Міністерство освіти і науки України
Харківський національний університет радіоелектроніки

Факультет _____Комп'ютерних наук_____
(повна назва)

Кафедра _____Інформаційних управляючих систем_____
(повна назва)

КВАЛІФІКАЦІЙНА РОБОТА
Пояснювальна записка

рівень вищої освіти _____другий (магістерський)_____

_Дослідження методів виявлення шиллінг-атак в рекомендаційній системі____
_музичних релізів_____
(тема)

Виконав:
студент _2_ курсу, групи _ІУСТм-21-1____
_____Владислав ДЕНИСЕНКО_____
(власне ім'я, прізвище)

Спеціальність __122 Комп'ютерні_науки____
_____
(код і повна назва спеціальності)
Тип програми ____освітньо-професійна_____
(освітньо-професійна або освітньо-наукова)
Освітня програма _Інформаційні управляючі системи та технології_____
(повна назва освітньої програми)

Керівник __проф. каф. ІУС Сергій ЧАЛИЙ___
(посада, власне ім'я, прізвище)

Допускається до захисту

Зав. кафедри _____ ___Костянтин ПЕТРОВ__
(підпис) (власне ім'я, прізвище)

2022 р.

Харківський національний університет радіоелектроніки

Факультет Комп'ютерних наук

Кафедра Інформаційних управляючих систем

Рівень вищої освіти другий (магістерський)

Спеціальність 122 Комп'ютерні науки
(код і повна назва)

Тип програми освітньо-професійна
(освітньо-професійна або освітньо-наукова)

Освітня програма Інформаційні управляючі системи та технології
(повна назва)

ЗАТВЕРДЖУЮ:

Зав. кафедри _____
(підпис)

« 21 » листопада 20 22 р.

ЗАВДАННЯ

НА КВАЛІФІКАЦІЙНУ РОБОТУ

студентові Денисенку Владиславу Петровичу
(прізвище, ім'я, по батькові)

1. Тема роботи Дослідження методів виявлення шиллінг-атак в рекомендаційній системі музичних релізів

затверджена наказом університету від 14 листопада 2022 р. № 1490Ст

2. Термін подання студентом роботи до екзаменаційної комісії 26 грудня 2022 р.

3. Вихідні дані до роботи технології побудування рекомендацій, методи створення рекомендацій, методи виявлення шиллінг-атак, сучасні рішення програмної реалізації алгоритмів виявлення шиллінг-атак

4. Перелік питань, що потрібно опрацювати в роботі 4.1 Вступ 4.2 Дослідження методів побудови рекомендацій в умовах шилінг-атак в рекомендаційних системах 4.2.1 Аналіз рекомендаційних систем 4.2.2 Дослідження шилінг-атак в музичних рекомендаційних системах 4.2.3 Дослідження методів виявлення атак користувачів на рейтинги в рекомендаційних системах 4.2.4 Постановка задачі дослідження 4.3 Виявлення шилінг-атак з використанням темпоральних знань 4.3.1 Використання темпоральних правил у процесі виявлення атак користувачів рекомендаційної системи 4.3.2 Вдосконалений метод виявлення шилінг-атак з використанням темпоральних правил 4.3 Технологія виявлення атак користувачів на рейтинги з використанням темпоральних знань 4.4 Експериментальна перевірка отриманих результатів 4.5 Висновки

КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Терміни виконання етапів роботи	Примітка
1	Узгодження завдання	10.10.2022	виконано
2	Підготовча робота	11.10-20.11.2022	виконано
3	Отримання завдання кваліфікаційної роботи	21.11.2022	виконано
4	Аналіз завдання, літератури та аналогів з теми кваліфікаційної роботи	22.11-24.11.2022	виконано
5	Структурне проєктування	25.11-28.11.2022	виконано
6	Аналіз алгоритму виявлення шиллінг-атак	29.11.2022	виконано
7	Розробка удосконаленого алгоритму виявлення шиллінг-атак	30.11-1.12.2022	виконано
8	Розробка прототипу системи	2.12-5.12.2022	виконано
9	Проведення експериментів	5.12-12.12.2022	виконано
10	Оформлення пояснювальної записки	13.12-16.12.2022	виконано
11	Оформлення графічної частини та презентаційних матеріалів комп'ютерного захисту	17.12-22.12.2022	виконано
12	Представлення на рецензування	23.12.2022	виконано
13	Представлення кваліфікаційної роботи в ЕК	26.12.2022	виконано

Дата видачі завдання _21__ _листопада_2022 р.

Студент _____
(підпис)

Керівник роботи _____ ___проф. каф. ІУС Сергій Чалий_
(підпис) (посада, власне ім'я, прізвище)

РЕФЕРАТ

Пояснювальна записка до магістерської кваліфікаційної роботи містить: 80 с., 4 розділи, 22 рис., 3 табл., 20 джерел, 1 додаток.

ЕЛЕКТРОННА КОМЕРЦІЯ, КОЛЛАБОРАТИВНА ФІЛЬТРАЦІЯ, МАШИННЕ НАВЧАННЯ, МУЗИЧНІ РЕЛІЗИ, ПЕРСОНАЛІЗАЦІЯ РЕКОМЕНДАЦІЙ, РЕКОМЕНДАЦІЙНІ СИСТЕМИ, ШИЛЛІНГ-АТАКИ, МАШИННЕ НАВЧАННЯ

Існуючі методи виявлення шиллінг-атак використовують дані рейтингів для виявлення неточних профілів користувачів та подальшого захисту від атак, пов’язаних із маніпуляцією рейтингами.

Однак використанню результатів неявного зворотнього зв’язку не приділяється достатньо уваги. В той же час в музичних рекомендаційних системах порівняння результатів рейтингів (явний зворотній зв’язок) та прослуховування музики (неявний зворотній зв’язок), де є можливість більш точно виявляти атаки, оскільки в таких онлайн-системах фіксуються дії з прослуховування музики користувачем в онлайн-режимі.

У даній роботі розглянуто методи виявлення шиллінг-атак в рекомендаційних системах музичних релізів для побудови релевантних рекомендацій в умовах некоректно виставлених рейтингів. Результатом даної роботи є удосконалений алгоритм виявлення шиллінг-атак в рекомендаційних системах музичних релізів.

ABSTRACT

Explanatory Note to master qualification work contains 80 pages, 4 sections, 22 pictures, 3 tables, 20 sources, 1 addendum.

COLLABORATIVE FILTERING, E-COMMERCE, MACHINE LEARNING, MUSIC RELEASES, PERSONALIZATION OF RECOMMENDATIONS, RECOMMENDER SYSTEMS, SHILLING ATTACKS

Existing methods for detecting shilling attacks use rating data to identify inaccurate user profiles and further protect against rating manipulation attacks.

However, not enough attention has been paid to the use of implicit feedback results. At the same time, music recommender systems compare the results of ratings (explicit feedback) and listening to music (implicit feedback), where it is possible to detect attacks more accurately, since such online systems record the user's listening to music online.

This paper considers methods for detecting shilling attacks in recommender systems for music releases to build relevant recommendations in the face of incorrect ratings. The result of this work is an improved algorithm for detecting shilling attacks in recommender systems for music releases.

СКОРОЧЕННЯ ТА УМОВНІ ПОЗНАКИ

ІП – інформаційне перевантаження;

ІС – інформаційна система;

КФ – коллаборативна фільтрація.

ЗМІСТ

ВСТУП

Рекомендаційні системи набирають велику популярність в цифровому просторі через різноманітний набір уподобань та вимог споживачів, які постійно потребують уваги. Використання рекомендаційних систем має сенс у галузі електронної комерції, зокрема великих інтернет-магазинах та сервісах трансляції розважального потокового контенту. Успішна рекомендаційна система може слугувати ключем до стрімкого розвитку підприємства та за допомогою надання ініціативи для подальшого користування інформаційною системою значно збільшує активну користувацьку базу, що приносить постійний дохід до бізнесу.

Одним з відомих методів підбору прогнозу в рекомендаційній системі є колаборативна фільтрація – метод, у якому за визначеними вподобаннями користувачів прогнозуються потенційні невідомі вподобання певного користувача з метою надання рекомендації вподобання, схожого на інші вподобання, притаманні користувачу. В системах рекомендацій музичних релізів це зазвичай підбір схожих виконавців, альбомів, пісень тощо.

Метод колаборативної фільтрації може піддаватися певним маніпуляціям, які робляться або з метою агресивного просування чи дискредитації певного продукту, або з метою умисного заподіювання шкоди, що суттєво знижує якість рекомендацій і ставить під загрозу розвиток підприємства. Одним із таких методів маніпуляції є шиллінг-атака – створення фіктивних користувачів системи, які масово ставлять позитивні або негативні оцінки конкретному предмету.

Сам факт суттєвості шкоди, нанесеної цією атакою, показує, наскільки серйозними можуть бути наслідки для системи рекомендацій, що в більш комерційних системах може завдати неабияких збитків бізнесам, що стали жертвами шиллінг-атак, тому є доцільним розгляд потенційних методів боротьби з ними.

Мета даної роботи – розглянути методи виявлення шиллінг-атак в системі рекомендацій музикальних релізів.

1 ДОСЛІДЖЕННЯ МЕТОДІВ ПОБУДОВИ РЕКОМЕНДАЦІЙ В УМОВАХ ШИЛЛІНГ-АТАК В РЕКОМЕНДАЦІЙНИХ СИСТЕМАХ

1.1 Аналіз рекомендаційних систем

Сучасні інформаційні системи мають опрацьовувати велику кількість елементів, що може мати розміри, що вираховуються в сотнях мільйонів або потенційно мільярдів елементів. Рекомендації при цьому мають бути дуже точними і персоналізованими для користувача і контексту. Це означає, що в кінцевому підсумку доведеться генерувати і використовувати безліч функцій про елемент, користувача, контекст і перехресне поєднання цих категорій з додатковою статистикою, такою як підрахунки і співвідношення. Багато з цих функцій є категоричними з потенційно дуже високою кардинальністю. В таких реаліях доволі важко вкластися в ті часові обмеження, які вам надаються для створення рекомендацій.

На додаток до оперативності, вимоги до затримки можуть також виходити від третіх сторін. Наприклад, це стосується платформ для розміщення реклами в режимі реального часу. Під час участі в торгах у Criteo є лише кілька мілісекунд, щоб вибрати продукти для показу користувачеві (з мільярдів кандидатів). Якщо вимога не виконана, Criteo втрачає ставку (і потенційно гроші).

З великого каталогу потенційна рекомендаційна система має вибрати десяток, щоб показати їх користувачеві. Якщо є, скажімо, 200 мілісекунд для здійснення підбору рекомендації, можна витратити лише 0,2 на одного кандидата, щоб вирішити, чи є він кращим вибором, ніж інші кандидати, що робить це складною технічною проблемою.

Вирішення цієї проблеми полягає в тому, щоб уникнути оцінювання всіх потенційних кандидатів, які є у існуючому каталозі. Один з простих способів

зробити це - впровадити бізнес-правила, засновані, наприклад, на популярності. Можна, наприклад, сказати, що треба оцінити тільки найпопулярніші елементи за останні 2 тижні або тільки елементи з тих самих каналів або тієї самої теми.

Користувачі систем електронної комерції приділяють значну увагу включенню рекомендацій щодо персоналізації. Персоналізація починається на домашній сторінці, яка складається з груп відео, розташованих горизонтальними рядами. Кожен рядок має заголовок, який передає передбачуваний змістовний зв'язок між предметами в цій групі. Більша частина нашої персоналізації базується на тому, як ми вибираємо рядки, як ми визначаємо, які елементи включати в них, і в якому порядку ці елементи розміщувати.

