TMDB.
3. Перевірка чи в системі уже є даний фільм.
4. Об’єднання старої інформації про фільм з новою.
5. Додавання нового фільму в базу даних.

3.5 Розробка алгоритму модуля редагування записів бази даних фільмів

Основною задачею даного модуля – надавання інтерфейсу редагування записів бази даних фільмів для клієнтської частини. Оскільки в розділі 3.1 було обрано REST архітектуру для серверної частини, клієнт буде взаємодії з сервером через HTTP запити. Для цього сервер повинен підтримувати усі основні HTTP методи для кожної сутності бази даних, що були описані в розділі 3.2. В таблиці 3.2 перелічено основні HTTP запити, які повинен підтримувати сервер, для надання клієнту можливості взаємодіяти з записами в базі даних. Усі вхідні та вихідні об’єкти будуть приставлені у вигляді JSON об’єктів.

Таблиця 3.10 – Основні HTTP запити модуля редагування записів БД

Назва	Метод	Шлях	Вхідні дані	Вихідні дані
Отримання сутності по індикатору	GET	/:id	Індифікатор сутності	Об’єкт сутності
Отримання усіх сутностей	GET	/	Фільтр	Об’єкти сутностей

Продовження таблиці 3.10

Назва	Метод	Шлях	Вхідні дані	Вихідні дані
Додавання сутності	POST	/	Об’єкт сутності	Об’єкт сутності
Оновлення усіє сутності	PUT	/:id	Індифікатор та об’єкт сутності	Об’єкт сутності
Оновлення окремих полів сутності	PATCH	/:id	Індифікатор та об’єкт сутності	Об’єкт сутності
Видалення сутності	DELETE	/:id	Індифікатор сутності	

3.6 Розробка алгоритму модуля прогнозування успішності кінофільму

Модуль прогнозування успішності кінофільму безпосередньо відповідає за надання прогнозу доходу фільму на онові ознак фільму. На рисунку 3.5 зображено схему алгоритму прогнозування фільму.

Рисунок 3.5 – Схема алгоритму модуля прогнозування успішності кінофільму

Опишемо алгоритм:

1. Отримання даних фільму.
2. Виділення ознак, що були визначені в розділі 2.
3. Нормалізація числових даних.
4. Тонізація текстових даних, так як модель прогнозування може працювати тільки чи чистовий значеннями.
5. Завантаження моделі прогнозування з розділу 2.
6. Застосування моделі для отримання прогнозу.
7. Інтерпретація прогнозу.

3.7 Висновок до розділу 3

У розділі наведено основні кроки роботи інформаційної технології прогнозування успішності кінофільму. Обрано клієнт-серверну архітектуру для програмного забезпечення прогнозування успішності кінофільму. Розбито програмний засіб на модулі, а саме: модуль взаємодії з базою даних, модуль імпортування кінофільмів, модуль редагування записів в базі даних, модуль прогнозування успішності кінофільму. Описано роботу та алгоритм кожного модуля. Спроектовано базу даних для програмного забезпечення прогнозування успішності кінофільмів.

4 ПРОГРАМНА РЕАЛІЗАЦІЯ ІНФОРМАЦІЙНОЇ ТЕХНОЛОГІЇ ПРОГНОЗУВАННЯ УСПІШНОСТІ КІНОФІЛЬМУ

4.1 Обґрунтування вибору мови програмування

Серверна частина буде розроблена на мові програмування *Python* 3.10, використовуючи фреймворк *FastAPI*. *Python* було обрано так як він надає широкий спектр потужних інструментів та бібліотек для розробки серверної частини додатка, обробки зображень, роботи з великим обсягом даних, навчання, роботи з багатовимірними матрицями, тренування та використання моделей машинного навчання. *FastAPI* – це сучасний, швидкий веб-фреймворк для створення *API*-інтерфейсів з *Python* 3.6+ на основі стандартних підказок типів *Python* [24]. Основні особливості FastAPI:

Висока продуктивність, нарівні з *NodeJS* та *Go*;

Збільшення швидкості розробки приблизно на 200% - 300 %;

Зменшення кількості помилок приблизно на 40 %, спричиненою розробником;

Розробка та навчання моделей машинного навчання відбувалася з використанням бібліотек *TensorFlow* та *Keras* - високопродуктивні бібліотеки для написання моделей машинного навчання. TensorFlow - відкрита програмна бібліотека для машинного навчання цілій низці задач, розроблена компанією *Google* для задоволення її потреб у системах, здатних будувати та тренувати нейронні мережі для виявляння та розшифровування образів та кореляцій, аналогічно до навчання й розуміння, які застосовують люди. Keras надає зручний та мінімалістичний інтерфейс над *TensorFlow* та *TensorFlow* 2.0, що пришвидшує написання час моделей та полегшує процес дослідження моделей машинного навчання.

Клієнтська частина буде розроблена на мові програмування *Dart*, використовуючи фреймворк *Flutter*. *Flutter* було обрано, оскільки він надає

програмісту можливість створювати красиві та нативні програми, задіяючи мінімальні зусилля. За допомогою *Flutter* можна створювати кросплатформені програми, використовуючи одну кодову базу [25]. Flutter підтримує наступні платформи:

- Web;
- Windows;
- Linux;
- MacOS;
- iOS;
- Android.

4.2 Обґрунтування вибору середовища програмування

Для розробки серверної частини інформаційної технології проказування успішності кінофільму, було обрано середовище «*PyCharm*». *PyCharm* забезпечує інтелектуальне завершення та підказки при написанні коду, перевірку коду, оперативне виділення помилок та підказки для виправлення помилок, а також автоматичний рефакторинг коду і широкі можливості навігації. Колекція інструментів *PyCharm* з коробки включає вбудований відладчик і тестовий прогін; Профілювальник Python; вбудований термінал; інтеграція з основними засобами контролю версій і вбудованими інструментами для бази даних; можливості віддаленої розробки з віддаленими інтерпретаторами; вбудований ssh-термінал; і інтеграція з *Docker* і *Vagrant*. *PyCharm* інтегрується з *IPython Notebook*, має інтерактивну консоль *Python* і підтримує Anaconda, а також кілька наукових пакетів, включаючи *Matplotlib* і *NumPy* [26]. Через вище вказані характеристики *PyCharm* є одним із провідних середовищ програмування на мові *Python*. Підтримка математичних модулів та *Anaconda* – безкоштовний дистрибутив пакетів для машинного навчання,

забезпечує комфортну розробки моделей для машинного навчання. На рисунку 4.1 зображено інтерактивне середовище розробки для мови програмування *Python – PyCharm.*

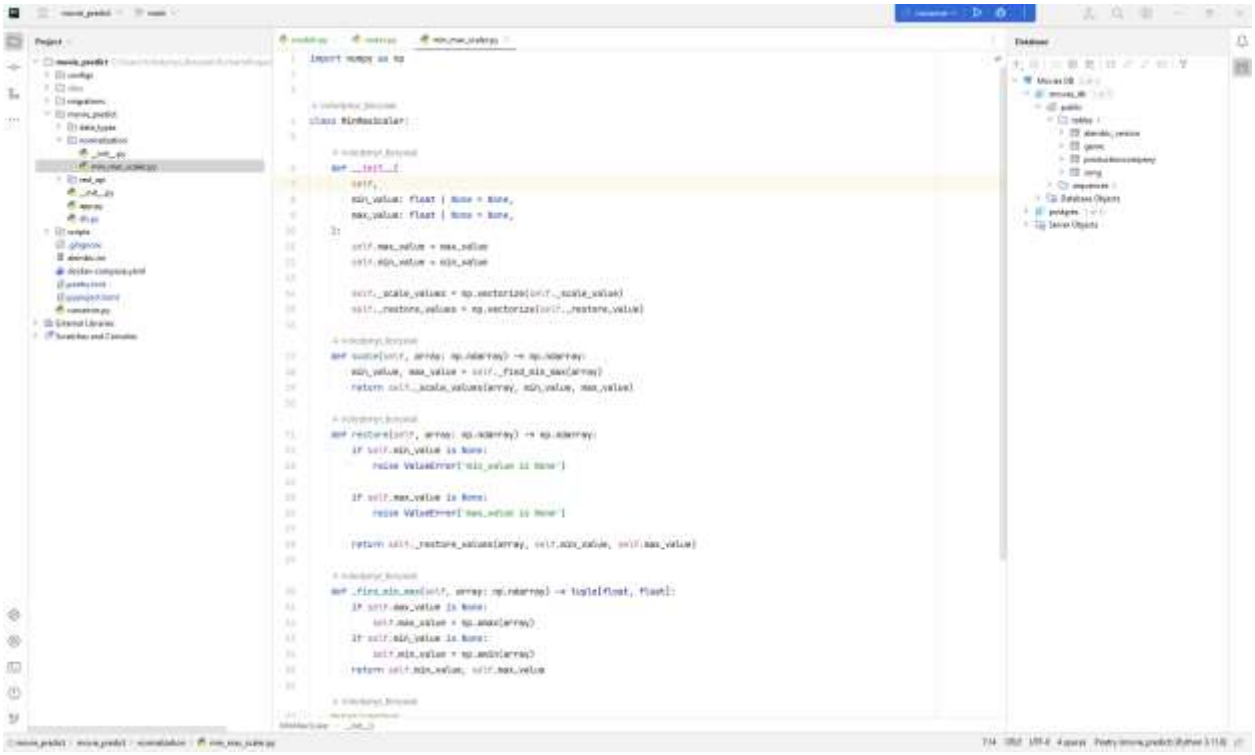


Рисунок 4.1 – середовище програмування *PyCharm*

Для розробки серверної частини інформаційної технології проказування успішності кінофільму, було обрано середовище «*Intellij Idea*» [27]. *Intellij Idea* має такі саме переваги, що і *PyCharm*, окрім підтримки математичних пакетів. Проте має підтримку мови програмування *Dart*, та фреймворку *Flutter*. Надає зручний інтерфейс для запуску емуляторів пристроїв, що надає можливість писати та тестувати програми на різних платформах, такі як:

- Windows.
- Linux.
- Android.
- iOS.

На рисунку 4.2 зображено інтерактивне середовище розробки – *Intellij Idea*.

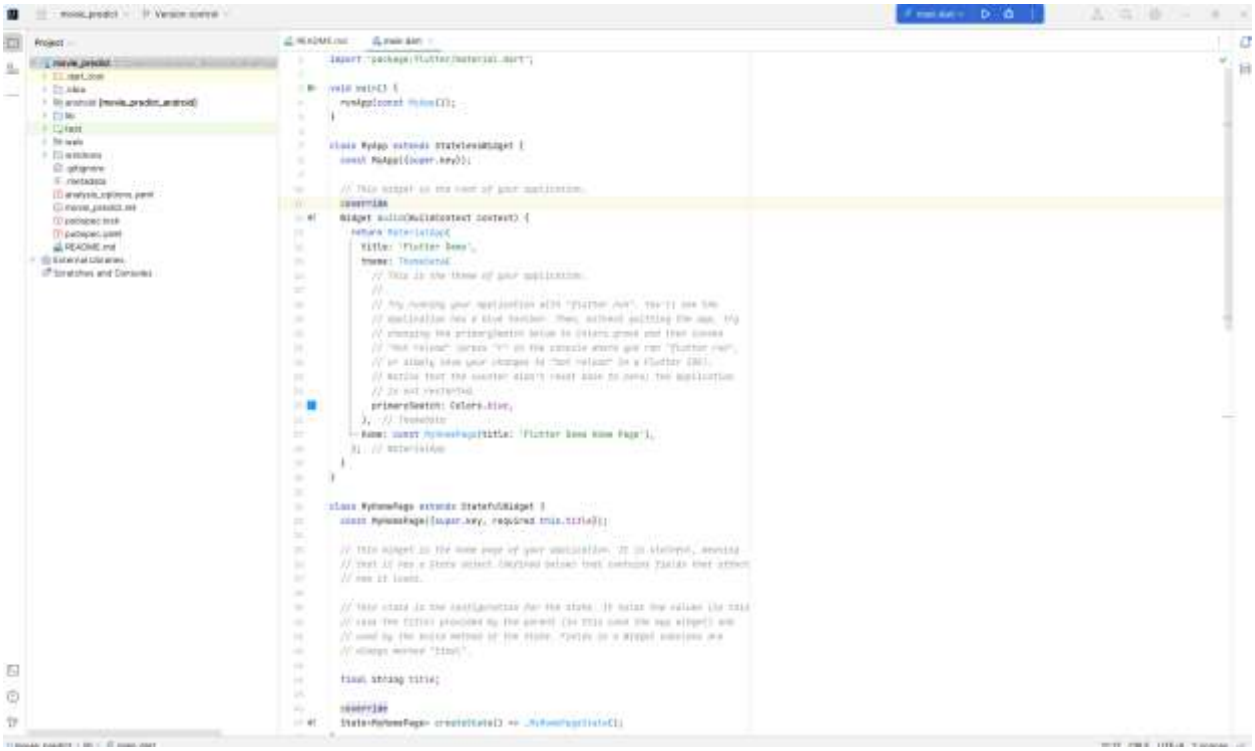


Рисунок 4.2 – середовище програмування *Intellij Idea*.

4.3 Програмна реалізація модулів інформаційної технології прогнозування успішності кінофільмів

4.3.1 Програмна реалізація модуля взаємодію з базою даних

Модуль взаємодії з базою даних реалізований на основі Python бібліотеки sqlmodel. Sqlmodel – надає зручний інтерфейс опису сутностей за допомогою класів та автоматичне створення SQL запитів на основі моделей. Інтерфейс Repo описує взаємодію з базою даних для кожної сутності. На рисунку 4.3 зображено фрагмент коду Repo.

```python
class Repo(Generic[_Model], ABC):

    def __init__(self, session: Session):
        self._session = session

    @abstractmethod
    def get(self, pk: int) -> _Model:
        pass

    @abstractmethod
    def get_many(self, *where) -> Iterable[_Model]:
        pass

    @abstractmethod
    def delete(self, pk: int) -> NoReturn:
        pass
```

Рисунок 4.3 – Фрагмент коду інтерфейсу Repo

На рисунку 4.4 зображено фрагмент реалізації інтерфейсу Repo для сутності Movie.

```python
class MovieRepo(Repo[Movie]):

    def get(self, pk: int) -> Movie:
        query = select(Movie).where(Movie.id == pk)
        movie = self._session.execute(query).first()
        return movie

    def get_many(self, *where) -> Iterable[Movie]:
        query = select(Movie).where(*where)
        return self._session.execute(query)

    def delete(self, pk: int) -> NoReturn:
        query = delete(Movie).where(Movie.id == pk)
        self._session.execute(query)
```

Рисунок 4.4 – Фрагмент коду класу MovieRepo

На рисунку 4.5 зображено реалізацію сутності Movie за допомогою sqlmodel.

```python
class Movie(SQLModel, table=True):
    id: int = Field(primary_key=True)
    budget: int
    genre_ids: list[ForeignKey('genre.id')] = []
    homepage: str | None
    imdb_id: str
    original_language: str
    original_title: str
    overview: str
    popularity: float
    production_companies: list[ForeignKey('production_company.id')] = []
    production_countries: list[ForeignKey('production_country.id')] = []
    release_date: dt.datetime
    runtime: int
    keywords: list[ForeignKey('keyword.id')] = []
    cast: list[ForeignKey('actor.id')] = []
    crew: list[ForeignKey('credit.id')] = []
    revenue: int | None
```

Рисунок 4.5 – Реалізація сутності Movie за допомогою sqlmodel.

4.3.2 Програмна реалізація модуля редагування записів в базі даних

Модуль редагування записів в базі даних реалізовано на основі фреймворку FastAPI, який надає можливості створювати обробники HTTP записів за допомогою декораторів. На рисунку 4.6 зображено фрагмент коду, що описує обробники HTTP запитів для редагування фільмів в базі даних.

```python
@router.put('/', response_model=MovieInDB)
async def create_movie(movie: CreateMovie) -> MovieInDB:
    movie_db = MovieInDB.parse_obj(movie)
    return await movie_db.insert()


volodymyrb
@router.get('/{movie_id}', response_model=MovieInDB)
async def get_movie(movie_id: PydanticObjectId) -> MovieInDB:
    return await MovieInDB.get(movie_id)


volodymyrb
@router.get('/', response_model=list[MovieInDB])
async def get_movies() -> list[MovieInDB]:
    return [movie async for movie in MovieInDB.find_all()]


volodymyrb
@router.patch('/{movie_id}', response_model=MovieInDB)
async def update_movie(movie_id: PydanticObjectId, movie: UpdateMovie) -> MovieInDB:
    movie_db = await MovieInDB.get(movie_id)
    await movie_db.set(movie.dict(exclude_none=True))
```

Рисунок 4.6 – Фрагмент коду для обробки запитів редагування фільмів.

4.3.3 Програмна реалізація модуля імпортування кінофільмів з *TMDB*

Модуль призначений для імпортування даних про фільми з сервісу *TMDB*, для того щоб не заповнювати усі інформація власноруч.

