

**Коротун О.В.**

Державний університет «Житомирська політехніка»

**Вакалюк Т.А.**

Державний університет «Житомирська політехніка»

**Кушнірчук О.М.**

Державний університет «Житомирська політехніка»

**Марцева Л.А.**

Державний університет «Житомирська політехніка»

## ТЕОРЕТИЧНІ АСПЕКТИ РОЗРОБКИ ІНТЕРАКТИВНОГО ВЕБСЕРВІСУ КІНОФІЛЬМІВ ІЗ МОЖЛИВІСТЮ ПРОГНОЗУВАННЯ КАСОВОГО УСПІХУ ФІЛЬМУ

*Кінотворці використовують приватні системи прогнозування касового успіху для розуміння того, чи варто створювати саме сьогодні фільми конкретних жанрів. Саме для цього створюється вебсервіс кінофільмів із можливістю прогнозування касового успіху фільму, щоб будь-який користувач міг побачити приблизний касовий успіх. Прогнозування касового успіху – це саме те, чого потребує кожний кінематограф, кожна кіностудія та кожен, хто цікавиться фільмами. Оскільки для виробників кінофільмів немає сенсу створювати те, що не принесе кошти, для глядачів немає сенсу витратити гроші на неякісну кінокартину. Застосування нейронної мережі для прогнозування касового успіху значно покращить досвід кінематографів та глядачів, яким кортять дізнатися приблизний касовий успіх фільму. Метою роботи є проектування архітектури та розробка алгоритмів роботи вебсервісу кінофільмів із можливістю прогнозування касового успіху. Прогнозування – процес передбачення майбутнього стану предмета чи явища на основі аналізу його минулого і сучасного, систематично оцінювана інформація про якісні й кількісні характеристики розвитку вибраного предмета чи явища в перспективі. Результатом прогнозування є знання про майбутнє і про ймовірний розвиток сьогочасних тенденцій конкретного явища-об'єкта в подальшому існуванні. У цьому дослідженні було проаналізовано функціональні та нефункціональні вимоги до інтерактивного вебсервісу з можливістю прогнозування касового успіху кінофільмів. Побудовано діаграму варіантів використання та на її основі розроблено діаграму класів для аналізу основної структури системи. Спроектовано та реалізовано базу даних відповідно до вимог реляційної моделі: забезпечення збереження та колективний доступ до переліку кінофільмів та інформації про них. Розроблено нейронну мережу для прогнозування та основні модулі для роботи вебсервісу.*

**Ключові слова:** прогнозування, нейронні мережі, вебсервіс, архітектура, алгоритм.

**Постановка проблеми.** Кожну секунду у світі проходять зйомки кінофільмів, що є доволі вигідною справою. Кінотворці використовують приватні системи прогнозування касового успіху для розуміння того, чи варто створювати нині фільми конкретних жанрів. Світ змінюється і попит також, як відомо, попит породжує пропозицію, а щоб запропонувати, необхідно знайти цей попит, відзнявши кінофільми різних жанрів, і на основі аналізу отриманих даних з'ясувати, які типи фільмів приносять найбільший прибуток. Саме для цього створюється вебсервіс кінофільмів із можливістю прогнозувати касовий успіх фільму, щоб будь-який користувач

міг побачити приблизний касовий успіх. Прогнозування касового успіху – це саме те, чого потребує кожен кінематограф, кожна кіностудія та кожен інший, хто цікавиться фільмами. Оскільки для виробників кінофільмів немає сенсу створювати те, що не принесе кошти, для глядачів немає сенсу витратити гроші на неякісну кінокартину.

Застосування нейронної мережі для прогнозування касового успіху значно покращить досвід кінематографів та глядачів, яким кортять дізнатися приблизний касовий успіх фільму.

**Аналіз останніх досліджень і публікацій.** Останнім часом дедалі більше науковців приді-

ляють увагу прогнозуванню в різних сферах: як у техніці, так і в економіці. Цьому питанню приділили свої праці такі науковці, як В. Волохата [6], О. Дзендзелок [7], О. Любич [4], В. Мартинюк [8], С. Реуцький [3], Н. Рилової [5] та ін. Окрім того, методикою застосування різних видів прогнозування займалися Ю. Бажал, Н. Громов, Л. Федулова та ін.

**Постановка завдання.** Метою роботи є проектування архітектури та розробка алгоритмів роботи вебсервісу кінофільмів із можливістю прогнозування касового успіху.

**Виклад основного матеріалу дослідження.** Прогнозування – процес передбачення майбутнього стану предмета чи явища на основі аналізу його минулого і сучасного, систематично оцінювана інформація про якісні й кількісні характеристики розвитку вибраного предмета чи явища в перспективі. Результатом прогнозування є знання про майбутнє і ймовірний розвиток сьогочасних тенденцій конкретного явища-об'єкта в подальшому існуванні.

З огляду на тематику кінофільмів прогнозування може бути як автоматизоване, так і ручне, в будь-якому з цих випадків потрібні попередні дані фільмів які вже відзначили прем'єру, яка доля їх настигла, скільки грошей було отримано, хто та як зіграв свою роль, який сценарій до фільму та хто його писав та невід'ємна частина – режисура.