Наприклад, використовується рядок Топ-10, як найкраще припущення про десять назв, які користувачеві, швидше за все, сподобаються. Персоналізація призначена для роботи з колективом, у якому, ймовірно, є різні люди з різними смаками.

Навіть для домогосподарства з однією людиною можна задовольнити коло інтересів і настроїв. Для досягнення цього, у рекомендаційній системи оптимізується як точність, так і різноманітність запропоновних товарів. Приклад для рекомендаційної системи Netflix наведено на рисунку 2.1.

Рисунок 1.1 – Різноманітність рекомендації на прикладі системи Netflix

Іншим важливим елементом персоналізації є обізнаність. Необхідно, щоб користувачі знали, як система адаптується до їхніх смаків.

Це не лише підвищує довіру до системи, але й заохочує користувачів залишати відгуки, що призведе до кращих рекомендацій.

Ще один спосіб підвищити довіру за допомогою компонента персоналізації – це надати пояснення, чому система вирішила порекомендувати певний предмет (фільм, шоу).

Його рекомендують не тому, що він відповідає потребам бізнесу, а тому, що він відповідає інформації, яка була отримана від користувачів: явні смакові уподобання й оцінки, історія переглядів або навіть рекомендацій друзів (рисунок 2.2).

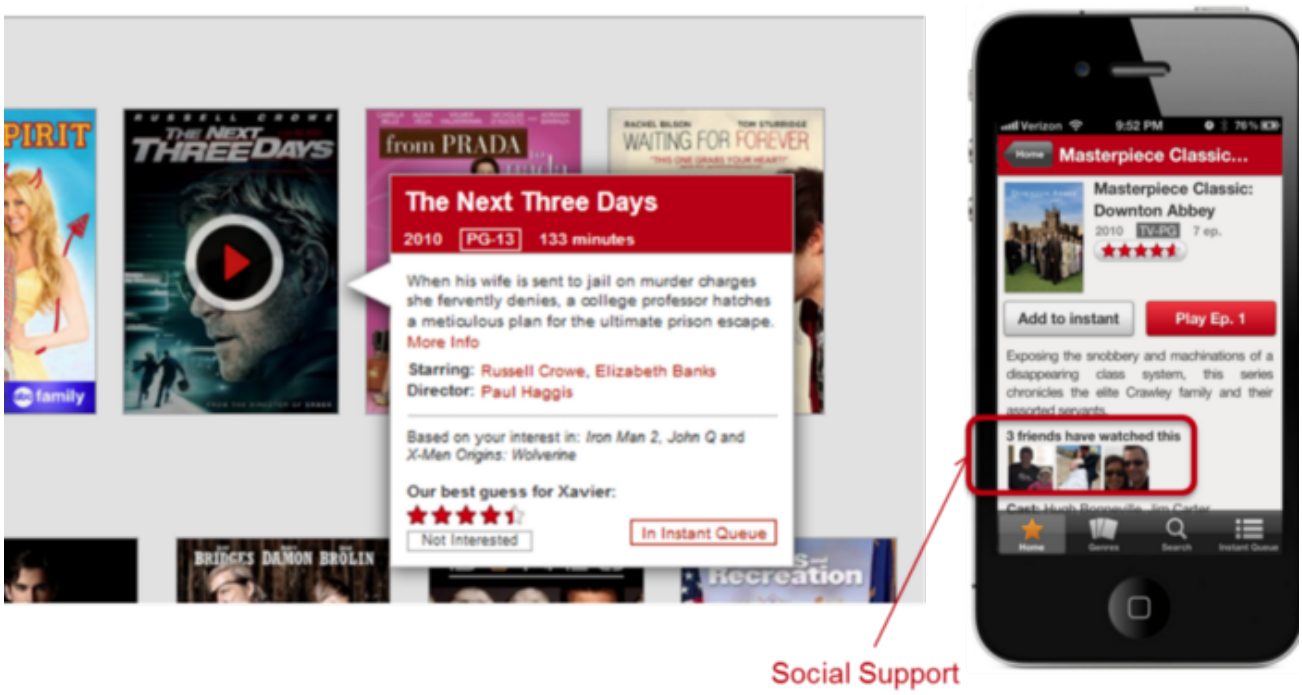

Рисунок 1.2 – Різноманітність рекомендації на прикладі системи Netflix

Що стосується друзів, то використовується функція підключення до Facebook. Знання про друзів не тільки дає інформацію для використання в алгоритмах персоналізації, але також дозволяє створювати рекомендації для різних рядків. Останні здебільшого покладаються на коло спілкування користувача.

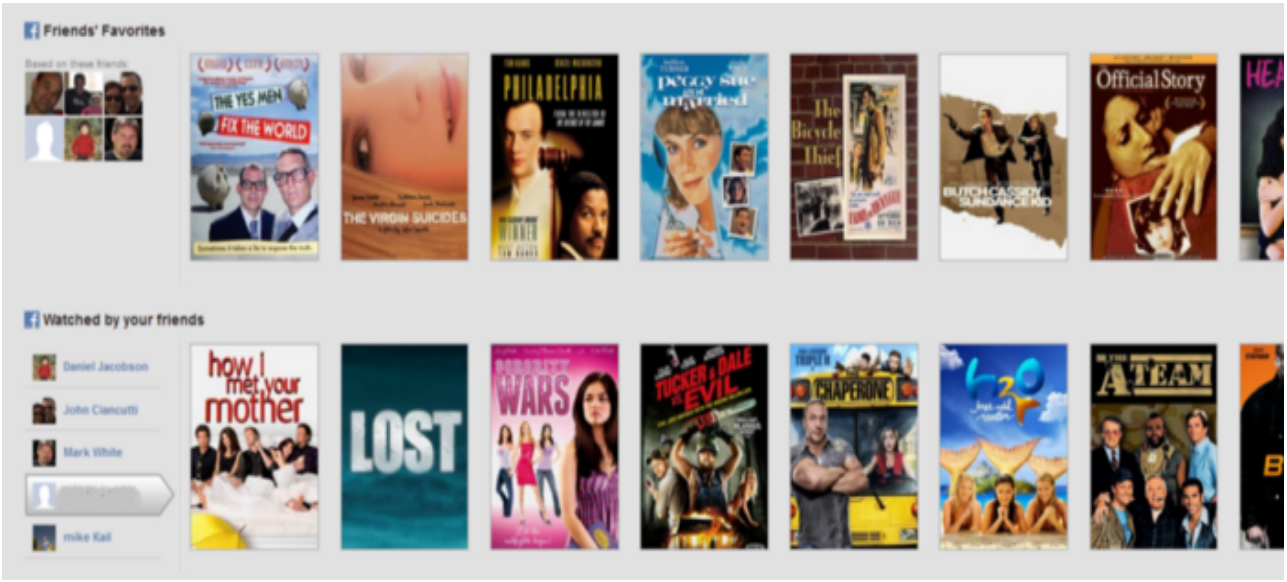

Рисунок 1.3 – Рекомендації для різних рядків на прикладі системи Netflix

Одним із варіантів персоналізації для фільмів є колекція рядків «жанр». Вони варіюються від знайомих категорій високого рівня, як-от «Комедії» та «Драми», до спеціально розроблених фрагментів, як-от «Фантастичні фільми про подорожі в часі 1980-х років».

Кожен рядок представляє 3 рівні персоналізації: сам вибір жанру, підмножина заголовків, вибраних у цьому жанрі, і рейтинг цих заголовків. Учасники настільки добре взаємодіють із цими рядками, що ми вимірюємо збільшення утримання учасників, розміщуючи найбільш підібрані рядки вище на сторінці, а не нижче.

Як і з іншими елементами персоналізації, свіжість і різноманітність береться до уваги, коли вирішується, які жанри показати з тисяч можливих.

Вона також використовується для генерації рядків «окремих жанрів» на основі подібності до заголовків, з якими учасник нещодавно взаємодіяв.

У більшості попередніх контекстів, будь то рядок Top10, жанри чи подібні рейтинги, вибір порядку розміщення елементів у рядку, має вирішальне значення для забезпечення ефективного персоналізованого досвіду.

Рисунок 1.4 – Жанровий підхід до персоналізації рекомендацій на прикладі системи Netflix

Мета рейтингування полягає в тому, щоб знайти найкраще можливе впорядкування набору елементів для учасника в певному контексті в режимі реального часу. Рейтинг розкладається на оцінку, сортування та фільтрацію наборів фільмів для представлення учаснику.

Система надає пояснення до вибору рядків, використовуючи приховані жанрові уподобання учасника – нещодавні відтворення, оцінки та інші взаємодії, або явні відгуки, надані при опитуванні. Також при опитуванні використовується рядок із додатковим чітким відгуком про переваги, якщо цього не вистачає.

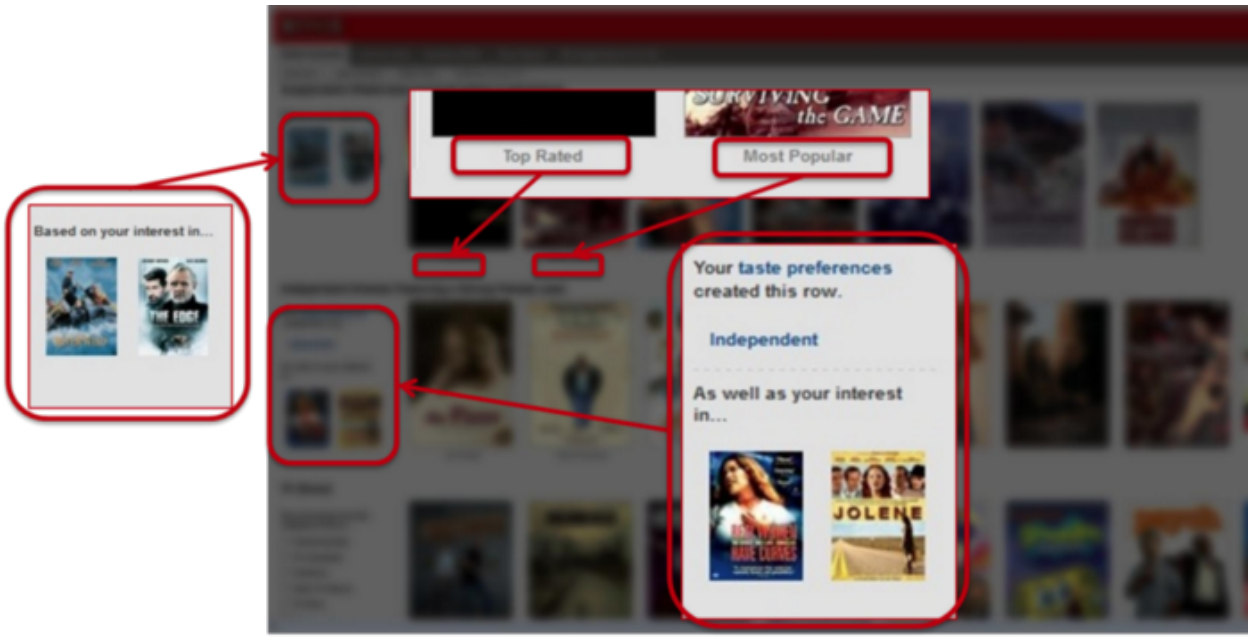

Рисунок 1.5 – Пояснення до вибору в рекомендаційній системі

Подібність також є важливим джерелом персоналізації послуг з рекомендації. Подібність розглядається у дуже широкому сенсі. Це може бути подібність між фільмами або між учасниками, і може бути в кількох вимірах, таких як метадані, рейтинги або дані перегляду.

Крім того, ці подібності можна поєднувати та використовувати як функції в інших моделях. Подібність використовується в багатьох контекстах, наприклад у відповідь на дії учасника, такі як пошук або додавання заголовка до черги.