Модуль імпортування фільмів реалізований у вигляді *HTTP* клієнта. Клієнт створює запити на сервіс *TMDB*, вказуючи Індифікатор фільму, що потрібно знайти. Далі модуль завантажує отриману інформація в базу даних. На рисунку 4.7 зображено фрагмент коду HTTP клієнта, а на рисунку 4.8 фрагмент коду завантаження результату в базу даних.

```
class TmdbClient(AsyncConnection[httpx.AsyncClient]):

    _BASE_URL = Url(
        host=CONFIG.TMDB.HOST,
        username=CONFIG.TMDB.TOKEN_ID,
        password=CONFIG.TMDB.TOKEN_VALUE,
    )

    _MOVE_URL = _BASE_URL.join_path(Path('movie'))

    new *
    async def close(self) → bool:
        await self._connected_conn.aclose()
        return True

    new *
    async def _create_conn(self) → httpx.AsyncClient:
        return httpx.AsyncClient()

    new *
    def get_movie(self, movie_pk: str) → Movie:
        movie_obj = self._connected_conn.get(
            url=self._MOVE_URL.update_query({
                'pk': movie_pk,
            }).build(),
        )

        if not movie_obj:
            raise ValueError(f'Movie {movie_obj} is not found')

        movie = Movie.parse_obj(movie_obj)
        return movie
```

Рисунок 4.7 – Фрагмент коду модуля імпортування кінофільмів з *TMDB*

```
new"
async def main():
    unique_production_companies = (
        load_production_companies(TRAIN_DATA_PATH)
        | load_production_companies(TEST_DATA_PATH)
    )
    unique_production_companies.add('other')

    await db.init_db()
    async with db.get_session_ctx() as session:
        for company_name in unique_production_companies:
            company = ProductionCompany(name=company_name)
            session.add(company)
        await session.commit()


if __name__ == '__main__':
    asyncio.run(main())
```

Рисунок 4.8 – Фрагмент модуля завантаження даних з *TMDB* в базу даних

4.3.4 Програмна реалізація модуля прогнозування успішності кінофільму

Безпосереднє прогнозування доходу фільму реалізовано в класі *Prediction*, який завантажує натреновану нейрону мережу за проганяє через неї вхідні дані кінофільму, на виході отримуємо прогнозований прибуток кінофільму. На рисунку 4.9 зображено фрагмент класу *Prediction*.

```
from decimal import Decimal

import keras
from keras import models

from configs import CONFIG
from movie_predict.rest_api.movie.models import Movie
from movie_predict.normalization.movie_normaliztion import prepare_movie


class Prediction:

    def __init__(self):
        self._model = self._load_model()

    def predict_revenue(self, movie: Movie) -> Decimal:
        tensor = prepare_movie(movie)
        prediction = self._model(tensor)
        return prediction[0]

    @staticmethod
    def _load_model() -> keras.Sequential:
        return models.load_model(str(CONFIG.MOVIE_PREDICTION.MODEL_PATH))
```

Рисунок 4.9 – Фрагмент модуля передбачення успішності кінофільму

Взаємодія з клієнтом реалізовано за допомогою обробника HTTP запитів - get_*predicted*_revenue. Даний обробник отримає на вхід індифікатор фільму, знаходить потрібний фільм в базі даних, застосовує клас Prediction, та повертає прогнозований прибуток кінофільму. На рисунку 4.10 зображено фрагмент реалізації обробника.

```
router = APIRouter(
    prefix='/prediction',
    tags=['Movie prediction'],
)


new *
@router.get('/revenue')
def get_predicted_revenue(
    movie_pk: int,
    session: AsyncSession = Depends(get_session),
) → dict:

    statement = select(Movie).where(Movie.id == movie_pk)
    response = await session.execute(statement)
    movie = response.scalar_one()

    prediction = Prediction()
    return {
        'result': prediction.predict_revenue(movie),
    }
```

Рисунок 4.10 – Фрагмент коду обробника HTTP запитів, для прогнозування доходу кінофільму

4.4 Тестування та аналіз результатів роботи програми прогнозування успішності кінофільму

Спочатку протестуємо в цілому роботу програми прогнозування успішності кінофільму. Вікно з результатом роботи програми зображено на рисунку 4.11.

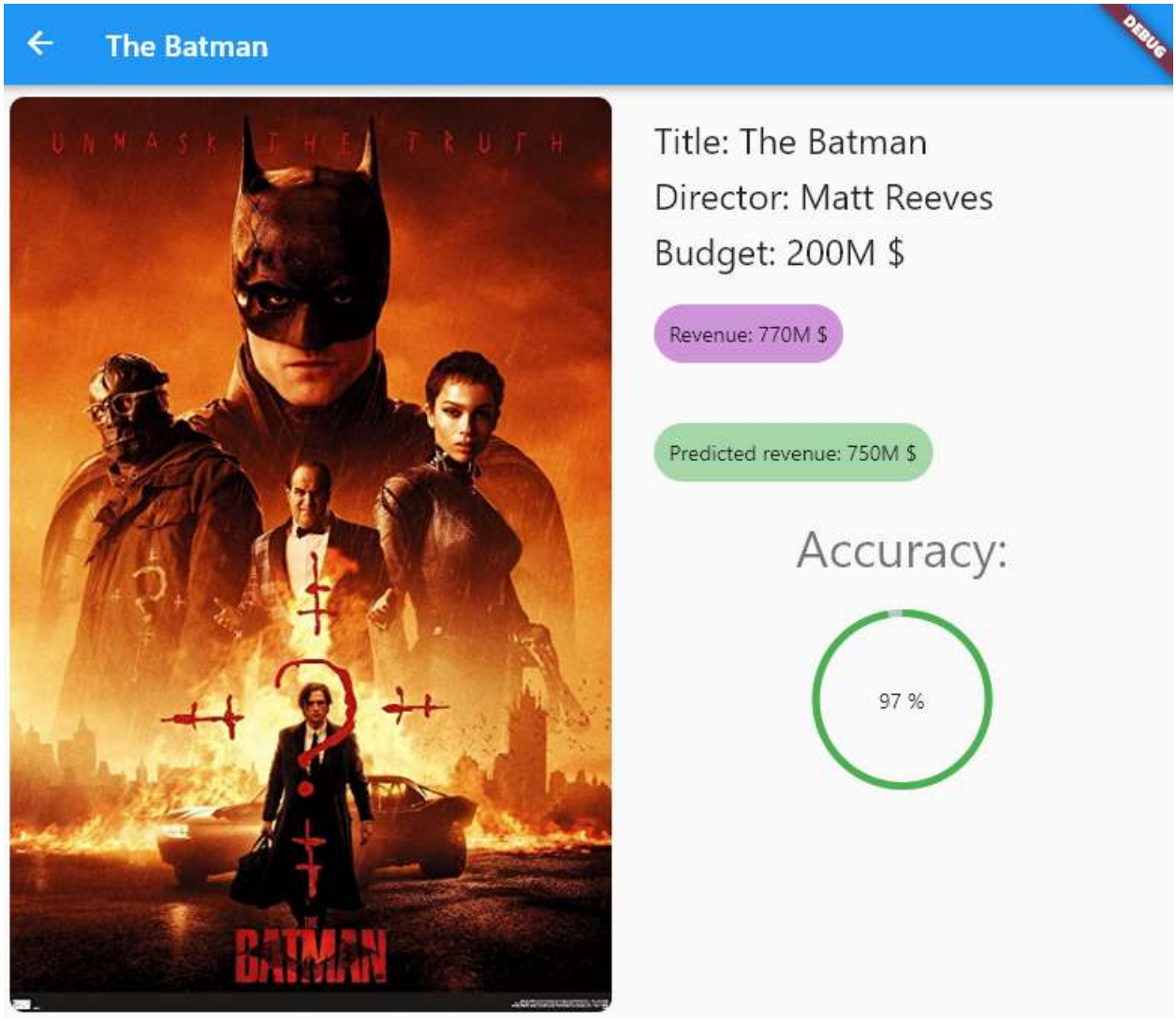


Рисунок 4.11 – Вікно програми з результатом прогнозування успішності кінофільму

З рисунка 4.5 видно, що програма для кінофільму “The Batman” з реальним доходом в 770 мільйонів доларів США, надала прогноз доходу в 750

мільйонів. Для даного випадку, точність прогнозу досягаю 97%, що цілком відповідає поставленим вимогам.

Для доведення факту досягнення поставленої в роботі мети – підвищення точності прогнозування успішності кінофільму – було проведено тестування роботи розробленої програми та аналогічних методів. Програми-аналоги не були розглянуті так як таких немає в публічному доступі. Тестова вибірка складає 600 кінофільмів з набору даних *TMDB*. У якості метрики, для порівняння точності різних методів прогнозування, обрано середню абсолютну похибку (*mae*) [*].

$$mae = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n}, \tag{4.1}$$

де y_i – передбачення; x_i – справжнє значення; n – загальна кількість тестів.

Результати тестування подані в таблиці 4.1.

Таблиця 4.1 – Результати тестування розробленої програми та аналогів

Метод прогнозування	Кількість кінофільмів у тестовій вибірці	Кількість прогонів	Точність прогнозування (на основі *mae*), %
Лінійна регресія	1000	4	61.96
Random forest	1000	4	54
Gradien Boosting	1000	4	85.37
Регресія на основі нейронної мережі (запропонований метод)	1000	4	94

Із таблиці 4.1 помітно, що розроблена програма характеризується вищою точністю проказування (94%), ніж аналоги (85%), а це означає, що

достовірність прогнозування успішності кінофільму у розробленій програмі підвищена на 9%, що свідчить про те, що мета роботи досягнута.

На рисунку 4.12 зображено фрагмент коду для обрахунку точності нейромережевого методу.

```
def validate():
    x_train, x_test, y_train, y_test, _, _, _ = samples

    k = 4
    num_val_samples = len(x_train) // k
    num_epochs = 100
    all_scores = []

    model: models.Sequential = models.load_model('trained_model.h5')
    for i in range(k):
        print(f'Processing # {i}')

        val_data = x_train[
            i * num_val_samples: (i + 1) * num_val_samples
        ]
        val_targets = y_train[
            i * num_val_samples: (i + 1) * num_val_samples
        ]

        val_mse, val_mae = model.evaluate(val_data, val_targets, verbose=0)
        all_scores.append(val_mae)

    print(all_scores)
    print(np.mean(all_scores))
```

Рисунок 4.12 – Фрагмент коду тестування нейромережевої моделі

Таким чином, після проведення порівняння розробленої програми прогнозування успішності кінофільму на основі нейромережі з аналогами, можна зробити висновок, що розроблена програма має більшу точність на 9%. Деякий ілюстративний матеріал до програми (у т.ч. скріншоти) подано в додатку В. Інструкцію користування розробленою програмою наведено у додатку Г.

4.5 Висновок до розділу 4

У розділі обґрунтовано вибір мови програмування *Python* для серверної частини, та *Dart* – для клієнтської частини. У результаті було розроблено програмне забезпечення прогнозування успішності кінофільму, створену мовою програмування *Python* із застосуванням безкоштовної бібліотек: *keras*, *tensorflow*, *fastapi*, *sqlmodel*, *urlx*, *iter-model*. Було проведено тестування програми нейромережевого проказування успішності кінофільму. Аналіз результатів тестування показує, що точність прогнозування складає 94%, у той час коли точність аналогів складає 85% при навчанні на однаковій навчальній множині у кількість 2400 кінофільмів. Тобто мета магістерської кваліфікаційної роботи досягнута – точність прогнозування успішності кінофільму підвищена на 9%.

5 ЕКОНОМІЧНА ЧАСТИНА

Науково-технічна розробка має право на існування та впровадження, якщо вона відповідає вимогам часу, як в напрямку науково-технічного прогресу та і в плані економіки. Тому для науково-дослідної роботи необхідно оцінювати економічну ефективність результатів виконаної роботи.

Магістерська кваліфікаційна робота з розробки та дослідження «Інформаційна технологія прогнозування успішності кінофільму» відноситься до науково-технічних робіт, які орієнтовані на виведення на ринок (або рішення про виведення науково-технічної розробки на ринок може бути прийнято у процесі проведення самої роботи), тобто коли відбувається так звана комерціалізація науково-технічної розробки. Цей напрямок є пріоритетним, оскільки результатами розробки можуть користуватися інші споживачі, отримуючи при цьому певний економічний ефект. Але для цього потрібно знайти потенційного інвестора, який би взявся за реалізацію цього проекту і переконати його в економічній доцільності такого кроку.

Для наведеного випадку нами мають бути виконані такі етапи робіт:

1) проведено комерційний аудит науково-технічної розробки, тобто встановлення її науково-технічного рівня та комерційного потенціалу;

2) розраховано витрати на здійснення науково-технічної розробки;

3) розрахована економічна ефективність науково-технічної розробки у випадку її впровадження і комерціалізації потенційним інвестором і проведено обґрунтування економічної доцільності комерціалізації потенційним інвестором.

5.1 Проведення комерційного та технологічного аудиту науково-технічної розробки

Метою проведення комерційного і технологічного аудиту дослідження за темою «Інформаційна технологія прогнозування успішності кінофільму» є

оцінювання науково-технічного рівня та рівня комерційного потенціалу розробки, створеної в результаті науково-технічної діяльності.

З появою Інтернету сильно зросла кількість інформації, з якої люди щодня стикаються. Це означає, що люди повинні орієнтуватися серед надзвичайно великої кількості доступних альтернатив, коли хочуть щось знайти. Наприклад, від вибору нового мобільного телефону або плеєра до пошуку кінофільму для вечірнього перегляду. З іншого боку виступають власники інтернет-магазинів і сервісів: вони зацікавлені в персональній рекламі і рекомендаціях кожному конкретному користувачеві, тому що такий підхід може істотно збільшити прибуток компаній. Як результат, в останні роки інтерес до розробки та вдосконалення існуючих рекомендаційних систем значно виріс.

Оцінювання науково-технічного рівня розробки та її комерційного потенціалу рекомендується здійснювати із застосуванням 5-ти бальної системи оцінювання за 12-ма критеріями, наведеними в табл. 4.1 [28].

Таблиця 5.1 – Рекомендовані критерії оцінювання науково-технічного рівня і комерційного потенціалу розробки та бальна оцінка

Бали (за 5-ти бальною шкалою)					
	0	1	2	3	4
Технічна здійсненність концепції					
1	Достовірність концепції не підтверджена	Концепція підтверджена експертними висновками	Концепція підтверджена розрахунками	Концепція перевірена на практиці	Перевірено працездатність продукту в реальних умовах
Ринкові переваги (недоліки)					
2	Багато аналогів на малому ринку	Мало аналогів на малому ринку	Кілька аналогів на великому ринку	Один аналог на великому ринку	Продукт не має аналогів на великому ринку
3	Ціна продукту значно вища за ціни аналогів	Ціна продукту дещо вища за ціни аналогів	Ціна продукту приблизно дорівнює цінам аналогів	Ціна продукту дещо нижче за ціни аналогів	Ціна продукту значно нижче за ціни аналогів

Продовження таблиці 5.1

4	Технічні та споживчі властивості продукту значно гірші, ніж в аналогів	Технічні та споживчі властивості продукту трохи гірші, ніж в аналогів	Технічні та споживчі властивості продукту на рівні аналогів	Технічні та споживчі властивості продукту трохи кращі, ніж в аналогів	Технічні та споживчі властивості продукту значно кращі, ніж в аналогів
5	Експлуатаційні витрати значно вищі, ніж в аналогів	Експлуатаційні витрати дещо вищі, ніж в аналогів	Експлуатаційні витрати на рівні експлуатаційних витрат аналогів	Експлуатаційні витрати трохи нижчі, ніж в аналогів	Експлуатаційні витрати значно нижчі, ніж в аналогів
Ринкові перспективи					
6	Ринок малий і не має позитивної динаміки	Ринок малий, але має позитивну динаміку	Середній ринок з позитивною динамікою	Великий стабільний ринок	Великий ринок з позитивною динамікою
7	Активна конкуренція великих компаній на	Активна конкуренція	Помірна конкуренція	Незначна конкуренція	Конкурентів немає
Практична здійсненність					
8	Відсутні фахівці як з технічної, так і з комерційної реалізації ідеї	Необхідно наймати фахівців або витрачати значні кошти та час на навчання наявних фахівців	Необхідне незначне навчання фахівців та збільшення їх штату	Необхідне незначне навчання фахівців	Є фахівці з питань як з технічної, так і з комерційної реалізації ідеї
9	Потрібні значні фінансові ресурси, які відсутні. Джерела фінансування ідеї відсутні	Потрібні незначні фінансові ресурси. Джерела фінансування відсутні	Потрібні значні фінансові ресурси. Джерела фінансування є	Потрібні незначні фінансові ресурси. Джерела фінансування є	Не потребує додаткового фінансування
10	Необхідна розробка нових матеріалів	Потрібні матеріали, що використовуються у військово промисловому комплексі	Потрібні дорогі матеріали	Потрібні досяжні та дешеві матеріали	Всі матеріали для реалізації ідеї відомі та давно використовуютьс я у виробництві

Продовження таблиці 5.1

11	Термін реалізації ідеї більший за 10 років	Термін реалізації ідеї більший за 5 років. Термін окупності інвестицій більше 10-ти років	Термін реалізації ідеї від 3-х до 5-ти років. Термін окупності інвестицій більше 5-ти років	Термін реалізації ідеї менше 3-х років. Термін окупності інвестицій від 3-х до 5-ти років	Термін реалізації ідеї менше 3-х років. Термін окупності інвестицій менше 3-х років
12	Необхідна розробка регламентних документів та отримання великої кількості дозвільних документів на виробництво та реалізацію продукту	Необхідно отримання великої кількості дозвільних документів на виробництво та реалізацію продукту, що вимагає значних коштів та часу	Процедура отримання дозвільних документів для виробництва та реалізації продукту вимагає незначних коштів та часу	Необхідно тільки повідомлення відповідним органам про виробництво та реалізацію продукту	Відсутні будь- які регламентні обмеження на виробництво та реалізацію продукту

Результати оцінювання науково-технічного рівня та комерційного потенціалу науково-технічної розробки потрібно звести до таблиці.