Алгоритми машинного навчання можна описати як навчання цільової функції  $f$ , яка найкращим чином співвідносить вхідні змінні  $X$  і вихідну змінну  $Y$ ,  $Y = f(X)$ .

Що собою являє функція  $f$ , невідомо. Якби знали, то використовували б її безпосередньо, а не намагалися навчити за допомогою різних алгоритмів.

Найбільш поширеним завданням у машинному навчанні є передбачення значень  $Y$  для нових значень  $X$ . Це називається прогностичним моделюванням, і наша мета – зробити передбачення якомога точнішим.

Нейронні мережі допомагають людству класифікувати та класифікувати інформацію. Класифікація включає передбачення, до якого класу належить об'єкт. Деякі класифікатори є двійковими, внаслідок чого приймається рішення «так / ні». Інші – багатокласні, здатні класифікувати об'єкт на одну з декількох категорій. Класифікація є дуже поширеним випадком машинного навчання – алгоритми класифікації використовуються для вирішення таких проблем, як фільтрування спаму по електронній пошті, категоризація

документів, розпізнавання мови, розпізнавання зображень, розпізнавання рукописного тексту та прогнозування. Розглянемо різновиди нейронних мереж.

**Нейронні мережі прямого поширення** (feed forward neural networks, FF або FFNN) і **перцептрони** (perceptrons, P) – дуже прямолінійні, вони передають інформацію від входу до виходу. Нейронні мережі нерідко описуються у вигляді листового торта, де кожен шар складається із вхідних, прихованих або вихідних нейронів. Нейрони одного шару не пов'язані між собою, а сусідні шари зазвичай повністю пов'язані. Найпростіша нейронна мережа має дві вхідних нейрони і один вихідний і може використовуватися в якості моделі логічних зв'язків. FFNN зазвичай навчається за методом зворотного поширення помилки, в якому мережа отримує безліч вхідних і вихідних даних [2]. Цей процес називається навчанням з учителем, і він відрізняється від навчання без учителя тим, що в другому випадку вихідні дані немає з чим звірити і нейронна мережа оцінює правильність даних самостійно і це зазвичай призводить до значної кількості помилок. Вищезазначена помилка є різницею між правильними даними і прогнозованими даними. Якщо в мережі є достатня кількість прихованих нейронів, вона теоретично здатна змоделювати взаємодію між вхідними і вихідними даними. На практиці такі нейронні мережі використовуються рідко, але їх часто комбінують з іншими різновидами мереж для отримання нових.



Рис. 1. Структура нейронних мереж прямого поширення

**Згорткові нейронні мережі** (convolutional neural networks, CNN) і **глибинні згорткові нейронні мережі** (deep convolutional neural networks, DCNN) – сильно відрізняються від інших видів мереж. Зазвичай вони використовуються для обробки зображень, рідше для аудіо. Типовим способом застосування CNN є класифікація зображень: якщо на зображенні присутній кіт, мережа видасть «кіт», якщо собака – «собака». Такі мережі зазвичай використовують «сканер», що не парсить усі дані за раз. Наприклад, якщо у вас є зображення  $200 \times 200$ , ви не будете одразу обробляти всі 40 тисяч пікселів. Замість цього, мережа зчитає квадрат розміру  $20 \times 20$  (зазвичай із лівого верхнього кута), потім

зрушиться на 1 піксель і зчитує новий піксель тощо. Потім ці вхідні дані передаються через згорткові шари, в яких не всі вузли з'єднані між собою. Ці шари мають властивість стискатися з глибиною, причому часто використовуються ступені двійки: 32, 16, 8, 4, 2, 1. На практиці до кінця CNN прикріплюють FFNN для подальшої обробки даних. Такі мережі називаються глибокими (DCNN) [2].

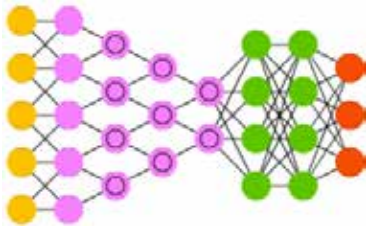


Рис. 2. Структура згорткових нейронних мереж

**Розгорткові нейронні мережі** (deconvolutional networks, DN), так звані зворотні графічні мережі, є зворотними до згорткових нейронних мереж. Уявіть, що ви передаєте мережі слово «кіт», а вона генерує картинку з котами, схожими на реальні зображення котів. DNN теж можна об'єднувати з FFNN. Варто зауважити, що здебільшого мережі передається не рядок, а якийсь бінарний вектор: наприклад,  $\langle 0, 1 \rangle$  – це кіт,  $\langle 1, 0 \rangle$  – собака, а  $\langle 1, 1 \rangle$  – і кіт, і собака [2].

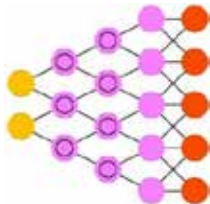


Рис. 3. Структура розгорткової нейронної мережі

Унікальна сила нейронних мереж полягає в здатності динамічно створювати складні функції прогнозування та імітувати людське мислення таким чином, що не може зробити жоден інший алгоритм. Є багато проблем класифікації, нейронні мережі вирішують такі проблеми найкращим чином.

