

БАКАЛАВРСЬКА РОБОТА

БР. ІІ - 54.00.00.000 ІІЗ

Група ІІ-21-4

Артиш Микола

2025

Івано-Франківський національний технічний університет нафти і газу

Факультет інформаційних технологій

Кафедра інженерії програмного забезпечення

Артиш Микола Володимирович

(прізвище, ім'я, по батькові)

УДК 004
(індекс)

БАКАЛАВРСЬКА РОБОТА

Веб-застосунок для прогнозування успішності розважальних релізів

(назва роботи)

Інженерія програмного забезпечення

(назва освітньої програми)

121 - Інженерія програмного забезпечення

(шифр і назва спеціальності)

Робота містить результати власних досліджень. Використання ідей, результатів і текстів інших авторів мають посилання на відповідне джерело

Здобувач освітнього рівня Артиш М.В.
(підпис, ініціали та прізвище здобувача)

Науковий керівник Юрчишин Володимир Миколайович, д.т.н., проф.
(підпис, прізвище, ім'я, по батькові, науковий ступінь, вчене звання керівника)

Допущено до захисту
Завідувач кафедри

доц. Бандура В.В.
(посада) (підпис) (дата) (ініціали та прізвище)

Івано-Франківськ – 2025

6. Консультанти розділів проекту (роботи)

Розділ	Консультант	Підпис, дата	
		Завдання видав	Завдання прийняв

7. Дата видачі завдання 28 квітня 2025 р.

Керівник _____

(підпис)

Завдання прийняв до виконання _____

(підпис)

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів дипломного проекту (роботи)	Строк виконання етапів проекту	Примітка
1	Аналіз предметної області розробки додатків прогнозування подій	04.05.2025	виконано
2	Алгоритмічна реалізація веб-застосунку прогнозування успішності релізів	15.05.2025	виконано
3	Моделювання структури веб-застосунку прогнозування	21.05.2025	виконано
4	Реалізація імпорту даних для системи прогнозування	28.05.2025	виконано
5	Реалізація інтерфейсу користувача веб-застосунку для прогнозування успішності релізів	03.06.2025	виконано
6	Оформлення пояснювальної записки дипломної роботи завідувачем кафедри	10.06.2025	виконано

Студент – дипломник _____

(підпис)

Керівник роботи _____

(підпис)

АНОТАЦІЯ

Бакалаврська робота містить 76 сторінок, 26 рисунків, список використаних джерел із 41 найменуваннями, 1 додаток.

Метою дипломної роботи є проектування та реалізація веб-застосунку для прогнозування успішності розважальних релізів, зокрема кінопродукції, на основі сучасних методів аналізу даних.

Об'єкт дослідження - процес прогнозування успішності релізів у сфері розважального контенту.

Предмет дослідження - моделі, методи та технології створення веб-застосунку для прогнозування успішності кінопродукції.

В першому розділі проведено аналіз предметної області, визначено переваги прогнозування релізів та виявлено обмеження сучасних інструментів.

В другому розділі сформовано архітектуру системи, змодельовано процеси та доведено доцільність реалізації запропонованого рішення.

В третьому розділі реалізовано імпорт і підготовку даних, розроблено користувацький інтерфейс та підтверджено працездатність прототипу системи.

Висновок: розроблено прототип веб-застосунку, який поєднує сучасні методи обробки даних, інтерфейсну доступність та структуроване моделювання процесу прогнозування успішності розважального контенту

КЛЮЧОВІ СЛОВА: ПРОГНОЗУВАННЯ, ВЕБ-ЗАСТОСУНОК, КІНОПРОДУКЦІЯ, УСПІШНІСТЬ РЕЛІЗУ, ОБРОБКА ДАНИХ, REST API, МОДЕЛЮВАННЯ, ІНФОРМАЦІЙНА СИСТЕМА.

ANNOTATION

The bachelor's thesis contains 76 pages, 26 figures, a list of used sources with 41 names, 1 appendix.

The method of the thesis is the design and implementation of a web application for predicting the success of entertainment releases, in particular film productions, based on modern data analysis methods.

The object of the study is the process of predicting the success of releases in the field of entertainment content.

The subject of the study is models, methods and technologies for creating a web application for predicting the success of film productions.

In the first section, an analysis of the subject area is conducted, the advantages of forecasting releases are determined and the limitations of modern tools are identified.

In the second section, an architectural system is formed, processes are modeled and the feasibility of implementing the proposed solution is proven.

In the third section, data import and preparation are implemented, a user interface is developed and the operability of the system prototype is confirmed.

Conclusion: a prototype of a web application has been developed that compensates for modern data processing methods, interface accessibility, and structured modeling of the process of predicting the effectiveness of entertainment content.

KEYWORDS: FORECASTING, WEB APPLICATION, FILM PRODUCTION, RELEASE SUCCESS, DATA PROCESSING, REST API, MODELING, INFORMATION SYSTEM.

ЗМІСТ

ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ	9
ВСТУП.....	10
РОЗДІЛ 1. АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ РОЗРОБКИ ДОДАТКІВ ПРОГНОЗУВАННЯ ПОДІЙ.....	14
1.1. Мета розробки веб-застосунку для прогнозування успішності розважальних релізів	14
1.2. Виклики та рішення процесу прогнозування успішності кінопродукції.....	15
1.2.1. Мета та механізм прогнозування успішності фільмів	15
1.2.2. Переваги та перспективи	16
1.3. Актуальний стан прогнозування успішності кінорелізів. Опис існуючих платформ	16
1.3.1. Професійні та закриті платформи для студій.....	18
1.3.2. Академічні дослідження та наукові розробки в області прогнозування успішності	23
1.3.3. Платформи для відстеження та оцінки фільмів	23
1.3.4. Ключова відмінність пропонованого рішення	28
РОЗДІЛ 2. АЛГОРИТМІЧНА РЕАЛІЗАЦІЯ ТА МОДЕЛЮВАННЯ СТРУКТУРИ ВЕБ-ЗАСТОСУНКУ ПРОГНОЗУВАННЯ УСПІШНОСТІ РЕЛІЗІВ.....	30
2.1. Архітектура та функціонування веб-застосунку прогнозування успішності релізів кіноіндустрії.....	30
2.1.1. Технічні вимоги та розгортання системи	30

					БР.ІІ – 54.00.00.000 ПЗ			
Змн.	Арк.	№ докум.	Підпис	Дата	Веб-застосунок для прогнозування успішності розважальних релізів Пояснювальна записка	Літ.	Арк.	Аркуші
Розроб.	Артиш М.В.						6	
Перевір.	Юрчишин В.М.							
Реценз.								
Н. Контр.	Піх М.М.							
Затверд.	Бандура В.В.					ІФНТУНГ Ш-21-4		

2.1.2. Механізм обробки даних та навчання системи	31
2.1.3. Специфікація системних вимог прототипу	31
2.2. Дослідження доцільності розробки веб-застосунку	32
2.2.1. Технічна доцільність	32
2.2.2. Економічна доцільність	33
2.2.3. Операційна доцільність	34
2.3. Функціональне моделювання системи за допомогою діаграми варіантів використання.....	35
2.4. Розробка діаграми послідовності взаємодії адміністратора з системою	38
2.5. Розробка алгоритму (діаграми активності) взаємодії з системою прогнозування	40
2.6. Моделювання інформаційних потоків у системі з використанням діаграми потоків даних (DFD)	43

РОЗДІЛ 3. ПІДГОТОВКА ДАНИХ ТА РЕАЛІЗАЦІЯ ІНТЕРФЕЙСУ КОРИСТУВАЧА ВЕБ-ЗАСТОСУНКУ ДЛЯ ПРОГНОЗУВАННЯ УСПІШНОСТІ РЕЛІЗІВ КІНОІНДУСТРІЇ.....

3.1. Стратегії імпорту даних з REST API.....	50
3.1.1. Реалізація імпорту даних для системи прогнозування	51
3.1.2. Припущення щодо набору даних	52
3.2. Очищення даних у процесі розробки системи прогнозування	52
3.2.1. Основні підходи до очищення даних.....	53
3.2.2. Застосування методу очищення даних у пропонованій системі	55
3.3. Демонстрація функціональності та валідація системи прогнозування успішності фільмів.....	55
3.3.1. Етапи автентифікації та навігації.....	55
3.3.2. Введення та управління даними	58
3.3.3. Взаємодія з базою даних та прогнозування	59

3.3.4. Валідація точності прогнозування	62
3.4. Фаза тестування в життєвому циклі розробки програмного забезпечення.....	64
3.5. Перспективи майбутнього вдосконалення системи прогнозування успішності фільмів.....	67
ВИСНОВКИ	70
ПЕРЕЛІК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ.....	72
ДОДАТКИ	
БІБЛІОГРАФІЧНА ДОВІДКА	

					БР.ІІІ – 54.00.00.000 ІІЗ	Арк.
						8
Змн.	Арк.	№ докум.	Підпис	Дата		

ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ

AI – Artificial Intelligence – Штучний інтелект

ASP.NET – Active Server Pages .NET – Технологія для створення веб-додатків від Microsoft

C# – C Sharp – Мова програмування C#

CSV – Comma-Separated Values – Значення, розділені комами

DB – Database – База даних

DFD – Data Flow Diagram – Діаграма потоку даних

ID – Identifier – Ідентифікатор

IK – User Interface (UI) – Інтерфейс користувача

JSON – JavaScript Object Notation – Об'єктна нотація JavaScript

ML – Machine Learning – Машинне навчання

NaN – Not a Number – Не число

N/A – Not Applicable / Not Available – Не застосовується / Недоступно

NA – Not Available – Недоступно (як стандартизоване нульове значення)

NLP – Natural Language Processing – Обробка природної мови

REST – Representational State Transfer – Передача стану представлення

SDLC – Software Development Life Cycle – Життєвий цикл розробки програмного забезпечення

UAT – User Acceptance Testing – Тестування прийняття користувачем

UML – Unified Modeling Language – Уніфікована мова моделювання

					БР.ІІІ – 54.00.00.000 ПЗ	Арк.
						9
Змн.	Арк.	№ докум.	Підпис	Дата		

ВСТУП

Сфера розважальних медіа, зокрема кіноіндустрія, відіграє важливу роль у сучасному інформаційному просторі, формуючи не лише культурні тренди, а й вагомі економічні потоки. Виробництво та просування кінопродукції вимагають значних фінансових, часових і людських ресурсів, тому вкрай актуальним є завдання прогнозування її потенційного успіху до моменту релізу. Таке прогнозування дозволяє зменшити ризики інвестування, оптимізувати маркетингові кампанії, підвищити прибутковість проєктів та сприяти ухваленню обґрунтованих рішень у стратегічному плануванні.

На сучасному етапі розвитку інформаційних технологій спостерігається активна інтеграція інструментів аналізу великих даних, штучного інтелекту та машинного навчання у прикладні області, пов'язані з прогнозуванням поведінки аудиторії, трендів і комерційних результатів. Водночас більшість доступних інструментів або розробляються для внутрішнього використання великими компаніями, або мають обмежену функціональність, не враховують динаміку змін у вподобаннях глядачів та не забезпечують прозорого механізму формування прогнозу. У відкритому доступі відчутно бракує адаптивних, зручних та інтерпретованих систем, які дозволяли б здійснювати обґрунтовані прогнози на основі відкритих джерел даних.

Актуальність роботи

Проблема прогнозування комерційної успішності розважального контенту, зокрема фільмів, набуває дедалі більшої актуальності у контексті цифрової трансформації індустрії та росту обсягів даних. Класичні експертні методи втрачають ефективність у швидкоплинному інформаційному середовищі. Натомість аналітика великих даних та моделі машинного навчання дозволяють формувати об'єктивні та адаптивні прогнози, що

					БР.ІІІ – 54.00.00.000 ПЗ	Арк.
						10
Змн.	Арк.	№ докум.	Підпис	Дата		

враховують широкий спектр факторів. Однак ринок таких рішень є або вузькоспеціалізованим (для великих кіностудій), або фрагментованим. Відтак, потреба у відкритому та функціональному веб-застосунку для прогнозування успіху релізів, який може бути адаптований до різних користувацьких сценаріїв, є очевидною.

У цьому контексті розробка веб-застосунку для прогнозування успішності розважальних релізів є перспективним напрямом, що поєднує міждисциплінарні підходи — інженерію програмного забезпечення, аналітику даних, економіку розваг і поведінкову інформатику. Запропонована система має на меті створення зручного та функціонального інструменту, що дозволяє користувачеві вводити або імпортувати параметри майбутнього релізу, аналізувати їх у контексті історичних даних та отримувати ймовірнісну оцінку комерційної успішності продукту.

Дипломна робота містить три логічно взаємопов'язані розділи. У першому розділі здійснено аналіз предметної області, вивчено сучасні методи прогнозування та проаналізовано наявні інструменти. У другому розділі представлено архітектуру веб-застосунку, функціональне моделювання та обґрунтування доцільності створення системи. Третій розділ присвячено реалізації ключових компонентів прототипу, включно з обробкою даних, побудовою інтерфейсу та тестуванням точності прогнозів.

Таким чином, тема дипломної роботи є актуальною як з прикладної, так і з дослідницької точки зору. Результати дослідження можуть бути застосовані як у професійній практиці, так і у навчальних цілях, а також закладати основу для подальших наукових розробок у сфері прогнозної аналітики у медіаіндустрії.

Метою дипломної роботи є проектування та реалізація веб-застосунку для прогнозування успішності розважальних релізів, зокрема кінопродукції, на основі сучасних методів аналізу даних.

					БР.ІІІ – 54.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		11

Завдання дослідження

1. Провести аналіз предметної області прогнозування успішності релізів.
2. Дослідити сучасні програмні рішення та наукові підходи до даної проблематики.
3. Побудувати архітектуру веб-застосунку та обґрунтувати технічну доцільність розробки.
4. Створити функціональні моделі системи з використанням UML-діаграм.
5. Реалізувати механізми імпорту, очищення та обробки даних.
6. Розробити інтерфейс користувача та налаштувати взаємодію з базою даних.
7. Визначити перспективи розвитку та вдосконалення застосунку.

Об'єкт дослідження - процес прогнозування успішності релізів у сфері розважального контенту.

Предмет дослідження - моделі, методи та технології створення веб-застосунку для прогнозування успішності кінопродукції.

Методи дослідження

- системний та порівняльний аналіз;
- функціональне моделювання (UML, DFD);
- методи збирання, очищення та нормалізації даних;
- машинне навчання для побудови прогнозної моделі;
- програмна реалізація (frontend + backend веб-технології);
- тестування програмного забезпечення.

Наукова новизна

Наукова новизна дипломної роботи полягає у створенні прототипу веб-застосунку, який поєднує сучасні методи обробки даних, інтерфейсну доступність та структуроване моделювання процесу прогнозування успішності розважального контенту. Розроблене рішення орієнтоване на

					БР.ІІІ – 54.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		12

відкритість, адаптивність та потенціал подальшої інтеграції зі сторонніми джерелами даних.

Практичне значення

Результати роботи можуть бути використані як основа для створення повноцінного онлайн-сервісу прогнозування для продюсерських центрів, незалежних студій, аналітичних агенцій, освітніх цілей, а також як платформа для подальших досліджень у галузі кіноаналітики та прогнозної інформатики.

Бакалаврська робота містить 76 сторінок, 26 рисунків, 3 розділи список використаних джерел із 41 найменуванням, 1 додаток.

					БР.ІІІ – 54.00.00.000 ПЗ	Арк.
						13
Змн.	Арк.	№ докум.	Підпис	Дата		

РОЗДІЛ 1. АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ РОЗРОБКИ ДОДАТКІВ ПРОГНОЗУВАННЯ ПОДІЙ

1.1. Мета розробки веб-застосунку для прогнозування успішності розважальних релізів

Метою даного дослідження є розробка веб-застосунку, призначеного для прогнозування потенційної успішності розважальних релізів, зокрема кінофільмів. Основна мотивація створення цього додатку полягає в задоволенні інформаційних потреб кіноманів, які прагнуть отримати об'єктивну оцінку якості майбутніх кінопродуктів ще до їхнього виходу в прокат, незалежно від зіркового складу акторів.

Кіноіндустрія є однією з наймасштабніших глобальних галузей, що генерує значний економічний вплив та забезпечує мільярди робочих місць по всьому світу. Її диверсифікована структура охоплює численні національні та регіональні кіноіндустрії, що відрізняються залежно від географії, мовних особливостей та культурних домінант. З плином часу ця галузь зазнала значних технологічних трансформацій, інтегруючи передові досягнення у сферах кінематографії (наприклад, камери високої роздільної здатності), постпродакшну (монтаж, спецефекти, комп'ютерна графіка), а також дистрибуції.

Розроблена система являє собою інтегроване рішення, що поєднує функціональність веб-дodatка, бібліотеки Python для аналізу даних та математичні алгоритми для розрахунку прогнозів. Застосунок розроблено з урахуванням простоти використання, що дозволяє будь-якому користувачеві з мінімальними технічними знаннями ввести назву фільму та отримати рекомендацію щодо його перегляду.

Технічна реалізація проекту виконана за допомогою Visual Studio 2019 для розробки клієнтської та серверної частин веб-застосунку. Етапи

					БР.ІІІ – 54.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		14

очищення та імпорту даних були реалізовані в середовищі розробки Spyder IDE. Завдяки застосованим методам аналізу даних та моделюванню, точність прогнозування успішності релізів досягає 95%. Цей високий показник точності свідчить про надійність системи як інструменту для прийняття обґрунтованих рішень щодо перегляду фільмів.

1.2. Виклики та рішення процесу прогнозування успішності кінопродукції

Незважаючи на постійне прагнення кожної кіноіндустрії до оптимізації продуктивності та подолання ринкових бар'єрів, однією з ключових проблем залишається недостатній рівень довіри аудиторії до майбутніх кінорелізів. Традиційні маркетингові та промоційні стратегії, включаючи випуск трейлерів, тизерів, розміщення реклами на телебаченні, у друкованих виданнях та соціальних мережах, а також організацію соціальних заходів, часто виявляються недостатньо ефективними. Як наслідок, навіть високоякісні фільми можуть не досягати очікуваних комерційних результатів, що свідчить про розрив між маркетинговими зусиллями та реальним сприйняттям аудиторії.

1.2.1. Мета та механізм прогнозування успішності фільмів

Цей проект спрямований на вирішення зазначеної проблеми шляхом підвищення довіри глядачів до майбутніх кінорелізів. Розроблений веб-застосунок реалізує механізм прогнозування успішності фільму ("хіт" або "провал"), надаючи потенційним глядачам обґрунтовану інформацію для прийняття рішення щодо перегляду.