Таблиця 5.2 – Результати оцінювання науково-технічного рівня і комерційного потенціалу розробки експертами

Критерії	Експерт (ПІБ, посада)		
	1	2	3
	Бали:		
1. Технічна здійсненність концепції	5	5	5
2. Ринкові переваги (наявність аналогів)	4	5	4
3. Ринкові переваги (ціна продукту)	4	4	4
4. Ринкові переваги (технічні властивості)	5	4	5
5. Ринкові переваги (експлуатаційні витрати)	1	1	1
6. Ринкові перспективи (розмір ринку)	3	3	3
7. Ринкові перспективи (конкуренція)	1	2	1
8. Практична здійсненність (наявність фахівців)	3	3	3
9. Практична здійсненність (наявність фінансів)	3	4	4
10. Практична здійсненність (необхідність нових матеріалів)	4	4	4
11. Практична здійсненність (термін реалізації)	4	4	4
12. Практична здійсненність (розробка документів)	4	4	5
Сума балів	41	43	43
Середньоарифметична сума балів $СБ_c$	42,3		

За результатами розрахунків, наведених в таблиці 5.2, зробимо висновок щодо науково-технічного рівня і рівня комерційного потенціалу розробки. При цьому використаємо рекомендації, наведені в табл. 5.3 [28].

Таблиця 5.3 – Науково-технічні рівні та комерційні потенціали розробки

Середньоарифметична сума балів СБ ,розрахована на основі висновків експертів	Науково-технічний рівень та комерційний потенціал розробки
41…48	Високий
31…40	Вище середнього
21…30	Середній
11…20	Нижче середнього
0…10	Низький

Згідно проведених досліджень рівень комерційного потенціалу розробки за темою «Інформаційна технологія прогнозування успішності кінофільму» становить 42,3 бала, що, відповідно до таблиці 4.3, свідчить про комерційну важливість проведення даних досліджень (рівень комерційного потенціалу розробки високий).

Перед переглядом кінофільмів компаніями має забезпечуватись інформування глядачів про кінофільми, їхній бюджет, прогнозований успіх та касовий збір. Перед тим, як замовляти білети до кінотеатру користувач може перевірити, чи потрібно йому витрачати гроші та час на прем’єру конкретного фільму, чи ні. Прогнозування успіху – це саме те, чого потребує користувач, який буде використовувати розроблений веб-сервіс. Основні відмінності від конкурентів полягають в наступному: 1) конкурентів у відкритому доступі майже немає; 2) покращений інтерфейс для користувача; 3) збільшена точність прогнозування; 4) забезпечена можливість додавання власних даних про кінофільм, та отримувати на основі цього прогнозування, в той час як у конкурентів можна переглянути тільки вже існуючі кінофільми; 5) програмний продукт також відображає, який параметр найбільше повипливав

на прогноз.

Використані програмні засоби:

IDE - PyCharm, мова програмування Python, Google Cloud Platform - для розгортання серверної частини та бази даних, Firebase - для хостингу клієнтської частини.

5.2 Розрахунок узагальненого коефіцієнта якості розробки

Окрім комерційного аудиту розробки доцільно також розглянути технічний рівень якості розробки, розглянувши її основні технічні показники. Ці показники по-різному впливають на загальну якість проектної розробки.

Узагальнений коефіцієнт якості ($B_н$) для нового технічного рішення розрахуємо за формулою [29]:

$$B_н = \sum_{i=1}^{k} \alpha_i \cdot \beta_i , \qquad (5.1)$$

де k – кількість найбільш важливих технічних показників, які впливають на якість нового технічного рішення;

α_i – коефіцієнт, який враховує питому вагу *i-го* технічного показника в загальній якості розробки. Коефіцієнт α_i визначається експертним шляхом і при цьому має виконуватись умова $\sum_{i=1}^{k} \alpha_i = 1$;

β_i – відносне значення *i-го* технічного показника якості нової розробки.

Відносні значення β_i для різних випадків розраховуємо за такими формулами:

- для показників, зростання яких вказує на підвищення в лінійній залежності якості нової розробки:

$$\beta_i = \frac{I_{нi}}{I_{ai}} , \qquad (5.2)$$

де $I_{нi}$ та $I_{на}$ – чисельні значення конкретного *i*-го технічного показника

якості відповідно для нової розробки та аналога;

- для показників, зростання яких вказує на погіршення в лінійній залежності якості нової розробки:

$$\beta_i = \frac{I_{ai}}{I_{ні}}; \qquad (5.3)$$

Використовуючи наведені залежності можемо проаналізувати та порівняти техніко-економічні характеристики аналогу та розробки на основі отриманих наявних та проектних показників, а результати порівняння зведемо до таблиці 5.4.

Таблиця 5.4 – Порівняння основних параметрів розробки та аналога

Показники (параметри)	Одиниця вимірю-вання	Аналог	Проектований пристрій	Відношення параметрів нової розробки до аналога	Питома вага показника
Задіяні операційні системи	-	2	2	1	0,15
Дружність інтерфейсу	бал	6	8	1,33	0,1
Кількість базових характеристик розпізнавання	шт.	32	64	2	0,3
Точність обробки прогнозу	%	85	94	1,11	0,25
Функціональність	шт.	3	6	2	0,2

Узагальнений коефіцієнт якості ($B_н$) для нового технічного рішення складе:

$$B_н = \sum_{i=1}^{k} \alpha_i \cdot \beta_i = 1 \cdot 0{,}15 + 1{,}33 \cdot 0{,}1 + 2 \cdot 0{,}3 + 1{,}11 \cdot 0{,}25 + 2 \cdot 0{,}2 = 1{,}56.$$

Отже за технічними параметрами, згідно узагальненого коефіцієнту якості розробки, науково-технічна розробка переважає існуючі аналоги приблизно в 1,56 рази.

5.3 Розрахунок витрат на проведення науково-дослідної роботи

Витрати, пов’язані з проведенням науково-дослідної роботи на тему «Інформаційна технологія прогнозування успішності кінофільму», під час планування, обліку і калькулювання собівартості науково-дослідної роботи групуємо за відповідними статтями.

5.3.1 Витрати на оплату праці

До статті «Витрати на оплату праці» належать витрати на виплату основної та додаткової заробітної плати керівникам відділів, лабораторій, секторів і груп, науковим, інженерно-технічним працівникам, конструкторам, технологам, креслярам, копіювальникам, лаборантам, робітникам, студентам, аспірантам та іншим працівникам, безпосередньо зайнятим виконанням конкретної теми, обчисленої за посадовими окладами, відрядними розцінками, тарифними ставками згідно з чинними в організаціях системами оплати праці.

Основна заробітна плата дослідників

Витрати на основну заробітну плату дослідників ($З_о$) розраховуємо у відповідності до посадових окладів працівників, за формулою [28]:

$$З_о = \sum_{i=1}^{k} \frac{M_{ni} \cdot t_i}{T_p}, \qquad (5.4)$$

де k – кількість посад дослідників залучених до процесу досліджень;

M_{ni} – місячний посадовий оклад конкретного дослідника, грн;

t_i – число днів роботи конкретного дослідника, дн.;

T_p – середнє число робочих днів в місяці, T_p=21 дні.

$$З_о = 14830{,}00 \cdot 34 / 21 = 24010{,}48 \; грн.$$

Проведені розрахунки зведемо до таблиці 5.5.

Таблиця 5.5 – Витрати на заробітну плату дослідників

Найменування посади	Місячний посадовий оклад, грн	Оплата за робочий день, грн	Число днів роботи	Витрати на заробітну плату, грн
Керівник проекту	14830,00	706,19	34	24010,48
Консультант (менеджер сфери прокату фільмів)	12720,00	605,71	15	9085,71
Інженер-програміст	12100,00	576,19	20	11523,81
Фахівець з аналітично-математичних досліджень	12000,00	571,43	20	11428,57
Консультант-аналітик цифрових обчислюваних систем	13800,00	657,14	15	9857,14
Лаборант	6800,00	323,81	20	6476,19
Всього				72381,90

Основна заробітна плата робітників

Витрати на основну заробітну плату робітників ($З_р$) за відповідними найменуваннями робіт НДР на тему «Інформаційна технологія прогнозування успішності кінофільму» розраховуємо за формулою:

$$З_р = \sum_{i=1}^{n} C_i \cdot t_i , \qquad (5.5)$$

де C_i – погодинна тарифна ставка робітника відповідного розряду, за виконану відповідну роботу, грн/год;

t_i – час роботи робітника при виконанні визначеної роботи, год.

Погодинну тарифну ставку робітника відповідного розряду C_i можна визначити за формулою:

$$C_i = \frac{М_М \cdot К_i \cdot К_с}{Т_р \cdot t_{зм}} , \qquad (5.6)$$

де $М_М$ – розмір прожиткового мінімуму працездатної особи, або мінімальної місячної заробітної плати (в залежності від діючого законодавства), приймемо $Мм$=6700,00 грн;

$К_i$ – коефіцієнт міжкваліфікаційного співвідношення для встановлення тарифної ставки робітнику відповідного розряду (табл. Б.2, додаток Б) [28];

$К_с$ – мінімальний коефіцієнт співвідношень місячних тарифних ставок робітників першого розряду з нормальними умовами праці виробничих об'єднань і підприємств до законодавчо встановленого розміру мінімальної заробітної плати.

$Т_р$ – середнє число робочих днів в місяці, приблизно $Т_р$ = 21 дн;

$t_{зм}$ – тривалість зміни, год.

$С_1$ = 6700,00 · 1,10 · 1,7 / (21 · 8) = 74,58 *грн.*
$З_{р1}$ = 74,58 · 8,00 = 596,62 *грн.*

Таблиця 5.6 – Величина витрат на основну заробітну плату робітників

Найменування робіт	Тривалість роботи, год	Розряд роботи	Тарифний коефіцієнт	Погодинна тарифна ставка, грн	Величина оплати на робітника грн
Встановлення допоміжного офісного обладнання	8,00	2	1,10	74,58	596,62
Монтаж робочого місця розробника системи прогнозування	12,00	2	1,10	74,58	894,93
Інсталяція програмного забезпечення	5,00	5	1,70	115,26	576,28
Встановлення цифрових обчислювальних систем	3,00	4	1,50	101,70	305,09
Відлагодження інтерполяційних модулів	7,00	5	1,70	115,26	806,79
Тренування цифрової експериментальної моделі	4,50	4	1,50	101,70	457,63
Формування бази даних прогнозного аналізу	16,00	3	1,35	91,53	1464,43
Інші допоміжні роботи	10,00	3	1,35	91,53	915,27
Всього					6017,04

Додаткова заробітна плата дослідників та робітників

Додаткову заробітну плату розраховуємо як 10 … 12% від суми основної заробітної плати дослідників та робітників за формулою:

$$З_{дод} = (З_о + З_р) \cdot \frac{Н_{дод}}{100\%}, \quad (5.7)$$

де $Н_{дод}$ – норма нарахування додаткової заробітної плати. Приймемо 10%.

$З_{дод}$ = (72381,90 + 6017,04) · 10 / 100% = 7839,89 грн.

5.3.2 Відрахування на соціальні заходи

Нарахування на заробітну плату дослідників та робітників розраховуємо як 22% від суми основної та додаткової заробітної плати дослідників і робітників за формулою:

$$З_н = (З_о + З_р + З_{дод}) \cdot \frac{Н_{зп}}{100\%} \quad (5.8)$$

де $Н_{зп}$ – норма нарахування на заробітну плату. Приймаємо 22%.

Зн = (72381,90 + 6017,04 + 7839,89) · 22 / 100% = 18972,54 грн.

5.3.3 Сировина та матеріали

До статті «Сировина та матеріали» належать витрати на сировину, основні та допоміжні матеріали, інструменти, пристрої та інші засоби і предмети праці, які придбані у сторонніх підприємств, установ і організацій та витрачені на проведення досліджень за темою «Інформаційна технологія прогнозування успішності кінофільму».

Витрати на матеріали (*М*), у вартісному вираженні розраховуються окремо по кожному виду матеріалів за формулою:

$$M = \sum_{j=1}^{n} Н_j \cdot Ц_j \cdot К_j - \sum_{j=1}^{n} В_j \cdot Ц_{вj}, \quad (5.9)$$

де $Н_j$ – норма витрат матеріалу *j*-го найменування, кг;

n – кількість видів матеріалів;

$Ц_j$ – вартість матеріалу *j*-го найменування, грн/кг;

$К_j$ – коефіцієнт транспортних витрат, ($К_j$ = *1,1 ... 1,15*);

$В_j$ – маса відходів *j*-го найменування, кг;

$Ц_{вj}$ – вартість відходів *j*-го найменування, грн/кг.

$М_1$ = 3,00 · 265,00 · 1,1 - 0,000 · 0,00 = 874,50 грн.

Проведені розрахунки зведемо до таблиці 5.7.

Таблиця 5.7 – Витрати на матеріали

Найменування матеріалу, марка, тип, сорт	Ціна за 1 кг, грн	Норма витрат, кг	Величина відходів, кг	Ціна відходів, грн/кг	Вартість витраченого матеріалу, грн
Папір канцелярський офісний ECONOMIC (A4-500)	265,00	3,00	-	-	874,50
Папір для заміток ECONOMIC (A5)-60	162,00	4,00	-	-	712,80
Начиння канцелярське DATUM FX	22,00	3,00	-	-	72,60
Органайзер офісний DATUM Office	150,00	3,00	-	-	495,00
Картридж для принтера HP-210A	1250,00	2,00	-	-	2750,00
Диск оптичний VEKO-10 (CD-R)	22,00	1,00	-	-	24,20
Диск оптичний VEKO-W (CD-RW)	23,00	1,00	-	-	25,30
FLASH-пам'ять Kingstar (32 ГБ) Class 10	340,00	1,00	-	-	372,00
FLASH-пам'ять Kingstar (64 ГБ) Class 10 A	680,00	1,000	-	-	748,00
Всього					6072,10

5.3.4 Розрахунок витрат на комплектуючі

Витрати на комплектуючі *($К_в$)*, які використовують при проведенні НДР на тему «Інформаційна технологія прогнозування успішності кінофільму» відсутні.

5.3.5 Спецустаткування для наукових (експериментальних) робіт

До статті «Спецустаткування для наукових (експериментальних) робіт» належать витрати на виготовлення та придбання спецустаткування необхідного для проведення досліджень, також витрати на їх проектування, виготовлення, транспортування, монтаж та встановлення.

Балансову вартість спецустаткування розраховуємо за формулою:

$$B_{спец} = \sum_{i=1}^{k} Ц_i \cdot C_{пр.і} \cdot K_i \ , \qquad (5.10)$$

де $Ц_i$ – ціна придбання одиниці спецустаткування даного виду, марки, грн;

$C_{пр.і}$ –кількість одиниць устаткування відповідного найменування, які придбані для проведення досліджень, шт.;

K_i – коефіцієнт, що враховує доставку, монтаж, налагодження устаткування тощо, (K_i = *1,10…1,12);*

k – кількість найменувань устаткування.

$B_{спец}$ = *40250,00 · 1 · 1,11 = 44677,50 грн.*

Отримані результати зведемо до таблиці 5.8.

Таблиця 5.8 – Витрати на придбання спецустаткування по кожному виду

Найменування устаткування	Кількість, шт	Ціна за одиницю, грн	Вартість, грн
Мультимедійна проекційна система	1	40250,00	44677,50
Всього			44677,50

5.3.6 Програмне забезпечення для наукових (експериментальних) робіт

До статті «Програмне забезпечення для наукових (експериментальних) робіт» належать витрати на розробку та придбання спеціальних програмних засобів і програмного забезпечення, (програм, алгоритмів, баз даних) необхідних для проведення досліджень, також витрати на їх проектування, формування та встановлення.