Ціль системи полягає в тому, щоб максимізувати задоволеність учасників, що добре співвідноситься з максимальним споживанням товарів та послуг, наприклад відеоконтенту. Тому виконується оптимізація рекомендаційних алгоритмів, щоб надавати найвищі бали тим товарам, які учасники найімовірніше виберуть.

При оптиомізації алгоритмів використовуються конкурси, наприклад Netflix Prize.

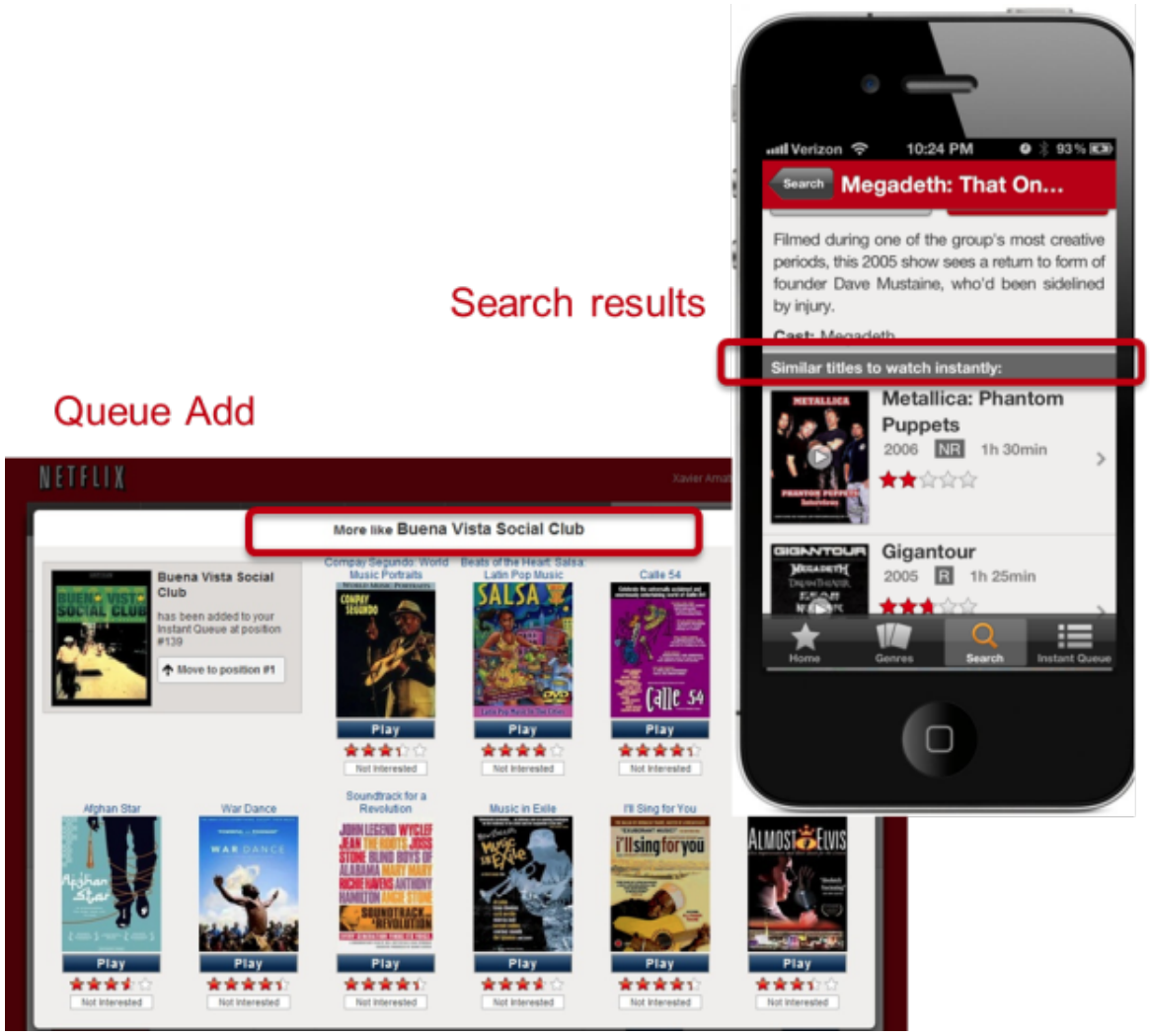

Рисунок 1.6 – Подібність у рекомендаційній системі

У 2006 році було оголошено Netflix Prize, визначний конкурс машинного навчання та аналізу даних для прогнозування рейтингу фільмів. Було запропоновано 1 мільйон доларів тому, хто покращить на 10% точність існуючої рекомендаційної системи під назвою Cinematch. Мета конкурсу полягала в тому, щоб знайти нові шляхи покращення рекомендацій. При проведенні конкурсу були використані проксі-запитання, які було легше оцінити та кількісно визначити: середньоквадратичну помилку (RMSE) прогнозованого рейтингу. Конкурс був спрямований на те, щоб подолати RMSE 0,9525, знизивши його до 0,8572 або менше.

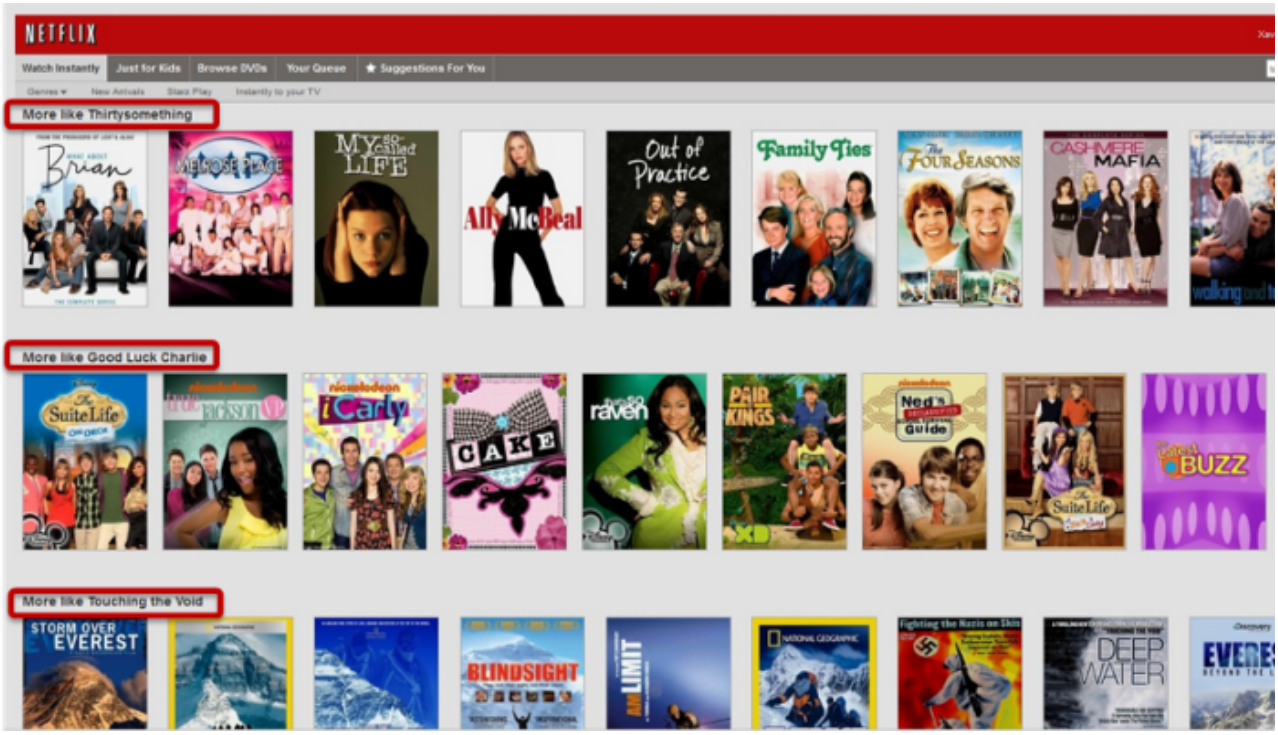

Рисунок 1.7 – Подібність у рекомендаційній системі по рядкам

Через рік змагань команда Korbell виграла першу премію з покращенням на 8,43%. Вони витратили понад 2000 годин роботи, щоб створити остаточну

комбінацію із 107 алгоритмів, яка принесла їм цю нагороду. Було використнао два основні алгоритми з найкращою продуктивністю в ансамблі:

– матрична факторізацію (SVD, сингулярне розкладання);

– обмежена машину Больцмана(RBM).

SVD сам по собі забезпечив 0,8914 RMSE, тоді як один RBM забезпечив конкурентоспроможний, але трохи гірший 0,8990 RMSE.

Лінійна суміш цих двох зменшила похибку до 0,88. Щоб застосувати ці алгоритми, в рекомендаційній системі довелося подолати деякі обмеження Наприклад таке, що вони створені для обробки 100 мільйонів оцінок замість понад 5 мільярдів в реальній системі. Також що вони не створені для адаптації до додавання учасників більше оцінок.

Тому мета Netflix Prize, точне передбачення рейтингу фільму, є лише одним із багатьох компонентів ефективної системи рекомендацій, яка оптимізує задоволення від учасників. Також потрібно брати до уваги такі фактори, як контекст, популярність назви, інтерес, докази, новизна, різноманітність і свіжість.

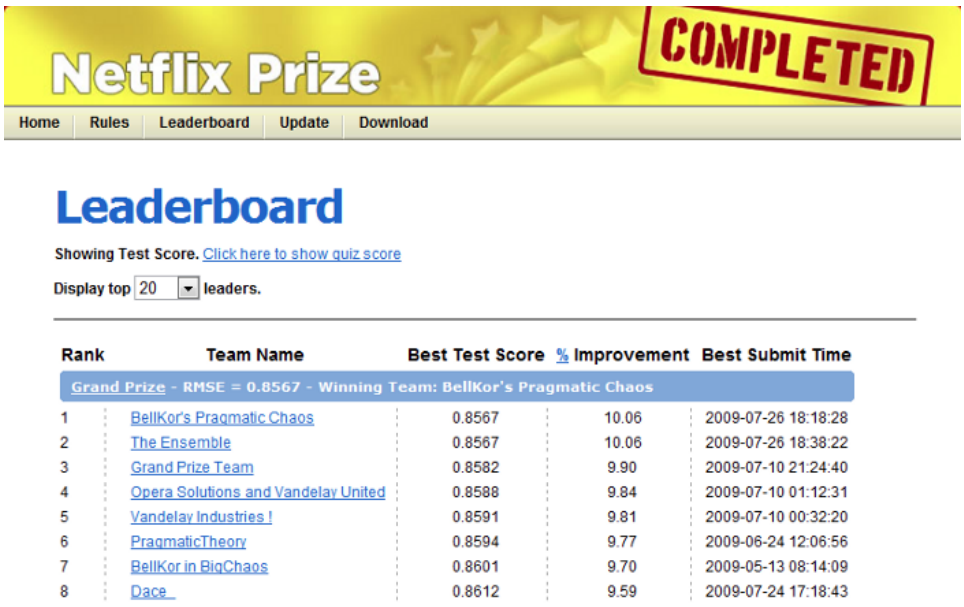

Рисунок 1.8 – Удосконалення алгоритмів побудови рекомендацій на конкурсі Netflix Prize

Підтримка всіх різних контекстів, у яких даються рекомендації, потребує ряду алгоритмів, налаштованих відповідно до потреб цих контекстів.

Метою рекомендаційних систем є надання людині ряду привабливих товарів для вибору. Зазвичай це досягається шляхом вибору деяких предметів і сортування їх у порядку очікуваного задоволення (або користі). Оскільки найпоширенішим способом представлення рекомендованих елементів є певна форма списку, наприклад, різні рядки на Netflix, потрібна відповідна модель рейтингу, яка може використовувати різноманітну інформацію, щоб створити оптимальний рейтинг елементів для кожного наших членів.