Основна мета цього додатку полягає у прогнозуванні касових показників майбутніх фільмів шляхом аналізу історичних даних про успішність залученого акторського складу, режисерів, продюсерів та

					БР.ІІІ – 54.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		15

музичних режисерів. Прогнозування здійснюється за допомогою простого арифметичного рівняння, що дозволяє інтегрувати попередні показники кожного члена творчої команди. Система враховує внесок кожного ключового учасника – від головного актора до музичного режисера – оскільки успішність фільму є результатом злагодженої роботи всієї команди.

На поточний момент додаток орієнтований на індійську Боллівудську індустрію. Цей вибір зумовлений доступністю та якістю наявних даних, що дозволяє досягти високої точності прогнозування. Поточна точність прогнозування становить 95% на основі доступного, хоча і обмеженого, набору даних.

1.2.2. Переваги та перспективи

Надаючи глядачам можливість самостійно оцінити потенційну успішність фільму, додаток дозволяє їм приймати обґрунтовані рішення, потенційно економлячи їхні кошти. Більше того, у довгостроковій перспективі, розширення використання таких прогностичних систем може зменшити навантаження на маркетингові та PR-відділи кінокомпаній, дозволяючи їм менше зосереджуватися на агресивному управлінні громадською думкою та більше на створенні якісного контенту. Це відкриває нові можливості для оптимізації маркетингових бюджетів та підвищення загальної ефективності кіноіндустрії.

1.3. Актуальний стан прогнозування успішності кінорелізів. Опис існуючих платформ

Наразі в кіноіндустрії відсутні комплексні інтегровані додатки, що дозволяють ефективно прогнозувати комерційну успішність кінофільмів до їхнього офіційного виходу в прокат. Існуючі механізми оцінки здебільшого базуються на ретроспективному аналізі та суб'єктивних оцінках. Зокрема,

					БР.ІІІ – 54.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		16

доступні численні платформи, де кінокритики публікують свої рецензії, а глядачі обмінюються особистими враженнями від перегляду. Однак, всі ці відгуки та рецензії стають доступними лише після релізу фільму, що не дозволяє потенційній аудиторії отримати об'єктивну інформацію для прийняття рішення про перегляд на початковому етапі.

Відсутність інструментів попереднього прогнозування призводить до інформаційної невизначеності серед потенційних глядачів. Численні суперечливі відгуки, що публікуються на різних онлайн-платформах після виходу фільму, часто дезорієнтують аудиторію.

Сучасні маркетингові стратегії кіноіндустрії значною мірою зосереджені на створенні ажіотажу та підвищеної зацікавленості навколо майбутніх релізів. Це досягається шляхом інтенсивних промоційних кампаній, що включають:

- Акцент на провідних акторів (головний актор та акторка), їхню популярність та попередні успіхи.
- Широкомасштабні маркетингові заходи, такі як продаж сувенірної продукції (мерчандайзу).
- Управління інформаційним полем через співпрацю з друкованими та електронними засобами масової інформації.
- Організація рекламних акцій та спеціальних подій, що створюють відчуття ексклюзивності та неминучості перегляду.

Ці інтенсивні рекламні кампанії ефективно формують високий рівень очікування серед громадськості, спонукаючи людей долучатися до загального "ажіотажу" навколо фільму. Однак, такий підхід, хоч і сприяє початковим касовим зборам, не гарантує довгострокової успішності та часто призводить до розчарування аудиторії, якщо реальна якість фільму не відповідає створеним очікуванням. Це підкреслює необхідність розробки більш об'єктивних та науково обґрунтованих методів прогнозування успішності кінорелізів.

					БР.ІІІ – 54.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		17

Але існують платформи та інструменти для прогнозування успішності фільмів і ця сфера активно розвивається, особливо завдяки прогресу в галузі штучного інтелекту та аналізу великих даних. Проте важливо розрізняти їхню природу та доступність.

1.3.1. Професійні та закриті платформи для студій

Великі кіностудії та інвестиційні компанії активно використовують власні або сторонні високотехнологічні аналітичні системи для прогнозування касових зборів, оцінки ризиків та оптимізації маркетингових стратегій. Прикладами таких інструментів є наступні.

1. Cinelytic. Ця платформа використовує штучний інтелект для аналізу сценаріїв, акторського складу, жанрів та інших факторів, щоб прогнозувати комерційний успіх фільму та потенційну реакцію аудиторії. Відомо, що Warner Bros. зверталася до них для прогнозування успішності своїх фільмів.

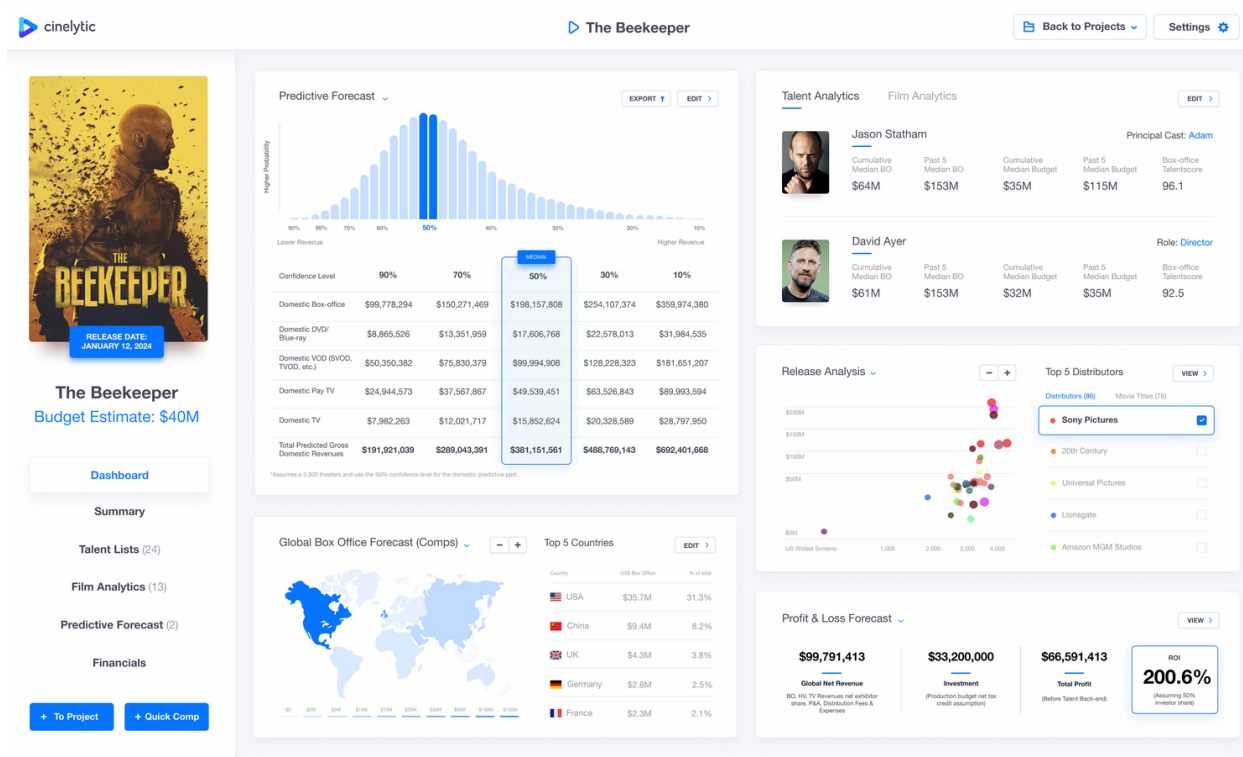


Рисунок 1.1 – Платформа Cinelytic

Cinelytic може оцінювати потенційні доходи фільму з різних каналів дистрибуції (кінотеатральний прокат, домашнє відео, кабельне телебачення, стрімінгові платформи) з різними рівнями впевненості. Платформа допомагає оцінити фінансові ризики та потенційну рентабельність інвестицій (ROI) проекту, надаючи сценарії ймовірності беззбитковості. Користувачі можуть моделювати різні сценарії, змінюючи такі параметри, як акторський склад, режисер, жанр, бюджет, дата релізу, стратегія дистрибуції, щоб побачити, як це вплине на прогностні показники.

Cinelytic призначена для професіоналів кіноіндустрії – студій, фінансистів, дистриб'юторів, продюсерських компаній. Вона допомагає їм:

- Приймати швидші та більш обґрунтовані рішення щодо "зеленого світла" (давати дозвіл на виробництво), придбання прав та стратегій релізу.
- Знижувати ризики та оптимізувати інвестиції, надаючи об'єктивні дані замість інтуїтивних припущень.
- Підвищувати ефективність та економити час, автоматизуючи аналіз даних, який раніше займав дні чи тижні.

Таким чином, Cinelytic позиціонується як потужний інструмент, що трансформує традиційні, часто інтуїтивні, процеси прийняття рішень у кіноіндустрії на основі даних та прогностної аналітики, що сприяє більш успішному комерційному результату.

2. ScriptBook. Ще одна платформа, яка спеціалізується на аналізі сценаріїв за допомогою ШІ для оцінки їхнього комерційного потенціалу. ScriptBook – це інноваційна платформа, що використовує штучний інтелект (ШІ) для глибокого аналізу кіносценаріїв та прогнозування їхнього комерційного та критичного успіху. Заснована у 2015 році, ця бельгійська компанія спрямована на революцію у кіновиробництві, надаючи об'єктивні, засновані на даних, інсайти, які допомагають студіям, продюсерам, дистриб'юторам та іншим зацікавленим сторонам приймати обґрунтовані рішення ще на ранніх стадіях розробки фільму.

					БР.ІІІ – 54.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		19

ScriptBook використовує складні алгоритми NLP для "читання" та розуміння тексту сценарію. Це дозволяє системі розкласти сценарій на окремі елементи, виявляти сюжетні лінії, характери, емоційний тон, структуру та інші ключові атрибути.

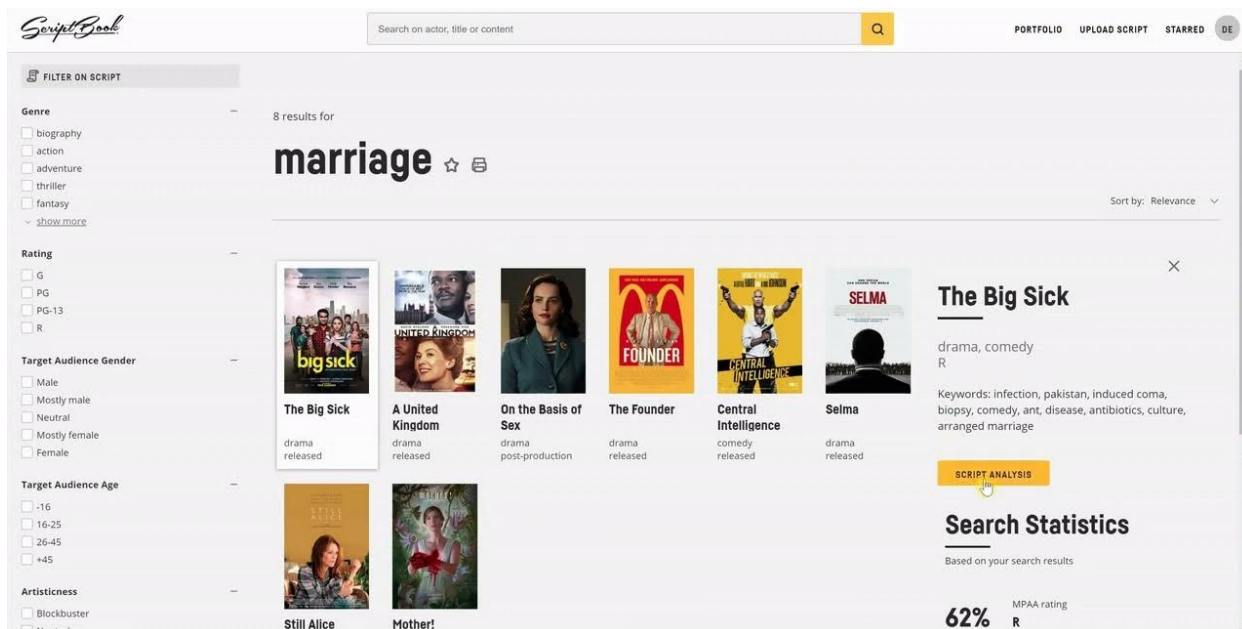


Рисунок 1.2 – Платформа ScriptBook

Платформа аналізує сценарій за більш ніж 400 параметрами, що охоплюють наративні елементи, психологію персонажів, динаміку сюжету та інші важливі характеристики. На відміну від простого аналізу ключових слів, ScriptBook фокусується на "метриці оповідання" (storytelling metric), намагаючись зрозуміти якість та потенціал самої історії. Метою ШІ-аналізу є усунення людської упередженості та суб'єктивності, яка часто присутня у традиційних рецензіях чи оцінках сценаріїв.

Система використовує великий набір історичних даних про сценарії та їхні фактичні касові збори для навчання своїх моделей. Вона може прогнозувати потенційний дохід фільму, враховуючи, які характеристики сюжетної лінії позитивно чи негативно впливали на касу в минулому.

					БР.ІІ – 54.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		20

Заявлена точність у прогнозуванні фінансових показників становить близько 87%.

ScriptBook позиціонує себе як потужний інструмент, що трансформує традиційний, часто суб'єктивний, процес оцінки сценаріїв у кіноіндустрії, замінюючи його на науково обґрунтовану, керовану даними, предиктивну аналітику, що зрештою має призвести до більш успішних комерційних та критичних релізів

3. Largo.ai. Також використовує ШІ для аналізу сценаріїв, демографічних даних аудиторії, акторського складу та режисерів для прогнозування успіху.

Largo.ai – це швейцарська компанія, яка розробила передову платформу на основі штучного інтелекту (ШІ) для оптимізації процесів створення та дистрибуції контенту в кіноіндустрії та рекламній сфері. Її місія полягає в тому, щоб зробити виробництво контенту більш керованим даними, зменшуючи ризики та збільшуючи потенційний прибуток для студій, продюсерів, дистриб'юторів, сценаристів, фінансистів та навіть акторів. Largo.ai здатна "читати" сценарії та виявляти сотні параметрів, включаючи жанрову рецептуру, емоційну інтенсивність (вимірюючи такі емоції, як радість, довіра, здивування, смуток, страх, гнів, відраза та нейтральність по ходу сюжету), ритм та темп оповідання. Це дозволяє зрозуміти "кінематографічну ДНК" контенту.

Largo.ai може прогнозувати потенційні доходи фільму як від кінопрокату, так і від стрімінгових платформ для різних територій. Точність цих прогнозів, за заявками компанії, може сягати 80%. Largo.ai створює "кінематографічну ДНК" кожного персонажа в сценарії та пропонує акторів, які найкраще підходять для цих ролей, враховуючи їхні попередні успіхи, вартість та відповідність характеру. Це допомагає студіям та продюсерам оптимізувати кастинг у межах бюджету. Аналізує запропоновані шорт-листи

					БР.ІІІ – 54.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		21

акторів та надає оцінки ризиків для кожного з них, виходячи з їхнього потенційного впливу на успіх проекту.

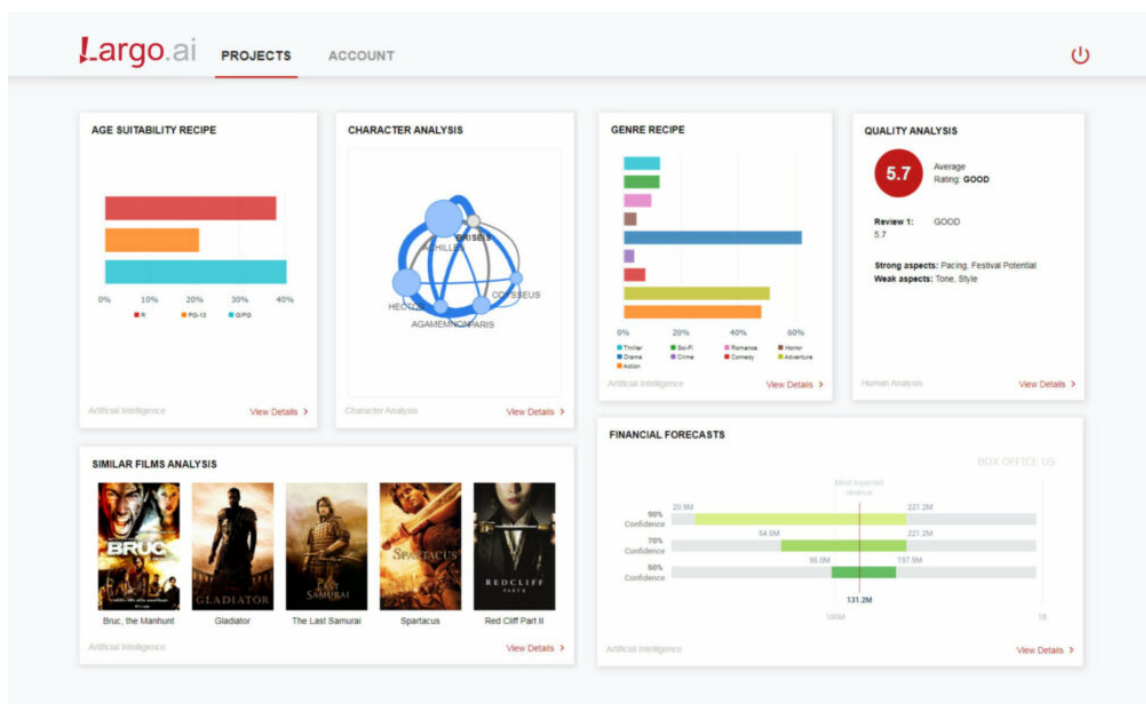


Рисунок 1.3 – Вигляд платформи Largo.ai

Largo.ai позиціонує себе як "рецепт успіху" для аудіовізуальної індустрії. Її основна мета – заповнити прогалину в використанні даних між фазами виробництва та дистрибуції, де традиційно переважали інтуїтивні рішення. Платформа допомагає:

- Знизити ризики шляхом надання точних прогнозів та аналізу потенційних проблем ще на ранніх стадіях.
- Збільшити прибутковість оптимізуючи стратегії кастингу, маркетингу та дистрибуції для максимізації доходів.
- Прискорити процеси автоматизуючи аналіз даних та надаючи інсайти в режимі реального часу.
- Демократизувати доступ до даних надаючи інструменти, подібні тим, що використовуються великими стрімінговими платформами, ширшому колу учасників індустрії.

Завдяки своїй здатності аналізувати величезні обсяги даних, виявляти закономірності та надавати прогнози, Largo.ai прагне зробити процес створення фільмів та реклами більш ефективним, науково обґрунтованим та, зрештою, більш успішним.