Балансову вартість програмного забезпечення розраховуємо за формулою:

$$B_{прг} = \sum_{i=1}^{k} Ц_{iпрг} \cdot C_{прг.і} \cdot K_i \ , \qquad (5.11)$$

де $Ц_{iпрг}$ – ціна придбання одиниці програмного засобу даного виду, грн;

$C_{прг.і}$ – кількість одиниць програмного забезпечення відповідного найменування, які придбані для проведення досліджень, шт.;

$K_і$ – коефіцієнт, що враховує інсталяцію, налагодження програмного засобу тощо, ($K_і$ = *1,10…1,12);*

k – кількість найменувань програмних засобів.

$B_{прг}$ = *8410,00 · 1 · 1,1 = 9251,00 грн.*

Отримані результати зведемо до таблиці 5.9.

Таблиця 5.9– Витрати на придбання програмних засобів по кожному виду

Найменування програмного засобу	Кількість, шт	Ціна за одиницю, грн	Вартість, грн
ОС Windows	1	8410,00	9251,00
Прикладний пакет Microsoft Office	1	7790,00	8569,00
Програмний засіб IDE – PyCharm	1	6580,00	7238,00
Google Cloud Platform	1	6830,00	7513,00
Firebase	1	5935,00	6528,50
Всього			39099,50

5.3.7 Амортизація обладнання, програмних засобів та приміщень

В спрощеному вигляді амортизаційні відрахування по кожному виду обладнання, приміщень та програмному забезпеченню тощо, розраховуємо з використанням прямолінійного методу амортизації за формулою:

$$A_{обл} = \frac{Ц_б}{T_в} \cdot \frac{t_{вик}}{12}, \quad (5.12)$$

де $Ц_б$ – балансова вартість обладнання, програмних засобів, приміщень тощо, які використовувались для проведення досліджень, грн;

$t_{вик}$ – термін використання обладнання, програмних засобів, приміщень під час досліджень, місяців;

$T_в$ – строк корисного використання обладнання, програмних засобів, приміщень тощо, років.

$A_{обл}$ = *(28300,00 · 2) / (3 · 12) = 1572,22 грн.*

Проведені розрахунки зведемо до таблиці 5.10.

Таблиця 5.10 – Амортизаційні відрахування по кожному виду обладнання

Найменування обладнання	Балансова вартість, грн	Строк корисного використання, років	Термін використання обладнання, місяців	Амортизаційні відрахування, грн
Програмно-аналітичний комплекс	28300,00	3	2	1572,22
Графічно-обчислювальний комплекс обробки даних	26400,00	4	2	1100,00
Програмні засоби реалізації Python	6890,00	2	2	574,17
Обладнання виводу інформації	10250,00	5	2	341,67
Місце оператора спеціалізоване	9100,00	4	2	379,17
Офісна оргтехніка	9600,00	5	2	320,00
Дослідницька лабораторія	300000,00	25	2	2000,00
Всього				6287,22

5.3.8 Паливо та енергія для науково-виробничих цілей

Витрати на силову електроенергію (B_e) розраховуємо за формулою:

$$B_e = \sum_{i=1}^{n} \frac{W_{yi} \cdot t_i \cdot Ц_e \cdot K_{вni}}{\eta_i}, \quad (5.13)$$

де W_{yi} – встановлена потужність обладнання на визначеному етапі розробки, кВт;

t_i – тривалість роботи обладнання на етапі дослідження, год;

$Ц_e$ – вартість 1 кВт-години електроенергії, грн; (вартість електроенергії визначається за даними енергопостачальної компанії), приймемо $Ц_e = 6,20$ грн;

$K_{вni}$ – коефіцієнт, що враховує використання потужності, $K_{вni} < 1$;

η_i – коефіцієнт корисної дії обладнання, $\eta_i < 1$.

$B_e = 0,28 \cdot 240,0 \cdot 6,20 \cdot 0,95 / 0,97 = 416,64$ *грн.*

Проведені розрахунки зведемо до таблиці 5.11.

Таблиця 5.11 – Витрати на електроенергію

Найменування обладнання	Встановлена потужність, кВт	Тривалість роботи, год	Сума, грн
Програмно-аналітичний комплекс	0,28	240,0	416,64
Графічно-обчислювальний комплекс обробки даних	0,32	200,0	396,80
Мультимедійна проекційна система	0,32	200,0	396,80
Обладнання виводу інформації	0,35	40,0	86,80
Місце оператора спеціалізоване	0,46	200,0	570,40
Офісна оргтехніка	0,80	24,0	119,04
Всього			1986,48

5.3.9 Службові відрядження

До статті «Службові відрядження» дослідної роботи на тему «Інформаційна технологія прогнозування успішності кінофільму» належать витрати на відрядження штатних працівників, працівників організацій, які працюють за договорами цивільно-правового характеру, аспірантів, зайнятих розробленням досліджень, відрядження, пов’язані з проведенням випробувань машин та приладів, а також витрати на відрядження на наукові з’їзди, конференції, наради, пов’язані з виконанням конкретних досліджень.

Витрати за статтею «Службові відрядження» розраховуємо як 20…25% від суми основної заробітної плати дослідників та робітників за формулою:

$$B_{св} = (З_о + З_р) \cdot \frac{Н_{св}}{100\%}, \qquad (5.15)$$

де $Н_{св}$ – норма нарахування за статтею «Службові відрядження», приймемо $Н_{св}$ = 25%.

$B_{св}$ = (72381,90 + 6017,04) · 25 / 100% = 19599,74 грн.

5.3.10 Витрати на роботи, які виконують сторонні підприємства, установи і організації

Витрати за статтею «Витрати на роботи, які виконують сторонні підприємства, установи і організації» розраховуємо як 30…45% від суми основної заробітної плати дослідників та робітників за формулою:

$$B_{сп} = (З_о + З_р) \cdot \frac{Н_{сп}}{100\%}, \quad (5.16)$$

де $Н_{сп}$ – норма нарахування за статтею «Витрати на роботи, які виконують сторонні підприємства, установи і організації», приймемо $Н_{сп}$= 40%.

$В_{сп}$ = (72381,90 + 6017,04) · 40 / 100% = 31359,58 грн.

5.3.11 Інші витрати

До статті «Інші витрати» належать витрати, які не знайшли відображення у зазначених статтях витрат і можуть бути віднесені безпосередньо на собівартість досліджень за прямими ознаками.

Витрати за статтею «Інші витрати» розраховуємо як 50…100% від суми основної заробітної плати дослідників та робітників за формулою:

$$І_в = (З_о + З_р) \cdot \frac{Н_{ів}}{100\%}, \quad (5.17)$$

де $Н_{ів}$ – норма нарахування за статтею «Інші витрати», приймемо $Н_{ів}$ = 100%.

$І_в$ = (72381,90 + 6017,04) · 100 / 100% = 78398,94 грн.

5.3.12 Накладні (загальновиробничі) витрати

До статті «Накладні (загальновиробничі) витрати» належать: витрати, пов’язані з управлінням організацією; витрати на винахідництво та раціоналізацію; витрати на підготовку (перепідготовку) та навчання кадрів; витрати, пов’язані з набором робочої сили; витрати на оплату послуг банків;

витрати, пов’язані з освоєнням виробництва продукції; витрати на науково-технічну інформацію та рекламу та ін.

Витрати за статтею «Накладні (загальновиробничі) витрати» розраховуємо як 100…150% від суми основної заробітної плати дослідників та робітників за формулою:

$$B_{нзв} = (З_о + З_р) \cdot \frac{Н_{нзв}}{100\%}, \quad (5.18)$$

де $Н_{нзв}$ – норма нарахування за статтею «Накладні (загальновиробничі) витрати», приймемо $Н_{нзв}$ = 150%.

$В_{нзв}$ = (72381,90 + 6017,04) · 150 / 100% = 117598,42 грн.

Витрати на проведення науково-дослідної роботи на тему «Інформаційна технологія прогнозування успішності кінофільму» розраховуємо як суму всіх попередніх статей витрат за формулою:

$$B_{заг} = З_о + З_р + З_{дод} + З_н + М + К_в + В_{спец} + В_{прг} + А_{обл} + В_е + В_{св} + В_{сп} + І_в + В_{нзв}. \quad (5.19)$$

$В_{заг}$= 72381,90 +6017,04 +7839,89 +18972,54432 +6072,10 +0,00 +44677,50 + 39099,50 + 6287,22 +1986,48 +19599,74 +31359,58 +78398,94 +117598,42 = 450290,86 грн.

Загальні витрати *ЗВ* на завершення науково-дослідної (науково-технічної) роботи та оформлення її результатів розраховується за формулою:

$$ЗВ = \frac{B_{заг}}{\eta}, \quad (5.20)$$

де η - коефіцієнт, який характеризує етап (стадію) виконання науково-дослідної роботи, приймемо η=0,85.

ЗВ = 450290,86 / 0,85 = 529753,95 грн.

5.4 Розрахунок економічної ефективності науково-технічної розробки при її можливій комерціалізації потенційним інвестором

В ринкових умовах узагальнюючим позитивним результатом, що його може отримати потенційний інвестор від можливого впровадження результатів тієї чи іншої науково-технічної розробки, є збільшення у потенційного інвестора величини чистого прибутку.

Результати дослідження проведені за темою «Інформаційна технологія прогнозування успішності кінофільму» передбачають комерціалізацію протягом 4-х років реалізації на ринку.

В цьому випадку основу майбутнього економічного ефекту будуть формувати:

ΔN – збільшення кількості споживачів яким надається відповідна інформаційна послуга у періоди часу, що аналізуються;

Показник	1-й рік	2-й рік	3-й рік	4-й рік
Збільшення кількості споживачів, осіб	1000	1500	1100	650

N – кількість споживачів яким надавалась відповідна інформаційна послуга у році до впровадження результатів нової науково-технічної розробки, приймемо 5800 осіб;

$Ц_б$ – вартість послуги у році до впровадження інформаційної системи, приймемо 3600,00 грн;

$\pm\Delta Ц_о$ – зміна вартості послуги від впровадження результатів, приймемо 1897,80 грн.

Можливе збільшення чистого прибутку у потенційного інвестора $\Delta П_i$ для кожного із 4-х років, протягом яких очікується отримання позитивних результатів від можливого впровадження та комерціалізації науково-технічної розробки, розраховуємо за формулою [28]:

$$\Delta П_i = (\pm \Delta Ц_o \cdot N + Ц_o \cdot \Delta N)_i \cdot \lambda \cdot \rho \cdot (1 - \frac{\vartheta}{100}), \quad (5.21)$$

де λ – коефіцієнт, який враховує сплату потенційним інвестором податку на додану вартість. У 2022 році ставка податку на додану вартість складає 20%, а коефіцієнт $\lambda = 0,8333$;

ρ – коефіцієнт, який враховує рентабельність інноваційного продукту). Приймемо ρ =30%;

ϑ – ставка податку на прибуток, який має сплачувати потенційний інвестор, у 2022 році $\vartheta = 18\%$;

Збільшення чистого прибутку 1-го року:

$\Delta П_1$ =(1897,80·5800,00+5497,80·1000)·0,83·0,3·(1-0,18/100%)=3369999,07 грн.

Збільшення чистого прибутку 2-го року:

$\Delta П_2$ =(1897,80·5800,00+5497,80·2500)·0,83·0,3·(1-0,18/100%)=5053810,27 грн.

Збільшення чистого прибутку 3-го року:

$\Delta П_3$ =(1897,80·5800,00+5497,80·3600)·0,83·0,3·(1-0,18/100%)=6288605,16 грн.

Збільшення чистого прибутку 4-го року:

$\Delta П_4$ =(1897,80·5800,00+5497,80·4250)·0,83·0,3·(1-0,18/100%)=7018256,68 грн.

Приведена вартість збільшення всіх чистих прибутків *ПП*, що їх може отримати потенційний інвестор від можливого впровадження та комерціалізації науково-технічної розробки:

$$ПП = \sum_{i=1}^{T} \frac{\Delta П_i}{(1+\tau)^t}, \quad (5.22)$$

де $\Delta П_i$ – збільшення чистого прибутку у кожному з років, протягом яких виявляються результати впровадження науково-технічної розробки, грн;

T – період часу, протягом якого очікується отримання позитивних результатів від впровадження та комерціалізації науково-технічної розробки, роки;

τ – ставка дисконтування, за яку можна взяти щорічний прогнозований рівень інфляції в країні, τ =0,18;

t – період часу (в роках) від моменту початку впровадження науково-технічної розробки до моменту отримання потенційним інвестором додаткових чистих прибутків у цьому році.

$$ПП = 3369999,07/(1+0,18)^1+5053810,27/(1+0,18)^2+6288605,16/(1+0,18)^3+$$
$$+7018256,68/(1+0,18)^4=2855931,41+3629567,85+3827439,25+3619938,72=$$
$$=13932877,23 \text{ грн.}$$

Величина початкових інвестицій PV, які потенційний інвестор має вкласти для впровадження і комерціалізації науково-технічної розробки:

$$PV = k_{інв} \cdot ЗВ, \qquad (5.23)$$

де $k_{інв}$ – коефіцієнт, що враховує витрати інвестора на впровадження науково-технічної розробки та її комерціалізацію, приймаємо $k_{інв}$=1,5;

$ЗВ$ – загальні витрати на проведення науково-технічної розробки та оформлення її результатів, приймаємо 529753,95 грн.

$PV = k_{інв} \cdot ЗВ$ = 1,5 · 529753,95 = 794630,92 грн.

Абсолютний економічний ефект $E_{абс}$ для потенційного інвестора від можливого впровадження та комерціалізації науково-технічної розробки становитиме:

$$E_{абс} = ПП - PV \qquad (5.24)$$

де $ПП$ – приведена вартість зростання всіх чистих прибутків від можливого впровадження та комерціалізації науково-технічної розробки, 13932877,23 грн;

PV – теперішня вартість початкових інвестицій, 794630,92 грн.

$E_{абс} = ПП - PV$ = 13932877,23 - 794630,92 = 13138246,30 грн.

Внутрішня економічна прибутковість інвестицій $E_в$, які можуть бути вкладені потенційним інвестором у впровадження та комерціалізацію науково-технічної розробки:

$$E_в = \sqrt[Tж]{1 + \frac{E_{абс}}{PV}} - 1, \quad (5.25)$$

де $E_{абс}$ – абсолютний економічний ефект вкладених інвестицій, 13138246,30 грн;

PV – теперішня вартість початкових інвестицій, 794630,92 грн;

$T_ж$ – життєвий цикл науково-технічної розробки, тобто час від початку її розробки до закінчення отримування позитивних результатів від її впровадження, 4 роки.

$$E_в = \sqrt[Tж]{1 + \frac{E_{абс}}{PV}} - 1 = (1+13138246,30/794630,92)^{1/4} = 1,05.$$

Мінімальна внутрішня економічна прибутковість вкладених інвестицій $\tau_{мін}$

$$\tau_{мін} = d + f, \quad (5.26)$$

де d – середньозважена ставка за депозитними операціями в комерційних банках; в 2022 році в Україні d =0,12;

f – показник, що характеризує ризикованість вкладення інвестицій, приймемо 0,3.

$\tau_{мін}$ = 0,12+0,3 = 0,42 < 1,05 свідчить про те, що внутрішня економічна прибутковість інвестицій $E_в$, які можуть бути вкладені потенційним інвестором у впровадження та комерціалізацію науково-технічної розробки вища мінімальної внутрішньої прибутковості. Тобто інвестувати в науково-дослідну роботу за темою «Інформаційна технологія прогнозування успішності кінофільму» доцільно.

Період окупності інвестицій $T_{ок}$ які можуть бути вкладені потенційним інвестором у впровадження та комерціалізацію науково-технічної розробки:

$$T_{ок} = \frac{1}{E_в}, \tag{5.27}$$

де $E_в$ – внутрішня економічна прибутокність вкладених інвестицій; $T_{ок}$ = 1 / 1,05 = 0,96 р; $T_{ок}$ < 3-х років, що свідчить про комерційну привабливість науково-технічної розробки і може спонукати потенційного інвестора профінансувати впровадження даної розробки та виведення її на ринок.

5.4 Висновок до розділу 5

Згідно проведених досліджень рівень комерційного потенціалу розробки за темою «Інформаційна технологія прогнозування успішності кінофільму» становить 42,3 бала, що, свідчить про комерційну важливість проведення даних досліджень (рівень комерційного потенціалу розробки високий).

При оцінюванні за технічними параметрами, згідно узагальненого коефіцієнту якості розробки, науково-технічна розробка переважає існуючі аналоги приблизно в 1,56 рази.

Також термін окупності становить 0,96 р., що менше 3-х років, що свідчить про комерційну привабливість науково-технічної розробки і може спонукати потенційного інвестора профінансувати впровадження даної розробки та виведення її на ринок.