При пошуку функції ранжирування, яка оптимізує споживання, очевидною базою є популярність товару. У середньому учасник, швидше за все, дивитиметься те, що дивиться більшість інших. Однак популярність є протилежністю персоналізації: вона створить однакове впорядкування предметів для кожного учасника. Таким чином, мета рекомендаційної системи полягає в тому, щоб знайти персоналізовану функцію рейтингу, яка є кращою за популярність товару, щоб ми могли краще задовольнити учасників із різними смаками (рисунок 1.9).

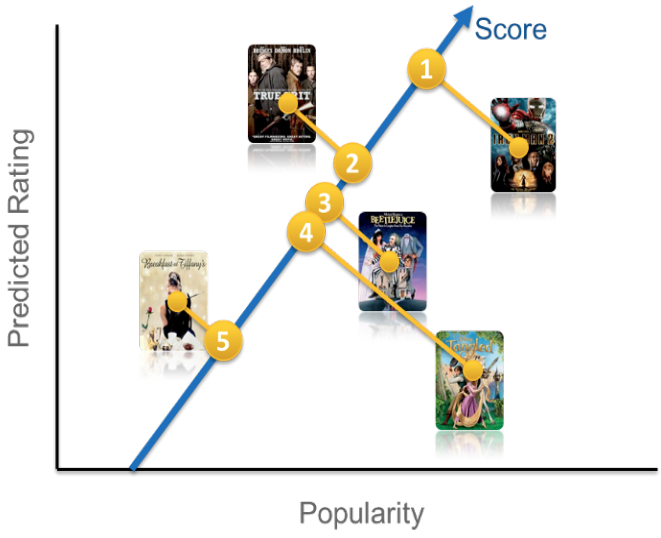

Рисунок 1.9 – Функція ранжування

Одним із очевидних способів підійти до цього є використання передбачуваного учасником рейтингу кожного товару як доповнення до популярності товару. Використання прогнозованих оцінок самостійно як функції ранжирування може призвести до того, що будуть рекомендовані предмети, які є надто нішевими або незнайомими, і може виключити елементи, які учасник хотів би переглянути, навіть якщо вони можуть не оцінити їх високо.

Щоб компенсувати це, замість використання популярності чи прогнозованого рейтингу окремо, доцільно створити рейтинги, які врівноважують обидва ці аспекти.

Є багато способів, якими можна побудувати функцію ранжирування, починаючи від простих методів підрахунку балів, попарних уподобань і оптимізації всього рейтингу. Для ілюстрації почнемо з дуже простого підходу до оцінки, вибравши нашу функцію рейтингу як лінійну комбінацію популярності та прогнозованого рейтингу. Це дає рівняння у вигляді frank(u,v) = w1 p(v) + w2 r(u,v) + b, де u=користувач, v=відео, p=популярність і r=передбачуваний рейтинг. Це рівняння визначає двовимірний простір, як показано нижче.

Коли відома така функція, можна передати набір відео через цю функцію та відсортувати їх у порядку спадання відповідно до оцінки. Тоді виникає питання, у простій двовимірній моделі визначити, чи є популярність більшою чи менш важливою, ніж прогнозований рейтинг?

До цього можна принаймні два підходи:

– вибір ваг;

– з використанням машинного навчання.

Можна вибрати простір можливих ваг і дозволити учасникам вирішити, що має сенс після багатьох тестів A/B. Ця процедура може зайняти багато часу і бути не дуже рентабельною.

Інша можлива відповідь передбачає формулювання цього як проблеми машинного навчання: необхідно вибрит позитивні та негативні приклади з

історичних даних і дозволити алгоритму машинного навчання визначити ваги, які оптимізують мету.

Це сімейство проблем машинного навчання відоме як "Вчимося ранжувати » і займає центральне місце в прикладних сценаріях, таких як пошукові системи або націлювання реклами. Однак важливою відмінністю у випадку ранжованих рекомендацій є важливість персоналізації: не очікується глобального поняття релевантності , а скоріше потрібно знайти шляхи оптимізації персоналізованої моделі.

Приклад покращення рейтингу завдяки додаванню різних функцій і оптимізації алгоритму машинного навчання наведено на рисунку 2.9.

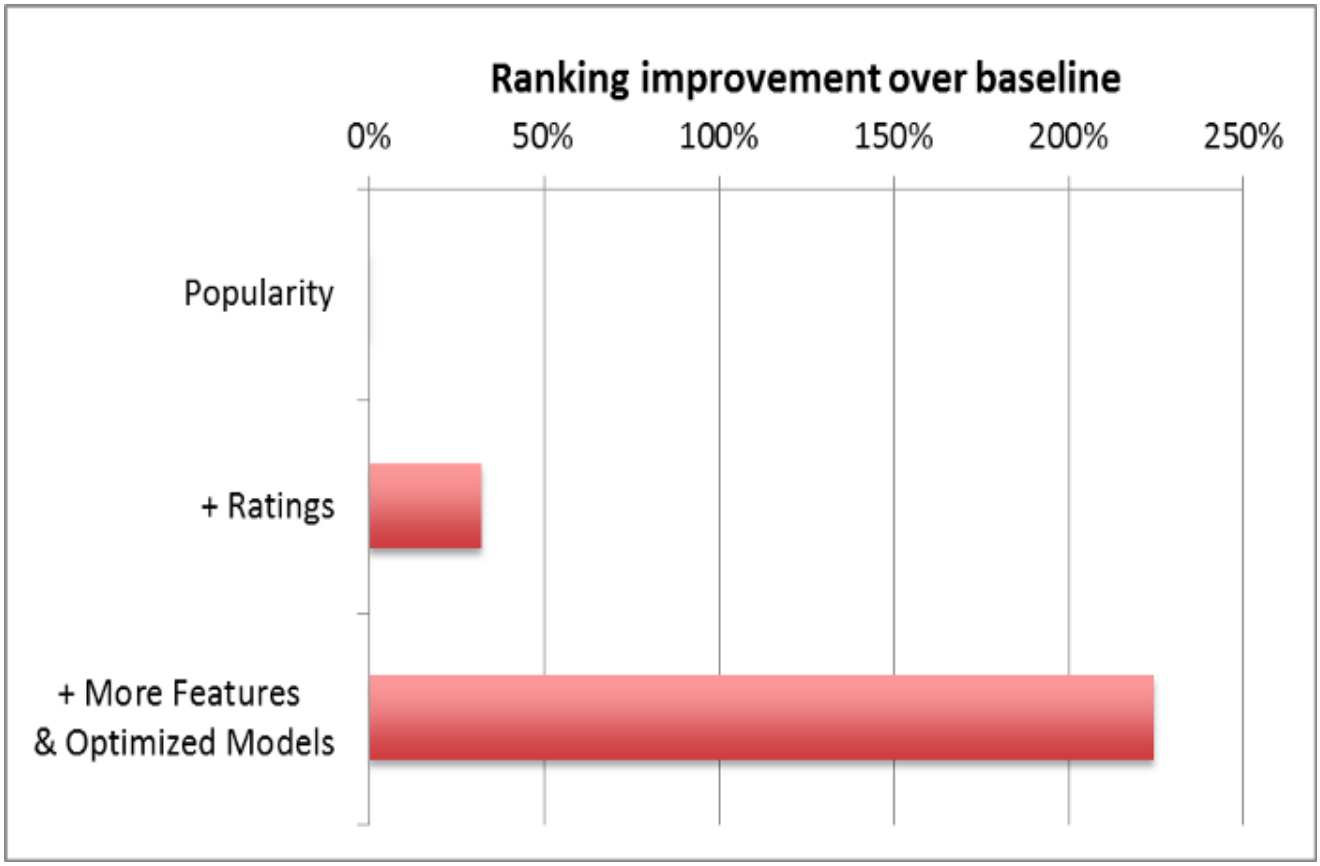

Рисунок 1.10 – Покращення рейтингу при комбінуванні функцій

Багато контрольованих методів класифікації можна використовувати для ранжування.

Типові варіанти включають:

– логістичну регресію;

– опорні векторні машини;

– нейронні мережі;

– методи на основі дерева рішень, такі як Gradient Boosted Decision Trees (GBDT).

За останні роки з’явилася велика кількість алгоритмів, спеціально розроблених для навчання ранжируванню, таких як RankSVM або RankBoost.

Попереднє обговорення алгоритмів ранжирування підкреслює важливість як даних , так і моделей у створенні оптимального персоналізованого досвіду для наших учасників. Джерела даних, які можуть бути використані для оптимізації рекомендацій:

– оцінки товарів від учасників;

– презентації товарів;

– соціальні дані;

– метадані товарів: актори, режисер, жанр, батьківський рейтинг і відгуки.

Є багато способів обчислити популярність товарів. Ми можемо обчислювати його в різних часових діапазонах, наприклад, щогодини, щодня або щотижня. Або ми можемо згрупувати учасників за регіоном або іншими показниками подібності та обчислити популярність у цій групі.

Щодня рекоменадаційна система ми отримує кілька мільйонів потокових відтворень , які включають такий контекст, як тривалість, час доби та тип пристрою. Користувачі додають мільйони товарів до своїх черг .

Кожен елемент у каталозі товарів має багаті метадані : актори, режисер, жанр, батьківський рейтинг і відгуки.

Презентації дають можливість визначити, які предмети рекомендувати, і можна подивитися, як це рішення вплинуло на дії користувача. Також можна

спостерігати за взаємодією учасників із рекомендаціями: прокручування, наведення курсора, клацання або час, проведений на певній сторінці.

Соціальні дані є останнім джерелом функцій персоналізації; можна обробляти те, що дивилися або оцінювали пов’язані друзі.

Рекомендаційні системи музичних релізів допомагають користувачеві знайти музику, найбільш властивих вподобанням користувача. Можливості такої системи грають велику роль як у розширенні музичних вподобань користувача, так і в комерційному просуванні таких сервісів, що позиціюють себе як великі музикальні каталоги з функціоналом щоденного підбору найактуальніших пісень для прослуховування.

Найбільш поширеними приклади успішного комерційного використання рекомендаційних систем музичних релізів є стримінгові сервіси – сайти та застосунки, що надають користувачам можливість необмеженого прослуховування музичних композицій онлайн в потоковому форматі за щомісячною підпискою. Прикладами стримінгових сервісів є такі платформи, як Apple Music, Deezer, Spotify, Tidal, YouTube Music тощо. Такі сервіси часто включають в себе функціонал автоматизованого підбору музичних релізів та генерації плейлистів, індивідуально підібрані під особисті вподобання користувачів-підписників на такі сервіси. Станом на 2021 рік стримінгові сервіси принесли доходу до музичної індустрії на суму в 18,3 мільярди доларів, що складає 64 відсотки доходів в музичній індустрії за цей період.

Також використання рекомендаційних системи музичних релізів можна спостерінати в більш нішевих сервісах для любителів музики, які збирають статистику вподобань для генерації рекомендацій автоматично або з даних, наданих особисто користувачем. Прикладами таких сервісів є сайти Last.fm та RateYourMusic.

На рисунку 1.11 зображена структура процесів типової інформаційної системи, в якій реалізована рекомендаційна система музичних релізів.

Користувач має змогу зареєструватися в системі та почати огляд доступної колекції музичних релізів, через перегляд повного каталогу, або через вибір за авторами та рекомендаціями. Під час прослуховування користувач ставить оцінку музичному релізу в залежності від реалізації в певній інформаційній системі: деякі системи використовують прості метрики на кшталт «подобається-не подобається», інші використовують рейтинги від 1 до 5, або від 0 до 100.

Для підбору рекомендацій в рекомендаційних системах частіше за все використовують такі найпоширеніші методи: метод фільтрації на основі вмісту, метод колаборативної фільтрації та гібридний метод.