Проаналізувавши кілька різновидів нейронних мереж можна вирішити, яку мережу необхідно використати для прогнозування касового успіху. Згорткові та розгорткові нейронні мережі не підходять для прогнозування. Оскільки нейронні мережі прямого поширення передбачають можливість прогнозування даних, було вибрано саме цей різновид мереж. До цієї нейронної мережі можна застосувати навчання з вчителем, також

вона здатна вмещувати велику кількість вхідних та прихованих шарів, що надалі дасть змогу прогнозувати максимально точно. Чим більше параметрів можна надати для аналізу, тим точнішим є результат.

Аналіз наукових джерел показав, що відсутні думки відомих математиків чи економістів, тому можемо оперувати лише поглядами глядачів, які поділилися своїми думками. Тому зрозуміло, що можливою причиною є складний процес виявлення успіху кінокартин, незначна швидкість відбору необхідної інформації, складний аналіз того, що саме людей цікавить у різні періоди їхнього життя.

Розглянемо одне з рішень глядачів. Глядач використав базу даних з інформацією від співтовариства про 5000 кінофільмів. БД присутня у відкритому доступі на вебресурсі kaggle.com. Знаючи лише кілька фактів про фільм до його виходу, прогнозуюча модель може показати доволі якісний та точний прогноз, наприклад, для того щоб кінотеатри заздалегідь вирішили, чи варто показувати фільм протягом тривалого періоду.

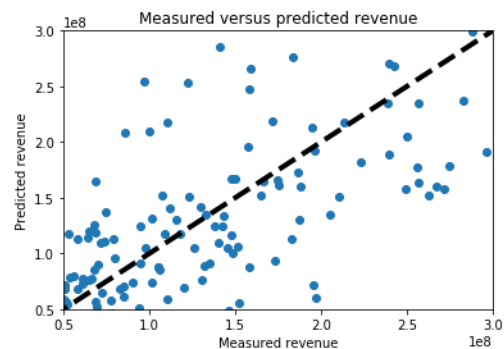


Рис. 4. Прогнозування касових зборів

Глядач-програміст повідомляє: «Для моделі прогнозування рейтингу фільму я прогнав дані, використавши «пошукову сітку» Гіперпараметр та регресорну бібліотеку XGBoost. Я спробував кілька інших бібліотек для роботи з «пошуковими сітками», включаючи випадкові регресори лісу та неймовірно продуктивну нейромережу. «Пошукова сітка» значно покращила результати XGBoost – бібліотеки, яка рекомендується за свою швидкість та точність». Модель прогнозує касові збори кінофільмів із точністю в 77%.

Під час пошуку аналогів застосування нейронних мереж для прогнозування касових зборів кінофільмів було знайдено лише один аналог, від глядача в мережі Інтернет.

Шаблон проектування Model View Controller (MVC) визначає, що додаток складається з моделі

даних, представлення інформації та керування даними. Згідно з шаблоном проектування необхідно, щоб кожна частина шаблону була поділена на різні об'єкти. MVC здебільшого стосується інтерфейсу користувача / рівня взаємодії програми. Також може знадобитися рівень бізнес-логіки, можливо, рівень обслуговування та рівень доступу до даних [1].

**Model** містить лише дані додатку, вона не містить жодної логіки, яка описує, як представити дані користувачеві [1]. **View** відображає користувачеві дані моделі. Представлення даних знає, як отримати доступ до даних моделі, але не знає, що ці дані означають або що користувач може зробити, щоб ними маніпулювати [1]. **Controller** існує між видом та моделлю. Він прослуховує події, викликані представленням (або іншим зовнішнім джерелом), і виконує відповідну реакцію на ці події. Здебільшого реакція полягає у виклику методу на моделі. Оскільки представлення та модель пов'язані за допомогою механізму сповіщення, результат цієї дії автоматично відображається у представленні [1].

Для створення інтерактивного вебсервісу було вибрано таку архітектуру: вебсервер Apache, система керування базою даних MySQL, мову програмування PHP, набір інструментів (фреймворк) Yii2.

Визначимо функціональні вимоги вебсервісу:

- 1) авторизація користувачів у системі, а також їх реєстрація;
- 2) можливість збереження інформації в системі;
- 3) можливість працювати з інформацією (додавати, редагувати, видаляти);
- 4) можливість адміністратора вебсервісу створювати добірки кінофільмів для користувачів;
- 5) підтримка англійської, української та російської мов вебсервісом;

6) пошук необхідної користувачеві інформації на вебсайті;

7) запуск прогнозування касового успіху кінофільмів.

На рис. 5 зображено діаграму варіантів використання, як саме можна взаємодіяти із сервісом у ролі гостя та в ролі авторизованого користувача. Аналіз функціональних вимог дав нам змогу виділити 4 ролі в проєкті (гість, авторизований користувач, модератор та адміністратор), що забезпечать функціонування вебсервісу. Відповідно, можливості в кожній ролі будуть свої (див. рис. 5 та 6).

