

ПОРІВНЯЛЬНИЙ АНАЛІЗ МОДИФІКАЦІЙ NCF З ІНТЕГРАЦІЄЮ ЖАНРОВИХ ОЗНАК ДЛЯ СИСТЕМИ РЕКОМЕНДАЦІЙ КІНОФІЛЬМІВ

Вінницький національний технічний університет

Анотація

Робота присвячена розробці аналітичної веб-системи рекомендацій кінофільмів на основі алгоритму Neural Collaborative Filtering (NCF). У межах дослідження проведено аналіз та порівняльну оцінку трьох модифікацій NCF, що різняться архітектурою та методами інтеграції жанрових ознак з метою визначення найбільш ефективного підходу до формування персоналізованих рекомендацій. Встановлено, що NCFWithGenres2 з окремою обробкою ембедінгів демонструє найвищу точність прогнозування. Розроблений підхід дозволяє створювати високоточні персоналізовані рекомендації та покращує якість користувацької взаємодії з контентом на стрімінгових платформах.

Ключові слова: системи рекомендації кінофільмів, Neural Collaborative Filtering (NCF), персоналізовані рекомендації, машинне навчання.

Annotation

The study focuses on the development of an analytical web-based movie recommendation system based on the Neural Collaborative Filtering (NCF) algorithm. The research includes an analysis and comparative evaluation of three NCF modifications that differ in architecture and methods of integrating genre features, with the aim of identifying the most effective approach for generating personalized recommendations. It was determined that NCFWithGenres2, which applies separate processing of embeddings, demonstrates the highest prediction accuracy. The proposed approach enables the creation of highly accurate personalized recommendations and enhances the quality of user interaction with content on streaming platforms.

Keywords: movie recommendation system, Neural Collaborative Filtering (NCF), personalized recommendations, machine learning.

Вступ

Кіно сьогодні посідає важливе місце в житті кожної людини, а стрімкий розвиток цифрових технологій зробив доступ до фільмів простішим, ніж будь-коли раніше. Сучасні стрімінгові платформи, такі як Netflix, Amazon Prime Video, Disney+ та інші, надають користувачам тисячі фільмів, серіалів і підбірок, доступних з будь-якого пристрою, підключеного до мережі Інтернет. Однак зі зростанням обсягів контенту актуальною стає проблема пошуку кінострічок, які найбільше відповідають індивідуальним інтересам та вподобанням користувача.

Інтеграція інформаційних технологій у створення рекомендаційних систем для підбору кіно має значний потенціал для покращення користувацького досвіду. Персоналізовані рекомендації дозволяють не лише економити час, але й отримувати більш релевантні пропозиції відповідно до жанрових уподобань, настрою та історії переглядів. Крім того, такі системи можуть аналізувати тенденції перегляду, рейтинги та популярність контенту, що є цінним інструментом як для користувачів, так і для представників кіноіндустрії.

Створення ефективної рекомендаційної моделі є складним завданням, оскільки потребує врахування численних факторів та прихованих патернів взаємодії користувача з контентом. Ця робота присвячена розробці системи рекомендацій кінофільмів, побудованої на основі архітектури Neural Collaborative Filtering (NCF), яка здатна моделювати складні нелінійні залежності у взаємодіях. Метою роботи є розробка та порівняльний аналіз модифікацій NCF для створення системи, що забезпечує високоточні персоналізовані рекомендації та покращує якість користувацької взаємодії з платформою.

Постановка задачі

Розробити систему, яка за допомогою методів машинного навчання надасть користувачам

рекомендації кінофільмів на основі їхніх індивідуальних уподобань.

Для досягнення мети необхідно вирішити наступні завдання:

- провести огляд існуючих систем і провести аналіз проблеми;
- підготувати дані для навчання моделі рекомендацій;
- розробити модель машинного навчання для рекомендацій кінофільмів.

Результати дослідження

Для аналізу та розробки системи було використано набір даних «MovieLens Small Latest Dataset», отриманий з ресурсу Kaggle [1]. Цей датасет містить усю необхідну інформацію для побудови рекомендаційної системи, зокрема дані про користувачів, рейтинги, фільми та жанри.

У роботі використано малий варіант датасету, оскільки великий набір MovieLens є занадто об'ємним для мого апаратного забезпечення й потребує значно більших обчислювальних ресурсів для повноцінної обробки та навчання моделей.

Приклад даних з цього датасету показано на рисунку 1.