Рисунок 1.11 – Структура інформаційної системи, в якій реалізована рекомендаційна система музичних релізів

У методі фільтрації на основі вмісту рекомендації користувачам надаються лише на основі поведінки та даних одного користувача. Опис об'єкта, тобто метадані, що до нього прив’язані з метою опису його характеристик, і профіль конкретного користувача відіграють важливу роль у методі фільтрації на основі контенту. Якщо система рекомендує такі елементи, як веб-сторінки, новини, статті, або харчові заклади, скоріш за все в такій системі використовується метод фільтрації на основі вмісту. Процес рекомендації в цьому методі наступний:

– аналізується опис об'єктів, яким надає перевагу конкретний користувач, щоб визначити переваги, які можуть бути використані для опису цих об'єктів. Ці переваги зберігаються в профілі користувача;

– порівнюється кожен атрибут товару з профілем користувача таким чином, що тільки пов'язані товари, що мають високий ступінь схожості з профілем користувача, будуть рекомендовані цьому конкретному користувачеві.

У методі колаборативної фільтрації рекомендації надаються користувачеві на основі поведінки конкретного користувача та поведінки інших користувачів, пов’язаних між собою схожістю у вподобаннях. У цій техніці рекомендація активному користувачеві видається на основі порівняння уподобань та інтересів інших подібних користувачів, які оцінили схожі об'єкти, що були оцінені цим активним користувачем. Системи колаборативної фільтрації розглядають більш ніж один загальний об'єкт, щоб сформувати набір користувачів, які впливають на результати рекомендацій. Метод колаборативної фільтрації є найбільш поширеним підходом до побудови рекомендаційної системи, і особливо має сенс у використанні на спеціалізованих музикальних платформах, де система рекомендацій поєднується з функціоналом каталогування релізів та соціальною взаємодією між користувачами платформи, наприклад RateYourMusic чи Last.fm.

Гібридний метод фільтрації поєднує в собі два способи надання рекомендацій або пропозицій активним користувачам. Для об'єднання методів існують різні способи, такі як індивідуальна реалізація методів з подальшим

об'єднанням їх результатів, а також використання методу на основі моделей з методом фільтрації на основі пам'яті. Гібридний підхід використовується в сучасних стримінгових сервісах, зокрема Spotify та YouTube включно з YouTube Music, для надання персоналізованих підбірок пісень для прослуховування, що адаптуються під особисту поведінку користувача (наприклад, яку музику якого жанру та яких виконавців він прослуховує щовечора) та співвідносять цю поведінку з користувачами зі схожими вподобаннями (на основі яких формуються рекомендації за схожістю в жанрах або авдиторії, що прослуховує певний набір виконавців).

Через відкриту природу систем рекомендацій для спільної фільтрації вони мають вразливість до атак з боку зловмисників шляхом вкидання профілів, що складаються з упереджених оцінок. Для збільшення зусиль, необхідних для створення профілів, доводиться вводити деякі додаткові заходи, наприклад, код верифікації, який необхідно заповнювати перед кожним виставленням оцінки, або збільшення вартості створення облікового запису користувача. Ці методи зменшують кількість профілів атаки, але також перешкоджають участі користуваців в цих системах, таким чином зменшуючи їхню залученість. Існує ймовірність того, що зловмисник може розпочати атаку, поки в рекомендаційних системах можна виставляти рейтинги.

Також в випадку з музичними рекомендаціями може грати велику роль грошовий стимул, бо при розподілі грошей, отриманих за прослуховування музики певних виконавців, найбільше отримують ті, у кого є найбільше прослуховувань і часто найвищий рейтинг в системі. Існують тіньові індустрії, які пропонують послуги «накрутки» рейтингів та прослуховувань, які вводяться з метою потенційного збільшення «привабливості» музичного релізу в рекомендаційній системі та отримання потенційного органічного росту популярності в рекомендаційній системі. Окремі особи можуть бути зацікавлені в просуванні або зниженні позиції певних елемениів шляхом маніпулювання системою рекомендацій.

Більшість атак можуть бути реалізовані наступним чином. Зловмисник отримує різні облікові записи в системі та створює профіль користувача для кожного облікового запису, які називаються профілями атаки. В рамках кожного зі створених профілів зловмисник маніпулює рекомендаціями, оцінюючи або рекомендуючи певний цільовий елемент. Для того, щоб замаскуватися і виглядати в системі як справжні користувачі, профілі атаки міститимуть рейтинги для нецільових об'єктів. Ці рейтинги можуть бути обрані різними способами, як випадковим чином, так і більш інтелектуально, якщо зловмисник має попередні знання про рейтинги в системі. Зловмисник може маніпулювати системою, щоб домогтися бажаної поведінки рекомендацій.

Нещодавні дослідження показали, що навіть простих атак достатньо, щоб маніпулювати поведінкою найбільш часто використовуваних алгоритмів рекомендацій. Існує кілька небезпек атак в рекомендаційних системах. Атаки можуть завдати різних втрат незахищеним системам в залежності від мети зловмисників. По-перше, це несправедливе представлення користувачів в рекомендаційних системах. По-друге, рекомендаційні системи можуть не видавати належні рекомендації користувачам. Тим самим потенційно може бути зіпсована репутація рекомендаційних систем, що стали жертвами такої атаки. За певних умов багато профілів атак можуть призвести до виводу системи з ладу. Недобросовісним користувачам складно запобігти вкиданню в систему фейкових даних (профілів). Для забезпечення надійності рекомендаційних систем необхідно точно виявляти й видаляти профілі атак.

1.2 Дослідження шиллінг-атак в музичних рекомендаційних системах

Шиллінг-атака – це особливий тип атаки, де профіль користувача-зловмисника вставляється в існуючий набір даних для зміни результатів роботи

рекомендаційної системи. Впроваджені профілі явно оцінюють елементи таким чином, що рейтинг цільового елемента або збільшується, або зменшується. Шиллінг-атаки повністю використовують основну характеристику рекомендаційної системи: рекомендаційні системи повинні дозволяти користувачам взаємодіяти з системою за допомогою різних операцій, таких як виставлення оцінок і перегляд каталогу.

Таким чином, рекомендаційна система отримує достатній зворотній зв'язок від користувачів для навчання своїх моделей і надання рекомендацій. Відкритість рекомендаційних систем до даних, що вводяться користувачами, робить можливим введення фальшивих профілів користувачів для здійснення шиллінг-атак.

Крім того, шиллінг-атаки розглядають рекомендаційну систему жертви як чорний ящик і потребують лише історичних даних про взаємодію користувача з об'єктом (наприклад, рейтинги), які зазвичай доступні з публічних сторінок реальних користувачів у системі.

Спочатку ми введемо термінологію шиллінг-атак:

– мета атаки;

– push-атаки;

– nuke-атаки;

– бюджет атаки;

– розмір атаки;

– розмір профілю;

– знання про атаку.

Мета атаки. Мета атакуючої сторони в шиллінг-атаках може бути досить комплексною. У даній роботі ми розглядаємо переважно цільову атаку, тобто вплив здійснюється на один об'єкт. Це найпоширеніший випадок шиллінг-атак на рекомендаційні системи через серйозні наміри збільшити продажі музичних релізів в умовах жорсткої бізнес-конкуренції.

Push-атаки вказують на необхідність просування одного або декількох цільових елементів, тобто цільові елементи повинні бути рекомендовані моделлю рекомендаційної системи-жертви більше, ніж вони були до атаки.

Nuke-атаки вказують на те, що один або кілька об'єктів повинні бути знижені в рейтингу.

Бюджет атаки. Атаки на системи рекомендацій є дорогими. При розробці практичного методу шиллінг-атаки на рекомендаційну систему ми повинні розглядати бюджет атаки з двох точок зору:

Розмір атаки – кількість підроблених профілів користувачів. Кількість профілів, що вводяться, і кількість елементів, що оцінюються на один профіль, значно впливає на охоплення атаки. Кількість впроваджених профілів, також відома як розмір атаки, повинна бути достатньо великою, щоб мати певний вплив на систему.

Розмір профілю – це кількість ненульових оцінок в одному фальшивому профілі користувача. У деяких сучасних роботах розмір профілю не враховується, та в системах рекомендацій музичних релізів вартість обходу перевірки є відносно невеликою. Однак, реалізація перевірок все ще є доцільною для захисту від низькокваліфікованих атак з невеликою кількістю оцінок, оскільки існують випадки, коли системи експлуатуються за допомогою шиллінг-атак з великою кількістю фальшивих профілів, що дають навіть одиничні оцінки.

Знання про атаку. Обсяг знань, доступних зловмиснику, є критично важливим фактором при розробці методів шиллінг-атаки. В цілому, найбільш бажані знання для покращення результатів атаки пов'язані з відгуками користувачів та моделлю системи рекомендацій жертв.

Відгуки користувачів – це набір даних, який використовується для навчання моделі системи рекомендацій жертв. Зловмисник може мати повні або часткові знання про відгуки користувачів. У цій роботі ми припускаємо, що зловмисник має повну інформацію про відгуки користувачів, тобто зловмисник знає, хто і що оцінює, а також точні значення оцінок. Це розумне припущення,

оскільки оцінки користувачів у рекомендаційних системах зазвичай доступні всім користувачам (наприклад, користувач може бачити оцінки всіх інших користувачів) і, таким чином, є вразливими для зловмисників.

Цільовий елемент вказує на елемент, за допомогою якого зловмисник хоче реалізувати свою зловмисну мету.

Елементи-заповнювачі – це елементи, які мають ненульові рейтинги у введених фейкових профілях користувачів. Заповнювачі в кожному фальшивому профілі користувача, як правило, різні.

Неоцінені предмети – предмети, яким не присвоєно жодних оцінок у введених фальшивих профілях користувачів.

Вибрані предмети – це кілька предметів, обраних людиною для спеціальної обробки. Не всі моделі атак враховують обрані предмети. Однією з можливих причин використання обраних елементів є вплив на користувачів всередині сегмента, тобто на користувачів, які надали перевагу обраним елементам.

Шиллінг-атаки можна розділити на атаки з високим рівнем знань та атаки з низьким рівнем знань. Атаки з низьким рівнем знань є більш практичними та мають більше шансів мати реальний вплив, але ефективність таких атак також низька. З іншого боку, атаки з високим рівнем знань можуть мати значний вплив на продуктивність рекомендаційних систем, але їх важче реалізувати. Оскільки вимога специфічних знань робить атаку з високим рівнем знань малоймовірною, помірно ефективна атака з низьким рівнем знань представляє набагато суттєвішу загрозу для реальних додатків. Залежно від вибору об'єктів атаки існують різні моделі атак, які можна класифікувати як стандартні або замасковані.

На рисунку 1.12 зображені найпоширеніші типи шиллінг-атак.

Стандартні атаки не мають на меті непомітність в системі рекомендацій. Більшість алгоритмів виявлення має вищі шанси виявлення профілів, що вводяться за допомогою цих атак.

Випадкова атака, також відома як атака RandomBot, є найпростішою формою шиллінг-атаки. У цій моделі елементи, оцінені профілем атаки,

обираються випадковим чином, за винятком цільового елемента. Рейтинги цих елементів визначаються навколо середнього значення по системі в цілому. Цільовий об'єкт отримує максимальний або мінімальний рейтинг в залежності від того, чи є це push або nuke-атака.

Деякі з таких атак мають на меті підірвати довіру до системи рекомендацій, так званий "випадковий вандалізм". Бувши найпростішою атакою, вона також є найменш ефективною. Випадкова атака, як правило, більш ефективна в порушенні працездатності рекомендаційної системи, а не в просуванні цільового об'єкта.