Виділимо нефункціональні вимоги вебсервісу:

1) сприйняття: час, потрібний для освоєння вебсайту, – від 1 до 2 хвилин, для досвідчених менше 1 хвилини; час відповіді вебсайту на дії користувача не мають перевищувати 1 секунди, а для складних не більше – 2 секунд; інтерфейс вебсайту має бути інтуїтивно зрозумілим та зручним для користувача та не вимагати від нього додаткової підготовки;

2) надійність: вебсервіс має бути доступний протягом усієї доби;

3) продуктивність: велика кількість користувачів має змогу переглядати вебресурс одночасно.

Зобразимо діаграму класів вебсервісу кінофільмів із можливістю прогнозування касового успіху (рис. 7).

Також реалізуємо базу даних movies, використовуючи СКБД MySQL. БД складається із 18 таблиць, що вміщують усі необхідні дані для коректної роботи вебсервісу. Схема цієї бази даних зображена на рис. 8.

Опишемо діаграму активності використання вебсервісу користувачами (рис. 9):

– під час переходу на сайт користувач може пройти реєстрацію, авторизуватися в обліковий запис, або переглянути перелік кінофільмів, які є на вебсервісі;

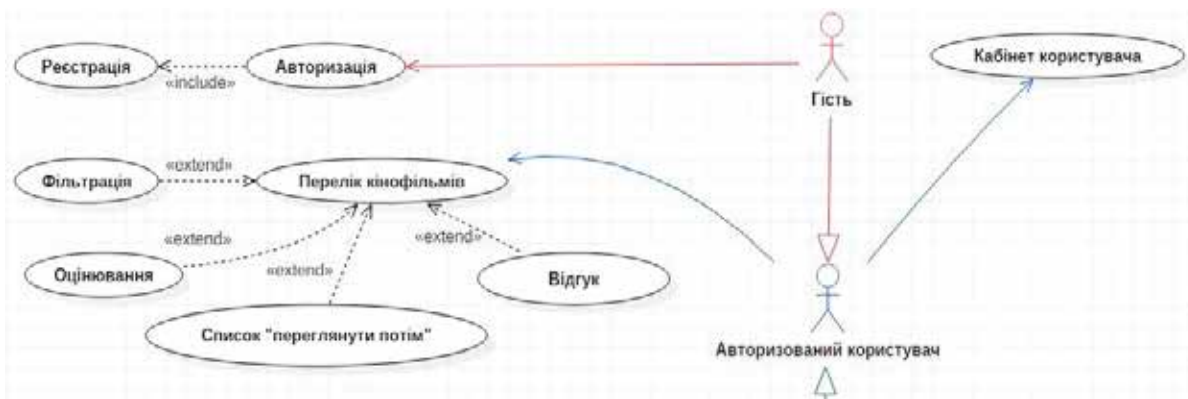


Рис. 5. Діаграма використання вебсервісу гостем та авторизованим користувачем

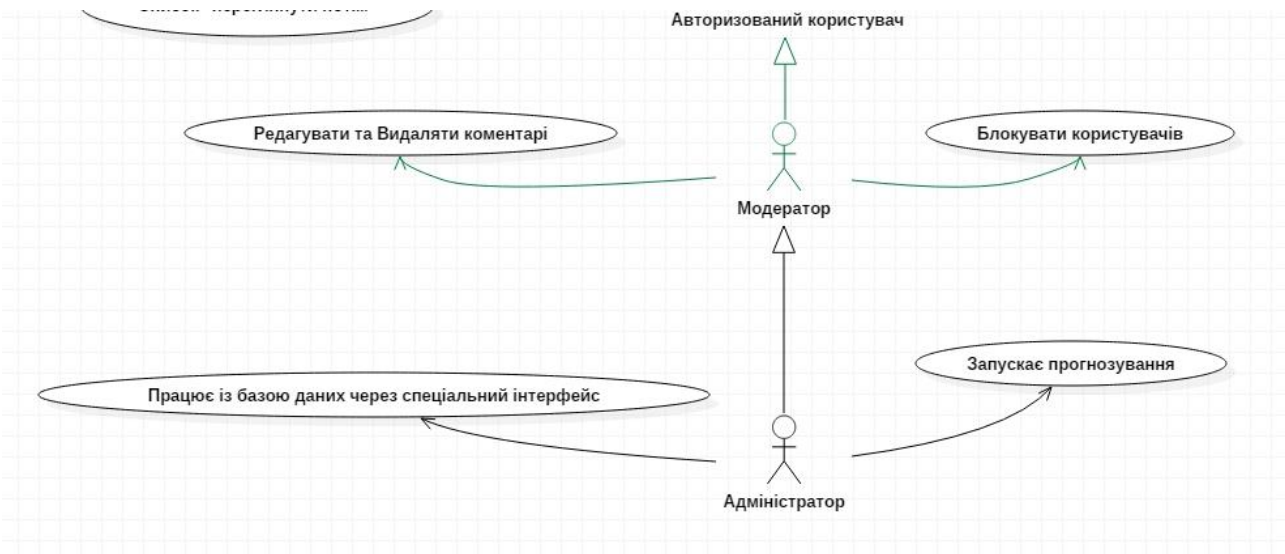


Рис. 6. Діаграма варіантів використання користувачами вебсервісу фільмів

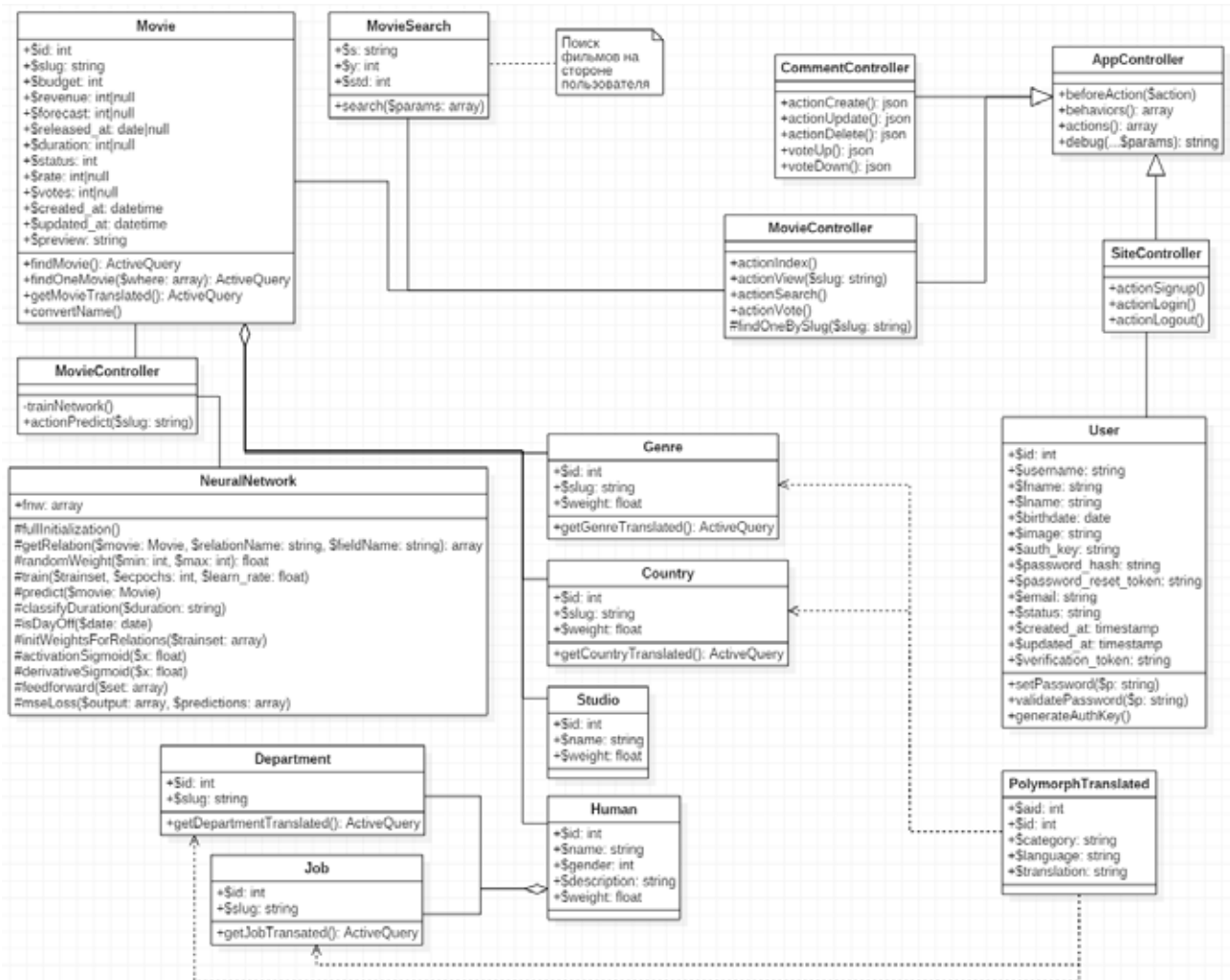


Рис. 7. Діаграма класів вебсервісу кінофільмів

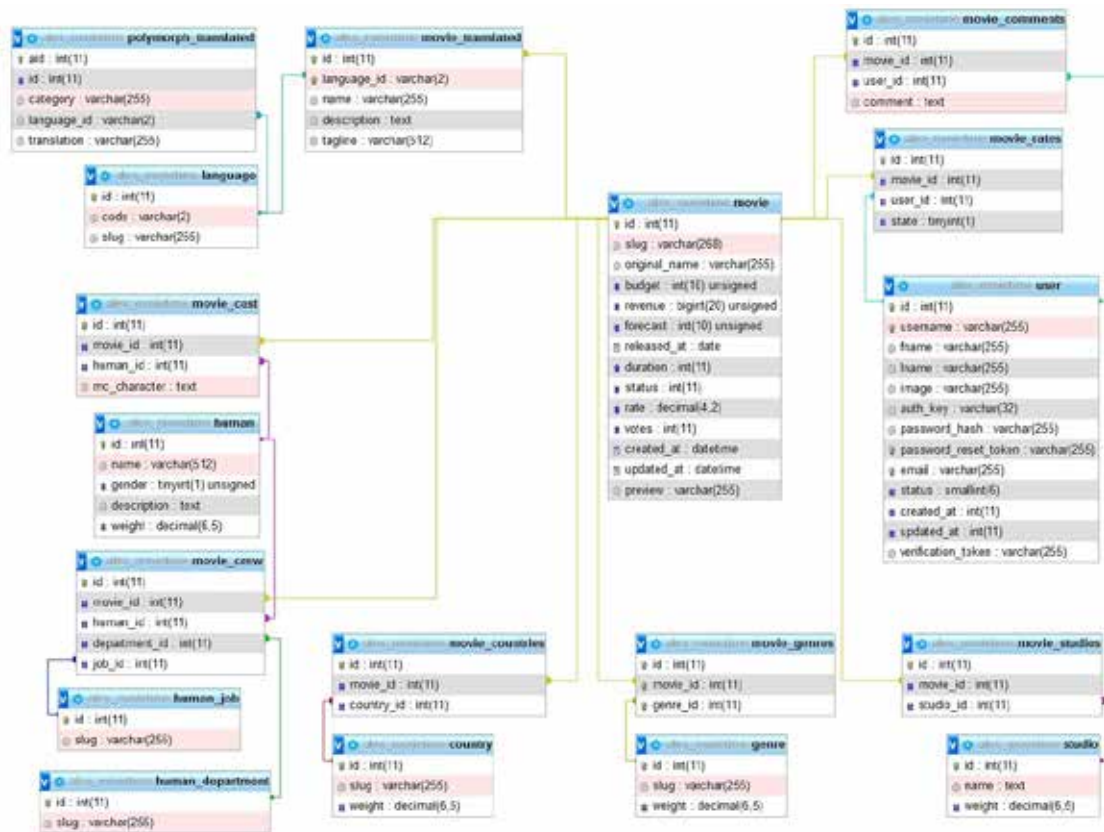


Рис. 8. Схема бази даних

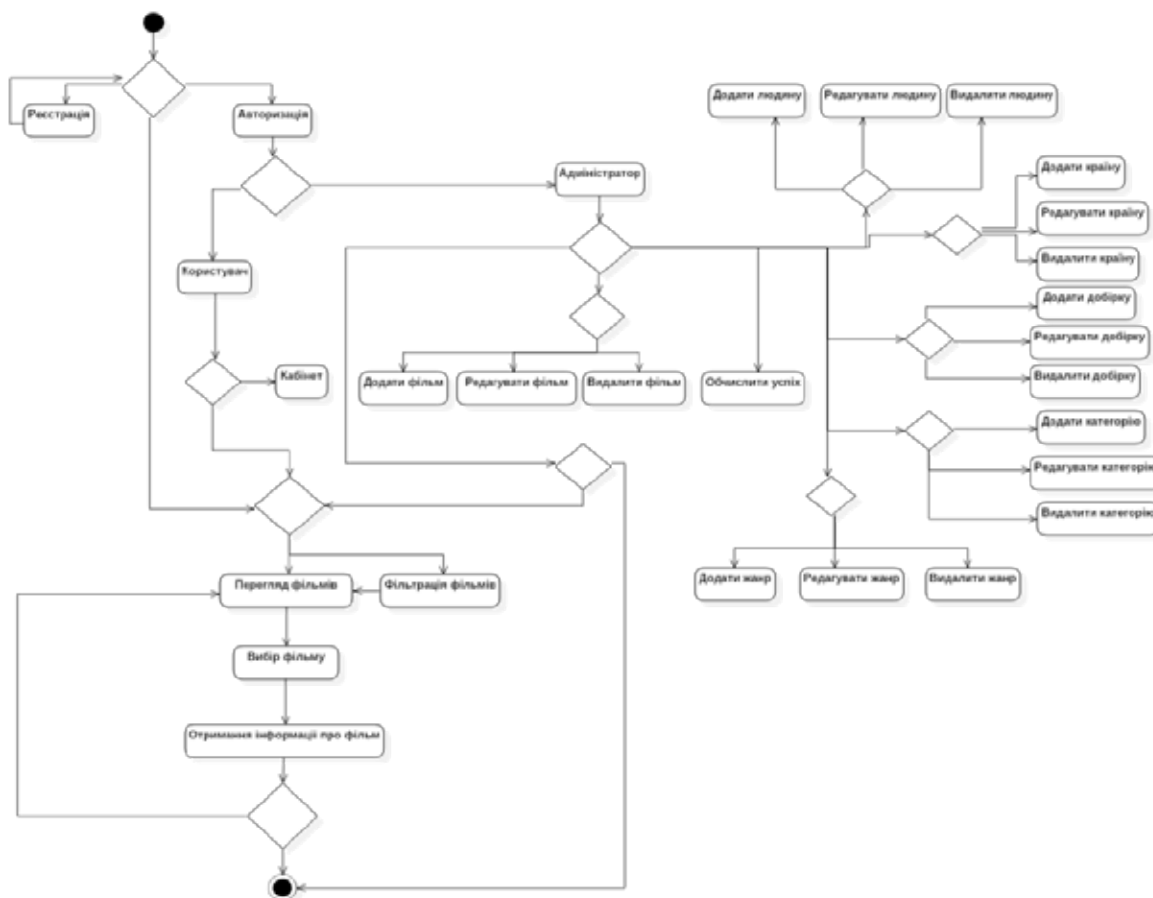


Рис. 9. Діаграма активності використання вебсервісу користувачами

- якщо користувач авторизується, він отримає доступ до свого кабінету, змогу залишати відгуки та ставити рейтинг фільму;

- адміністратор може створювати добірки з фільмами, які є в базі даних, редагувати їх або ж знищувати;

- адміністратор може запустити систему прогнозування касового успіху фільмів.

Зобразимо діаграму послідовності додавання нового фільму та подальшого обрахування успіху цього ж фільму (див. рис. 10).