Отже можна зробити висновок про доцільність проведення науково-дослідної роботи за темою «Інформаційна технологія прогнозування успішності кінофільму».

ВИСНОВКИ

При виконанні магістерської кваліфікаційної роботи розв’язано задачу розробки інформаційної технології та програмного забезпечення розв'язання задачі прогнозування успішності кінофільму основі власної структури регресійної нейронної мережі.

В першому розділі проведено аналіз предметної області прогнозування успішності кінофільму, а саме – сформульовано постановку задачі, проведено огляд відомих методів розв'язання задачі прогнозування успішності кінофільму. Виходячи з аналізів предметної області, дало можливість визначити основні вимоги до системи прогнозування успішності кінофільму. Найперспективнішим було визначено метод на основі штучних нейронних мереж. Проведено аналіз існуючих методів, так виявлено, що можна підвищити точність прогнозування успішності кінофільмів.

У другому розділі магістерської кваліфікаційної роботи було проаналізовано ознаки кінофільмів, та визначено, ті які найбільше вливають на прибуток кінофільму. Запропоновано власну структура нейронної мережі на базі регресійної моделі. Кінцева модель складає: шар вхідних даних, два прихований шар, та один вихідний шар з 1 нейроном, для отримання скалярної величини на виході.

У третьому розділі наведено основні кроки роботи інформаційної технології прогнозування успішності кінофільму. Обрано клієнт-серверну архітектуру для програмного забезпечення прогнозування успішності кінофільму. Розбито програмний засіб на модулі, а саме: модуль взаємодії з базою даних, модуль імпортування кінофільмів, модуль редагування записів в базі даних, модуль прогнозування успішності кінофільму. Описано роботу та алгоритм кожного модуля. Спроектовано базу даних для програмного забезпечення прогнозування успішності кінофільмів.

У четвертому розділі обґрунтовано вибір мови програмування *Python* для серверної частини, та *Dart* – для клієнтської частини. У результаті було

розроблено програмне забезпечення прогнозування успішності кінофільму.

Було проведено тестування програми нейромережевого проказування успішності кінофільму. Аналіз результатів тестування показує, що точність прогнозування складає 94%, у той час коли точність аналогів складає 85% при навчанні на однаковій навчальній множині у кількість 5000 кінофільмів. Тобто мета магістерської кваліфікаційної роботи досягнута – точність прогнозування успішності кінофільму підвищена на 9%.

У п’ятому розділі було виконано розрахунок витрат на розробку та виготовлення нового технічного рішення. При оцінюванні за технічними параметрами, згідно узагальненого коефіцієнту якості розробки, науково-технічна розробка переважає існуючі аналоги приблизно в 1,56 рази. Термін окупності становить 0,96 р., що менше 3-х років, що свідчить про комерційну привабливість науково-технічної розробки і може спонукати потенційного інвестора профінансувати впровадження даної розробки та виведення її на ринок.

ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. В. М. Борисюк, С. В. Барабан, «Виділення ознак кінофільмів, що впливають на успішність кінофільму», в Матеріали конференції «Молодь в науці: дослідження, проблеми, перспективи (МН-2023)», Вінниця, 2023, [Електронний ресурс] - режим доступу: https://conferences.vntu.edu.ua/index.php/mn/mn2023/paper/view/16839/1404.

2. L. Zhang, J. Luo, and S. Yang, “Forecasting box office revenue of movies with bp neural network,” Expert Systems with Applications, 2009.

3. V. Biramane, H. Kulkarni, A. Bhave, and P. Kosamkar, “Relationships between classical factors, social factors and box office collections,”, 2016.

4. R. Masrury, M. Saputra, A. Alamsyah, and M. Primantari, “A comparative study of hollywood movie successfulness prediction model,” 2019.

5. S. S. Mohanbir, J. Eliashberg, «A Parsimonious Model for Forecasting Gross Box-Office Revenues of Motion Pictures», 1996.

6. J. Eliashberg et al, “The motion picture industry: critical issues in practice, current research, and new research directions”, Marketing Science, 2006.

7. Gagniuc, A. Paul, «Markov Chains: From Theory to Implementation and Experimentation», John Wiley & Sons, 2017. - 1–235 с.

8. D. A. Edwards, R. Buckmire, «A differential equation model of North American cinematic box-office dynamics», IMA Journal of management Mathematics, 2001.

9. R. Parimi and D. Caragea, “Pre-release box-office success prediction for motion pictures,”, 2013.

10. L. Bottou, O. Bousquet, «The Tradeoffs of Large-Scale Learning», Cambridge: MIT Press, 2012. - 51–368 с.

11. Офіційна сторінка TMDB [Електронний ресурс] – режим доступу: https://www.themoviedb.org/.

12. Word cloud [Електронний ресурс] – режим доступу: https://en.wikipedia.org/wiki/Tag_cloud.

13. *Wordcloud* документація [Електронний ресурс] – режим доступу: http://amueller.github.io/word_cloud/.

14. A Rajaraman, J. D. Ulman, «Mining of Massive Datasets», 2011. – 1-17 с.

15. Офіційна сторінка *elif*5 [Електронний ресурс] – режим доступу: https://eli5.readthedocs.io/en/latest/.

16. Коефіцієнт детермінації [Електронний ресурс] – режим доступу: https://en.wikipedia.org/wiki/Coefficient_of_determination

17. Середня квадратична похибка [Електронний ресурс] – режим доступу: https://www.probabilitycourse.com/chapter9/mean_squared_error.php

18. Chollet Francois, "Deep Learning with Python", 2018. — 111 с.

19. Dense layer [Електронний ресурс] – режим доступу: https://keras.io/api/layers/core_layers/dense/

20. BatchNormalization layer [Електронний ресурс] – режим доступу: https://keras.io/api/layers/normalization_layers/batch_normalization/

21. Activation layer [Електронний ресурс] – режим доступу: https://keras.io/api/layers/activations/

22. Dropout layer [Електронний ресурс] – режим доступу: https://keras.io/api/layers/regularization_layers/dropout/

23. "Порівняння архітектурних стилів API" [Електронний ресурс] – режим доступу: https://www.altexsoft.com/blog/soap-vs-rest-vs-graphql-vs-rpc/

24. Офіційна документація FastAPI [Електронний ресурс] – режим доступу: https://fastapi.tiangolo.com/

25. Офіційна документація Flutter [Електронний ресурс] – режим доступу: https://docs.flutter.dev/

26. Офіційна сторінка PyCharm [Електронний ресурс] – режим доступу: https://www.jetbrains.com/pycharm/

27. Офіційна сторінка IntelliJ IDEA [Електронний ресурс] – режим доступу: https://www.jetbrains.com/idea/

28. Методичні вказівки до виконання економічної частини магістерських кваліфікаційних робіт / Уклад. : В. О. Козловський, О. Й. Лесько, В. В. Кавецький. – Вінниця : ВНТУ, 2021. – 42 с.

29. Кавецький В. В. Економічне обґрунтування інноваційних рішень: практикум / В. В. Кавецький, В. О. Козловський, І. В. Причепа – Вінниця : ВНТУ, 2016. – 113 с.

ДОДАТКИ

Додаток А (обов'язковий)

Результат перевірки на плагіат в онлайн-системі UNICHECK

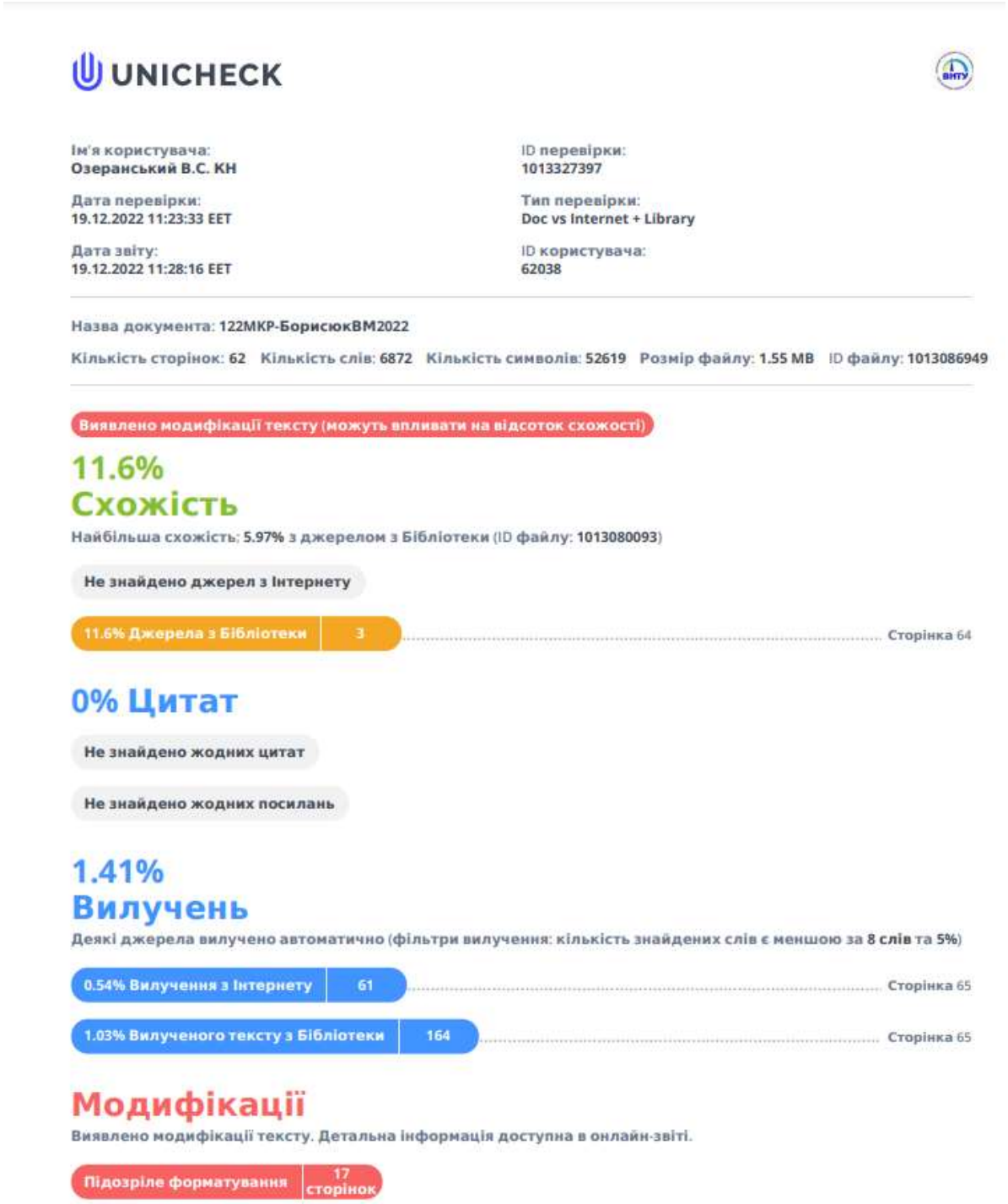


Рисунок А.1 - Результат перевірки на плагіат в онлайн-системі UNICHECK

Додаток Б (обов'язковий)

Лістинг програми

```
from sklearn.ensemble import RandomForestRegressor

RF_model = RandomForestRegressor(random_state =0, n_estimators=500,
max_depth=10)
RF_model.fit(X_train, y_train)

y_hat = RF_model.predict(X_test)
print ("R2 score:", r2_score(y_hat, y_test))
importances =
pd.DataFrame({'feature':X_train.columns,'importance':np.round(RF_model.feature
_importances_,3)})
importances =
importances.sort_values('importance',ascending=False).set_index('feature')
importances.plot.bar();
X_train, X_valid, y_train, y_valid = train_test_split(X, y, test_size=0.2)
params = {'num_leaves': 30,
        'min_data_in_leaf': 20,
        'objective': 'regression',
        'max_depth': 5,
        'learning_rate': 0.01,
        "boosting": "gbdt",
        "feature_fraction": 0.9,
        "bagging_freq": 1,
        "bagging_fraction": 0.9,
        "bagging_seed": 11,
        "metric": 'rmse',
        "lambda_l1": 0.2,
        "verbosity": -1}
lgb_model = lgb.LGBMRegressor(**params, n_estimators = 10000, nthread = 4,
n_jobs = -1)
lgb_model.fit(X_train, y_train,
       eval_set=[(X_train, y_train), (X_valid, y_valid)], eval_metric='rmse',
       verbose=1000, early_stopping_rounds=200)
from sklearn import ensemble
params = {'n_estimators': 500, 'max_depth': 4, 'min_samples_split': 2,
        'learning_rate': .01, 'loss': 'ls'}
clf = ensemble.GradientBoostingRegressor(**params)
predictions2 = clf.fit(X_train,y_train)
training_score = clf.score(X_train, y_train)
print(f"Training Score: {training_score}")
```

```
predictions2 = np.expand_dims(clf.predict(X_test), axis = 1)
MSE = mean_squared_error(y_test, predictions2)
r2 = clf.score(X_test, y_test)
print(f"MSE: {MSE}, R2: {r2}")

import datetime as dt

from sqlmodel import SQLModel, Field, ForeignKey


class Movie(SQLModel, table=True):
    id: int = Field(primary_key=True)
    budget: int
    genre_ids: list[ForeignKey('genre.id')] = []
    homepage: str | None
    imdb_id: str
    original_language: str
    original_title: str
    overview: str
    popularity: float
    production_companies: list[ForeignKey('production_company.id')] = []
    production_countries: list[ForeignKey('production_country.id')] = []
    release_date: dt.datetime
    runtime: int
    keywords: list[ForeignKey('keyword.id')] = []
    cast: list[ForeignKey('actor.id')] = []
    crew: list[ForeignKey('credit.id')] = []
    revenue: int | None

from abc import abstractmethod, ABC
from typing import Iterable, Generic, TypeVar, NoReturn

from sqlmodel import select, delete, Session, SQLModel

from .models import Movie


_Model = TypeVar('_Model', bound=SQLModel)


class Repo(Generic[_Model], ABC):

    def __init__(self, session: Session):
        self._session = session
```

```
    @abstractmethod
    def get(self, pk: int) -> _Model:
        pass

    @abstractmethod
    def get_many(self, *where) -> Iterable[_Model]:
        pass

    @abstractmethod
    def delete(self, pk: int) -> NoReturn:
        pass


class MovieRepo(Repo[Movie]):

    def get(self, pk: int) -> Movie:
        query = select(Movie).where(Movie.id == pk)
        movie = self._session.execute(query).first()
        return movie

    def get_many(self, *where) -> Iterable[Movie]:
        query = select(Movie).where(*where)
        return self._session.execute(query)

    def delete(self, pk: int) -> NoReturn:
        query = delete(Movie).where(Movie.id == pk)
        self._session.execute(query)
```

98

Додаток В (обов'язковий)

ІЛЮСТРАТИВНА ЧАСТИНА

ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ РОЗПІЗНАВАННЯ МАТЕМАТИЧНИХ ФОРМУЛ НА ОСНОВІ НЕЙРОННОЇ МЕРЕЖІ

Виконав: студент 2-го курсу,
групи 2КН-21м
спеціальності 122 «Комп'ютерні науки»
(шифр і назва напряму підготовки, спеціальності)
Борисюк В. М.
(прізвище та ініціали)

Керівник: к.т.н., доцент каф. КН
Барабан С. В.
(прізвище та ініціали)
«15» 12 2022 р.