Рисунок 1.12 – Типи шиллінг-атак

Простота виконання випадкових атак пояснюється тим, що вони не вимагають особливих знань. Все, що потрібно зловмиснику – це загальне середнє значення системи, яке можна легко обчислити емпіричним шляхом. Бувши найпростішою атакою, вона не дуже ефективна.

Усереднена атака подібна до випадкової атаки з точки зору процесу вибору елементів. Випадково вибрані елементи оцінюються на основі розподілу рейтингу окремих елементів. Кожному елементу-наповненню присвоюється середній рейтинг цього елемента. Ефективність цієї моделі пропорційна знанням зловмисника. Хоча єдина відмінність між випадковою атакою та середньою атакою – це рейтинги-наповнення, ефективність середньої атаки є набагато вищою.

Атака типу "вагон з оркестром" (Bandwagon Attack) – це тип атаки, при якому профілі, створені зловмисниками, наповнюються популярними елементами з високими рейтингами. Цільовому об'єкту присвоюється найвищий рейтинг. Таку атаку можна розділити на випадкову і середньостатистичну, в залежності від схеми оцінювання, яка використовується для товарів-наповнювачів. Атака також належить до категорії атак з низьким рівнем знань, оскільки зловмиснику потрібні лише загальнодоступні дані. Обернена атака типу "вагон з оркестром" (Reverse Bandwagon Attack) – це повна протилежність атаки типу "вагон з оркестром". Ця атака використовується для заподіяння шкоди рейтингу цільовим елементам, надаючи низькі рейтинги елементам з великою кількістю негативних відгуків і надаючи найменший рейтинг цільовому товару. Попри схожість атак, ефективність оберненої атаки дещо вища.

Атака "люблю-ненавиджу" – це високоефективна Nuke-атака. Тут зловмисник випадковим чином обирає елементи-наповнення і надає їм найвищі рейтинги, а цільовому елементу – найнижчі. Попри простоту цієї моделі, її ефективність доволі висока. Хоча вона була переважно розроблена для Nuke-атак, вона також може бути використана для Push-атаки.

Щоб залишитися непоміченими алгоритмами виявлення, зловмисники намагаються замаскувати сліди своїх атак. Багато моделей включають невеликі модифікації стандартних методів атаки для досягнення маскування. Хоча маскування може дещо зменшити вплив атаки, це краще, ніж бути виявленим.

Накладання шуму додає до кожної оцінки випадкове число, розподілене за нормальним розподілом Гауса, помножене на константу, для підмножини профілів, що накладаються.

Ступінь маскування залежить від константи, яка множиться. Вона може бути ефективно застосована до всіх стандартних методів атаки для маскування підпису. Оскільки накладання шуму впливає на схему рейтингу, можна помітити невелике, але помітне падіння ефективності атаки.

Зсув користувачів – це тактика маскування, при якій модифікується підмножина оцінок кожного профілю, що вводиться. Рейтинги цієї підмножини елементів або збільшуються, або зменшуються, щоб зменшити схожість між профілями атаки.

Для різних груп профілів, що вводяться, модифікуються різні підмножини оцінюваних елементів.

Зсув цілі зміщує рейтинг об'єкта-цілі на один рівень нижче, ніж найвищий можливий в Push-атаках.

В ядерних атаках рейтинг цілі зміщується на один рівень вище від найменш можливого рейтингу. Ця стратегія особливо часто використовується для ухилення від методів виявлення, які карають користувачів, що дають екстремальний рейтинг об'єктам.

Якщо цільовий об'єкт вже популярний, його буде важче проштовхнути, використовуючи маскування зі зміщенням цілі. У таких випадках слід використовувати інші методи затушовування.

Атака "середнє серед популярного" використовується для маскування усереднених атак.

Тут елементи-заповнювачі вибираються з найпопулярніших елементів з рівною ймовірністю. Цей метод набагато ефективніший, ніж випадковий вибір з усієї колекції елементів.

Змішана атака здійснюється шляхом одночасного використання декількох стандартних атак в рівних пропорціях.

Для того, щоб техніка виявлення була успішною, вона повинна мати можливість виявляти всі стандартні атаки. Різні методи атаки використовуються для того, щоб підштовхнути/підірвати один і той самий об'єкт. Це допомагає ухилитися від декількох методів виявлення.

Таблиця 1.1 – Опис типів шиллінг-атак, характерних для рекомендаційних систем музичних релізів

Стандартні атаки	
Назва атаки	**Опис атаки**
Усереднена атака	Обирається випадкова вибірка елементів, якій ставиться середня оцінка
Атака «Вагон з оркестром» + обернений варіант	Обирається вибірка найпопулярніших найкраще оцінених елементів, якій ставиться найвища оцінка. В оберненому варіанті обираються найгірше оцінені елементи, яким ставиться найнижча оцінка
Атака «Люблю-ненавиджу»	Обираються випадкові елементи, яким ставиться найвища оцінка, цілі ставиться найнижча оцінка
Замасковані атаки	
Зсув користувачів	Для профілів-нападників робиться зсув оцінок для зниження схожості
Зсув цілі	Ставиться оцінка, трохи вища/нижча за мінімальну/максимальну

Середнє серед популярного	Обфускація усередненої атаки шляхом голосування за найпопулярніші елементи

Модель усередненої атаки вимагає знання середнього рейтингу кожного елемента в системі рекомендацій. Зловмисники оцінюють елементи в наборі заповнювачів випадковим чином, використовуючи нормальний розподіл із середнім значенням, рівним середньому рейтингу оцінюваних елементів наповнювача, і стандартним відхиленням.

Впроваджуючи середню модель атаки, зловмисники маскуються і їх важче відрізнити від справжніх користувачів, а отже, вони мають більший вплив на рекомендації. Як і у випадковій моделі атаки, рейтинги об'єктів атаки встановлюються на рівні максимально або мінімально допустимих залежно від мети атаки. На додаток до моделей випадкової та усередненої атаки, було досліджено декілька більш складних моделей.

Зловмисники вибирають об'єкти, які багато користувачів оцінили як обрані об'єкти, щоб зробити профілі атак схожими на справжні профілі. Ці профілі мають велику ймовірність бути схожими на численні справжні профілі, оскільки об'єкти з високою видимістю оцінені на стільки ж великою кількістю користувачів.

Сегментна атака і групова атака з різними обраними наборами можуть розглядатися як групові атаки. Принцип, що лежить в основі групової атаки, полягає в тому, що найкращий спосіб збільшити економічну вигоду від атаки – це спрямувати свої зусилля на тих, хто вже схильний до вашого продукту. Іншими словами, зловмисника, який бажає просунути певний товар, швидше за все, буде цікавити не те, як часто він рекомендується всім користувачам, а те, як часто він рекомендується ймовірним користувачам.

Модель сегментної атаки призначена для таргетованого просування певного музичного релізу групі користувачів з легко прогнозованими або заздалегідь відомими уподобаннями.

1.3 Дослідження методів виявлення атак користувачів на рейтинги в рекомендаційних системах

Атрибути, які відрізняють шиллінгові профілі від справжніх профілів, вважаються атрибутами виявлення. Атрибути виявлення, які призначені для роботи незалежно від типу моделі атаки, відомі як загальні атрибути. Ефективність цих атрибутів змінюється залежно від різних моделей атак, що використовуються.

Відхилення рейтингу від середньої згоди (RDMA) – це міра відхилення рейтингу користувача за набором цільових елементів відносно інших користувачів у поєднанні з оберненою частотою рейтингу цих елементів.

Зважений показник відхилення від середньої згоди (WDMA) тісно пов'язаний з показником RDMA. Істотною відмінністю цього атрибуту є те, що він надає велику вагу відхиленням у рейтингу для рідкісних позицій. Експериментально встановлено, що WDMA дає більший інформаційний виграш.

Зважений ступінь згоди (WDA) фіксує кумулятивну різницю між оцінкою користувача та середньою оцінкою елемента, поділену на кількість оцінок для цього елемента. WDA емпірично збігається з чисельником RDMA.

Відхилення довжини (LengthVar) вимірює відмінність довжини профілю користувача від середньої довжини профілю. Тут довжина означає кількість елементів, оцінених даним профілем користувача. Деякі профілі атак мають тенденцію мати занадто багато оцінених елементів, що значно відхиляється від середньої довжини профілю користувача.

Проблема з використанням тільки загальних атрибутів полягає в тому, що іноді не вдається відрізнити зловмисні профілі від справжніх користувачів, особливо коли справжній користувач демонструє незвичайну поведінку. Атрибути, специфічні для атак, були розроблені для подолання цих недоліків. Ці атрибути виявлення виявляють розділи в профілях користувачів таким чином, що їх поведінка демонструє схожість з однією конкретною моделлю атаки.

Середня дисперсія (MeanVar) використовується для виявлення середньостатистичних атак. Він розділяє профілі атак на три частини: об'єкти з екстремальними рейтингами (цільові об'єкти), всі інші об'єкти в профілях з рейтингом (об'єкти-заповнювачі) і об'єкти без рейтингу. Цей атрибут працює шляхом обчислення середньої дисперсії між усіма елементами-заповнювачами та загальним середнім значенням. Низька дисперсія вказує на можливість середньої атаки.

Filler Mean Target Difference Model (FMTD) націлена на модель сегментованої атаки. Цей атрибут ґрунтується на різниці між рейтингами елементів у цільовому розділі та елементів у розділі заповнювача.

Filler Average Correlation (FAC) фокусується на виявленні випадкової моделі атаки.

Коли виконується випадкова атака, то рейтинги, що присвоюються елементам, вибираються випадковим чином. Цей атрибут обчислює кореляцію між оцінками в профілі та середніми оцінками елементів. Очікується, що кореляція буде низькою для випадкових атак.

Filler Mean Difference (FMD) використовує той факт, що елементи заповнювача мають середній рейтинг, подібний до загального середнього рейтингу системи в моделі випадкових атак. Якщо середні оцінки подібні, то профіль користувача потенційно може бути профілем випадкової атаки.

Більшість алгоритмів виявлення працюють, орієнтуючись на певну ознаку, яка спостерігається в шиллінг-атаках. Хоча обфускації вдається до певної міри

уникнути виявлення, деякі властиві якості повинні бути присутніми в атаці, щоб вона була ефективною.

На такі якості зазвичай націлені алгоритми виявлення, як в контрольованій класифікації, так і в неконтрольованих методах кластеризації.

Коли мова йде про ознаки, орієнтовані на користувача, основний поділ таких ознак виявлення походить від того, чи зосереджується алгоритм виявлення на пошуку профілів користувачів атаки або на самих об'єктах атаки. У випадку ознак, орієнтованих на користувача, поведінка користувача перевіряється на наявність аномалій, які можуть свідчити про те, що профіль є фейковим.

Для більшості профілів атак характерна схожість профілю користувача з багатьма його сусідами.

Розмір атаки, тобто кількість введених профілів атаки, відносно набагато менша, ніж вся сукупність користувачів. Ця різниця в розмірах, в поєднанні з високою схожістю між ними, виявляється корисним ресурсом для виявлення.

Коли справа доходить до ознак на основі елементів, більшість методів виявлення покладаються на набір елементів, оцінених кожним профілем, щоб перевірити, чи є він фальшивим профілем чи ні. З точки зору виявлення, ми можемо розділити елементи в профілі атаки на дві групи.

Вибрані елементи – це елементи, які використовуються для підтримки push/nuke-атаки цільового профілю. Як вибрані, так і заповнювачі потрапляють в цю категорію з точки зору фронту виявлення. Довжина профілю атаки, тобто кількість елементів, що оцінюються профілем атаки, зазвичай набагато більша, ніж у звичайного профілю.