Опишемо діаграму послідовності додавання нового фільму та обрахування для нього успіху:

- гість проходить авторизацію і входить в обліковий запис із правами адміністратора;

- переходить до панелі адміністратора, проходить ще одну аутентифікацію;

- переходить на сторінку додавання нових фільмів, створює такий, заповнюючи усі необхідні йому поля;

- переходить у місце обчислення касового успіху кінофільмів та натискає кнопку прогнозу фільму.

Для реалізації інтерактивного вебсервісу кінофільмів із можливістю прогнозування касового успіху фільмів було створено та реалізовано значну кількість класів. Основним класом є NeuralNetwork, який описує нейронну мережу.

Було створено нейронну мережу з 3 шарами у вигляді класичного перцептрона. Прихований шар містить 2 нейрони. Вихідний шар містить 1 нейрон (див. рис. 11).

Нейронна мережа складається із 4 вхідних параметрів: бюджет кінофільму, тривалість

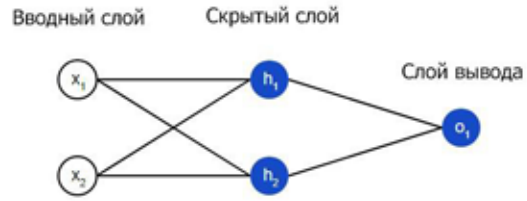


Рис. 11. Перцептрон