1.3.2. Академічні дослідження та наукові розробки в області прогнозування успішності

Велика кількість наукових робіт і дисертацій присвячена розробці моделей прогнозування успішності фільмів. Вони використовують різні методи машинного навчання (Random Forest, XGBoost, Support Vector Machine, нейронні мережі), аналіз настроїв соціальних мереж, історичні дані про касові збори, інформацію про акторів, режисерів, жанри, бюджет, а також дані з IMDb, Rotten Tomatoes, Metacritic тощо. Ці дослідження демонструють високу точність, але часто є прототипами або фокусуються на конкретних аспектах, а не на повноцінному користувацькому застосунку.

1.3.3. Платформи для відстеження та оцінки фільмів

Існують популярні сайти, такі як Letterboxd, Trakt.tv, IMDb, Rotten Tomatoes, які дозволяють користувачам оцінювати фільми, писати відгуки та відстежувати свої перегляди. Деякі з них можуть пропонувати персоналізовані рекомендації (щодо того, які фільми можуть сподобатися користувачеві), але це не є прямим прогнозуванням комерційної успішності фільму до його виходу.

IMDb (Internet Movie Database) — це найбільша та найавторитетніша онлайн-база даних, що містить вичерпну інформацію про кінофільми, телевізійні серіали, відеоігри, а також про акторів, режисерів, знімальні групи та інших учасників кіноіндустрії. Заснована у 1990 році Колом Нідхемом як фанатська база даних на Usenet, вона еволюціонувала в

					БР.ІІІ – 54.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		23

повноцінний веб-сайт у 1993 році, а з 1998 року є дочірньою компанією Amazon.

The image shows a screenshot of the IMDb website. On the left, the 'IMDb Charts' section displays a list of 'Most Popular Movies' as determined by users. The list includes titles like 'Thor: Love and Thunder (2022)', 'Pinocchio (2022)', 'Babylon (2022)', 'The Whale (2022)', 'The Little Mermaid (2023)', 'Barbarian (2022)', and 'Don't Worry Darling (2022)'. Each entry shows its rank, IMDb Rating (e.g., 6.5, 5.1, 9.1), and a 'Your Rating' star icon. On the right, there is a movie poster for 'NOPE' by Jordan Peele, featuring Keke Palmer, Daniel Kaluuya, and Brandon Scott Lee. Below the poster, it says 'ON DIGITAL NOW' and 'BUY'. Further down, there are sections for 'You Have Seen' (0/100) and 'IMDb Charts' with links to various categories like 'Box Office', 'Most Popular Movies', etc.

Рисунок 1.4 – Приклад IMDb рейтингу

IMDb розраховує середньозважену оцінку (зважена на кількість голосів), яка відображається на сторінці фільму. Цей метод розрахунку є пропрієтарним і не розкривається, щоб запобігти маніпуляціям, але він прагне забезпечити справедливе представлення оцінок.

IMDb агрегує дані з різних джерел: офіційні прес-релізи студій, інформація від виробничих компаній, а також значна частина даних надходить від спільноти користувачів. Користувачі можуть надсилати виправлення, доповнення та нові записи, які потім перевіряються та затверджуються редакторами IMDb. Це робить базу даних надзвичайно об'ємною та актуальною.

Letterboxd — це соціальна мережа, розроблена спеціально для кіноманів, яка дозволяє користувачам відстежувати, оцінювати, рецензувати та обговорювати фільми. Запущена у 2011 році в Окленді, Нова Зеландія, Метью Б'юкененом та Карлом фон Рандоу, платформа швидко здобула популярність завдяки своєму фокусу на спільноті та зручному інтерфейсу. Її часто називають "Goodreads для фільмів" через схожий функціонал для відстеження переглядів та взаємодії з іншими користувачами.

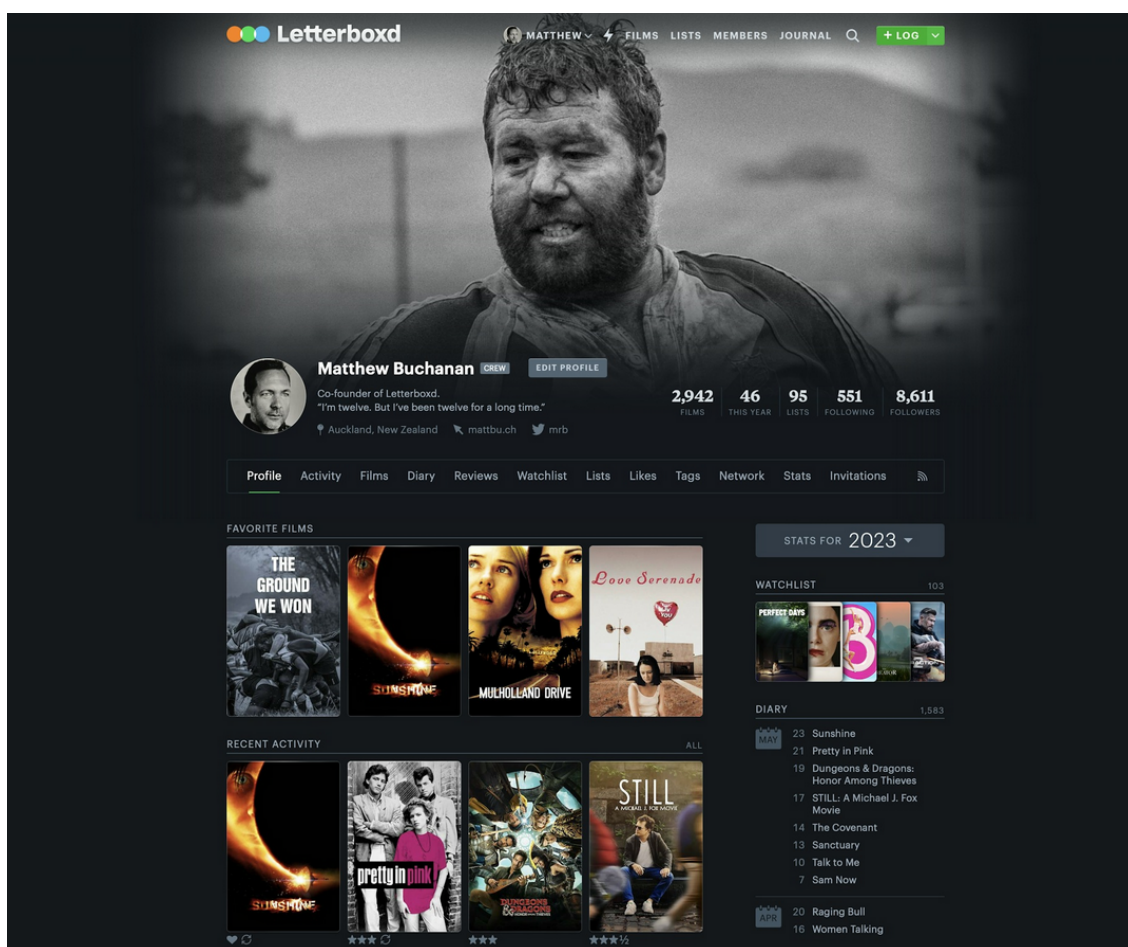


Рисунок 1.5 – Проект Letterboxd

Основна функція, яка дозволяє користувачам вести щоденник своїх переглянутих фільмів. При "логуванні" фільму можна зазначити дату перегляду, поставити оцінку (від 0.5 до 5 зірок), написати рецензію (з можливістю позначити спойлери) та додати теги.

					БР.ІІ – 54.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		25

Рецензії на Letterboxd часто відрізняються креативністю, гумором та особистим підходом, створюючи унікальну культуру спілкування навколо кіно. Можна писати як короткі замітки, так і розгорнуті есе. На відміну від IMDb, яка є більш енциклопедичною базою даних, Letterboxd акцентує увагу на спілкуванні та обміні враженнями між кіноманами.

Trakt.tv – це популярна веб-платформа та сервіс, призначений для відстеження переглянутих фільмів та серіалів, їх організації, відкриття нового контенту та взаємодії з спільнотою кіноманів. Вона слугує своєрідним "щоденником переглядів" та соціальною мережею для любителів кіно та телебачення.

Основні функції та можливості Trakt.tv:

- Користувачі можуть легко реєструвати всі фільми та епізоди серіалів, які вони переглянули. Це може робитися вручну, або автоматично через інтеграції з медіацентрами (наприклад, Kodi, Plex) або іншими додатками, що підтримують Scrobbling (автоматичний запис переглянутого).

- Trakt відстежує ваш прогрес по серіалах, показуючи, які епізоди ви вже подивилися, а які ще залишилися до перегляду.

- Можливість позначати фільми та серіали, які ви маєте у своїй колекції (наприклад, на Blu-ray, DVD або в цифровому форматі).

- Календар майбутніх прем'єр фільмів та виходу нових епізодів серіалів.

- Персоналізовані рекомендації фільмів та серіалів на основі вашої історії переглядів, оцінок та вподобань.

- Розширені фільтри для пошуку контенту за жанрами, роком, країною, мовою, студіями та іншими параметрами.

- Інтеграція з сервісами, такими як JustWatch, дозволяє знайти, на яких стрімінгових платформах (Netflix, Amazon Prime, HBO Max тощо) доступний певний фільм або серіал у регіоні.

					БР.ІІІ – 54.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		26

- Створення необмеженої кількості власних списків для організації контенту на свій розсуд (наприклад, "Фільми для Хелловіну", "Топ-10 серіалів").

- Можливість створювати спільні списки з друзями, що дозволяє колективно планувати перегляди.

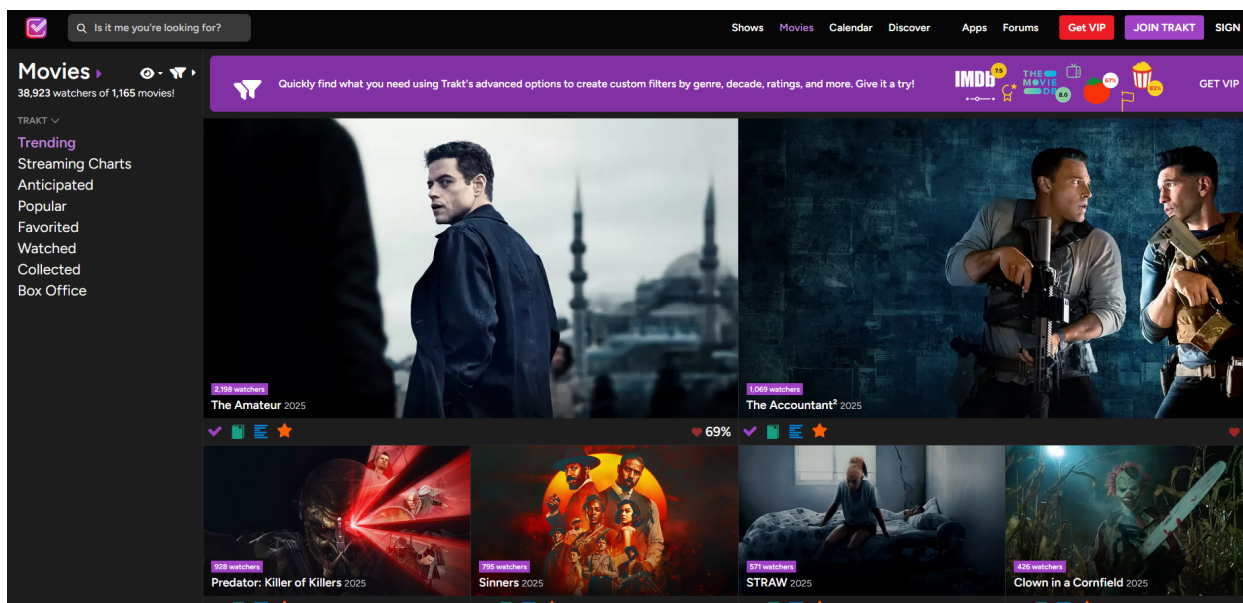


Рисунок 1.6 – Платформа Trakt.tv

Також є можливість оцінювати фільми та серіали, а також залишати коментарі та відгуки. Присутня детальна персоналізована статистика щодо ваших звичок перегляду (кількість переглянутих годин, улюблені жанри, актори, режисери, річні огляди тощо).

Trakt.tv надає широкий та добре документований API, який дозволяє розробникам інтегрувати функціональність Trakt.tv у власні додатки та сервіси. Це є однією з ключових переваг платформи, що дозволяє створювати сторонні клієнти, інструменти синхронізації та аналізу.

Завдяки відкритому API існує велика екосистема додатків та плагінів (наприклад, для Kodi, Plex, медіаплеєрів), які автоматично синхронізують перегляди з Trakt.tv.

									Арк.
									27
Змн.	Арк.	№ докум.	Підпис	Дата					

Trakt.tv використовує модель Freemium. Більшість основних функцій доступні безкоштовно, але деякі розширені можливості доступні лише користувачам з VIP-підпискою. Важливою особливістю політики Trakt.tv є те, що вони не продають дані користувачів та не розміщують сторонню рекламу, покладаючись на дохід від VIP-підписок.

Загалом, Trakt.tv є цінним інструментом для організації та збагачення досвіду перегляду фільмів та серіалів, а його API робить його потужною платформою для розробників, які створюють супутні програми.

1.3.4. Ключова відмінність пропонованого рішення

Відмінність пропонованого програмного рішення полягає в наступному:

1. Доступність для широкого загалу.

Згадані професійні платформи є комерційними продуктами з високою вартістю, призначеними для бізнесу, а не для рядового споживача, який просто хоче дізнатися, чи варто йому дивитися фільм.

2. Фокус на "довірі глядачів".

Даний проект більше зосереджений на наданні об'єктивної оцінки для кінцевого споживача, щоб він міг прийняти зважене рішення, а не просто потрапити під вплив маркетингового ажіотажу. Більшість існуючих платформ III працюють для кіностудій та інвесторів, допомагаючи їм мінімізувати ризики та оптимізувати прибутки, а не інформувати пересічного глядача.

3. Використання "простого арифметичного рівняння" та специфічного набору даних.

Це вказує на більш спрощений, але, можливо, більш доступний для розуміння кінцевим користувачем підхід, відмінний від складних нейронних мереж, які використовуються у комерційних системах.

					БР.ІІІ – 54.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		28

Таким чином, запропонований проект буде загальнодоступний, простий у використанні додаток для пересічного споживача, який прогнозує успішність фільму (хіт/провал) до його виходу, базуючись на аналізі ключових учасників команди (акторів, режисерів тощо) і буде доступним безкоштовно. Більшість існуючих рішень є або складними аналітичними інструментами для індустрії, або академічними прототипами, або рекомендаційними системами, які працюють з вже випущеними фільмами.

Отже, підкреслюючи відсутність доступних для широкої аудиторії, простих у використанні інструментів предиктивного аналізу комерційного успіху кінорелізів, орієнтованих на прийняття рішення кінцевим споживачем, даний проект вважається актуальною розробкою.

					БР.ІІІ – 54.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		29

РОЗДІЛ 2. АЛГОРИТМІЧНА РЕАЛІЗАЦІЯ ТА МОДЕЛЮВАННЯ СТРУКТУРИ ВЕБ-ЗАСТОСУНКУ ПРОГНОЗУВАННЯ УСПІШНОСТІ РЕЛІЗІВ

2.1. Архітектура та функціонування веб-застосунку прогнозування успішності релізів кіноіндустрії

Представлений веб-застосунок для прогнозування успішності розважальних релізів функціонує на основі двохетапного алгоритму усереднення.

На першому етапі здійснюється розрахунок індивідуального рейтингу для кожного члена знімальної групи (актора, режисера, продюсера, музичного директора) на основі історичних даних, що зберігаються у відповідному наборі даних. Отримані на цьому етапі рейтинги слугують вхідними параметрами для другого етапу.

На другому етапі застосунок інтегрує ці індивідуальні рейтинги за допомогою функції середнього значення для генерації остаточного вердикту щодо фільму: буде він "хітом" чи "провалом".

2.1.1. Технічні вимоги та розгортання системи

Для успішної компіляції та функціонування застосунку необхідно дотриматися таких мінімальних системних вимог:

- Мова програмування: Версія C# має бути 16 або вищою.
- Середовище розробки: Microsoft Visual Studio 2019 або новіша версія середовища.
- Серверний компонент: Застосунок розгортається на веб-сервері WampServer, що може бути розміщений як на настільному комп'ютері, так і на мобільному пристрої для обробки запитів.

					БР.ІІІ – 54.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		30

2.1.2. Механізм обробки даних та навчання системи

У випадку, якщо для деяких членів знімальної групи, що беруть участь у запиті користувача, відсутні дані в історичному наборі, система передбачає можливість ручного введення користувачем відповідного рейтингу для цих відсутніх елементів. Варто зазначити, що відсутність даних для окремих членів команди не впливає на кінцевий результат прогнозування. Після генерації прогнозу, отриманий результат (включаючи статус фільму та залучених членів знімальної групи) зберігається в базі даних. Це дозволяє системі постійно навчатися та адаптуватися до динаміки даних, що надходять з часом, підвищуючи точність майбутніх прогнозів.

2.1.3. Специфікація системних вимог прототипу

Проект розробки веб-застосунку для прогнозування успішності розважальних релізів охоплює чотири ключові етапи: імпорт даних, очищення даних, розробка та виробництво. Для етапу прототипу були визначені такі мінімальні апаратні та програмні вимоги:

Апаратні вимоги:

- Процесор: Intel Core i3-класу або еквівалентний.
- Оперативна пам'ять (RAM): Не менше 1 ГБ.
- Жорсткий диск (HDD): Мінімум 256 ГБ вільного простору.

Програмні вимоги:

- Операційна система: Windows 10, Windows 11.
- Середовище розробки: Visual Studio 2010 або вище.
- Система керування базами даних (СКБД): SQL Server 2008 або вище.

Зазначені вимоги є мінімальними, і використання більш сучасних версій програмного забезпечення та потужнішого апаратного забезпечення забезпечить оптимальну продуктивність та стабільність застосунку прогнозування релізів.

					БР.ІІІ – 54.00.00.000 ПЗ	Арк.
						31
Змн.	Арк.	№ докум.	Підпис	Дата		

2.2. Дослідження доцільності розробки веб-застосунку

Дослідження доцільності є критично важливим етапом у процесі розробки будь-якого проекту, зокрема інформаційної системи. Його метою є всебічний аналіз поточного стану, наявного досвіду та майбутніх перспектив для відповіді на ключові питання, що виникають у керівництва проекту. Це дослідження проводиться після чіткого визначення та розуміння поставленої проблеми. Воно має встановити, чи виправдовує запропонована система інвестиції часу та ресурсів, і чи зможе вона ефективно вирішити заявлені проблеми після її впровадження. Дослідження доцільності дозволяє менеджменту візуалізувати концепцію майбутньої системи та врахувати досвід попередніх аналогічних розробок.