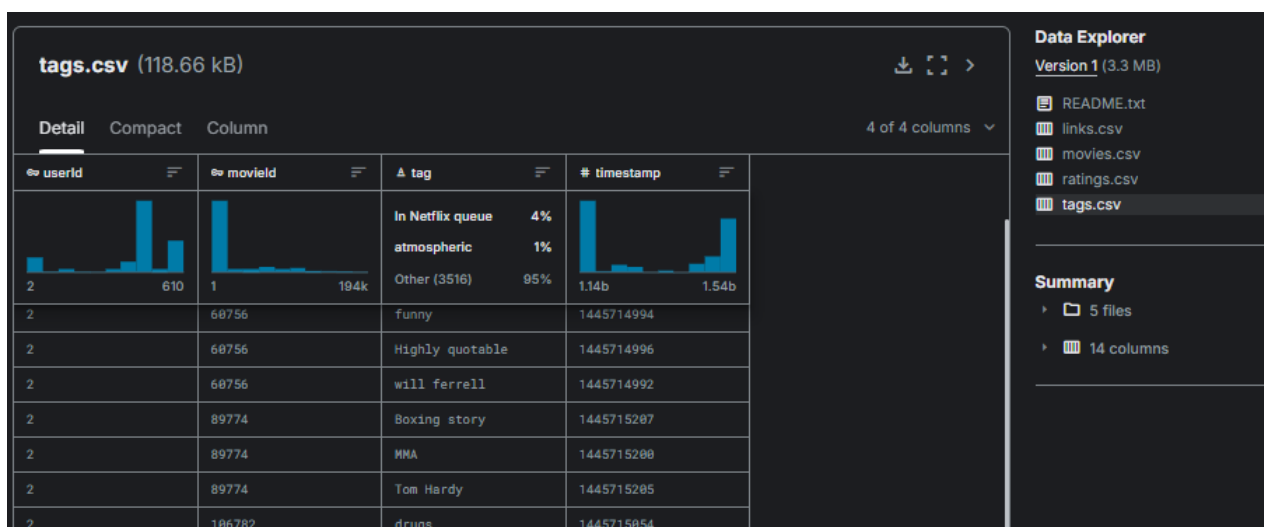


Рисунок 1 - Набір даних

Набір даних, використаний у цьому проєкті, — MovieLens Small Latest Dataset, що містить приблизно 105 000 рейтингів, 1 620 користувачів та 9 742 фільми. Датасет включає декілька таблиць, зокрема ratings, movies та tags, що містять ключову інформацію для побудови рекомендаційної системи: ідентифікатори користувачів та фільмів, рейтинги, часові позначки, назви фільмів і жанрову класифікацію.

Для підготовки даних було проведено попередню обробку таблиць movies та ratings датасету MovieLens Small. Код виконує очищення жанрової інформації, перетворюючи текстові жанри у списки та формуючи унікальний словник жанрів. Рейтинги було зведено у матрицю «фільм–користувач», де відсутні значення заповнено нулями. Надалі видалено фільми з менш ніж 10 оцінками та користувачів із менш ніж 50, після чого матрицю перетворено у розріджений формат CSR.

Додатково створено тензори жанрів та offsets для роботи з EmbeddingBag, а також сформовано датасет GenreDataset і функцію collate_fn. Після очищення отримано приблизно 100 000 структурованих рейтингів, придатних для навчання моделі рекомендацій.

У дослідженні було використано три моделі Neural Collaborative Filtering (NCF) [2], що різняться структурою та обробкою даних. Нейронне колаборативне фільтрування (NCF) — це сучасна рекомендаційна модель, яка замінює традиційну матричну факторизацію на глибокі нейронні мережі. Це дає змогу навчатися складним нелінійним залежностям у взаємодіях між користувачами та елементами. На відміну від моделювання переваг за допомогою простих скалярних добутків, NCF використовує багатошарову нейромережу для виявлення багатших латентних представлень користувачів та елементів, що забезпечує точніше прогнозування оцінок або іншої неявної інформації про взаємодії.

NCF слугує фундаментальною архітектурою [3,4], в якій векторні представлення (ембединги)

користувачів та елементів навчаються спільно через навчання за принципом «end-to-end» (від входу до виходу). Це робить NCF гнучким і потужним інструментом для побудови персоналізованих рекомендаційних систем.

Базова модель — це простий варіант, який використовує ембедінги для користувачів, фільмів і жанрів, об'єднуючи їх в один вектор. Жанри представляються як середнє значення їхніх ембедінгів, що забезпечує стабільну, але не дуже гнучку роботу з різними жанровими комбінаціями. Архітектура складається з кількох повнозв'язних шарів із активацією ReLU, що робить модель швидкою, легкою і мало схильною до перенавчання. Така модель добре підходить для початкових тестів або коли важлива швидкість, а не максимальна точність. Результати роботи базової моделі зазначені на рисунку 2.