фільму, день прем'єри (відбувся у вихідний день чи ні), та 18 жанрів. Кожному жанру на вхід подається значення 0 або 1 залежно від того, чи підпадає фільм під конкретний жанр.

Для обчислення похибки прогнозування було використано метод середньоквадратичної помилки. Нейронна мережа навчалась за алгоритмом оптимізації «стохастичний градієнтний спуск».

Навчена нейронна мережа показала похибку прогнозування у 0.193 за тисячу епох, меншого значення помилки досягти не вдалось через недостатність вхідних даних (див. рис. 12).

Epoch 0 loss: 0.225	Epoch 900 loss: 0.193
Epoch 10 loss: 0.194	Epoch 910 loss: 0.193
Epoch 20 loss: 0.193	Epoch 920 loss: 0.193
Epoch 30 loss: 0.193	Epoch 930 loss: 0.193
Epoch 40 loss: 0.193	Epoch 940 loss: 0.193
Epoch 50 loss: 0.193	Epoch 950 loss: 0.193
Epoch 60 loss: 0.193	Epoch 960 loss: 0.193
Epoch 70 loss: 0.193	Epoch 970 loss: 0.193
Epoch 80 loss: 0.193	Epoch 980 loss: 0.193
Epoch 90 loss: 0.193	Epoch 990 loss: 0.193
Epoch 100 loss: 0.193	Epoch 1000 loss: 0.193