Найбільш поширеними та важливими видами дослідження доцільності для прийняття рішень є технічна, економічна та операційна доцільність. Комплексний аналіз охоплює ідею продукту, ринковий попит та технічну можливість реалізації. Результатом дослідження є звіт, який дозволяє ідентифікувати та усунути потенційні проблеми до їх виникнення.

У контексті розробки веб-застосунку прогнозування для кіноіндустрії, ми розглянемо три ключові аспекти доцільності: технічний, економічний та операційний.

2.2.1. Технічна доцільність

Технічна доцільність полягає в оцінці необхідного рівня технічних знань, ресурсів та інфраструктури для успішної реалізації системи. Вона дозволяє розробникам визначити, чи потрібно здобувати додаткові навички, і як це вплине на часові та фінансові показники проекту. Це ітеративний процес, який включає оцінку наступних вимог:

- Підтримка оновлень: Система має бути здатною до адаптації та підтримки майбутніх оновлень і нових релізів.

					БР.ІІІ – 54.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		32

- Безпека, масштабованість та оптимальність: Застосунок повинен відповідати високим стандартам безпеки, бути масштабованим для зростаючих обсягів даних та користувачів, а також забезпечувати оптимальну продуктивність.

- Відгук та інтерактивність: Система повинна бути чутливою та оперативно реагувати на дії користувача.

Після задоволення всіх перерахованих вимог система вважається технічно доцільною.

Визначені технічні вимоги для прототипу:

- Операційна система: Windows 10/11.
- Мови програмування та фреймворки: ASP.NET з C# (.NET Framework 2010).
- Система управління базами даних: MS-SQL Server.
- Інструмент для документування: MS Word 2013.

2.2.2. Економічна доцільність

Економічна доцільність є одним з визначальних факторів життєздатності проекту, оскільки вона оцінює фінансову доцільність його реалізації. У даному проекті, розробка прогностичного веб-застосунку визнається фінансово доцільною, оскільки вона не потребує значних додаткових фінансових інвестицій, а прогнозований термін завершення проекту становить шість місяців.

В рамках економічної доцільності, поточні витрати порівнюються з майбутніми фінансовими вигодами від впровадження запропонованої системи. Розраховується показник повернення інвестицій (ROI). Проект вважається фінансово доцільним, якщо очікувані прибутки від інвестицій перевищують початкові вкладення. Економічна доцільність підсумовує, чи може проект бути реалізований за наявних ресурсів.

При оцінці економічної доцільності перевіряються наступні аспекти:

					БР.ІІІ – 54.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		33

- Вартість повного дослідження системи.
- Вартість встановлення та експлуатації системи.
- Вартість розробки, включаючи ліцензування інструментів.

Враховуючи, що очікуване повернення від цього проекту є значно вищим у порівнянні з необхідними інвестиціями, проект визнається економічно доцільним.

2.2.3. Операційна доцільність

Операційна доцільність визначає, наскільки запропонована система відповідатиме операційним вимогам та функціональності для кінцевих користувачів, а також оцінює необхідні ресурси (робоча сила, час) для її проектування, розробки та підтримки. Найбільш операційно доцільним вважається рішення, що вимагає мінімальних зусиль для реалізації та інтеграції.

Ключові критерії операційної доцільності включають:

1. Прийняття культури роботи.

Стиль виконання, планування та дизайн продукту повинні бути прийнятні всіма членами команди та зацікавленими сторонами.

2. Безперебійна робота.

Запропонована система не повинна створювати жодних проблем для кінцевого користувача за будь-яких обставин та функціонувати належним чином.

3. Відповідність вимогам клієнтів.

Система має задовольняти всі визначені вимоги та функціональність клієнтів.

У рамках даного дослідження доцільності, вимоги до часових ресурсів та кадрового забезпечення були задоволені, що підтверджує операційну доцільність запропонованого рішення.

					БР.ІІІ – 54.00.00.000 ПЗ	Арк.
						34
Змн.	Арк.	№ докум.	Підпис	Дата		

2.3. Функціональне моделювання системи за допомогою діаграми варіантів використання

Діаграма варіантів використання є фундаментальним компонентом у процесі моделювання програмного забезпечення, що забезпечує логічне розуміння загальної архітектури системи через призму її сценаріїв взаємодії. Вона відіграє ключову роль у визначенні та візуалізації всіх функціональних та нефункціональних вимог, які впливають на систему як внутрішньо, так і зовнішньо. Ця діаграма наочно демонструє способи, за допомогою яких користувач взаємодіє з системою.

Після ідентифікації вимог до системи, ми переходимо до її системного аналізу, що передбачає генерацію варіантів використання та ідентифікацію відповідних сутностей. Діаграма варіантів використання відображає всі можливі шляхи взаємодії користувача з системою. Стандартним підходом до розробки цієї діаграми є декомпозиція системи на дрібніші функціональні модулі, для кожного з яких створюються окремі варіанти використання. Такий підхід дозволяє усунути дублювання функціональності на діаграмі та забезпечити її чіткість.

У системному середовищі користувачі можуть виконувати різні ролі. Для забезпечення максимальної операційної доцільності та уникнення плутанини користувача, найбільш ефективним рішенням є надання чіткої та зрозумілої діаграми варіантів використання. Ця діаграма має бути динамічною і оновлюватися при кожній зміні функціональності або вимог до системи.

Діаграма варіантів використання ілюструє функціональні можливості, що надаються системою. У контексті діаграми варіантів використання, кінцевий користувач системи позначається як актор. Для поточної системи передбачено одного єдиного актора – Адміністратора.

					БР.ІІІ – 54.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		35

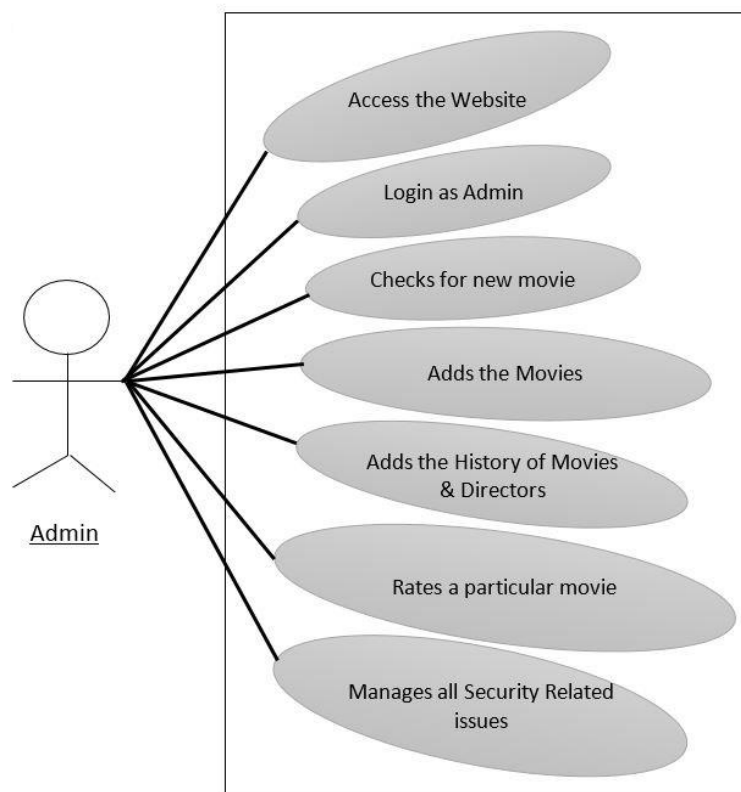


Рисунок 2.1 - Діаграма варіантів використання, що візуально представляє взаємодію актора "Адміністратор"

Діаграма (рис. 2.1) ідентифікує наступні варіанти використання, які може виконувати актор "Адміністратор":

- Access the Website (Доступ до веб-сайту): Цей варіант використання є базовою точкою входу до системи. Він передбачає можливість адміністратора отримати доступ до інтерфейсу веб-застосунку.

- Login as Admin (Вхід як адміністратор): Після доступу до веб-сайту, адміністратор повинен пройти процес автентифікації для отримання відповідних привілеїв та доступу до адміністративних функцій.

- Checks for new movie (Перевірка наявності нового фільму): Цей варіант використання передбачає функціональність для адміністратора, що дозволяє перевіряти наявність інформації про нові фільми, які ще не були додані до системи. Це може включати моніторинг зовнішніх джерел даних або внутрішні перевірки.

- Adds the Movies (Додавання фільмів): Після ідентифікації нового фільму, адміністратор має можливість ввести та зберегти його дані у системі. Це включає назву, рік випуску, жанр, бюджет, та іншу релевантну інформацію.

- Adds the History of Movies & Directors (Додавання історії фільмів та режисерів): Цей варіант використання є критично важливим для прогностичної моделі. Він вказує на те, що адміністратор відповідає за внесення історичних даних про попередні фільми, в яких брали участь актори, режисери, продюсери, а також дані про касові збори цих фільмів. Це дозволяє системі накопичувати базу для розрахунку ефективності.

- Rates a particular movie (Оцінювання конкретного фільму): Цей варіант використання може стосуватися введення або корекції адміністратором оцінок для вже існуючих фільмів. У контексті прогностичної моделі, це може бути функція для калібрування або доповнення даних, що впливають на розрахунок "рейтингу" членів знімальної групи, як було згадано в попередньому описі двохетапного алгоритму.

- Manages all Security Related issues (Керування всіма питаннями, пов'язаними з безпекою): Цей комплексний варіант використання охоплює всі аспекти управління безпекою системи. Він може включати управління обліковими записами користувачів (якщо такі будуть додані у майбутньому), контроль доступу, моніторинг загроз, оновлення системи безпеки та інші адміністративні завдання, спрямовані на захист даних та функціональності.

Діаграма підтверджує, що на поточному етапі система є переважно бек-офісним інструментом, орієнтованим на збір, обробку та управління даними, які є основою для прогностичної моделі. Відсутність інших акторів, крім "Адміністратора", свідчить про те, що інтерфейс для кінцевого користувача є публічною частиною, яка не потребує автентифікації або розширених

функцій взаємодії, і його функціональність не відображена на цій адміністративній діаграмі варіантів використання.

2.4. Розробка діаграми послідовності взаємодії адміністратора з системою

Представлена діаграма (рис. 2.2) послідовності (Sequence Diagram) ілюструє динамічну взаємодію між актором "Адміністратор" та елементами системи "Movie Success Prediction System" (Система прогнозування успішності фільмів) та "Main Page" (Головна сторінка). Ця діаграма деталізує послідовність повідомлень та дій, що виконуються для реалізації адміністративних функцій.

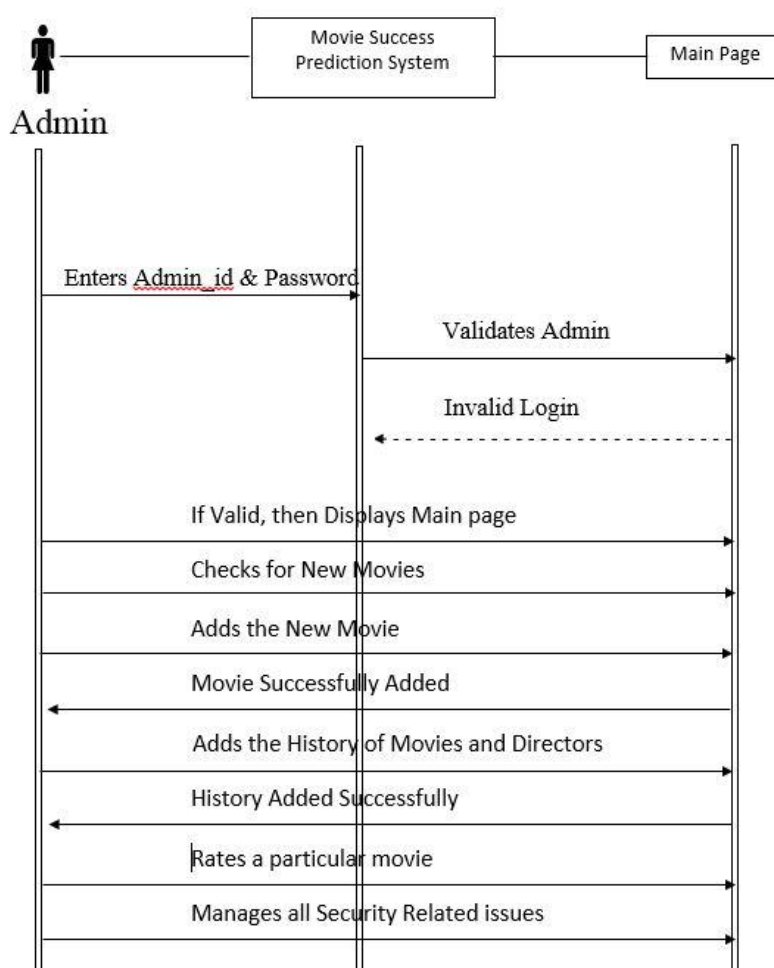


Рисунок 2.2 – Діаграма послідовності

Основні елементи діаграми:

- Актор: "Admin" – ініціатор всіх представлених операцій.
- Об'єкти:
 - "Movie Success Prediction System" (Система прогнозування успішності фільмів) – центральний компонент, що обробляє запити адміністратора.
 - "Main Page" (Головна сторінка) – компонент користувацького інтерфейсу, що відображає інформацію та взаємодіє з адміністратором.

Діаграма відображає наступну послідовність дій та повідомлень:

- Автентифікація Адміністратора:

1. Адміністратор вводить Admin_id & Password (логін та пароль) до системи прогнозування.
2. Система валідує Адміністратора (Validates Admin). Це внутрішній процес перевірки облікових даних.
3. У випадку недійсної спроби входу (Invalid Login), система повертає відповідне повідомлення адміністратору.
4. Якщо облікові дані валідні (If Valid), Система прогнозування успішності фільмів передає управління до Головної сторінки (Displays Main page), яка відображається адміністратору.

- Керування Фільмами:

1. Адміністратор ініціює запит до Головної сторінки для перевірки нових фільмів (Checks for New Movies).
2. Після цього, Адміністратор додає новий фільм (Adds the New Movie) через інтерфейс Головної сторінки, яка, ймовірно, взаємодіє з базовою Системою.
3. Система повідомляє Адміністратора про успішне додавання фільму (Movie Successfully Added).

- Керування історичними даними:

1. Адміністратор додає історію фільмів та режисерів (Adds the History of Movies and Directors). Це включає внесення даних про попередні

					БР.ІІІ – 54.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		39

показники фільмів та ефективність учасників знімальної групи, що є критично важливим для прогностичної моделі.

2. Система підтверджує успішне додавання історії (History Added Successfully).

- Оцінювання Фільмів:

Адміністратор може оцінювати конкретний фільм (Rates a particular movie). Це пов'язано з функцією введення або коригування даних для розрахунку "рейтингу" членів знімальної групи або для калібрування моделі.

- Керування Безпекою:

Адміністратор керує всіма питаннями, пов'язаними з безпекою (Manages all Security Related issues). Це узагальнене повідомлення, що охоплює широкий спектр адміністративних завдань, спрямованих на підтримання цілісності та захищеності системи.

Операції додавання нових фільмів та історичних даних про них, а також оцінювання фільмів, є безпосередньо пов'язаними з функціональністю прогнозування успішності. Функція управління безпекою підкреслює важливість захисту адміністративного доступу та даних системи.

Ця діаграма ефективно демонструє потік контролю та повідомлень для основних адміністративних сценаріїв, надаючи чітке уявлення про те, як адміністратор взаємодіє з системою для підтримки її функціональності та наповнення даних.

2.5. Розробка алгоритму (діаграми активності) взаємодії з системою прогнозування

Представлена блок-схема ілюструє послідовність дій та рішень у процесі взаємодії адміністратора з системою прогнозування успішності

					БР.ІІІ – 54.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		40

фільмів, починаючи від входу в систему і завершуючи виконанням адміністративних функцій.

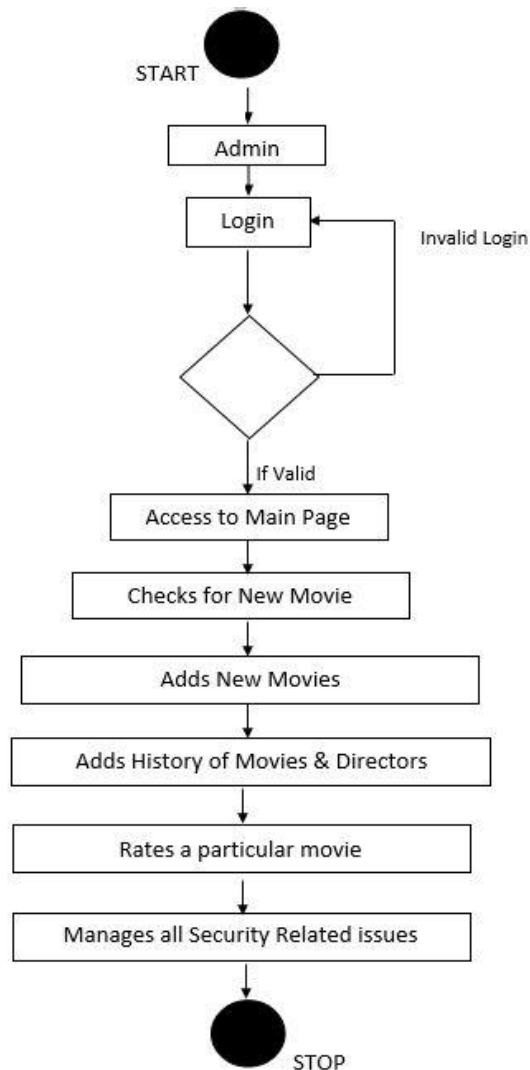


Рисунок 2.3 – Діаграма активності

Послідовність виконання алгоритму:

1. START (Початок): Алгоритм починається.
2. Admin - цей блок символізує початкову ініціацію процесу з боку адміністратора або його роль у системі.
3. Login - адміністратор виконує спробу входу до системи.
4. Рішення - системна перевірка валідності облікових даних для входу.

- If Valid - якщо облікові дані правильні, потік виконання переходить до наступного кроку.

- Invalid Login - якщо облікові дані невірні, потік повертається до кроку "Login", що означає необхідність повторної спроби входу.

5. Access to Main Page - після успішної автентифікації адміністратор отримує доступ до головного інтерфейсу системи.

6. Checks for New Movie - адміністратор перевіряє наявність інформації про нові фільми, які ще не внесені в базу даних.

7. Adds New Movies - адміністратор вносить дані про нові фільми до системи.

8. Adds History of Movies & Directors - адміністратор додає історичні дані про попередні фільми та їхніх режисерів (та інших членів знімальної групи), що є основою для прогностичної моделі.

9. Rates a particular movie - адміністратор виконує оцінювання певного фільму, що слугує для оновлення або калібрування даних, які використовуються в алгоритмі прогнозування.