ВНТУ - 2022 рік

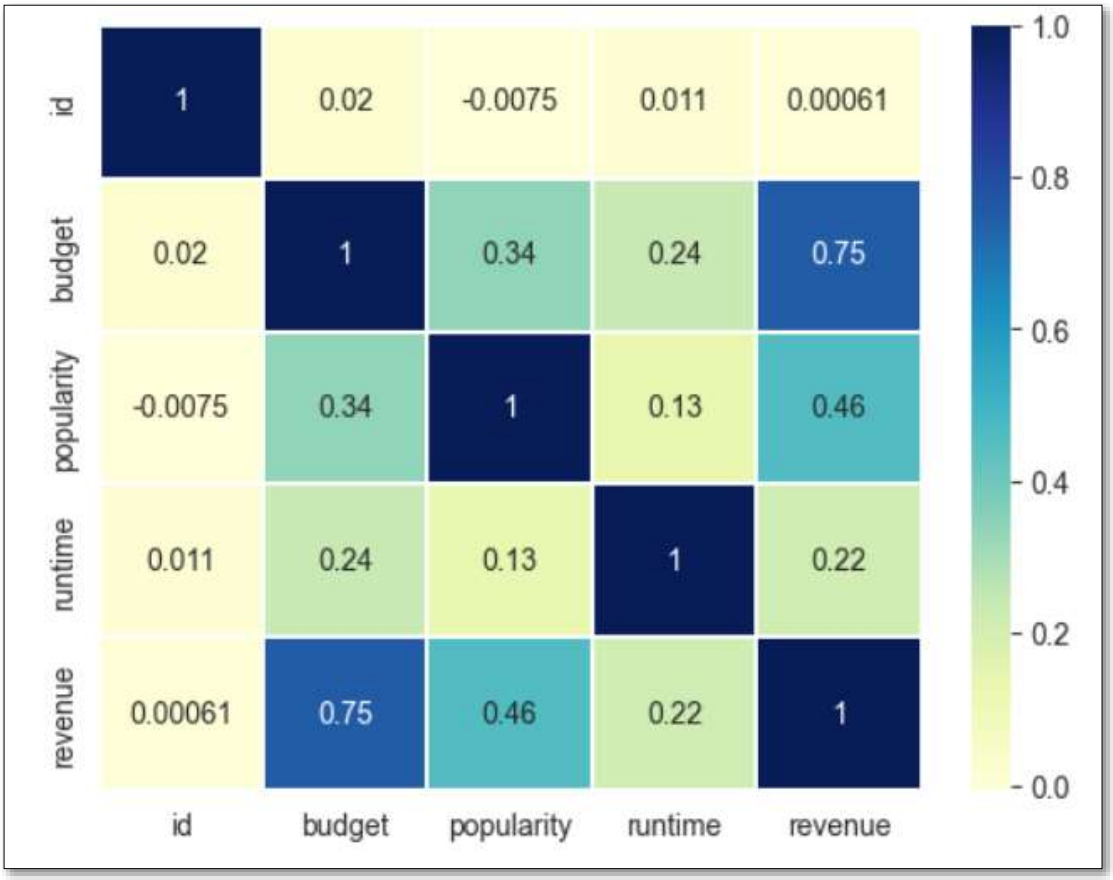


Рисунок В.1 – Графік кореляції ознак кінофільму до його доходу

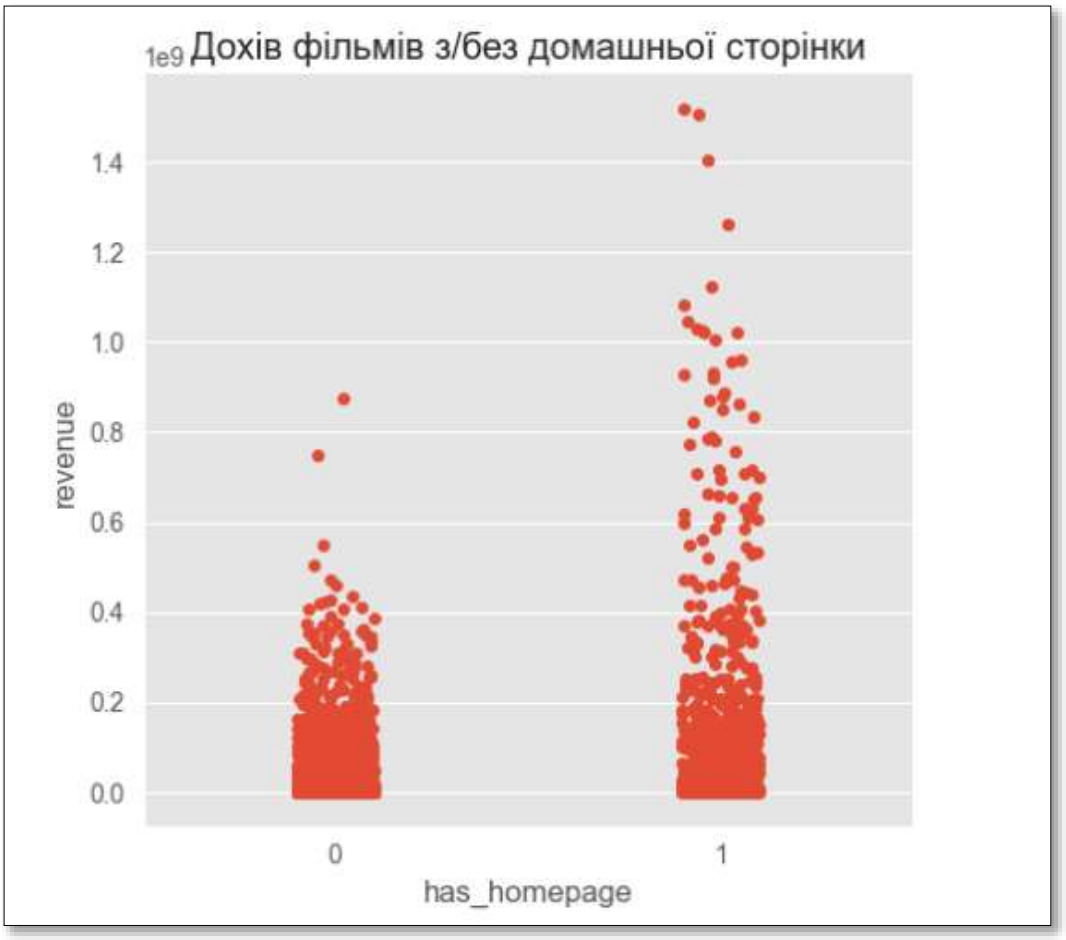


Рисунок В.2 – Графік розподілу доходу кінофільмів з домашньою і без неї

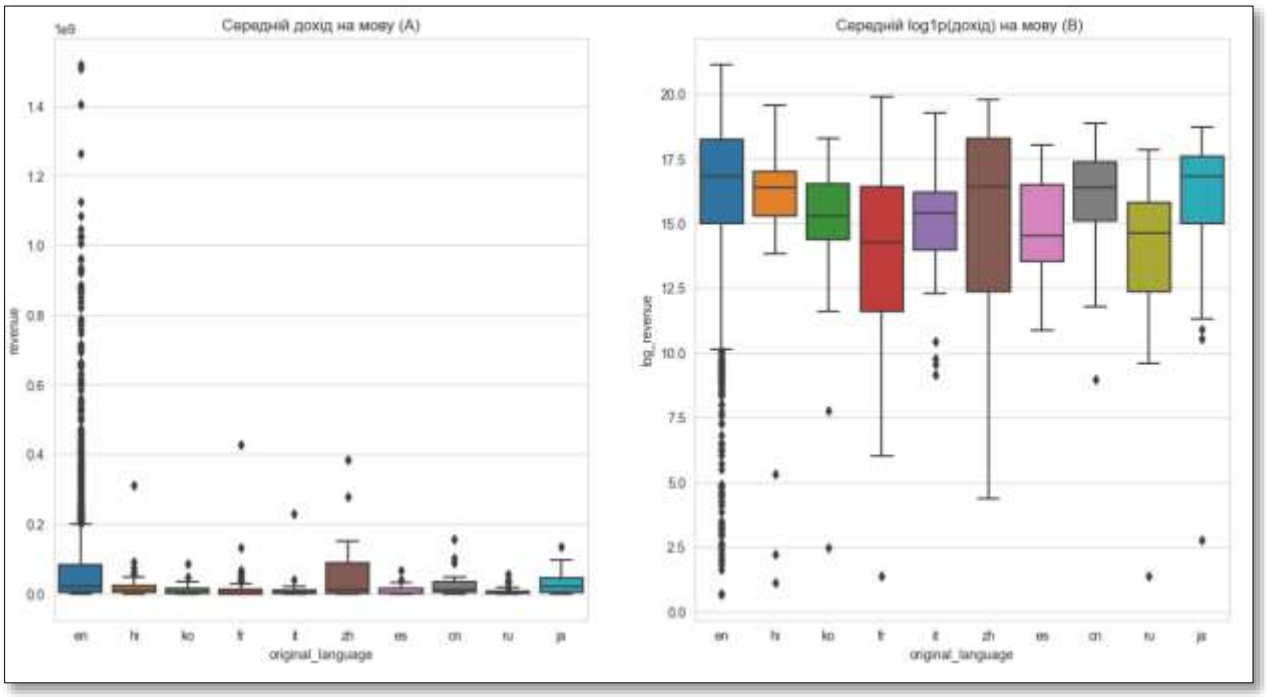


Рисунок В.3 – Графіки розподілу доходів для фільмів на різних мовах

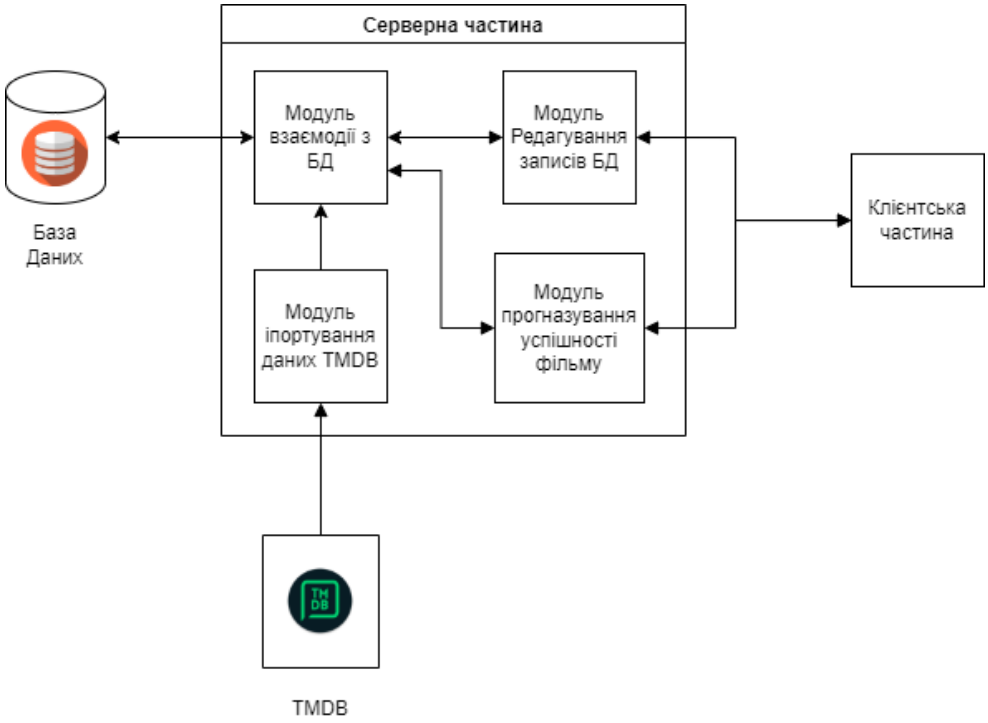


Рисунок В.4 – Загальна структура програмного забезпечення прогнозування успішності фільмів