Рис. 12. Помилка прогнозування досягла 0.193

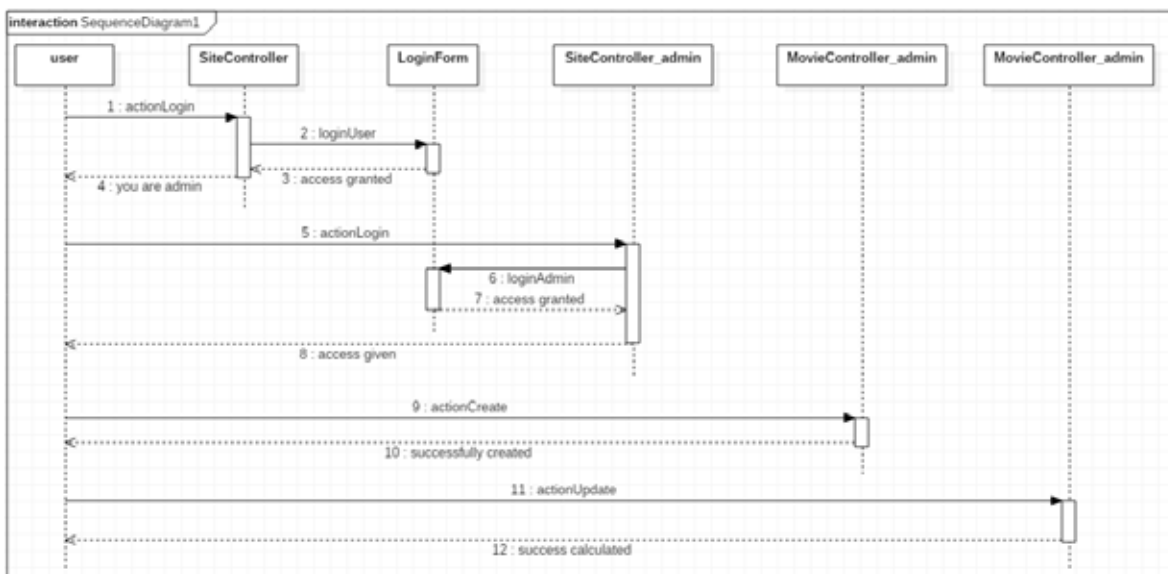


Рис. 10. Діаграма послідовності додавання нового фільму та обрахування для нього успіху

**Висновки.** У цьому дослідженні було проаналізовано функціональні та нефункціональні вимоги до інтерактивного вебсервісу з можливістю прогнозування касового успіху кінофільмів. Побудовано діаграму варіантів використання та на її основі розроблено діаграму класів для аналізу основної

структури системи. Спроектовано та реалізовано базу даних відповідно до вимог реляційної моделі: забезпечення збереження та колективний доступ до переліку кінофільмів та інформації про них. Розроблено нейронну мережу для прогнозування та основні модулі для роботи вебсервісу.

#### Список літератури:

1. MVC Design Pattern. URL: <https://www.geeksforgeeks.org/mvc-design-pattern/>. (дата звернення: 19.03.2021)
2. Нейронні мережі – шлях до глибинного навчання. URL: <https://codeguida.com/post/739>. (дата звернення: 19.03.2021)
3. Реуцький Є. Метод прогнозування характеристик точності вимірювальних каналів енергетичних систем. *Енергетика і автоматика*. 2016. № 3. С. 171–181.
4. Любіч О., Домрачев В. Застосування моделі ARIMA для прогнозування доходів зведеного бюджету України. *Економічно-математичне моделювання соціально-економічних систем: збірник наук. пр.* Вип. 16. Київ: МННЦІТС НАН та МОН України, 2011. С. 33–37.
5. Рилова Н., Оксаніч І. Синтез ARIMA-моделей для прогнозування коефіцієнтів виходу кондиційних напівпровідникових матеріалів. *Системи обробки інформації*. 2015. Вип. 5 (130). С. 102–107.
6. Волохата В. Управління залученими ресурсами банку на основі цінового прогнозування з використанням мультиплікативних моделей ARIMA. *Інвестиції: практика та досвід*. 2014. № 24. С. 72–78.
7. Дзендзелюк О., Костів Л., Рабик В. Побудова ARIMA-моделей часових рядів для прогнозування метеоданих на мові програмування R. *Електроніка та інформаційні технології*. 2013. Вип. 3. С. 211–219.
8. Мартинюк В. Прогнозування надходження податкових платежів до державного бюджету за допомогою використання ARIMA-моделі. *Наукові праці НДФІ*. 2011. № 2 (55). С. 46–55.

#### Korotun O.V., Vakaliuk T.A., Kushnirchuk O.M., Martseva L.A. THEORETICAL ASPECTS OF DEVELOPING AN INTERACTIVE WEB SERVICE OF MOVIES WITH THE ABILITY TO PREDICT THE BOX OFFICE SUCCESS OF THE FILM

*Filmmakers use private box office success prediction systems to understand whether to create films of specific genres today. This is why a web service of movies is created with the ability to predict the box office success of the film so that any user can see the approximate box office success. Predicting box office success is exactly what every cinema, every film studio, and everyone else who is interested in movies needs. Since there is no point for filmmakers to create something that will not bring money, it makes no sense for viewers to spend money on a low-quality film. The use of a neural network to predict box office success will greatly improve the experience of cinemas and viewers who want to know the approximate box office success of a film. The purpose of the work is to design the architecture and develop algorithms for the web service of movies with the ability to predict box office success. Forecasting – the process of predicting the future state of an object or phenomenon based on the analysis of its past and present, systematically evaluated information about the qualitative and quantitative characteristics of the development of the selected object or phenomenon in the future. The result of forecasting is knowledge about the future and the probable development of current trends of a particular phenomenon-object in the future. This study analyzed the functional and non-functional requirements for an interactive web service with the ability to predict the box office success of movies. The diagram of variants of use is constructed and on its basis, the diagram of classes for the analysis of the basic structure of the system is developed. The database was designed and implemented by the requirements of the relational model: ensuring the preservation and collective access to the list of films and information about them. A neural network for forecasting and basic modules for web service operation has been developed.*

**Key words:** forecasting, neural networks, web service, architecture, algorithm.