10. Manages all Security Related issues - адміністратор виконує функції, пов'язані з підтриманням безпеки системи. Це може включати управління обліковими записами, правами доступу, моніторинг тощо.

11. STOP - алгоритм завершується після виконання всіх адміністративних операцій.

Алгоритм чітко визначає послідовність, в якій адміністратор виконує завдання, починаючи з наповнення бази даних (додавання фільмів та історичних даних) і закінчуючи підтримкою системи (оцінювання, безпека). Блок-схема підкреслює, що адміністратор є центральною фігурою, яка виконує всі операції з керування даними та підтримки системи.

Після входу, блок-схема не демонструє складних розгалужень рішень у робочому процесі адміністратора, що може вказувати на те, що ці операції є досить рутинними і виконуються відповідно до встановленого порядку.

					БР.ІІІ – 54.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		42

2.6. Моделювання інформаційних потоків у системі з використанням діаграми потоків даних (DFD)

Діаграма потоку даних (Data Flow Diagram, DFD) є фундаментальним компонентом системної документації та аналізу, що надає структуроване візуальне представлення руху даних в інформаційній системі. Її призначення полягає у візуалізації процесів передачі та трансформації даних між різними компонентами системи. Завдяки своїй здатності відображати логіку обробки інформації, DFD часто називають логічною діаграмою потоку даних. Цей інструмент є особливо цінним для розробників та аналітиків, оскільки дозволяє швидко орієнтуватися у структурі системи та виявляти взаємозв'язки між її частинами у випадку виникнення складнощів.

DFD, також відома як діаграма бульбашок (bubble chart), використовується для уточнення вимог до системи та ідентифікації ключових перетворень, необхідних для її проектування. Вона допомагає розкрити сутність системи, її основні функції та те, як дані протікають через неї.

Зазвичай DFD розробляється в ієрархічній структурі, що складається з кількох рівнів деталізації:

- Рівень 0 (Контекстна діаграма): Це найвищий рівень абстракції, який представляє систему як єдиний процес ("бульбашку"), взаємодіючий із зовнішніми сутностями (акторами або іншими системами). Його основна мета — визначити межі системи та її зв'язки із зовнішнім середовищем. Цей рівень є ключовим для початкового розуміння системи та її контексту.

- Рівень 1: Деталізує центральний процес рівня 0, розбиваючи його на основні підпроцеси. Кожен підпроцес показує, як дані перетворюються всередині системи, а також взаємодію з сховищами даних.

- Рівень 2: Надає ще більшу деталізацію підпроцесів з рівня 1, розкриваючи їхню внутрішню логіку та більш дрібніші потоки даних.

					БР.ІІІ – 54.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		43

Процес розробки DFD є ітеративним: кожен наступний рівень базується на деталізації попереднього. Цей процес продовжується до досягнення необхідного рівня деталізації, який дозволяє аналітику отримати повну та вичерпну перспективу щодо функціональності системи.

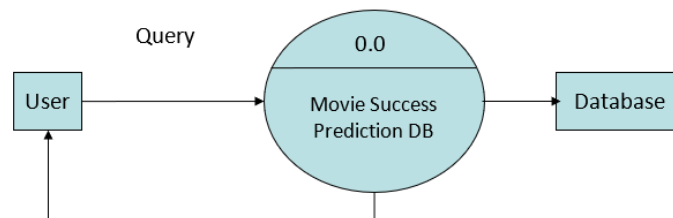


Рисунок 2.4 - Діаграма потоку даних рівня 0

Представлена діаграма потоку даних рівня 0 (рис. 2.4) надає високоабстраговане уявлення про систему та її взаємодію із зовнішнім середовищем. Цей рівень DFD зосереджується на визначенні меж системи та її основних зовнішніх інтерфейсів, не вдаючись у деталі внутрішніх процесів.

Розглянемо основні компоненти діаграми.

Центральний процес "Movie Success Prediction DB" це єдина "бульбашка" на діаграмі, що представляє всю систему прогнозування успішності фільмів як єдиний, цілісний процес.

Зовнішні Сутності:

- "User" - зовнішня сутність, яка взаємодіє з системою. З попередніх діаграм відомо, що це може бути як "Адміністратор" (для керуючих функцій), так і "Звичайний користувач" (для отримання прогнозу). На цьому контекстному рівні вони узагальнені як "User".

- "Database" - ще одна зовнішня сутність. Це сховище даних, з яким система взаємодіє. Важливо зазначити, що на контекстній діаграмі "Database", позначає зовнішнє сховище даних, а не внутрішні сховища, які

були б деталізовані на нижчих рівнях DFD. Це є зовнішня джерельна база даних, з якої система отримує історичні дані.

Потоки даних:

- "Query" - потік даних від "User" до "Movie Success Prediction DB". Це представляє вхідний запит від користувача до системи (наприклад, введення назви фільму для прогнозу).

- Потік від "Movie Success Prediction DB" до "User": Це потік даних, що йде від системи назад до користувача. У контексті прогностичної системи, це результат прогнозування (чи буде фільм хітом/провалом).

- Потік від "Movie Success Prediction DB" до "Database": Цей потік вказує на те, що система записує і оновлює дані у зовнішній базі даних. Це узгоджується з функціональністю додавання нових фільмів або збереження результатів прогнозування.

- Потік від "Database" до "Movie Success Prediction DB": Цей потік вказує на те, що система читає дані з зовнішньої бази даних. Це є критично важливим для отримання історичних даних, необхідних для прогностичного алгоритму.

Система приймає запити від користувачів, обробляє їх, взаємодіючи з зовнішньою базою даних (для читання історичних даних та запису результатів/нових даних), і повертає відповідний результат користувачу. На цьому рівні не розкриваються внутрішні процеси обробки даних, алгоритми прогнозування, або архітектура бази даних. Все це буде деталізовано на нижчих рівнях DFD (Рівень 1, Рівень 2). Ця контекстна діаграма є важливою відправною точкою для подальшої декомпозиції та деталізації системи.

Діаграма потоку даних рівня 1 (DFD Level 1) показана на рисунку 2.5 і деталізує центральний процес контекстної діаграми (Рівень 0). Вона розкриває основні підпроцеси та сховища даних, що входять до складу загальної системи, а також їхні взаємодії із зовнішніми сутностями. Ця

					БР.ІІІ – 54.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		45

діаграма показує, як саме дані перетворюються всередині системи для задоволення запиту користувача.

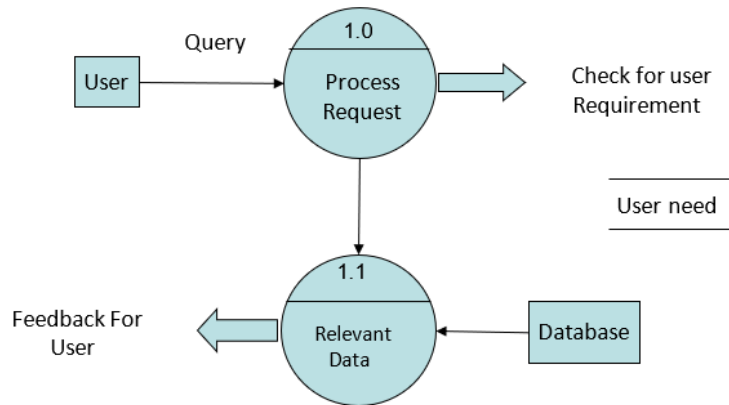


Рисунок 2.5 - Діаграма потоку даних рівня 1

Основні компоненти діаграми.

Зовнішні сутності - "User" - ініціатор запиту до системи та "Database" - зовнішнє сховище даних, яке взаємодіє з системою. Це та ж зовнішня база даних, що й на DFD рівня 0, яка містить історичні дані, необхідні для прогнозування.

Процес 1.0 "Process Request" відповідає за отримання запиту від користувача та його початкову обробку. Це може включати парсинг запиту, вилучення ключових параметрів (наприклад, назви фільму).

Процес 1.1 "Relevant Data" відповідає за збір, обробку та формування релевантних даних, необхідних для генерації прогнозу.

Потоки даних:

- "Query": дані від "User" до процесу "1.0 Process Request". Це початковий запит користувача до системи (наприклад, назва фільму, що цікавить).

- Потік від "1.0 Process Request" до "Check for user Requirement" (Перевірка вимог користувача): Цей потік репрезентує вихід даних або запит від процесу обробки запиту для верифікації або уточнення вимог користувача. "Check for user Requirement" - сховище даних яке містить інформацію про те, що саме потрібно перевіряти у запитах користувачів.

- Потік від "1.0 Process Request" до "1.1 Relevant Data" - це дані, які після початкової обробки запиту (процес 1.0) передаються для подальшої обробки та збору релевантних відомостей (процес 1.1).

- Потік від "Database" до "1.1 Relevant Data": Процес "1.1 Relevant Data" отримує необхідні історичні дані з зовнішньої "Database". Це підтверджує, що для формування прогнозу система звертається до накопиченої інформації.

- "Feedback For User": Потік даних від процесу "1.1 Relevant Data" назад до "User". Це представляє кінцевий результат прогнозування, який повертається користувачу у відповідь на його запит.

Отже, діаграма успішно декомпозує глобальний процес "Movie Success Prediction DB" з рівня 0 на два основні підпроцеси: "Process Request" та "Relevant Data". Процес 1.0 відповідає за прийом та початкову обробку запиту, тоді як процес 1.1 відповідає за збір та формування даних для відповіді.

Ця діаграма надає більш детальне уявлення про внутрішні потоки даних та основні кроки обробки запиту користувача в системі прогнозування успішності фільмів, розширюючи розуміння її функціональної архітектури.

Представлена діаграма є Діаграмою Потоків Даних Рівня 2 (DFD Level 2). Цей рівень деталізує один з процесів попереднього рівня (рівень 1), розбиваючи його на більш дрібніші підпроцеси, що дозволяє отримати ще більш глибоке розуміння логіки обробки даних у системі. Ця діаграма деталізує процес "1.0 Process Request" і "1.1 Relevant Data" з DFD Рівня 1, показуючи більш гранульований потік даних для обробки запиту користувача.

					БР.ІІІ – 54.00.00.000 ПЗ	Арк.
						47
Змн.	Арк.	№ докум.	Підпис	Дата		

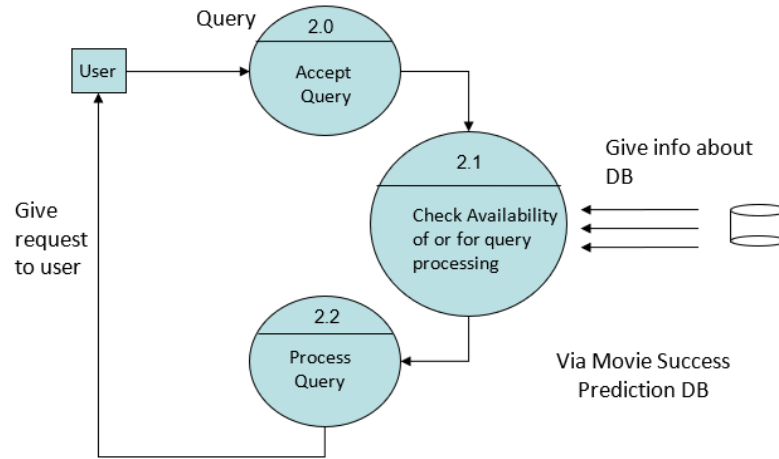


Рисунок 2.6 - Діаграма потоку даних рівня 2

Основні компоненти діаграми – зовнішня сутність "User": ініціатор запиту до системи та одержувач відповіді.

Опишемо процеси:

- 2.0 "Accept Query" - початковий процес відповідає за отримання вхідного запиту від користувача.

- 2.1 "Check Availability of or for query processing" - цей процес є критично важливим. Він виконує перевірку доступності необхідних даних або ресурсів для обробки запиту. Це може включати перевірку наявності фільму в базі, доступності історичних даних для акторів/режисерів, або навіть перевірку системних ресурсів.

- 2.2 "Process Query" - цей процес відповідає за фактичне виконання логіки прогнозування на основі отриманих даних та їх перетворення у кінцевий результат.

Потоки даних:

- "Query" - потік даних від "User" до процесу "2.0 Accept Query". Це є вхідним запитом користувача до системи.

					БР.ІІ – 54.00.00.000 ПЗ	Арк.
						48
Змн.	Арк.	№ докум.	Підпис	Дата		

- Потік від "2.0 Accept Query" до "2.1 Check Availability of or for query processing" - вказує на передачу прийнятого запиту для подальшої перевірки та обробки.

- Потік "Give info about DB" - потік даних від сховища даних до процесу "2.1 Check Availability of or for query processing". Це свідчить про те, що процес 2.1 отримує дані з бази для перевірки.

- Потік до сховища даних від "2.1 Check Availability of or for query processing" вказує на те, що процес 2.1 також може записувати або оновлювати інформацію в сховищі даних під час перевірки доступності.

- Потік "Via Movie Success Prediction DB": потік даних від процесу "2.1 Check Availability of or for query processing" до "2.2 Process Query". Це вказує на передачу релевантних даних до процесу обробки запиту.

- "Give request to user" - потік даних від процесу "2.2 Process Query" назад до "User" - це представляє результат прогнозування, який повертається користувачу у відповідь на його початковий запит.

Отже, діаграма розкриває внутрішні кроки, що відбуваються після отримання запиту: його прийняття, перевірка наявності необхідних даних та, власне, обробка. Процес 2.1 є центральним, оскільки він забезпечує доступність даних для подальшого прогнозування. Його взаємодія з сховищем даних підкреслює залежність прогностичної системи від якісних та повних вхідних даних.

Потоки даних вказують на логічну послідовність, де прийняття запиту передуює перевірці даних, а перевірка даних передуює їхній обробці. Ця діаграма поглиблює розуміння внутрішніх механізмів обробки запитів у системі прогнозування, показуючи, як дані рухаються через ключові етапи перед тим, як кінцевий результат буде надано користувачеві.

					БР.ІІІ – 54.00.00.000 ПЗ	Арк.
						49
Змн.	Арк.	№ докум.	Підпис	Дата		

РОЗДІЛ 3. ПІДГОТОВКА ДАНИХ ТА РЕАЛІЗАЦІЯ ІНТЕРФЕЙСУ КОРИСТУВАЧА ВЕБ-ЗАСТОСУНКУ ДЛЯ ПРОГНОЗУВАННЯ УСПІШНОСТІ РЕЛІЗІВ КІНОІНДУСТРІЇ

3.1. Стратегії імпорту даних з REST API

Імпорт даних є критично важливим етапом у розробці будь-якої аналітичної або прогностичної системи, оскільки він забезпечує наявність необхідної інформації для функціонування моделей. Існують кілька фундаментальних підходів до інтеграції даних у систему:

1. Імпорт з локальних джерел.

Цей метод передбачає завантаження даних із файлів, що зберігаються на локальному сховищі розробника або користувача. Зазвичай це файли у стандартизованих форматах, таких як Excel (.xlsx, .xls), SAS-формати, CSV (Comma-Separated Values) та інші. Для обробки таких файлів у Python, дані зчитуються та трансформуються у структуровані об'єкти, наприклад, датафрейми Pandas, що дозволяє подальшу маніпуляцію та аналіз.

2. Імпорт з веб-серверів через REST API.

У випадках, коли локальні дані є недостатніми, застарілими або не відображають динаміку реального часу (наприклад, для потокових даних з соціальних мереж, великих масивів даних з аналітичних платформ, або швидкозмінних ринкових показників), виникає необхідність інтеграції із зовнішніми серверами. REST (Representational State Transfer) API (Application Programming Interface) є архітектурним стилем для створення веб-сервісів, що дозволяє взаємодіяти з віддаленими серверами для обміну даними.

Для отримання даних з віддалених серверів використовуються стандартні HTTP-методи (протоколи). Серед них:

- GET - призначений для запиту та отримання даних від сервера.

					БР.ІІІ – 54.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		50

- POST - використовується для надсилання даних на сервер для створення нового ресурсу.

- PUT - застосовується для оновлення існуючого ресурсу на сервері.

- DELETE - використовується для видалення ресурсу з сервера.

У Python для взаємодії з REST API широко застосовується бібліотека Requests, яка спрощує виконання HTTP-запитів. Для отримання даних використовується метод GET. Обмін даними між системою та сервером, як правило, відбувається у форматі JSON (JavaScript Object Notation), що є легкочитаним та зручним для парсингу форматом. Після отримання відповіді від сервера (яка включає такі властивості, як заголовок, тіло, статус-код), дані JSON доступні для подальшої обробки. Для перетворення даних JSON у структурований формат, придатний для аналізу, зазвичай використовується метод .json() об'єкта відповіді, після чого дані конвертуються у датафрейми Pandas.

3.1.1. Реалізація імпорту даних для системи прогнозування

У рамках розробленої системи прогнозування успішності фільмів, дані імпортувалися з веб-сервера BookMyShow.com, що є провідною платформою для продажу квитків та отримання інформації про розваги в Індії.

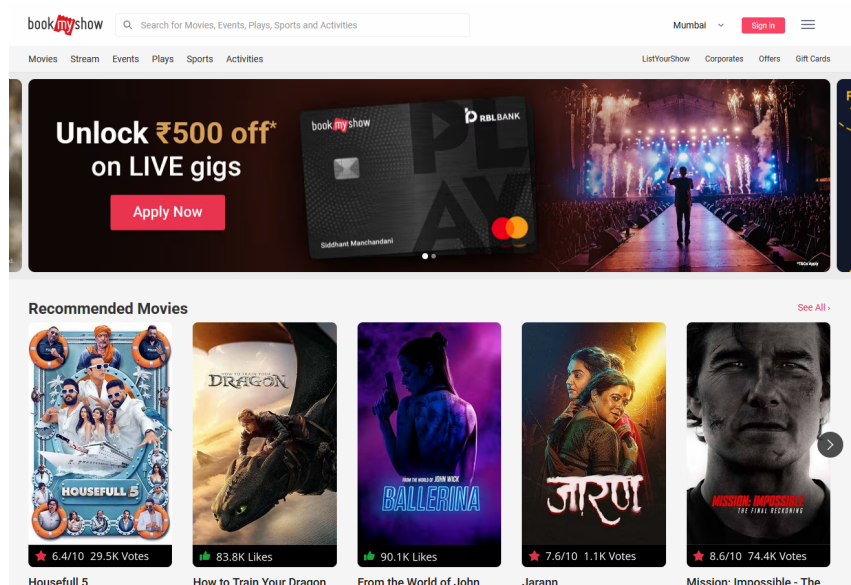


Рисунок 3.1 - Веб-ресурс BookMyShow.com для імпорту даних

										Арк.
										51
Змн.	Арк.	№ докум.	Підпис	Дата						

Для отримання рейтингів фільмів хінді в реальному часі за період з 2015 по 2019 роки, були застосовані описані вище методи імпорту даних через REST API. Це забезпечило достатній обсяг інформації для навчання та тестування прогностичної моделі.