Зловмисник зазвичай намагається збільшити схожість між профілем атаки і багатьма іншими профілями, оцінюючи кілька елементів-заповнювачів. Рейтинг, присвоєний елементу, підтримується ближче до середнього рейтингу елемента, щоб забезпечити максимальну схожість. Алгоритми виявлення зазвичай націлені на таку аномальну поведінку рейтингу.

Цільовий елемент – це елемент, рейтинг якого підвищується або знижується в результаті атаки. Концентрація користувачів, які оцінюють цільовий елемент, при виконанні атаки є аномально високою. Такі аномалії мають значний вплив на загальний рейтинг елемента.

Основною метою атаки є зміна думки користувачів про об'єкт атаки. Таку думку неможливо змінити, не надавши об'єкту атаки високого рейтингу у випадку push-атаки і мінімально можливого рейтингу у випадку nuke-атаки. Зазвичай, такі рейтинги сильно відхиляються від справжніх оцінок, які надаються об'єкту.

Існуючі методи виявлення шиллінг-атак орієнтовані в основному на формування шаблонів атак з використанням результатів явного зворотного зв'язку. Для побудови таких шаблонів використовуються статистичні методи та машинне навчання. Статистичні методи в першу чергу орієнтовані на виявлення атак першої групи.

Такі методи виявляють ключові характеристики профілю зловмисника, які відрізняють його від середньостатистичного користувача. Наприклад, нехарактерні високі або низькі рейтинги, які не збігаються з оцінками інших користувачів. Для оцінки відповідності використовуються такі показники, як ступінь схожості з найближчими сусідами, відхилення від середнього значення рейтингу з урахуванням кількості оцінок і кількості користувачів, які ці оцінки поставили.

Методи машинного навчання дозволяють розпізнавати атаки з усіх трьох груп, але вони дуже ресурсомісткі.

Існує три категорії алгоритмів виявлення атак: контрольовані, неконтрольовані та напівконтрольовані.

У першій категорії методи виявлення атак моделюються як проблема класифікації. Було проведено багато досліджень для використання керованого навчання для виявлення шиллінг-атак.

Ці керовані алгоритми потребують великої кількості маркованих користувачів для підвищення точності. Методи, засновані на класифікації, вимагають збалансованої кількості атакуючих і нормальних профілів для навчання класифікаторів.

Більшість алгоритмів раннього виявлення використовували сигнатури профілів атак. Ці методи вважалися менш точними, оскільки вони розглядали окремих користувачів та ігнорували сукупний вплив таких зловмисників. Крім того, ці алгоритми погано працюють, коли профілі атак завуальовані. Деякі з цих методів використовують класифікатори найближчих сусідів, методи дерева рішень, класифікатори на основі правил, класифікатори Байєса, класифікатори нейронних мереж або класифікатори на основі SVM.

У другій категорії підходи неконтрольованого виявлення вирішують ці проблеми шляхом навчання на немаркованому наборі даних. Ці методи вимагають набагато менше обчислювальних зусиль у порівнянні з контрольованими підходами. Перевагою цього є те, що ці методи полегшують навчання в режимі онлайн і підвищують точність виявлення. Значний інтерес дослідників зосереджений на виявленні профілів атак з використанням неконтрольованого підходу.

Деякі з методів використовують кластеризацію, методи асоціативних правил і статистичні підходи.

У третій категорії, напівконтрольовані підходи до виявлення використовують як немарковані, так і марковані профілі користувачів для багатокласового моделювання.

Для виявлення складних атак в роботі [2] використовується порівняння оцінок на фіксованих часових інтервалах: спочатку рейтинги сортуються за часом створення рейтингу, потім рейтинги розбиваються на інтервали, виявляються припустимі межі інтервалу, та визначаються підозрілі інтервали, які виходять за припустимі межі.

В роботі [3] пропонується динамічно змінювати тривалість часових інтервалів, на яких буде виявлятися профіль атаки. В цілому підходи першої групи орієнтовані на побудову шаблонів, що описують відмінності між фейковими користувачами та реальними в залежності від типу атаки. Отримані шаблони лише частково враховують зміну інтересів користувачів з часом. В той же час, на практиці пріоритети користувачів динамічно змінюються, і в результаті змінюються моделі їх поведінки і, як наслідок, відмінності з фейковими користувачами, що не дозволяє оперативно виявити атаку на рейтинги.

В роботі [4] при виявленні атак було запропоновано враховувати темпоральні правила [5] для опису динаміки продажів або рейтингів. Темпоральні правила відображають наступні властивості послідовностей подій, які формуються інформаційно-керуючою системою та описують послідовності станів об'єкта управління:

–кожна подія містить інформацію про дію, що виконується з об'єктом управління, а також контекст виконання цієї дії; контекстна інформація представлена атрибутами події;

–атрибути події відповідають атрибутам артефактів, що входять до складу об'єкта контролю;

–кожна з подій, пов'язаних з одним об'єктом контролю, має однаковий фіксований набір атрибутів;

–події представлені в дискретному часі; кожна подія має мітку часу, яка відображає момент її появи;

–події впорядковані в часі відповідно до зміни поведінки об'єкта контролю у вигляді трас;

–існує множина впорядкованих послідовностей подій, що відповідають різним екземплярам кожного об'єкта контролю;

–множина послідовностей подій є "розгорткою" часових знань про його поведінку

Такі правила можуть задавати як неявні обмеження на вибір споживача [6], так і умови цього вибору. Підхід, заснований на порівнянні індивідуальних часових правил покупок і рейтингів, орієнтований в першу чергу на шиллінг-атаки другої групи, оскільки користувачі постійно оцінюють такі об'єкти. Однак під час атак першої групи рейтинги можуть виставлятися нерегулярно. У таких випадках відсутні тимчасові правила виставлення рейтингів на окремих часових інтервалах, що не дозволяє порівняти зміну продажів і рейтингів та виявити атаку.

В роботі [7] пропонується представити динаміку формування рекомендацій за заданий період у вигляді багатошарового часового графа. Однак, така графова модель не забезпечує порівняння процесу відбору та представлення рейтингів.

Для врахування результатів неявного зворотного зв'язку в межах обраного часового періоду доцільно описати процеси вибору продукту та виставлення його рейтингів на декількох послідовних інтервалах, пов'язавши їх часовими правилами.

Таким чином, для виявлення шиллінг-атак на основі виявлення розбіжностей між процесами вибору об'єктів та виставлення оцінок можуть бути використані темпоральні моделі, що враховують динаміку інтересів користувача.

1.4 Постановка задачі дослідження

Існуючі методи виявлення шиллінг-атак використовують дані рейтингів для виявлення неточних профілів користувачів та подальшого захисту від атак, пов’язаних із маніпуляцією рейтингами.

Однак використанню результатів неявного зворотнього зв’язку не приділяється достатньо уваги. В той же час в музичних рекомендаційних

системах порівняння результатів рейтингів (явний зворотній зв’язок) та прослуховування музики (неявний зворотній зв’язок), де є можливість більш точно виявляти атаки, оскільки в таких онлайн-системах фіксуються дії з прослуховування музики користувачем в онлайн-режимі.

Метою даної роботи є дослідження методів виявлення шиллінг-атак в рекомендаційних системах музичних релізів для побудови релевантних рекомендацій в умовах некоректно виставлених рейтингів.

Об’єкт дослідження роботи є шиллінг-атаки, предметом дослідження – методи виявлення шиллінг-атак в рекомендаційних системах музичних релізів.

Задачами дослідження є огляд методів виявлення шиллінг-атак, притаманній предметній галузі, удосконалення методу виявлення шиллінг-атак на основі порівняння змін рейтингів та часу прослуховування, визначення практичної сфери застосування удосконаленого методу та експериментальна перевірка удосконаленого методу виявлення шиллінг-атак в рекомендаційних системах музичних релізів.

2 ВИЯВЛЕННЯ ШИЛЛІНГ-АТАК З ВИКОРИСТАННЯМ ТЕМПОРАЛЬНИХ ЗНАНЬ

2.1 Використання темпоральних правил у процесі виявлення атак користувачів рекомендаційної системи

Зміни вподобань користувачів рекомендаційної системи з тим чи іншим елементом знаходять своє відображення в процесах встановлення її рейтингів. Для опису темпоральної послідовності подій цих процесів запропоновано адаптовані темпоральні правила двох типів: "Next" та "Future". Кожне з цих правил задає порядок у часі для пари фактів Φ_m та Φ_s, що відображають встановлення його рейтингу.

Факт стає істинним, коли відбуваються задані події, такі як вибір заданого об'єкта в заданий час τ; вибір декількох екземплярів елемента на заданій підмножині рейтингів.

Кожне темпоральне правило $r_{m,s}^{(j)}$ визначає відносний темпоральний порядок типу "раніше-пізніше".

Правило $r_{m,s}^{(j)}$ визначає, що після факту Φ_m рейтингу i_j на інтервалі $\Delta\tau_m$, факт Φ_s рейтингу i_j на інтервалі $\Delta\tau_s$, буде істинним.

Таким чином, такі правила можуть задавати темпоральні відношення між інтервалами або моментами часу, а також між підмножинами фактів, впорядкованих у часі.

Правило "Next" використовує темпоральний оператор X, який пов'язує дві послідовні події вибору/оцінки [13].

При виконанні цього правила між фактами Φ_m і Φ_s не може існувати істинних проміжних фактів.

Правило типу "Future" використовує темпоральний оператор F, який з'єднує дві непослідовні події. Між фактами Φ_m і Φ_s в композиції F-правила повинен бути хоча б один проміжний факт.

Узагальнений вигляд правила $r_{m,s}^{(j)}$ має вигляд:

$$r_{m,s}^{(j)} = \Phi_m(X \vee F)\Phi_s. \quad (2.1)$$

Послідовність правил Π_R $r_{m,k}^{(j)}$ описує темпоральну впорядкованість рейтингів об'єкта ij за період T:

$$\Pi_R = \langle r_{1,2}^{(j)}, \dots, r_{1,\mathrm{K}}^{(j)}, \dots, r_{k,\mathrm{k+1}}^{(j)}, \dots, r_{K-1,\mathrm{K}}^{(j)}: \forall \mathrm{k} \Delta\tau_k \in T \rangle. \quad (2.2)$$

Правило виразу $r_{1,2}^{(j)}$ містить темпоральний оператор X, оскільки пов'язує факти Φ_1 та Φ_2 присвоєння рейтингів на двох суміжних інтервалах $\Delta\tau_1$ та $\Delta\tau_2$.

Залежність $r_{1,3}^{(j)}$ є прикладом правила з темпоральним оператором F, що зв'язує факти Φ_1 та Φ_3.

Крім перевизначення фактів, адаптація темпоральних правил полягає у встановленні їх ваг з урахуванням динаміки інтересів користувача.

Вага $w_{m,s}^{(j)}$ правила $r_{m,s}^{(j)}$ задається через нормовану різницю між кількістю покупок або середнім значенням рейтингів продукту i_j на інтервалах $\Delta\tau_m$ і $\Delta\tau_s$.

Послідовність встановлення рейтингів музичних релізів представлена впорядкованою множиною нормованих ваг правил Π_W:

$$\Pi_W = \langle w_{1,2}, \dots, w_{1,S}, \dots, w_{m,m+1}, \dots, w_{S-1,S}: \forall m \forall s r_{m,s}^{(j)} = \text{true} \rangle. \quad (2.3)$$

На основі послідовності ваг для будь-якого часового інтервалу $\Delta\tau_s$ можна оцінити $W_s^{(j)}$, зміну інтересу користувача до теми i_j з часом.

2.2 Вдосконалений метод виявлення шиллінг-атак з використанням темпоральних правил

Загальну схему удосконаленого методу представлено на рисунку 2.1.

Метод передбачає порівняння зміну у рейтингах та при прослуховування музичних релізів.

За результатами порівняння скорочується перелік атакуючих користувачів, що дає змогу визначити атаку.