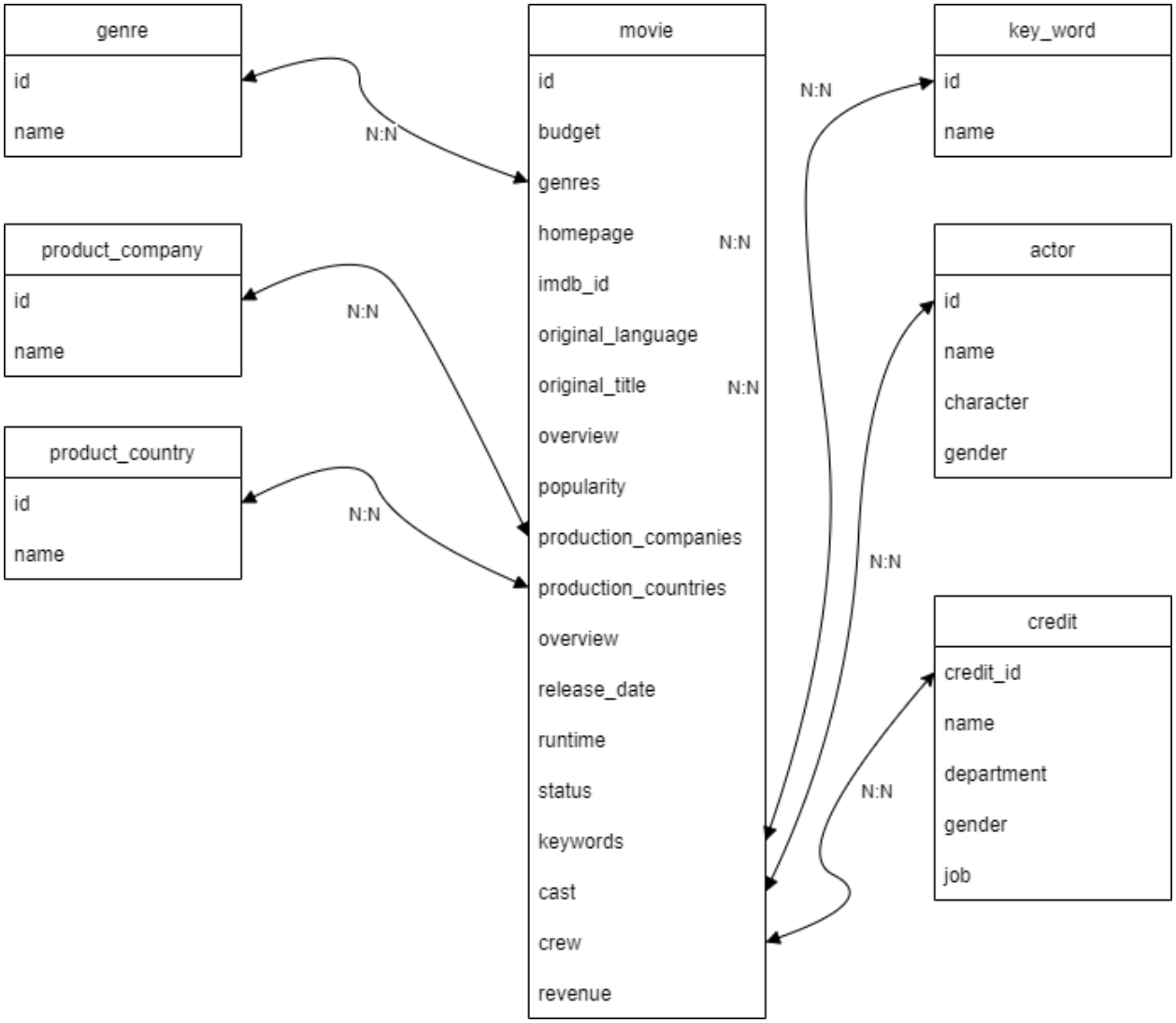


Рисунок В.5 - Загальна схема бази-даних кінофільмів

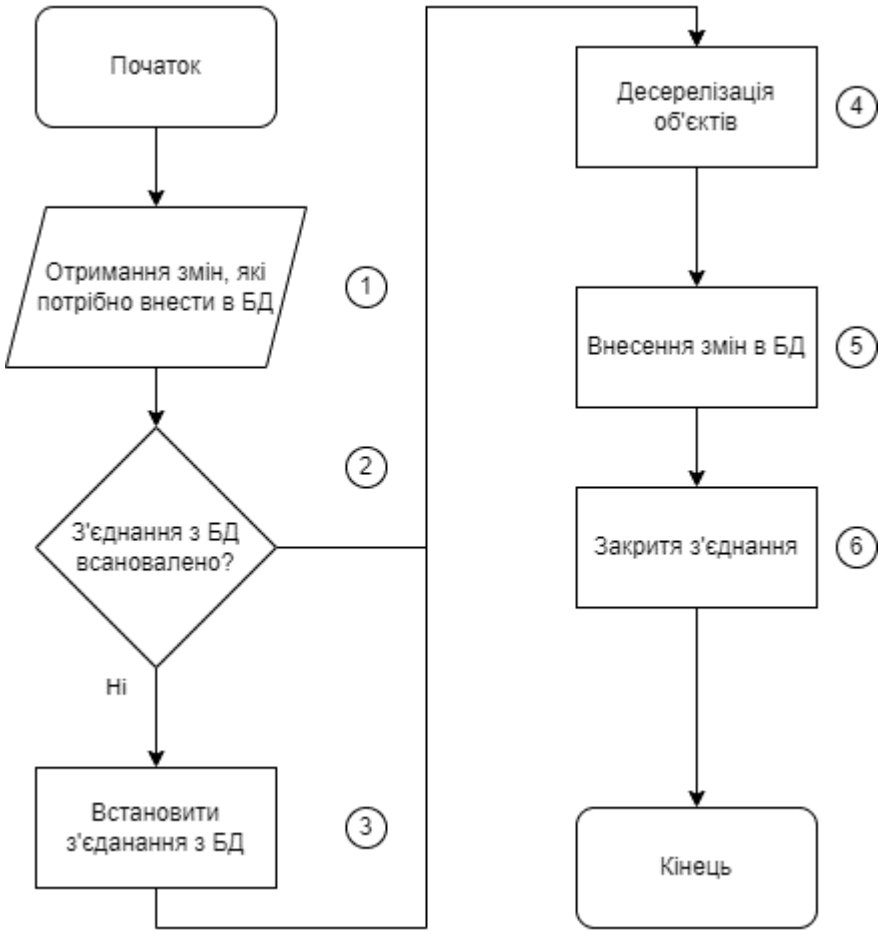


Рисунок В.6 - Схема алгоритму модуля взаємодії з базою даних

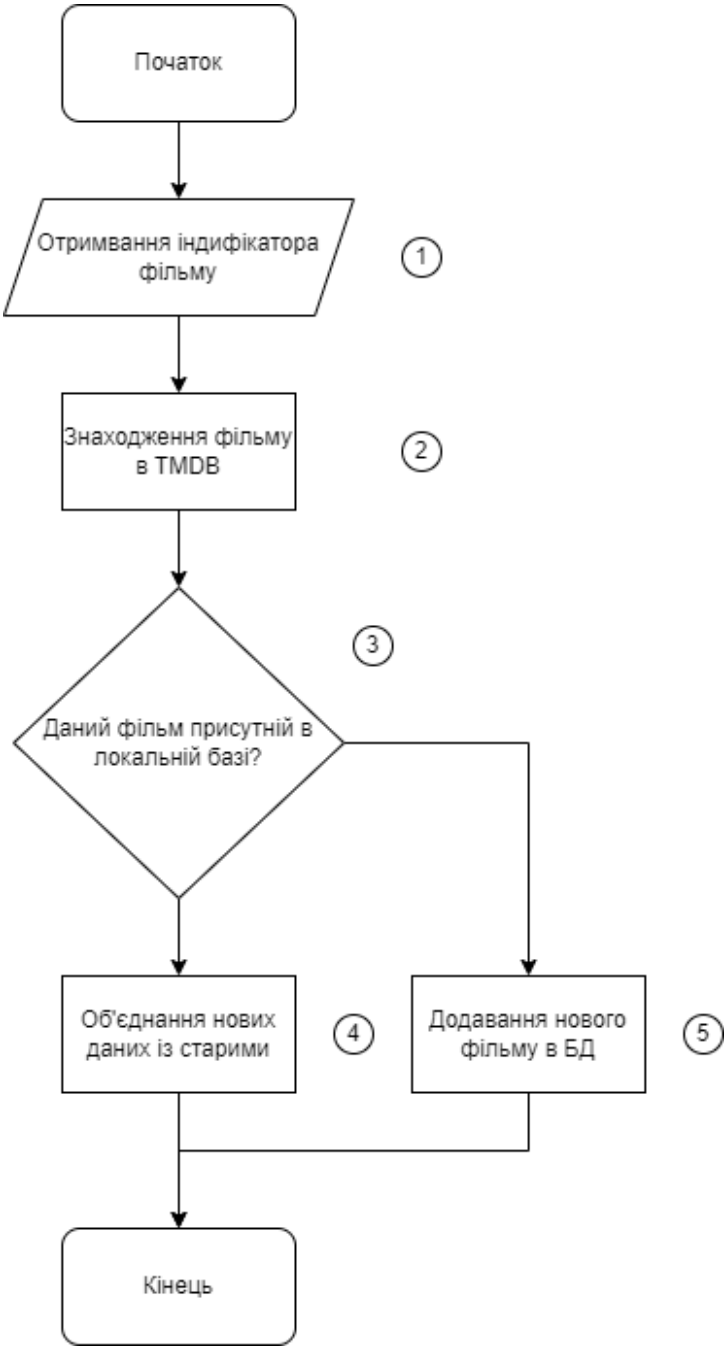


Рисунок В.7 – Схема алгоритму модуля імпортування фільмів з TMDB

Рисунок В.8 – Схема алгоритму модуля прогнозування успішності кінофільму

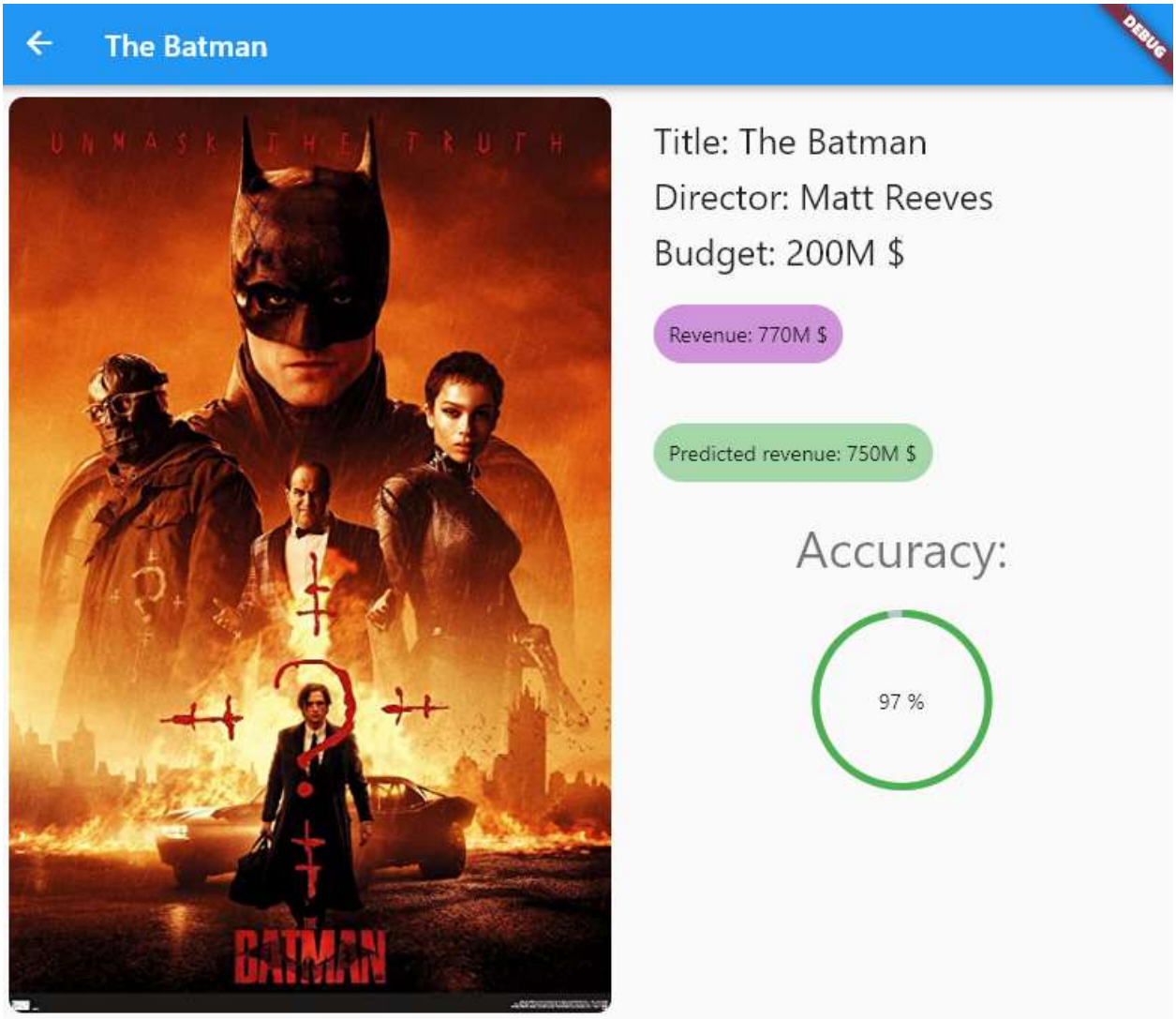


Рисунок В.9 – Вікно програми з результатом прогнозування успішності кінофільму

Додаток Г (довідниковий)

Інструкція користувача

Для використання програмного забезпечення, прогнозування успішності кінофільму, користувачу потрібно виконати наступні кроки:

1. Запустити ярлик програми – рисунок Г.1.

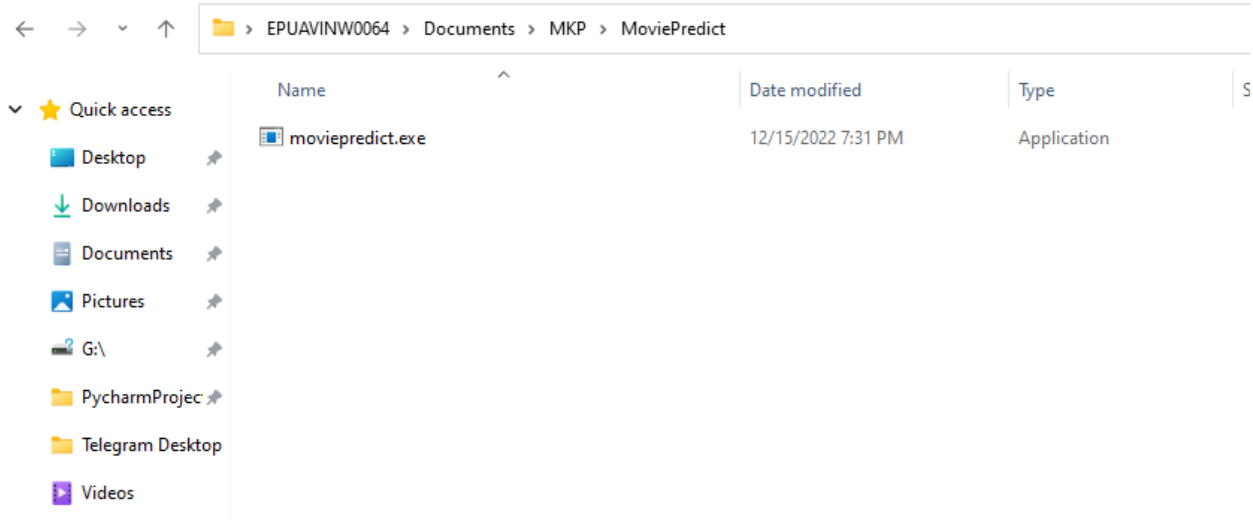


Рисунок Г.1 – Зображення ярлика програми

2. Відкриється головне вікно програмного забезпечення, на якому користувач може побачити увесь список кінофільмів- - рисунок Г.2.

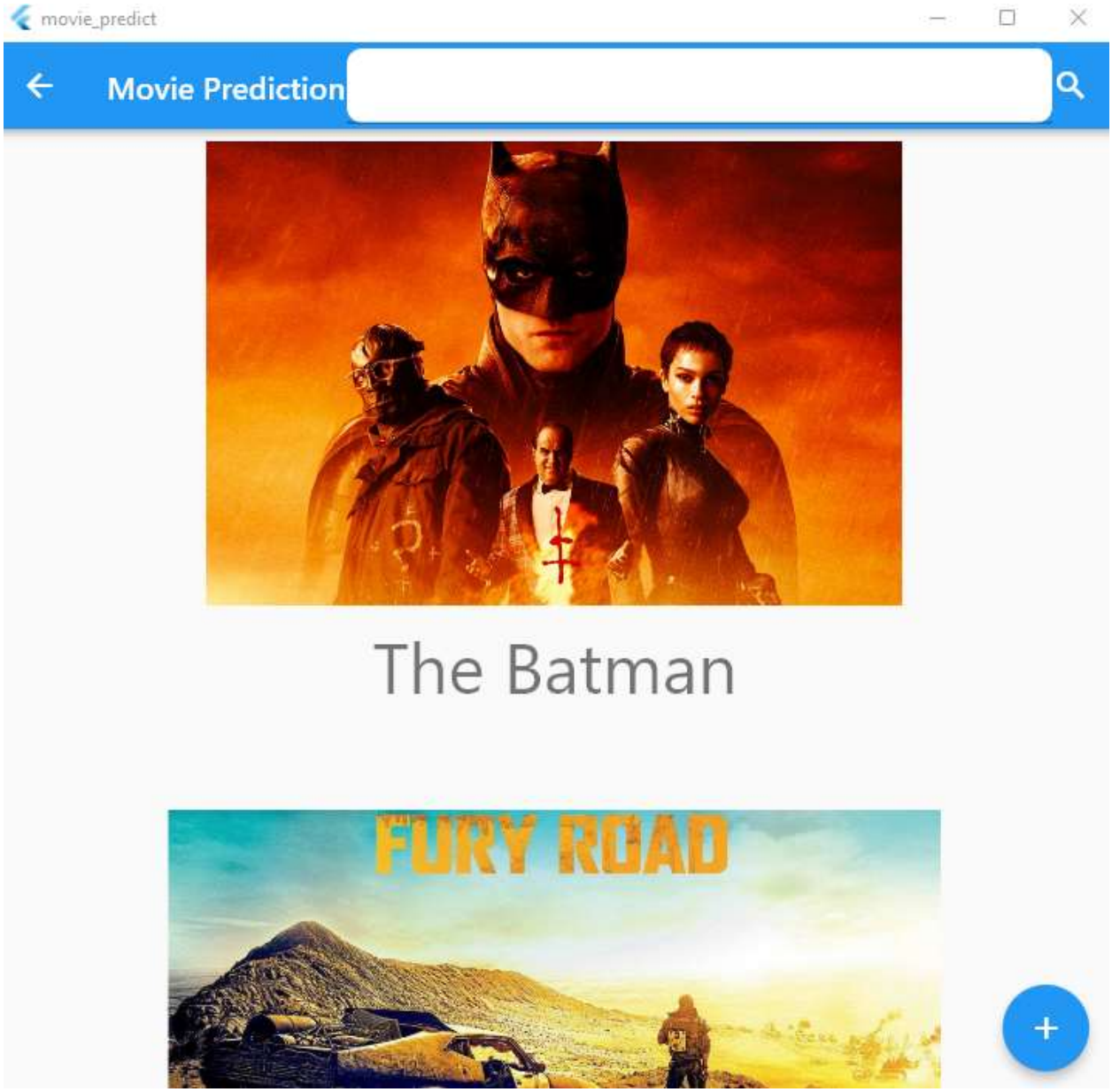


Рисунок Г.2 – Головне вікно програми прогнозування успішності кінофільму

3. Для пошуку фільмів в середині програми користувач може скористатись пошуковим полем у верхній частині вікна – рисунок Г.3.

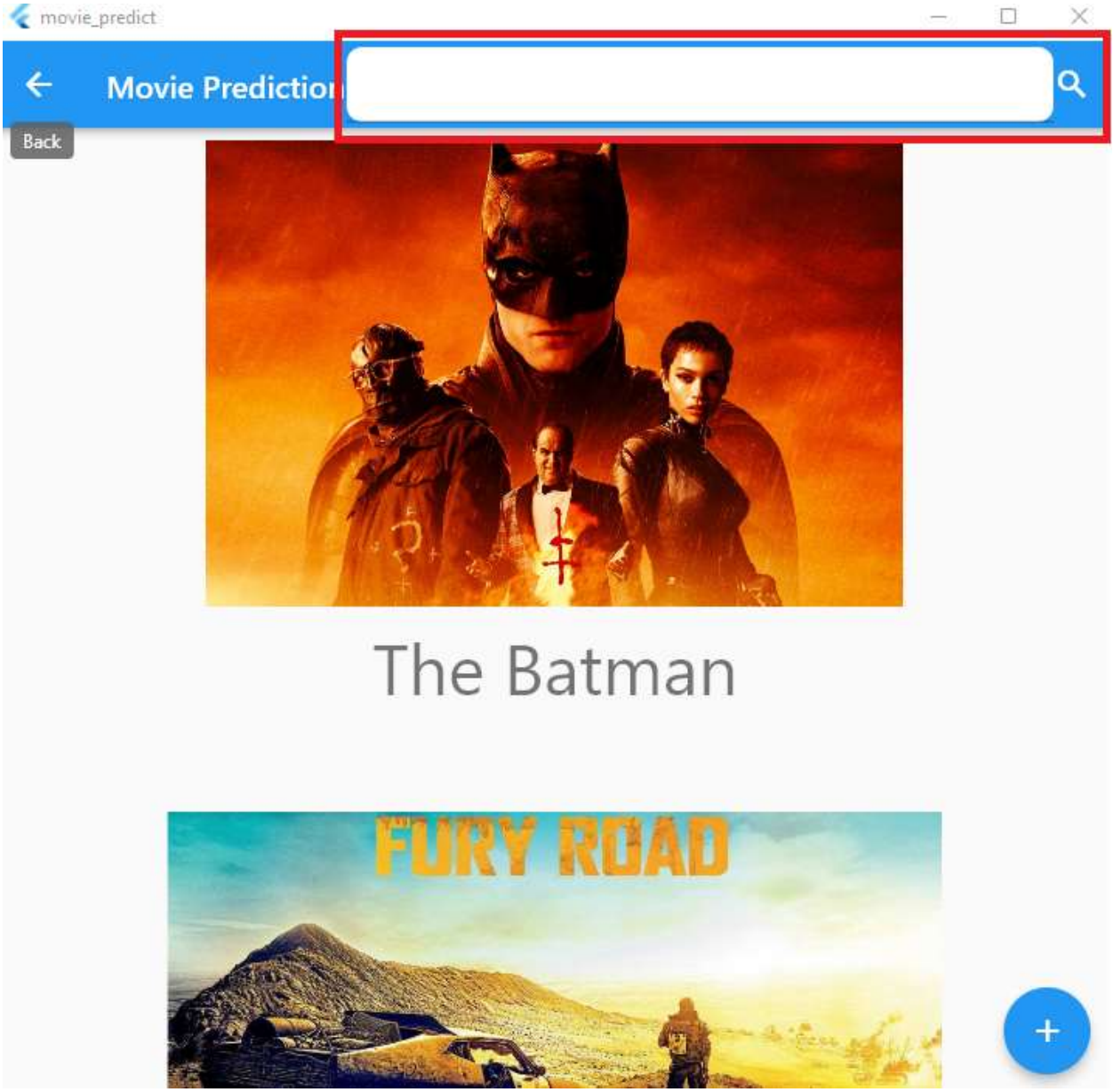


Рисунок Г.3 – Пошукове поле вводу

4. Щоб відрити вікно з детальної інформацією про кінофільм та отримати прогноз щодо доходу кінофільму, потрібно натиснути на постер фільму. Причому, якщо інформація про реальний прибуток кінофільму присутня, виведеться індикатор, який відобразить точність прогнозу. Вікно з прогнозом доходу кінофільму на його детальною інформацією зображено на рисунку Г.4.