3.1.2. Припущення щодо набору даних

Наявний набір даних є відносно невеликим і складається з 165 записів про фільми, випущені у період з 2015 по 2019 роки. Для цілей навчання та тестування моделі було прийнято наступне розподілення:

- Навчальний набір (Training Set): 140 записів фільмів за період з 2015 по 2018 роки. Цей набір використовується для навчання прогностичного алгоритму.

- Тестовий/Прогностичний набір (Test/Prediction Set): 25 записів фільмів за 2019 рік. Цей набір використовується для оцінки точності та ефективності навченої моделі у прогнозуванні успішності нових, раніше невідомих фільмів.

Цей підхід дозволяє імітувати реальні умови, коли система прогнозує успішність фільмів на основі минулих даних.

3.2. Очищення даних у процесі розробки системи прогнозування

Очищення даних (Data Cleansing/Data Cleaning) є одним з найбільш критичних та трудомістких етапів у життєвому циклі розробки програмного забезпечення, особливо для систем, що базуються на аналізі даних та машинному навчанні. Цей процес вимагає значних зусиль від інженерів з даних, оскільки сирі дані, отримані з різних джерел, часто містять інформаційні аномалії та несумісності. До типових проблем належать:

1. Пропущені значення: Позначення NaN (Not a Number), N/A (Not Applicable) або порожні комірки.

					БР.ІІІ – 54.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		52

2. Неправильні типи даних: Наприклад, числовий атрибут, представлений як рядок; або логічний тип даних, де записано текстове значення.

3. Неузгоджені формати: Різне представлення однієї й тієї ж інформації (наприклад, дати у різних форматах).

4. Дублікати записів.

5. Викиди (Outliers): Значення, що значно відхиляються від загальної тенденції даних.

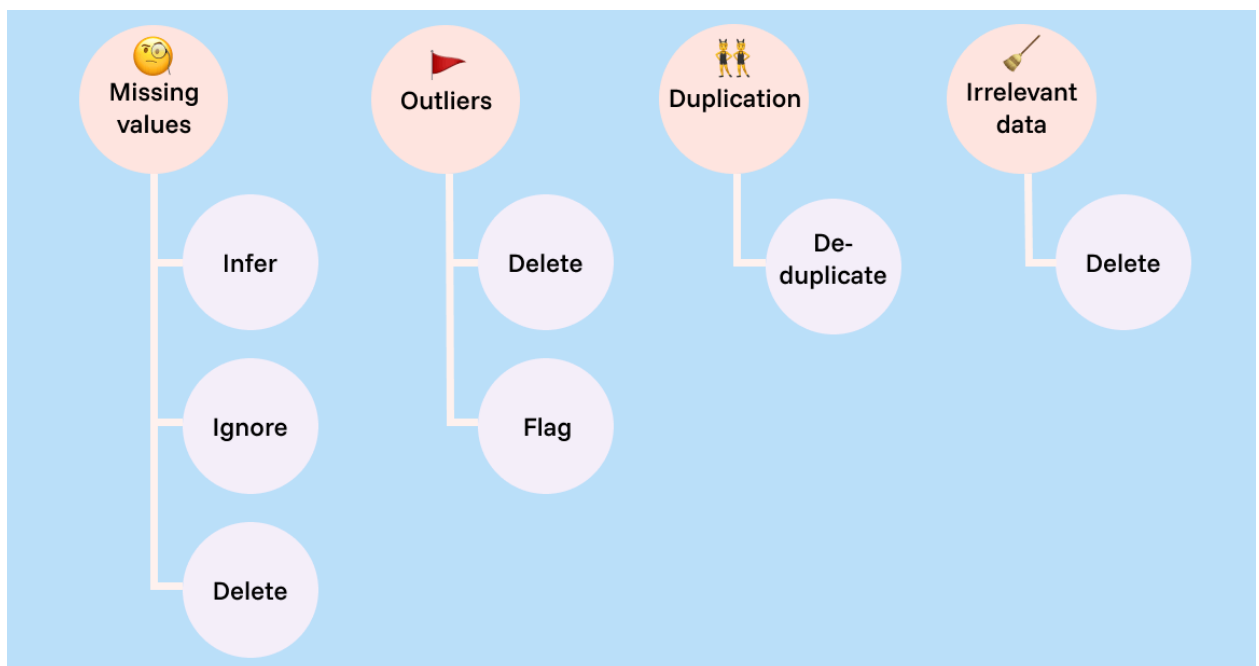


Рисунок 3.2 – Процес очищення даних

Неусунення цих неефективностей може призвести до значного зниження продуктивності системи, некоректних результатів аналізу та, як наслідок, помилкових прогнозів.

3.2.1. Основні підходи до очищення даних

При очищенні даних застосовуються два основні підходи для роботи з аномаліями:

					БР.ІІІ – 54.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		53

1. Заміщення значень (imputation).

Цей підхід передбачає заміну "сміттєвих" або пропущених значень (таких як NaN, N/A) на більш продуктивні або репрезентативні дані.

Заміщення на "реальні" значення.

У деяких випадках розробник може замінити некоректні дані на актуальні "живі" значення. Однак цей метод несе ризик впливу на загальну дистрибуцію даних та може призвести до спотворення результатів, особливо якщо замінені значення не є типовими для набору даних.

Заміщення на статистичні агрегати.

Більш безпечним та поширеним методом є заміна пропущених значень на статистичні показники, такі як медіанне значення або середнє значення (mean) відповідного атрибута. Медіана є більш стійкою до викидів порівняно із середнім значенням. Незважаючи на це, навіть заміщення медіаною може впливати на статистичні властивості даних, особливо якщо пропущених значень багато, або якщо замінене значення знаходиться далеко від справжнього розподілу, потенційно створюючи "штучні" викиди або впливаючи на реальні викиди.

2. Нульове Значення / Симетризація (Nullification / Standardization)

Цей підхід передбачає встановлення пропущених або некоректних значень у нульове (або стандартизоване нульове) значення, наприклад, замінюючи NaN або N/A на уніфіковане позначення NA (Not Available).

Цей метод вважається одним з найбільш надійних для очищення даних, оскільки він не змінює реальний розподіл даних і не впливає на викиди. При такій заміні вихідні дані зберігають свою оригінальну структуру, а алгоритми можуть бути налаштовані на ігнорування або специфічну обробку цих нульових значень.

Для ефективної маніпуляції даними, включаючи доступ до окремих комірок, стовпців та застосування симетризації, використовуються потужні

					БР.ІІІ – 54.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		54

бібліотеки Python, такі як Pandas (для роботи з датафреймами) та NumPy (для числових операцій та роботи з масивами).

3.2.2. Застосування методу очищення даних у пропонованій системі

У розробленій системі прогнозування успішності фільмів був застосований другий підхід – нульове значення / симетризація. Пропущені або некоректні значення в комірках були стандартизовані за допомогою бібліотек Pandas та NumPy. Цей вибір обґрунтовується тим, що прогностична модель приймає множинні вхідні параметри, і наявність стандартизованих нульових значень не вплине на генерацію кінцевого результату. Такий підхід забезпечує збереження оригінальної структури даних та мінімізує ризики внесення спотворень, що є важливим для точності прогнозування.

3.3. Демонстрація функціональності та валідація системи прогнозування успішності фільмів

Цей розділ представляє візуальне підтвердження функціональності розробленого веб-застосунку для прогнозування успішності фільмів, використовуючи серію скріншотів виводу системи. Аналіз цих зображень дозволяє оцінити як інтерфейс користувача (ІК), так і точність прогнозів.

3.3.1. Етапи автентифікації та навігації

На рисунку 3.3 представлений скріншот, що відображає початкову сторінку входу (Login Page) веб-застосунку під назвою "MOVIE PREDICTION SYSTEM" (Система Прогнозування Фільмів).

Центральне місце займає блок, що містить форму для входу. Вона має два поля для введення даних:

- Admin ID: Поле для введення ідентифікатора адміністратора.
- Password: Поле для введення пароля.

					БР.ІІІ – 54.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		55

- Login (кнопка): Кнопка для ініціювання процесу автентифікації після введення облікових даних.

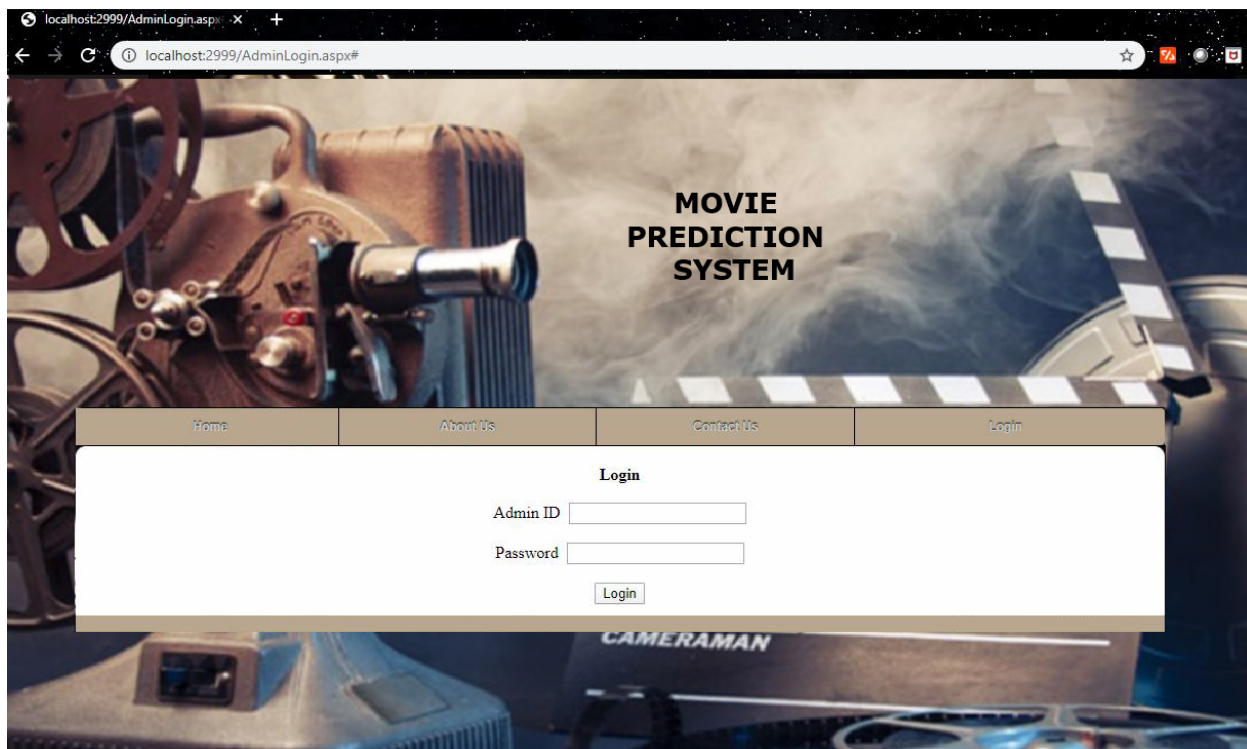


Рисунок 3.3 - Початкова сторінка додатку. Відображає стартовий ІК застосунку, який є точкою входу для користувачів

Даний скріншот відповідає етапу "Початкова сторінка додатку" та "Введення правильного імені користувача та пароля". Він є першою точкою взаємодії для користувачів, які прагнуть отримати адміністративний доступ до системи. Основна мета цієї сторінки — забезпечити безпечну автентифікацію адміністраторів перед наданням їм можливості керувати даними та функціоналом прогностичної системи

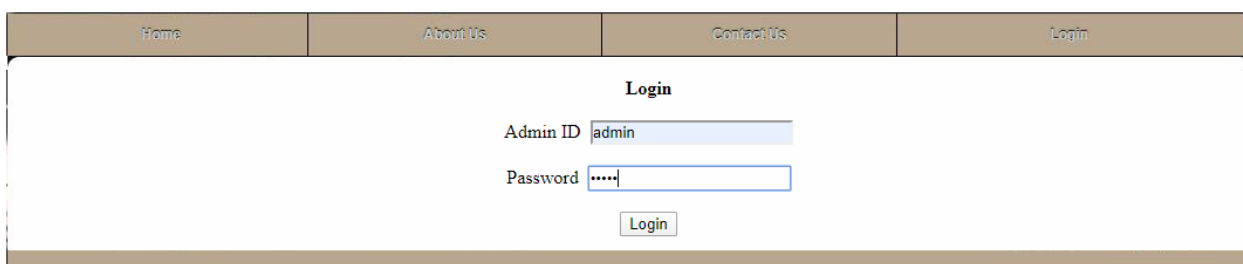


Рисунок 3.4 - Введення правильного імені користувача та пароля

Рисунок 3.4 ілюструє процес введення облікових даних для адміністратора.

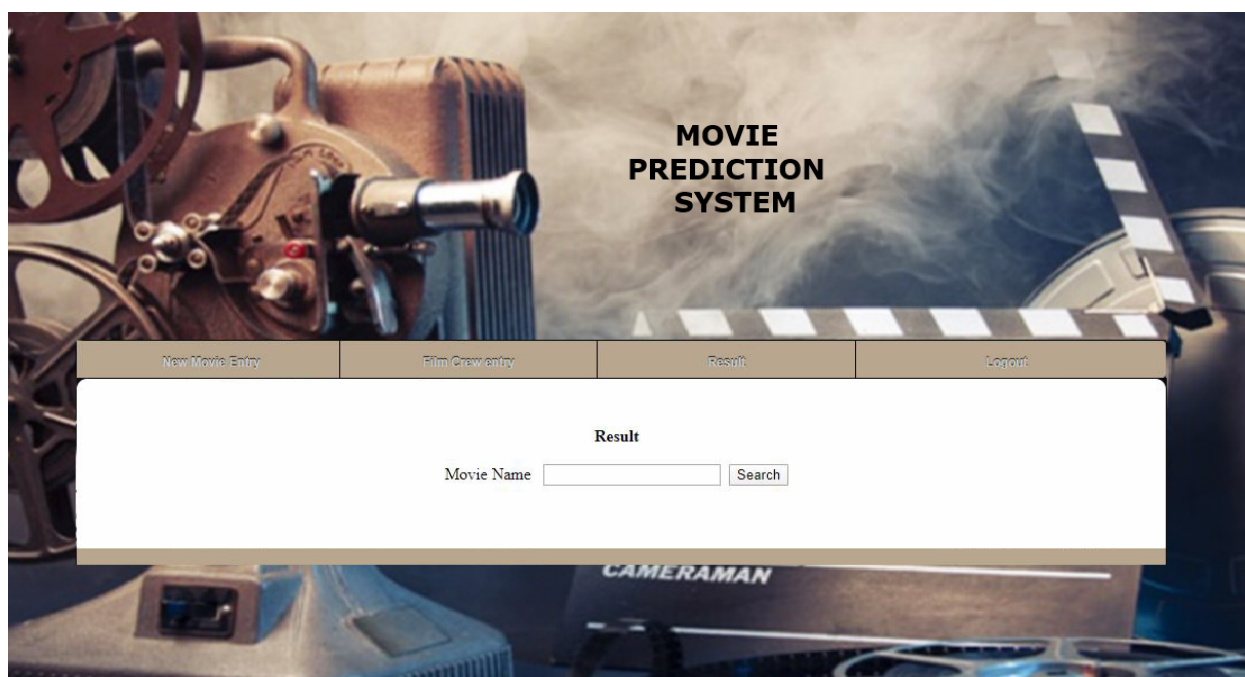


Рисунок 3.5 - Успішна аутентифікація

Рисунок 3.5 демонструє підтвердження успішного входу в систему, що відкриває доступ до адміністративних функцій.

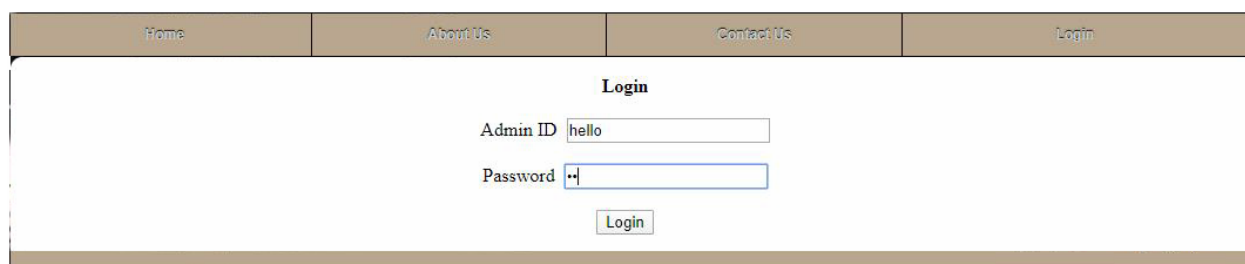
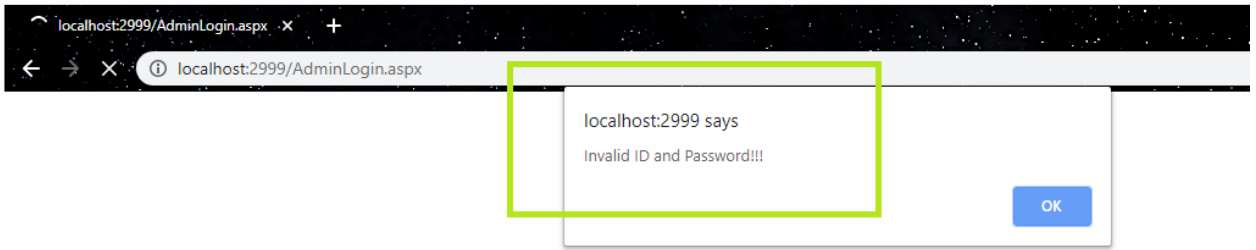


Рисунок 3.6 - Введення неправильних облікових даних.

На рисунку 3.6 показано сценарій, коли адміністратор вводить некоректні дані для входу. На рисунку 3.7 відображається повідомлення про відмову в доступі, підтверджуючи коректну роботу системи безпеки.



Waiting for localhost...

Рисунок 3.7 – Видача помилки про неправильно введені дані аутентифікації

3.3.2. Введення та управління даними

На рисунку 3.8 представлено інтерфейс для внесення детальної інформації про майбутній кінофільм. Це включає ключові параметри, які будуть використані для прогнозування.