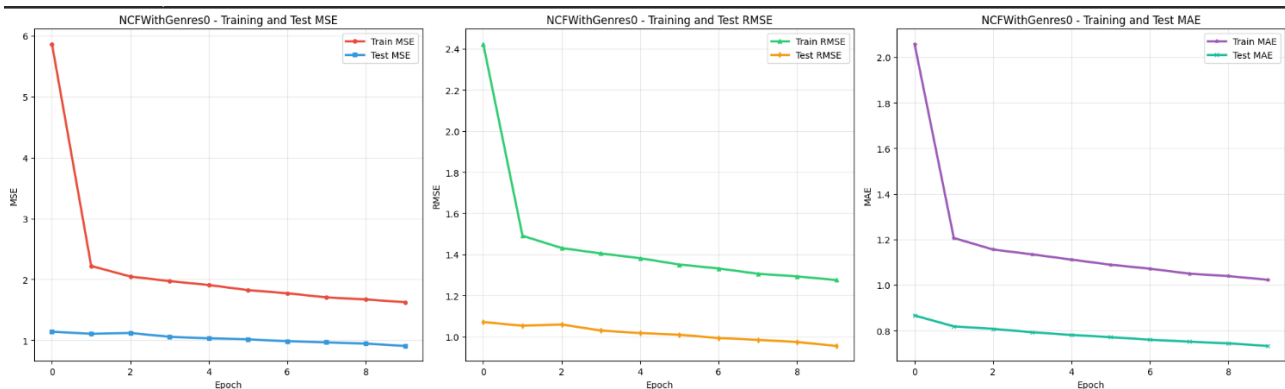


Рисунок 2 – Результати першої моделі

Розширена версія — це глибша й більш точна модель із більшими ембедінгами та складнішою архітектурою. Вона містить додаткові шари та Dropout, що допомагає уникати перенавчання і краще узагальнювати дані. Завдяки цьому модель здатна виявляти складніші взаємозв'язки між користувачем, фільмом і жанрами, забезпечуючи вищу точність прогнозів. Проте вона повільніша, потребує більше ресурсів і часу на навчання, особливо на великих датасетах. Результати роботи розширеної версії зазначені на рисунку 3.

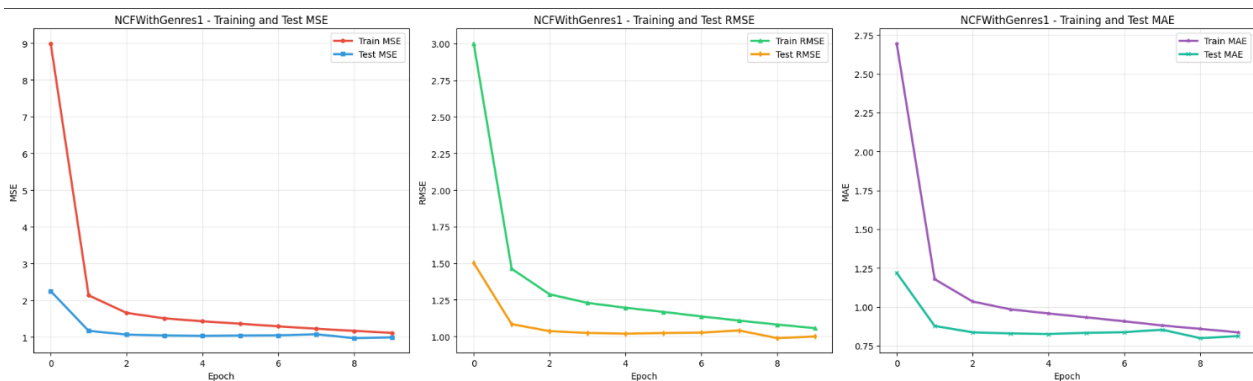


Рисунок 3 – Результати 2 моделі

Найбільш просунута модель реалізує підхід із окремою обробкою кожного типу ознак. Ембедінги користувачів, фільмів і жанрів спочатку проходять через окремі підмережі, що дозволяє зберігати їхню унікальну семантику. Після цього оброблені вектори об'єднуються та проходять через фінальні шари для передбачення рейтингу. Така архітектура гнучкіша, краще відловлює складні взаємозв'язки і придатна для розширення моделі новими ознаками, хоча є повільнішою і складнішою в налаштуванні.