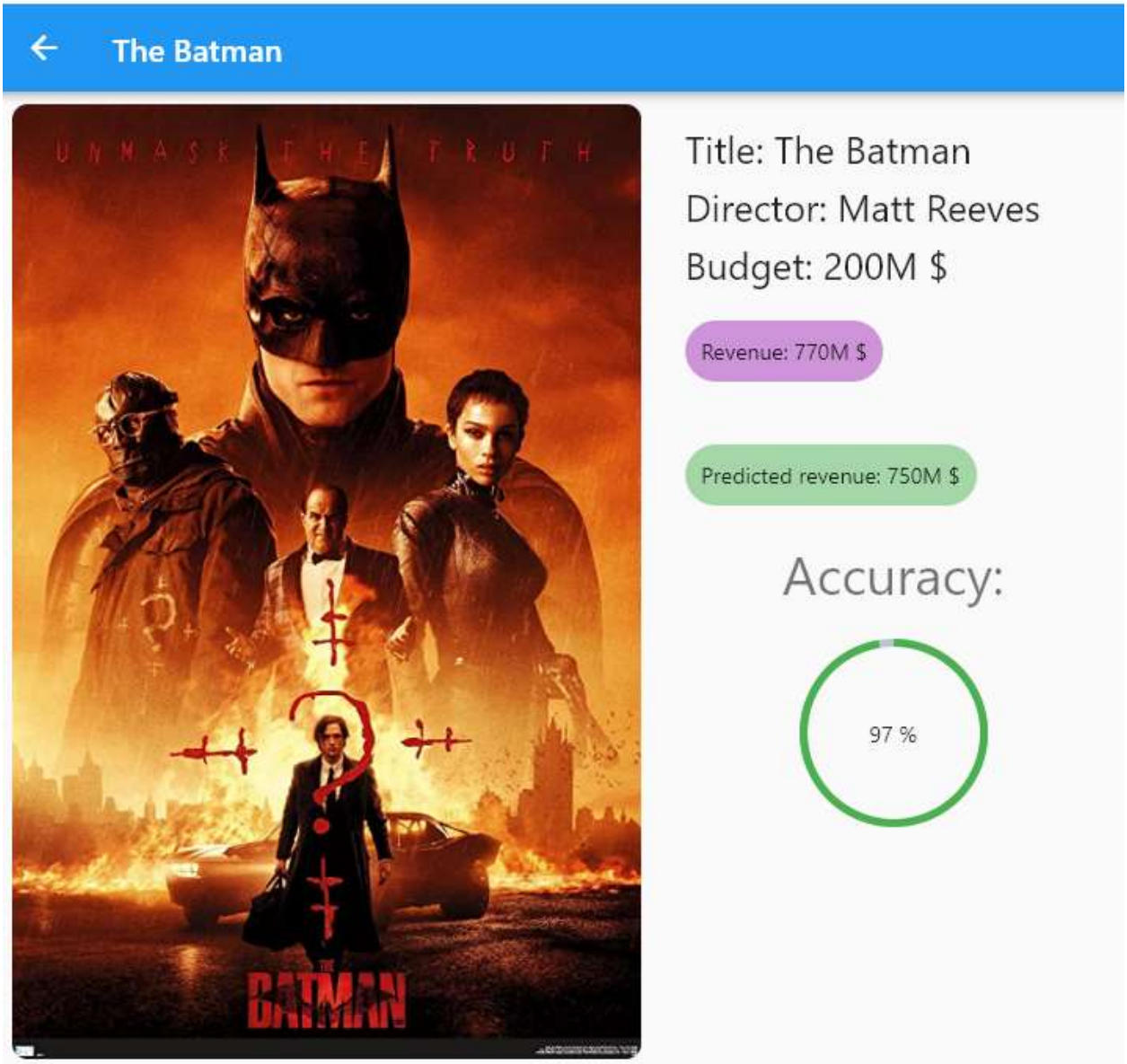


Рисунок Г.4 – Вікно з прогнозом доходу та детальною інформацією для кінофільму “The Batman”

5. Щоб добавити новий кінофільм, потрібно натиснути відповідну кнопку на головному вікні – рисунок Г.5. Тоді відриється вікно з формою, за допомогою якої можна добавити новий кінофільм – рисунок Г.6.

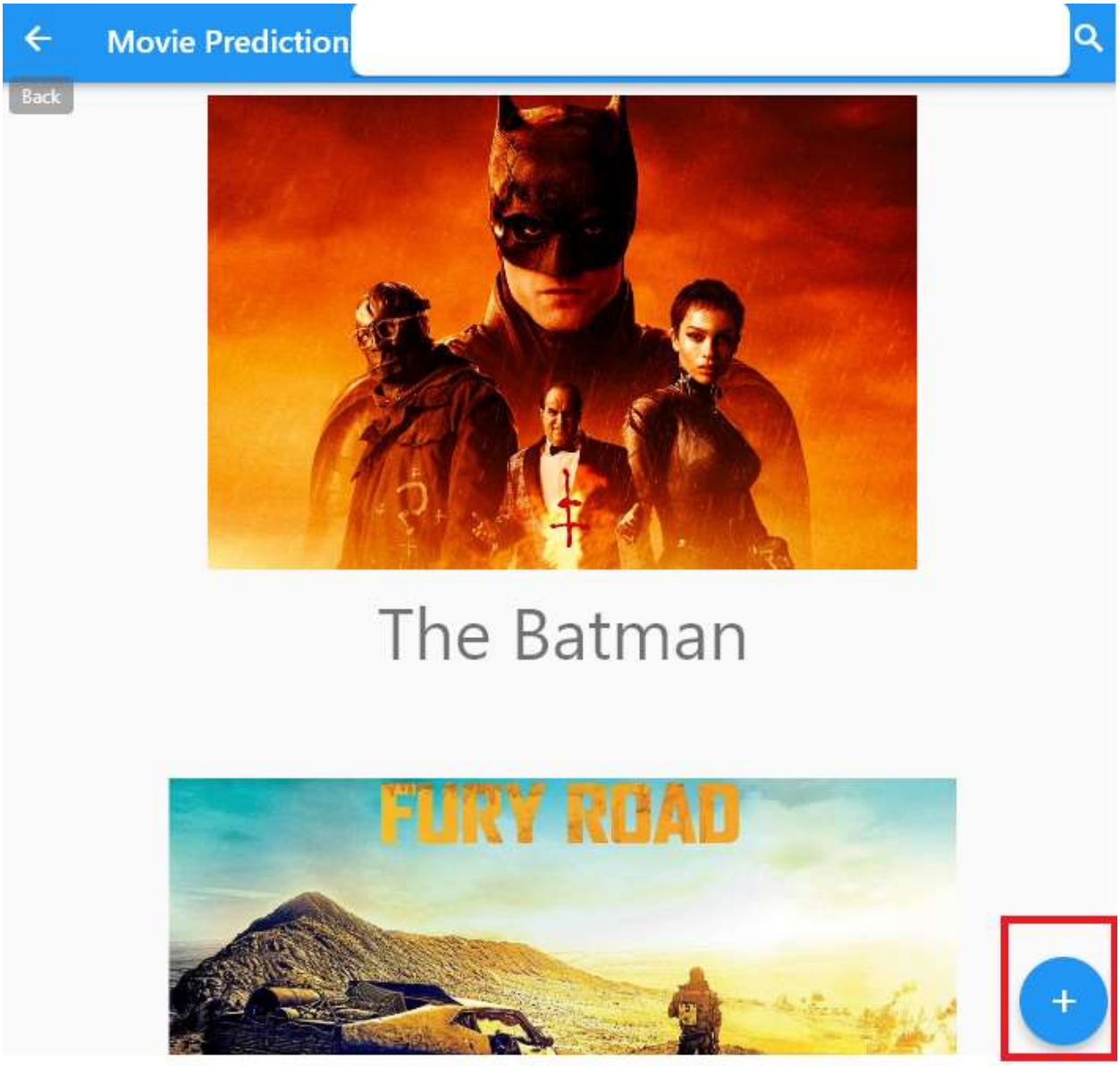


Рисунок Г.4 – Кнопка для переходу на форму для добавлення нового кінофільму

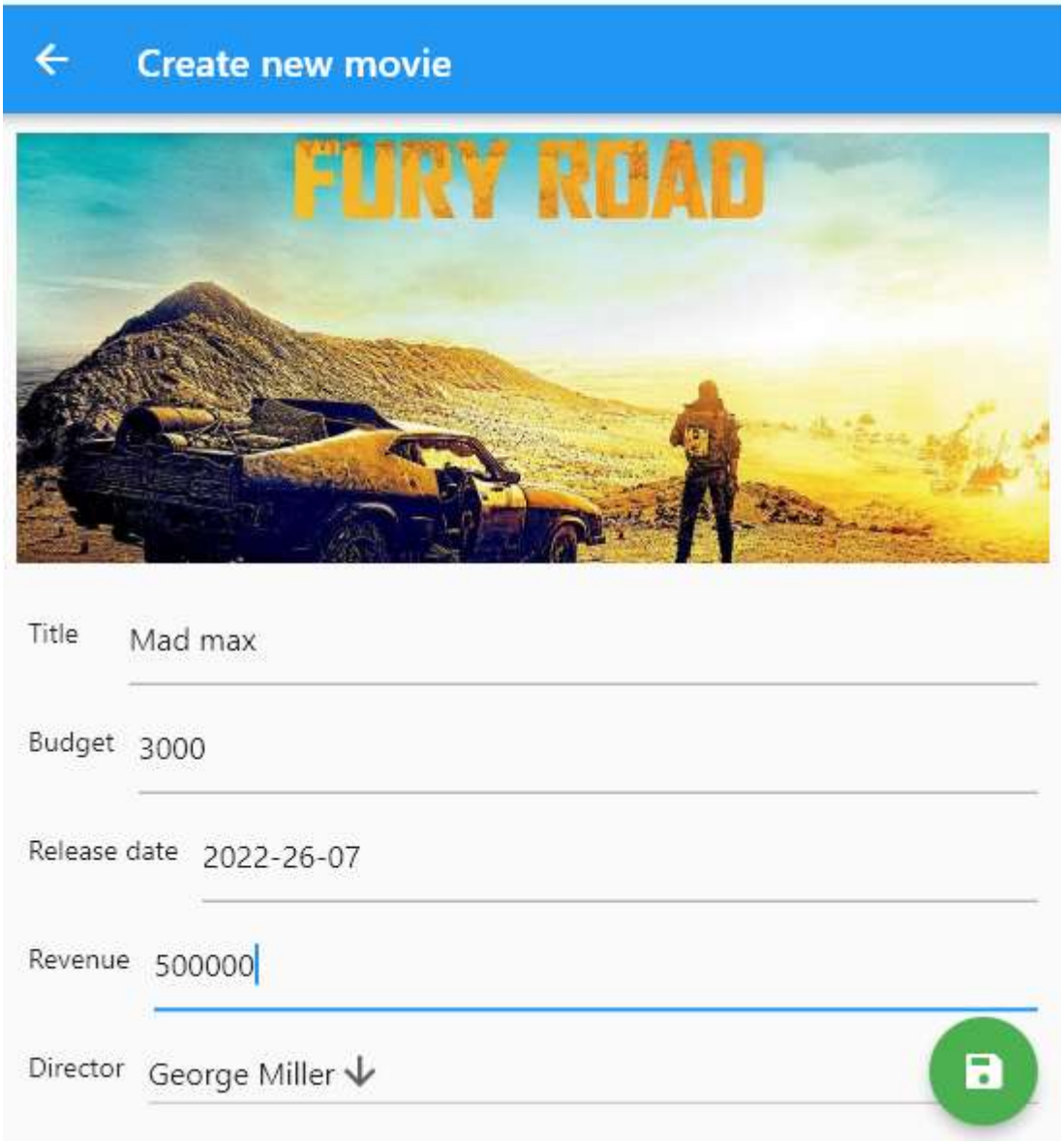


Рисунок Г.5 – Форма для додавання нового кінофільму

6. Для збереження кінофільму потрібно натиснути відповідно кнопку в правому нижньому куті – рисунок Г.6.

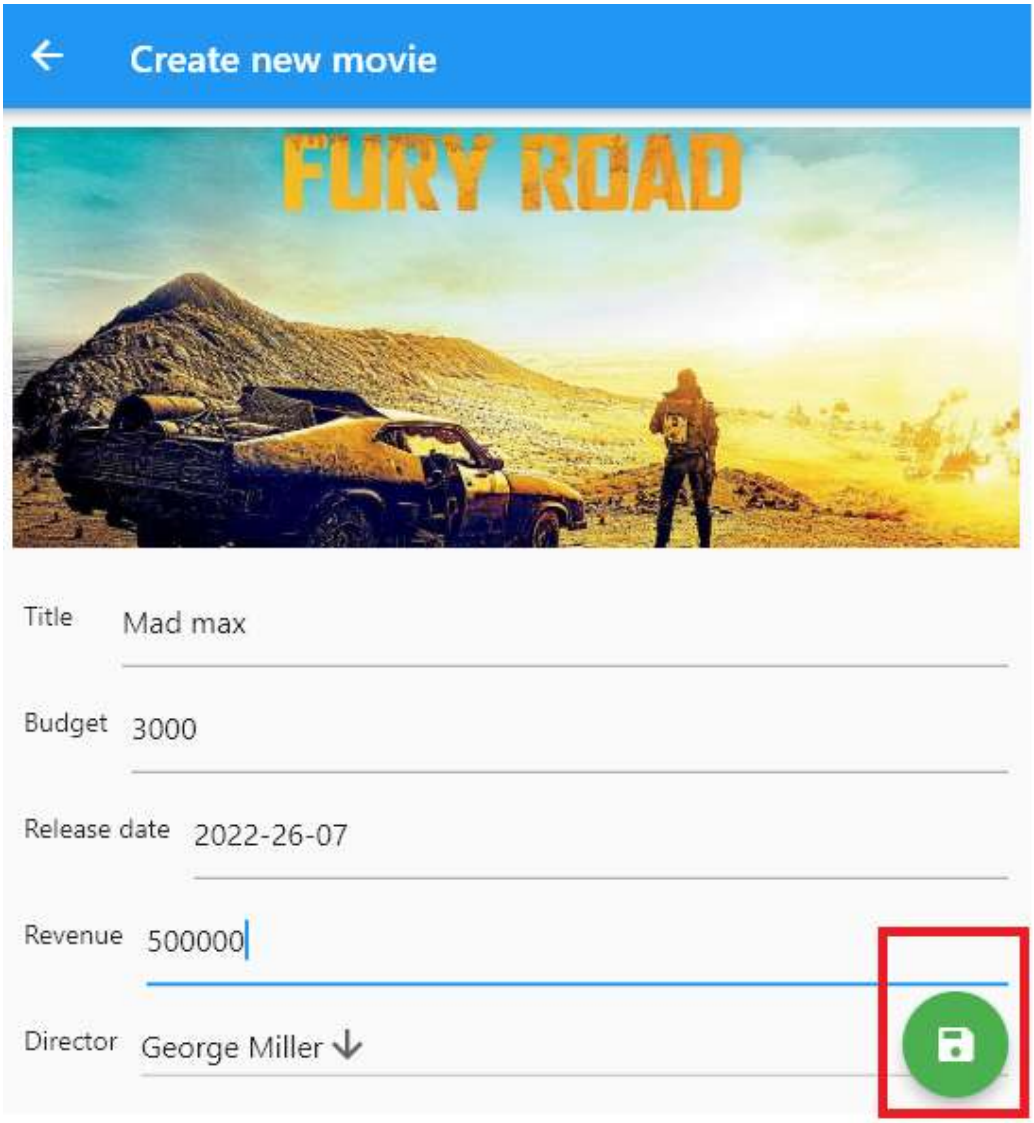


Рисунок Г.6 – Кнопка збереження кінофільму