New Movie Entry	Film Crew entry	Result	Logout																																										
<p>New Movie Entry</p> <p>Movie ID <input type="text" value="1006"/></p> <p>Movie Name <input type="text"/></p> <p>Producer <input type="text" value="--Select--"/></p> <p>Director <input type="text" value="--Select--"/></p> <p>Actor <input type="text" value="--Select--"/></p> <p>Actress <input type="text" value="--Select--"/></p> <p>Music Director <input type="text" value="--Select--"/></p> <p>Writer <input type="text" value="--Select--"/></p> <p>Market Budget <input type="text"/> Crores</p> <p>Date</p> <table border="1" style="margin-left: auto; margin-right: auto;"> <thead> <tr> <th colspan="7">May 2025</th> </tr> <tr> <th>Sun</th> <th>Mon</th> <th>Tue</th> <th>Wed</th> <th>Thu</th> <th>Fri</th> <th>Sat</th> </tr> </thead> <tbody> <tr> <td>29</td> <td>30</td> <td>31</td> <td>1</td> <td>2</td> <td>3</td> <td>4</td> </tr> <tr> <td>5</td> <td>6</td> <td>7</td> <td>8</td> <td>9</td> <td>10</td> <td>11</td> </tr> <tr> <td>12</td> <td>13</td> <td>14</td> <td>15</td> <td>16</td> <td>17</td> <td>18</td> </tr> <tr> <td>19</td> <td>20</td> <td>21</td> <td>22</td> <td>23</td> <td>24</td> <td>25</td> </tr> </tbody> </table>				May 2025							Sun	Mon	Tue	Wed	Thu	Fri	Sat	29	30	31	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
May 2025																																													
Sun	Mon	Tue	Wed	Thu	Fri	Sat																																							
29	30	31	1	2	3	4																																							
5	6	7	8	9	10	11																																							
12	13	14	15	16	17	18																																							
19	20	21	22	23	24	25																																							

Рисунок 3.8 - Сторінка введення нового фільму

Рисунок 3.9 демонструє функціональність для додавання інформації про членів знімальної групи, дані про яких відсутні в системі (наприклад, дебютанти або ті, хто повернувся після тривалої перерви). Ця можливість забезпечує гнучкість системи у разі неповної інформації про учасників проекту.

Рисунок 3.9 - Введення нової кіногрупи

3.3.3. Взаємодія з базою даних та прогнозування

Лістинг 3.1 ілюструє процес отримання системою даних з бази даних для подальших розрахунків, що є ключовим етапом у роботі прогностичної моделі.

Лістинг 3.1. Підключення до бази даних

```
// (файл MovieEntry.aspx.cs, частина обробника події Page_Load)
using System;
using System.Data.SqlClient; // Для роботи з SQL Server
using System.Web.UI; // Для роботи з ASP.NET Web Forms
public partial class MovieEntry : System.Web.UI.Page
{
    // Створення об'єкта підключення до бази даних SQL Server
    // Рядок підключення: Data Source= (LocalDB)\MSSQLLocalDB;
    AttachDbFilename=C:\Users\MintaDeva\...
```

```

        SqlConnection con = new SqlConnection(@"Data
Source=(LocalDB)\MSSQLLocalDB;AttachDbFilename=C:\Users\MintaDeva\Docu
ments\MovieDB.mdf;Integrated Security=True;Connect Timeout=30");
        // Захищений метод, який виконується під час завантаження
сторінки
        protected void Page_Load(object sender, EventArgs e)
        {
            // Перевірка, чи сторінка завантажується вперше (не є
зворотним надсиланням даних на сервер)
            if (!IsPostBack)
            {
                // Якщо це не PostBack, очистити текстове поле
                TextBox10
                TextBox10.Text = "";
            }
            // Відкриття підключення до бази даних
            con.Open();

            // Створення SQL-команди для вибору максимального id з
таблиці NewMovieEntry
            // "select top 1 id from NewMovieEntry order by id desc"
            // Ця команда поверне найбільше значення id з таблиці
NewMovieEntry, що фактично є останнім доданим id
            SqlCommand cmd = new SqlCommand("select top 1 id from
NewMovieEntry order by id desc", con);
            // Виконання команди та отримання єдиного скалярного
значення (id)
            object count = cmd.ExecuteScalar();
            // Перетворення отриманого значення (object) у ціле число
(int)
            int c = Convert.ToInt32(count);
            // Перевірка, чи отримане id не дорівнює нулю
            if (c != 0)
            {
                // Якщо id не нульове, збільшити його на 1
                c = c + 1;
                // Відобразити нове id у текстовому полі TextBox9
                TextBox9.Text = Convert.ToString(c);
                // Зробити текстове поле TextBox9 доступним лише для
читання
                TextBox9.ReadOnly = true;
            }
            else // Якщо c == 0 (це означає, що таблиця NewMovieEntry
порожня або id дорівнює 0)
            {
                // Встановити id нового запису як 1
                c = 1;
            }
        }
    }
}

```

					БР.ІІІ – 54.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		60

```

// Відобразити id 1 у текстовому полі TextBox9
TextBox9.Text = Convert.ToString(c);
// Зробити текстове поле TextBox9 доступним лише для
читання
TextBox9.ReadOnly = true;
    }
con.Close();
}
}

```

На рисунку 3.10 показано показники продуктивності серверної частини. Надає інформацію про ефективність роботи серверних компонентів системи, що є важливим для моніторингу та оптимізації.

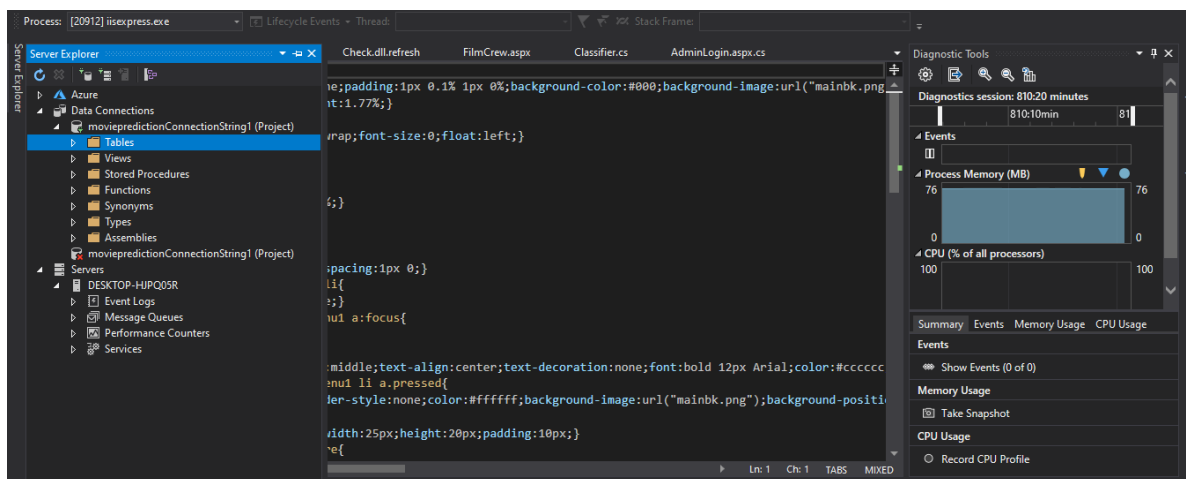


Рисунок 3.10 - Показники продуктивності серверної частини

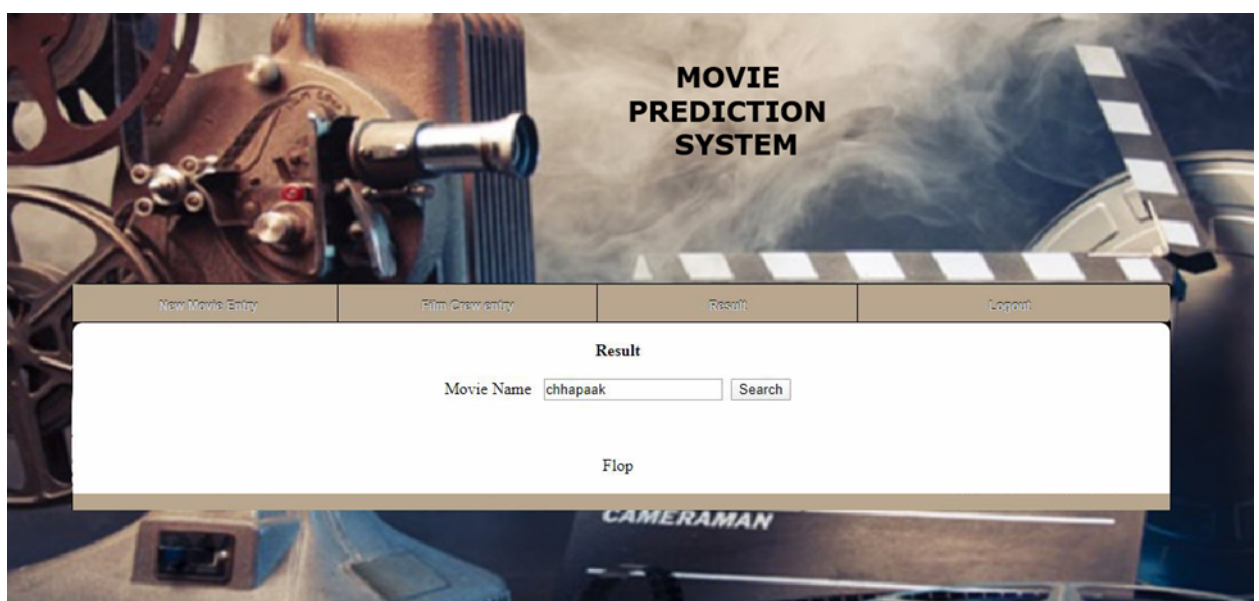


Рисунок 3.11 - Прогнозований результат

Рисунок 3.11 відображає інтерфейс, де користувач вводить назву фільму для отримання прогнозу його успішності.

3.3.4. Валідація точності прогнозування

Рисунок 3.12 представляє реальний касовий та критичний успіх фільму, який вийшов у 2020 році і не входив до навчального набору даних (який обмежений до 2019 року). Фільм отримав рейтинг 4.6 зірки, що є нижчим за порогове значення 5 зірок, встановлене системою для визначення "провального" фільму. Це свідчить про те, що для системи даний фільм був "майбутнім релізом". Результати прогнозування системи повністю збіглися з фактичними ринковими показниками, що підтверджує високу точність прогностичної моделі.



Рисунок 3.12 - Фактичний результат

					БР.ІІ – 54.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		62

Рисунок 3.13 демонструє застосування системи для прогнозування успішності іншого фільму, також випущеного у 2020 році (поза навчальним набором даних).

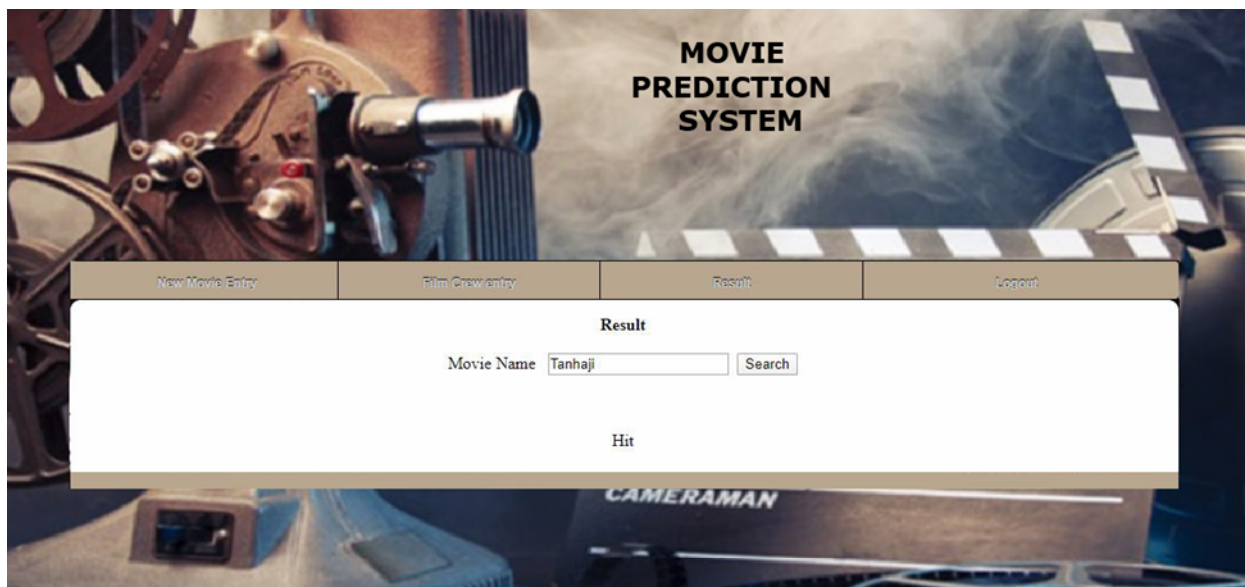


Рисунок 3.13 - Прогнозований результат (інший фільм)



Рисунок 3.14 - Результат у реальному часі (інший фільм)

					БР.ІІІ – 54.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		63

Рисунок 3.14 відображає фактичні ринкові показники цього другого фільму. Знову ж таки, прогнозований результат системи співпав з реальним результатом, що додатково підтверджує надійність та ефективність прогностичного алгоритму.

Ці скріншоти слугують емпіричним доказом функціональності застосунку, його здатності до валідації введених даних, управління інформацією про фільми та знімальні групи, а головне – демонстрації високої точності прогнозування комерційного успіху кінорелізів на основі історичних даних.

3.4. Фаза тестування в життєвому циклі розробки програмного забезпечення

Фаза тестування є невід'ємним та критично важливим компонентом життєвого циклу розробки програмного забезпечення (SDLC), що забезпечує впевненість розробників у функціональності, надійності та відповідності системи заданим вимогам. На цьому етапі застосовується комплекс різноманітних методик тестування, включно з модульним, інтеграційним та системним тестуванням. Ключовим завданням фази тестування є визначення, чи відповідає розроблений додаток (у даному випадку, веб-додаток на C#) встановленим специфікаціям та чи забезпечує він необхідну функціональність.

Для великомасштабних проектів успішне проходження всіх етапів тестування є обов'язковою умовою. Успішне тестування передбачає, що кожен компонент системи функціонує коректно, надає бажаний результат для будь-яких валідних вхідних даних, має ідеальний потік даних, забезпечує масштабованість, підтримує можливість оновлень та адекватно реагує на всі вхідні операції. Позитивний результат тестування свідчить про готовність

					БР.ІІІ – 54.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		64

системи до розгортання та є фундаментальним для подальшого розвитку проекту.

Існують різні фази тестування, які виконуються послідовно або ітеративно:

1. Модульне Тестування (Unit Testing)

Модульне тестування є початковою фазою тестування, що дозволяє виявити дефекти на найбільш гранулярному рівні. На цьому етапі систему декомпонують на найдрібніші самостійні робочі одиниці (модулі або компоненти коду). Кожна така одиниця тестується ізольовано, щоб переконатися у її коректній функціональності та відсутності помилок на базовому рівні реалізації. Модульне тестування, як правило, виконується самими розробниками під час написання коду модуля, що дозволяє оперативного виявляти та виправляти дефекти.

2. Інтеграційне Тестування (Integration Testing)

Після успішного модульного тестування окремих компонентів, інтеграційне тестування зосереджується на перевірці взаємодії та сумісності цих модулів. На цьому етапі окремі, попередньо протестовані модулі інтегруються у більші структурні одиниці, і їх спільна робота перевіряється. Основна мета інтеграційного тестування – виявити дефекти, що виникають через некоректну взаємодію між модулями, навіть якщо кожен з них окремо функціонує правильно. Це тестування також включає перевірку відповідності інтерфейсів системи встановленим вимогам.

3. Системне Тестування (System Testing)

Системне тестування є третьою фазою у життєвому циклі тестування, що відбувається після завершення модульного та інтеграційного тестування. На цьому етапі система тестується як єдине ціле, оцінюючи її поведінку у комплексному середовищі. Системне тестування не обмежується лише функціональністю, а також включає перевірку:

- Адаптивності до нових змін.

					БР.ІІІ – 54.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		65

- Масштабованості (здатність обробляти зростаючі навантаження).
- Сумісності з іншими системами та компонентами.
- Прийнятності нових оновлень.

Продуктивності з точки зору кінцевого користувача. Успішне проходження системного тестування означає, що всі компоненти проекту функціонують злагоджено та система готова до використання.

4. Валідаційне Тестування (Validation Testing)

Валідаційне тестування є критичним етапом, що оцінює відповідність розробленої системи початковим вимогам та очікуванням замовника. Ця фаза відбувається після успішного модульного, інтеграційного та системного тестування. Вона зосереджена на підтвердженні того, що система виконує свої функції згідно з визначеними специфікаціями та задовольняє потреби кінцевих користувачів. Для даної системи, відсутність невирішених питань після валідаційного тестування свідчить про її успішне проходження.

5. Тестування Виводу (Output Testing)

Тестування виводу є другою фазою фінального тестування, що відбувається після валідаційного тестування. Його мета полягає у перевірці, чи система генерує вивід у зазначеному форматі та чи відповідає цей вивід вимогам користувача. Розглядаються два аспекти виводу: відображення на екрані та друкований формат. Успішне проходження тестування виводу підтверджує, що система надає користувачам інформацію в очікуваній та придатній для використання формі.

6. Тестування Прийняття Користувачем (User Acceptance Testing, UAT)

Тестування прийняття користувачем є заключною та найбільш важливою фазою тестування. На цьому етапі система тестується реальними кінцевими користувачами (аудиторією), які будуть використовувати продукт. Їхня безперервна участь та контроль дозволяють виявити будь-які розбіжності між функціональністю системи та їхніми реальними потребами та очікуваннями. Будь-які зміни або оновлення, які виникають внаслідок

					БР.ІІІ – 54.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		66

відгуків кінцевих користувачів, інтегруються в систему. UAT є фінальним "іспитом" для системи перед її остаточним розгортанням.

3.5. Перспективи майбутнього вдосконалення системи прогнозування успішності фільмів

Поточна система прогнозування успішності фільмів демонструє високу точність, досягаючи 95% на тестовому наборі з 25 фільмів за допомогою простого арифметичного розрахунку. Однак, існує значний потенціал для подальшого вдосконалення її можливостей та розширення функціоналу, що дозволить підвищити її надійність, точність та комерційну цінність.

1. Розширення обсягу набору даних

Одним з першочергових напрямків удосконалення є значне розширення історичного набору даних. Наразі дані імпортуються до 2015 року. Розширення часового діапазону імпортованих даних (наприклад, включення інформації до 2010 року або раніше) дозволить створити більший пул даних. Збільшення обсягу навчальних даних є критично важливим для:

- Зниження ймовірності інформаційної недоступності. Більший обсяг даних зменшить кількість випадків відсутності інформації про конкретних членів знімальних груп або фільми.

- Підвищення репрезентативності. Великі набори даних забезпечують ширший спектр сценаріїв та взаємозв'язків, що дозволить моделі краще узагальнювати та виявляти приховані закономірності.