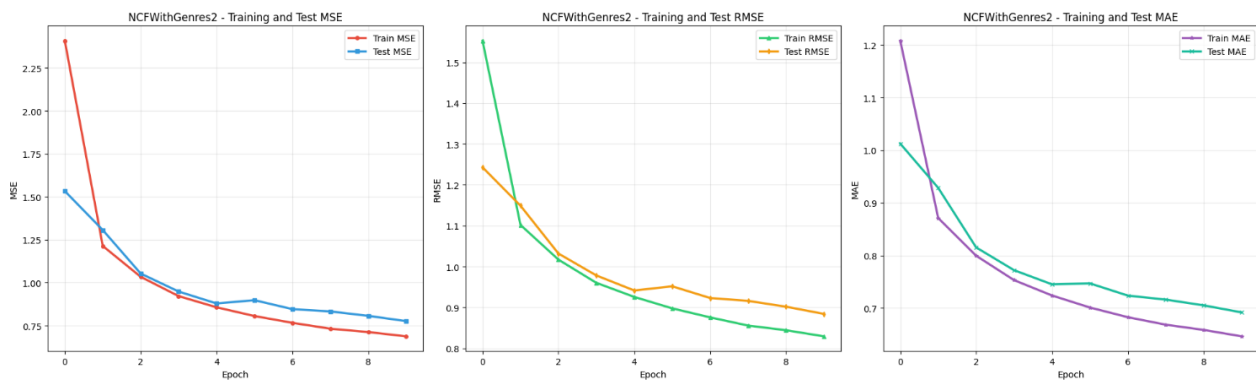


Рисунок 4 – Результати 3 моделі

Model	Train MSE	Test MSE	Train RMSE	Test RMSE	Train MAE	Test MAE
NCFWithGenres0	1.6260	0.9086	1.2751	0.9548	1.0228	0.7323
NCFWithGenres1	1.1158	0.9946	1.0563	0.9998	0.8360	0.8118
NCFWithGenres2	0.6873	0.7760	0.8290	0.8839	0.6461	0.6916

Рисунок 5 – Результати моделей

Система рекомендацій фільмів, що використовує нейронну колаборативну фільтрацію (NCF) з інтеграцією жанрових характеристик, є потужним інструментом для персоналізації відкриття кіно для користувачів. Таким чином, система рекомендацій фільмів на основі NCF пропонує масштабований, керований даними підхід до персоналізованого виявлення фільмів, а також її здатність враховувати складні взаємодії між користувачами та фільмами робить її ефективним вибором для сучасних стрімінгових платформ.

Висновки

Під час виконання проекту була розроблена архітектура веб-системи рекомендацій кінофільмів. Проведено дослідження та порівняльний аналіз трьох модифікацій нейронної колаборативної фільтрації (NCF) з інтеграцією жанрових характеристик на основі датасету MovieLens Small. Експериментально встановлено, що найбільш просунута модель (NCFWithGenres2), яка використовує окрему обробку ембедінгів для користувачів, фільмів і жанрів, забезпечує найвищу точність прогнозування. Вона досягла показника RMSE: 0.8698, MAE: 0.6707 на тестовій вибірці, покращивши RMSE на 7.42% та MAE на 5.57% порівняно з базовою моделлю NCFWithGenres0, а також на 11.6% і 13.8% відповідно порівняно з покращеною версією NCFWithGenres1 перевершуючи базові версії. Запропонований підхід є ефективним інструментом для надання високоточних персоналізованих рекомендацій, що дозволяє суттєво покращити користувацький досвід на стрімінгових платформах.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Mehta, S. (n.d.). *MovieLens Small Latest Dataset*. Kaggle. [Електронний ресурс]. URL: <https://www.kaggle.com/datasets/shubhammehta21/movie-lens-small-latest-dataset>
2. Сверхун М. (2022) НЕЙРОМЕРЕЖЕВА ТЕХНОЛОГІЯ ФОРМУВАННЯ РЕКОМЕНДАЦІЙ КОНТЕНТУ ЗА ПРІОРИТЕТАМИ СПОЖИВАЧІВ. Proceedings of the International Conference. DOI: <https://doi.org/10.62731/mcnd-18.04.2025.002>
3. Ruihui Mu (2023) A Novel Neural Collaborative Filtering Recommendation Based on Side Information Fusion DOI: <https://doi.org/10.7546/CRABS.2023.01.09>
4. Haoyang Xia and Yuanyuan Wang (2025). Enhancing Neural Collaborative Filtering for Product Recommendation by Integrating Sales Data and User Satisfaction DOI: <https://doi.org/10.3390/electronics14163165>

Тарасовський Тарас Сергійович – студент групи 2ІСТ-24м, факультет інтелектуальних інформаційних технологій та автоматизації, Вінницький національний технічний університет, Вінниця, e-mail: tarasovskiy.taras01@gmail.com

Жуков Сергій Олександрович– к.т.н., доцент кафедри системного аналізу та інформаційних технологій, Вінницький національний технічний університет, e-mail: sazhukov@gmail.com

Tarasovskiy Taras S. - student of Faculty of Intelligent Information Technologies and Automation, 2ICT-24m, Vinnytsia National Technical University, Vinnytsia, e-mail tarasovskiy.taras01@gmail.com

Zhukov Serhii O. - Ph.D., Assistant Professor of the Department of Systems Analysis and Information Technologies, Vinnytsia National Technical University, Vinnytsia, e-mail: sazhukov@gmail.com