- Оптимізація навчання та тестування. Розширений набір даних надає більше гнучкості для розподілу даних на навчальні та тестові підмножини, що сприятиме більш ретельній валідації моделі. Гіпотеза: Хоча існуюча точність є високою, подальше розширення даних може вплинути на цей показник. Це може призвести до незначного зниження початкової точності,

					БР.ІІІ – 54.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		67

але одночасно забезпечить більшу надійність та стійкість моделі в довгостроковій перспективі.

2. Інтеграція функціоналу штучного інтелекту та машинного навчання

Після формування обширного та якісного набору даних, наступним кроком є включення передових алгоритмів машинного навчання (ML) та штучного інтелекту (AI). Це дозволить відійти від простого арифметичного розрахунку до складніших моделей, здатних виявляти нелінійні залежності та патерни. Інтеграція ML/AI методологій, таких як нейронні мережі, дерева рішень, або регресійні моделі, є майбутнім обчислень у сфері прогнозування та дозволить досягти значно вищого рівня точності та адаптивності системи до змінних ринкових умов.

3. Додавання функціональності обробки природної мови (NLP)

Поточна система базується виключно на рейтингових показниках. Для поглиблення аналізу та підвищення точності прогнозів, пропонується додати функціональність обробки природної мови (NLP). Це дозволить:

- Аналіз тональності (Sentiment Analysis) - збираючи та аналізуючи коментарі та відгуки глядачів з відкритих джерел (наприклад, соціальних мереж, кінофорумів), можна визначити їхні емоційні відгуки та настрої. Інтеграція цього показника в прогностичну модель додасть новий вимір до оцінки потенційного успіху фільму.

- Автоматична категоризація даних - NLP може бути використане для автоматичної обробки текстових описів фільмів та коментарів для формування більш точних характеристик жанрів або специфічних особливостей фільму, а також для прив'язки цих даних до відповідних членів знімальної групи.

- Збільшення точності - включення семантичного аналізу та інших виходів NLP до існуючих розрахунків дозволить суттєво підвищити прогностичну точність, оскільки дозволить врахувати суб'єктивні, але впливові фактори, що формують суспільне сприйняття.

					БР.ІІІ – 54.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		68

Отже, система прогнозування успішності фільмів демонструє значний потенціал для комерційного використання та задоволення ринкових вимог. Поточна точність у 95% на обмеженому тестовому наборі свідчить про міцну базову архітектуру та алгоритм. Однак, реалізація запропонованих удосконалень – розширення набору даних, інтеграція методів штучного інтелекту/машинного навчання та додавання функціоналу NLP – дозволить створити більш надійну, точну та адаптивну систему, яка буде здатна генерувати прогнози, що відповідають реальним ринковим результатам інших веб-сайтів. Це зробить систему цінним інструментом для кіноіндустрії.

					БР.ІІІ – 54.00.00.000 ПЗ	Арк.
						69
Змн.	Арк.	№ докум.	Підпис	Дата		

ВИСНОВКИ

У ході виконання дипломної роботи було комплексно досліджено теоретичні та практичні аспекти розробки веб-застосунку, призначеного для прогнозування успішності релізів у сфері розважального контенту, зокрема фільмів. Здійснений аналіз предметної області дозволив сформулювати мету створення системи — підвищення точності й доступності прогнозів стосовно комерційного успіху кінорелізів на основі аналізу релевантних історичних та поточних даних.

У першому розділі проаналізовано сучасний стан галузі прогнозування успішності кінопродукції. Виявлено, що існуючі платформи або є закритими (індустріального характеру), або мають обмежену функціональність і не надають користувачеві гнучкого механізму роботи з даними. Акцент зроблено на перевагах впровадження автоматизованих моделей, заснованих на сучасних алгоритмах обробки даних та машинного навчання. Окрему увагу приділено актуальним науковим підходам, зокрема аналізу кореляції між факторами успішності релізів та їхньою подальшою комерційною реалізацією.

У другому розділі описано архітектуру та структурну модель веб-застосунку, охоплено технічні, функціональні та експлуатаційні аспекти створення системи. Було обґрунтовано доцільність розробки рішення з огляду на технічні можливості, економічну вигідність та операційну ефективність. Створені діаграми варіантів використання, послідовності взаємодій, діаграми активності та потоку даних сприяли чіткому розумінню логіки функціонування системи, взаємодії користувача з інтерфейсом та обробки інформації в межах веб-застосунку.

Третій розділ був присвячений реалізації прикладної частини системи. У ньому детально розглянуто процес підготовки даних — від імпорту з REST API до очищення та валідації. Особливу увагу приділено забезпеченню якості

					БР.ІІІ – 54.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		70

даних, адже саме надійність вхідної інформації суттєво впливає на точність прогнозів. Описано етапи реалізації інтерфейсу користувача, функції автентифікації, введення та управління даними, а також механізми взаємодії з базою даних і обчислення прогнозів. Проведено тестування системи, що підтвердило її працездатність та відповідність заявленим функціональним вимогам.

У результаті дослідження було створено прототип веб-застосунку, здатного здійснювати базовий прогноз успішності кінорелізів на основі історичних даних. Система реалізована з урахуванням вимог масштабованості, розширюваності та подальшої адаптації до інших типів розважального контенту (музика, серіали, відеоігри). Здійснено оцінку потенційних напрямів удосконалення, зокрема інтеграції додаткових джерел даних, розширення функціоналу аналітики та впровадження моделей глибокого навчання.

Таким чином, досягнуто всіх поставлених у роботі цілей і завдань. Запропоноване рішення може бути основою для подальших наукових досліджень та практичних розробок у сфері прогнозування споживчого попиту в індустрії розваг.

					БР.ІІІ – 54.00.00.000 ПЗ	Арк.
						71
Змн.	Арк.	№ докум.	Підпис	Дата		

ПЕРЕЛІК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Pressman, R. S., & Maxim, B. R. (2019). Software Engineering: A Practitioner's Approach (9th ed.). McGraw-Hill Education.
2. Larman, C. (2004). Applying UML and Patterns: An Introduction to Object-Oriented Analysis and Design and Iterative Development (3rd ed.). Prentice Hall.
3. Офіційна документація Microsoft: <https://docs.microsoft.com/en-us/dotnet/csharp/> (для C#) та <https://docs.microsoft.com/en-us/aspnet/>
4. VanderPlas, J. (2016). Python Data Science Handbook: Essential Tools for Working with Data. O'Reilly Media.
5. Fielding, R. T. (2000). Architectural Styles and the Design of Network-based Software Architectures. Doctoral dissertation, University of California, Irvine.
6. Jurafsky, D., & Martin, J. H. (2009). Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition (2nd ed.). Prentice Hall.
7. Ambler, S. W. (2002). Agile Modeling: Effective Practices for eXtreme Programming and the Unified Process. John Wiley & Sons.
8. Asur, S., & Huberman, B. A. (2010). Predicting the Future with Social Media. Proceedings of the 2010 IEEE/ACM International Conference on Web Intelligence and Intelligent Agent Technology, Toronto, ON, Canada: IEEE. pp. 492-499.
9. Mestyán, M., & Thelwall, M. (2018). Box Office Prediction: A Review. Future Generation Computer Systems, 87, 592-601. Elsevier.
10. Sharda, R., & Delen, D. (2006). Predicting Box-Office Success of Movies with Neural Networks. Expert Systems with Applications, 30(2), 241-249. Elsevier.

					БР.ІІІ – 54.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		72

11. Zhang, T., Shi, Y., & Chen, Y. (2017). Movie Success Prediction Based on Social Media and Deep Learning. Proceedings of the 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Banff, AB, Canada: IEEE. pp. 2933-2938.
12. Du, W., Wang, C., Li, S., & Li, L. (2015). Movie Box Office Prediction with Sentiment Analysis Based on Micro-blogging. Proceedings of the 2015 IEEE International Conference on Service Operations and Logistics, and Informatics (SOLI), Chongqing, China: IEEE. pp.
13. Rahm, E., & Do, H. H. (2000). Data Cleaning: Problems and Current Approaches. IEEE Data Engineering Bulletin, 23(4), 3-13. IEEE.
14. Naumann, F., & Herschel, M. (2017). An Introduction to Data Quality. Morgan & Claypool Publishers. (Хоча це книга, вона часто цитується як авторитетне джерело).
15. Fan, W., & Geerts, F. (2012). Data Quality Challenges in a World of Big Data. Proceedings of the VLDB Endowment, 5(12), 1776-1777. VLDB Endowment.
16. Kang, J., Lary, T., & Jin, R. (2013). Data Cleaning for Big Data: An Overview. Proceedings of the 2013 IEEE International Conference on Big Data, Santa Clara, CA, USA: IEEE. pp. 1-8.
17. Zou, J., Zhao, L., Lu, R., & Wu, X. (2019). Data Cleaning in Machine Learning: An Overview. Proceedings of the 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC), Bari, Italy: IEEE. pp. 3703-3708.
18. Pautasso, C., Zimmermann, O., & Leymann, F. (2008). RESTful Web Services: The Big Picture. Proceedings of the 2008 IEEE International Conference on Web Services (ICWS), Beijing, China: IEEE. pp. 3-10.
19. Thakur, G. (2016). A Comprehensive Review of RESTful Web Services for Mobile Applications. Proceedings of the 2016 International Conference on

					БР.ІІІ – 54.00.00.000 ІІЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		73

- Communication and Electronics Systems (ICCES), Coimbatore, India: IEEE. pp. 1-6.
20. Vinoski, S. (2002). Web Services: Beware of Blind Faith. *IEEE Internet Computing*, 6(3), 85-87. IEEE.
21. Ambler, S. W. (2003). *Agile Modeling: Effective Practices for eXtreme Programming and the Unified Process*. John Wiley & Sons.
22. Binder, R. V. (2000). *Testing Object-Oriented Systems: Models, Patterns, and Tools*. Addison-Wesley Professional.
23. Black, R. (2009). *Managing the Testing Process: Practical Tools and Techniques for Managing and Implementing Software Testing (3rd ed.)*. John Wiley & Sons.
24. Ammann, P., & Offutt, J. (2016). *Introduction to Software Testing (2nd ed.)*. Cambridge University Press.
25. Runeson, P., & Höst, M. (2009). Guidelines for conducting and reporting case study research in software engineering. *Empirical Software Engineering*, 14(2), 131-164. Springer.
26. What is data cleaning and why is it important? Data cleaning guide - <https://funnel.io/blog/what-is-data-cleaning-your-complete-guide>
27. Movie Tickets, Plays, Sports, Events & Cinemas near Mumbai - BookMyShow Mumbai. - <https://in.bookmyshow.com/explore/home/mumbai>
28. Zheng, Yirui. “Predicting Movie Box Office Based on Machine Learning, Deep Learning, and Statistical Methods.” *Applied and Computational Engineering*, vol. 94, 21 Nov. 2024, pp. 20–32. EWA Publishing, Oxford, UK.
29. Agarwal, Manav; Venugopal, Shreya; Kashyap, Rishab; Bharathi, R. “A Comprehensive Study on Various Statistical Techniques for Prediction of Movie Success.” arXiv (preprint), 1 Dec. 2021.

					БР.ІІІ – 54.00.00.000 ІІЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		74

30. Sikana, Nadya, and Ronsen Purba. “Movie Success Prediction Based on Feature and Trailer Comments Using Ensemble+LSTM Model.” *Journal La Multiapp*, vol. 5, no. 5, Oct. 2024, pp. 595–608. Auricle Global Society of Education & Research.
31. Zhang, Yizhen, and Zejun Bai. “Prediction of Movies Popularity in Supervised Learning Techniques.” *Applied and Computational Engineering*, vol. 29, 26 Dec. 2023, pp. 142–147. EWA Publishing, Oxford, UK.
32. Jhalani, Ritu; Virwani, Harshita; Goyal, Divyanshi; Vashishtha, Somya. “Prediction of Movie Success Using Classification.” *Indian Journal of Computer Science*, vol. 3, no. 6, Nov. 2018. doi:10.17010/ijcs/2018/v3/i6/141443.
33. Unknown author. “Finding Nemo: Predicting Movie Performances by Machine Learning Methods.” *International Journal of Financial Studies*, vol. 13, no. 5, 2019, article 93. MDPI.
34. Singh, Mohit; Jain, Poonam; Pandey, Aaditya; Prasad, Ankit. “Movie Success Prediction System Using Python and Machine Learning Algorithms.” *IJRASET Journal for Research in Applied Science and Engineering Technology*, 25 Feb. 2025. ISSN 2321-9653.
35. Udandarao, Vikranth, and Pratyush Gupta. “Movie Revenue Prediction using Machine Learning Models.” arXiv (preprint), 19 May 2024.
36. Wang, Wei; Xiu, Jiapeng; Yang, Zhengqiu; Liu, Chen. “A Deep Learning Model for Predicting Movie Box Office Based on Deep Belief Network.” In: *Advances in Swarm Intelligence*, edited by Tan Y., Shi Y., Tang Q., Springer, Cham, 2018, pp. —. doi:10.1007/978-3-319-93818-9_51.
37. Sharda, R., and Delen, Dursun. “Predicting Box-Office Success of Motion Pictures with Neural Networks.” *Expert Systems with Applications*, vol. 30, no. 2, Aug. 2006, pp. 243–254.
38. Lash, Michael T., and Kang Zhao. “Early Predictions of Movie Success: the Who, What, and When of Profitability.” arXiv (preprint), 17 June 2015.

					БР.ІІІ – 54.00.00.000 ІІЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		75

39. Eliashberg, J., and Sharda, R. "Using Neural Networks and Classification Trees for Box-Office Forecasting." J. Media Econ., vol. 22, no. 1, 2009, pp. No pages provided.
40. Du, H.; Xu, H.; Yuan, Z.; Xu, Z. "Support Vector Machine and Microblog-Based Movie Box-Office Prediction." In: Proceedings of the International Conference on Information and Management Sciences, 2014,
41. Galvão, A.; Henriques, R. "Predicting Movie Profit through Neural Networks, Regression and Decision Trees." Applied Intelligence, vol. 48, 2018, pp.

					БР.ІІІ – 54.00.00.000 ІІЗ	Арк.
						76
Змн.	Арк.	№ докум.	Підпис	Дата		

ДОДАТКИ

Додаток А

Фрагменти програмних кодів побудови системи прогнозування успішності кінорелізів

Web.config

```
<?xml version="1.0" encoding="utf-8"?>
<configuration>
  <connectionStrings>
    <add name="MovieDBConnection"
      connectionString="Data
Source=(LocalDB)\MSSQLLocalDB;AttachDbFilename=|DataDirectory|\MovieDB
.mdf;Integrated Security=True;Connect Timeout=30"
      providerName="System.Data.SqlClient" />
  </connectionStrings>
  <system.web>
    <compilation debug="true" targetFramework="4.7.2" />
    <httpRuntime targetFramework="4.7.2" />
    <authentication mode="Forms">
      <forms loginUrl="AdminLogin.aspx" timeout="2880"/>
    </authentication>
    <authorization>
      <deny users="?" />
    </authorization>
  </system.web>
  <location path="AdminLogin.aspx">
    <system.web>
      <authorization>
        <allow users="*" />
      </authorization>
    </system.web>
  </location>
</configuration>
```

AdminLogin.aspx

```
<%@ Page Language="C#" AutoEventWireup="true"
CodeBehind="AdminLogin.aspx.cs"
Inherits="MoviePredictionSystem.AdminLogin" %>
```

```
<!DOCTYPE html>

<html xmlns="http://www.w3.org/1999/xhtml">
<head runat="server">
    <meta charset="utf-8" />
    <meta name="viewport" content="width=device-width, initial-
scale=1.0" />
    <title>MOVIE PREDICTION SYSTEM - БҮИД</title>
    <style>
        body {
            font-family: Arial, sans-serif;
            margin: 0;
            padding: 0;
            background-image: url('images/movie_background.jpg');
            background-size: cover;
            background-position: center;
            display: flex;
            justify-content: center;
            align-items: center;
            min-height: 100vh;
            color: #fff;
            text-align: center;
        }
        .container {
            background-color: rgba(0, 0, 0, 0.7);
            padding: 30px 50px;
            border-radius: 10px;
            box-shadow: 0 0 20px rgba(0, 0, 0, 0.5);
            max-width: 400px;
            width: 100%;
        }
        h1 {
            color: #fff;
            font-size: 2.5em;
            margin-bottom: 30px;
            text-shadow: 2px 2px 4px rgba(0,0,0,0.5);
        }
    </style>
</head>
<body>
    <div class="container">
        <h1>MOVIE PREDICTION SYSTEM</h1>
    </div>
</body>
</html>
```

```
.login-box {
    background-color: #fff;
    padding: 30px;
    border-radius: 8px;
    box-shadow: 0 0 15px rgba(0, 0, 0, 0.2);
    color: #333;
}

.login-box h2 {
    margin-top: 0;
    color: #333;
    font-size: 1.8em;
    margin-bottom: 25px;
}

.form-group {
    margin-bottom: 20px;
    text-align: left;
}

.form-group label {
    display: block;
    margin-bottom: 8px;
    font-weight: bold;
}

.form-group input[type="text"],
.form-group input[type="password"] {
    width: calc(100% - 20px);
    padding: 10px;
    border: 1px solid #ddd;
    border-radius: 5px;
    font-size: 1em;
}

.btn-login {
    background-color: #007bff;
    color: white;
    padding: 12px 25px;
    border: none;
    border-radius: 5px;
    cursor: pointer;
    font-size: 1.1em;
}
```

```

        transition: background-color 0.3s ease;
        width: 100%;
    }
    .btn-login:hover {
        background-color: #0056b3;
    }
    .message {
        margin-top: 15px;
        color: red;
        font-weight: bold;
    }
</style>
</head>
<body>
    <form id="form1" runat="server">
        <div class="container">
            <h1>MOVIE PREDICTION SYSTEM</h1>
            <div class="login-box">
                <h2>Login</h2>
                <div class="form-group">
                    <label for="txtAdminID">Admin ID:</label>
                    <asp:TextBox ID="txtAdminID"
runat="server"></asp:TextBox>
                </div>
                <div class="form-group">
                    <label for="txtPassword">Password:</label>
                    <asp:TextBox ID="txtPassword" runat="server"
TextMode="Password"></asp:TextBox>
                </div>
                <asp:Button ID="btnLogin" runat="server" Text="Login"
CssClass="btn-login" OnClick="btnLogin_Click" />
                <asp:Literal ID="litMessage" runat="server"
CssClass="message"></asp:Literal>
            </div>
        </div>
    </form>
</body>
</html>

```

БІБЛІОГРАФІЧНА ДОВІДКА

Тема дипломної роботи: “ Веб-застосунок для прогнозування успішності розважальних релізів ”

Обсяг пояснювальної записки: 76 аркушів.

Дата закінчення роботи: 10 червня 2025 р.

Підпис студента _____