Рисунок 2.1 – Узагальнена послідовність етапів методу

Метод включає наступні етапи.

ЕТАП 1. Встановлення залежностей щодо збільшення/зменшення часу прослуховування заданого музичного релізу i_j.

Крок 1.1. Відбір вхідних даних про час прослуховування для заданого музичного релізу. На даному етапі виконується фільтрація помилок у вхідних даних.

Крок 1.2. Розбивка вхідних даних та підсумування часу прослуховування згідно заданого рівня деталізації часу.

На даному кроці перелік музичних релізів розбивається на підмножини, що відображають операції в заданому інтервалі часу $\Delta\tau_s$, з метою побудови фактів присвоєння рейтингів на зазначених інтервалах. Ці факти містять в собі інформацію про середню оцінку даного елементу на інтервалах для користувачів.

Крок 1.3. Визначення змін у часі прослуховування музичного релізу $\Delta m_{2,s}$ (збільшення відображено як true, зменшення відображено як false) для пари послідовних інтервалів часу.

ЕТАП 2. Встановлення залежностей по збільшенню/зменшенню рейтингу заданого музичного релізу i_j.

Крок 2.1. Відбір вхідних даних про рейтинги для заданого музичного релізу i_j. На даному етапі виконується фільтрація помилок, знайдених у вхідних даних.

Крок 2.2. Розбивка вхідних даних та сумування рейтингів згідно заданого рівня деталізації часу.

Крок 2.3. Визначення змін у рейтингу музичного релізу $\Delta q_{2,k}$ (збільшення відбражено як true, зменшення відображено як false) для пари послідовних інтервалів часу.

ЕТАП 3. Формування переліку потенційних атакуючих на основі порівняння змін часу прослуховування музичних релізів та рейтингів.

Порівняння виконується згідно виразу:

$$r_k = \Delta m_{2,k} \oplus \Delta q_{2,k}. \quad (2.4)$$

До множини потенційних атакуючих належать користувачі, для яких виконується умова:

$$B = \left\{ u_s : u_s \in p_{1,k} \wedge r_k = true \right\}, \quad (2.5)$$

де u_s – користувач;

$p_{1,s}$ – запис про рейтинг музичного релізу;

ЕТАП 4. Формування уточненої підмножини B потенційних профілів атакуючих користувачів із множини A:

$$B^* = \left\{ u_s : \left(u_s \in B \right) \wedge \left(\forall k \, u_s \notin m_{1,k} \right) \right\}, \quad (2.6)$$

де $m_{1,s}$ – s-й запис про прослуховування музичного релізу.

Крок 4.1. Формування множини користувачів, що ставили рейтинги, але мають підозрілі патерни прослуховування певних музичних релізів..

Крок 4.2. Уточнення рейтингів, що є найбільш підозрілими серед множини користувачів та відсіювання найменш підозрілих рейтингів.

Можливості виявлення шиллінг-атак за допомогою зазначеного методу наведені на таблиці 2.1.

Удосконалення методу полягає у формуванні підмножини користувачів на першому кроці та уточненню рейтингів на другому кроці етапу 4.

Це дає можливість запускати метод ітеративно, щоб уточнити перелік фальшивих користувачів, що є потенційно атакуючими користувачами.

Даний метод може бути використаний для підтримки побудови рекомендацій в режимі онлайн.

Таблиця 2.1 – Можливості виявлення шиллінг-атак за допомогою вдосконаленого методу виявлення шиллінг-атак з використанням темпоральних правил

Тип атаки	Можливості виявлення атаки	Коментар
Випадкова	+	Невідповідність змін у рейтингах та часі прослуховування
Усереднена	–	Внаслідок усереднення рейтингу
На популярні музичні релізи	+–	«–» У випадку штучного зниження рейтингу популярних музичних релізів
З використанням популярних музичних релізів	–	При внесенні до популярних релізів рейтинг музичного релізу росте
Атака «Вагон з оркестром»	+	Невідповідність змін у рейтингах та часі прослуховування
Зворотня атака «Вагон з оркестром»	+	Невідповідність змін у рейтингах та часі прослуховування
З зондуванням додаткової інформації	–	Зміни рейтингу в цілому корелюються з часом прослуховування

3 ТЕХНОЛОГІЯ ВИЯВЛЕННЯ АТАК КОРИСТУВАЧІВ НА РЕЙТИНГИ З ВИКОРИСТАННЯМ ТЕМПОРАЛЬНИХ ЗНАНЬ

Основне визначення інформаційних технологій полягає в тому, що це застосування технологій для вирішення ділових або організаційних проблем у широкому масштабі. Незалежно від ролі, співробітник IT-відділу працює з іншими над вирішенням технологічних проблем, як великих, так і малих.

IT-відділ має три основних принципи відповідальності:

– управління IT;

– IT-операції;

– апаратне забезпечення та інфраструктура;

Управління IT стосується поєднання політик і процесів, які забезпечують ефективну роботу IT-систем і їх відповідність потребам організації.

IT-операції- це загальна категорія для щоденної роботи IT-відділу. Це включає надання технічної підтримки, обслуговування мережі, тестування безпеки та обов’язки з керування пристроєм.

Апаратне забезпечення та інфраструктура стосується всіх фізичних компонентів IT-інфраструктури. Цей аспект IT включає налаштування та технічне обслуговування такого обладнання, як маршрутизатори, сервери, телефонні системи та окремих пристроїв, таких як ноутбуки.

Робота більшості організацій сповільнилася б без функціональних IT-систем. Важко знайти компанію, яка хоча б частково не покладається на комп’ютери та мережі, які їх з’єднують. Підтримання стандартного рівня обслуговування, безпеки та зв’язку є величезним завданням, але це не єдиний пріоритет чи потенційна проблема, яка стоїть на їхніх тарілках.

Ключові задачі в сфері інформаційних технологій, пов'язані із перевантаженням даних, використанням бездротового зв'язку, хмарними

послугами, пропускною здатністю систем, а також атаками на інформаційні системи.

Перевантаження даних: підприємствам потрібно обробляти величезні обсяги даних. Це вимагає великої потужності процесора, складного програмного забезпечення та людських аналітичних навичок.

Використання мобільного та бездротового зв’язку: все більше роботодавців пропонують варіанти віддаленої роботи, для яких потрібні смартфони, планшети та ноутбуки з бездротовими точками доступу та можливістю роумінгу.

Хмарні послуги: більшість компаній більше не використовують власні «серверні ферми» для зберігання величезних обсягів даних. Зараз багато компаній працюють із хмарними службами — сторонніми хостинговими платформами, які зберігають ці дані.

Пропускна здатність для відеохостингу: рішення для відеоконференцій стають дедалі популярнішими, тому для їх належної підтримки потрібна більша пропускна здатність мережі.

Технологію виявлення атак користувачів за допомогою темпоральних знань наведено на рисунку 3.1. Технологія включає наступні кроки.

Етап 2. Введення вхідних даних.

На цьому кроці необхідно ввести дані про рейтинг Q, період T, суб'єкта ij; користувачів U та рівні деталізації часу.

Гранулярність часу залежить від формату міток часу в журналах рейтингів.

Етап 2. Розбиття періоду T на часові інтервали $\Delta\tau s$.

На цьому кроці вибирається рівень деталізації часу, який забезпечує найбільше значення ваг правил процесу продажу або всіх інтервалів з періоду T. Результатом цього кроку є набір інтервалів $\Delta\tau s$ для періоду T.

Етап 3. Реалізація розробленого методу виявлення шиллінг-атак.

Етап 4. Уточнення списку користувачів.

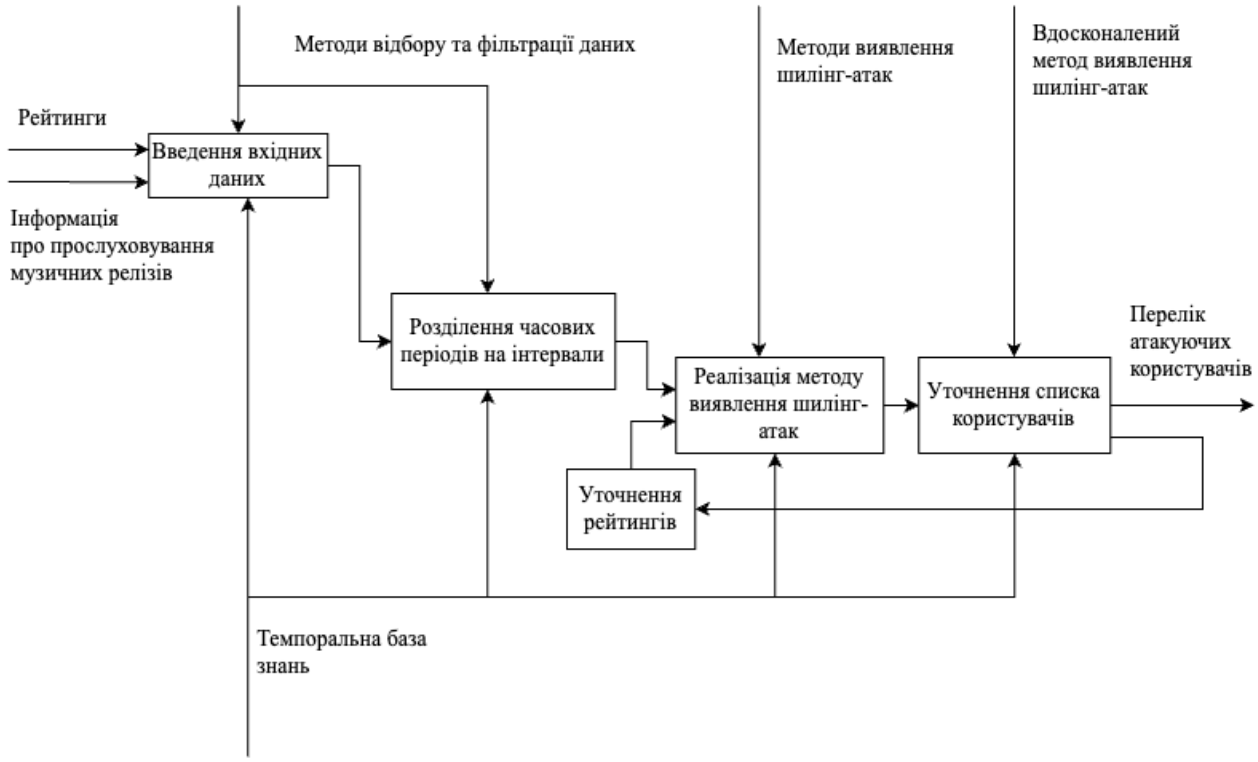

Рисунок 3.1 – Технологія виявлення шиллінг-атак

З множини U на n ітераціях видаляються користувачі, які не виставляли рейтинги на інтервалах Δτs зі знаком, оскільки ці користувачі не брали участі у фальсифікації оцінок. Результатом кроку є підмножина користувачів U(n), що містить потенційних зловмисників.

Проводиться перевірка умови завершення роботи алгоритму, згідно з якою кількість користувачів на поточній n-ній ітерації не змінилася порівняно з ітерацією n-1. При виконанні цієї умова виконується, робота алгоритму закінчується.

В іншому випадку виконується крок 5.

Етап 5. Уточнення множини Q рейтингів.

З цієї множини видаляються рейтинги користувачів, оскільки ці користувачі не є зловмисниками. Далі виконується перехід до виконання методу на кроці 3 алгоритму.

4 ЕКСПЕРИМЕНТАЛЬНА ПЕРЕВІРКА ОТРИМАНИХ РЕЗУЛЬТАТІВ

Експериментальна перевірка проведена на впорядкованому за часом наборі даних про рейтинги та продаж без інформації про абсолютні значення моментів часу.

Перевірено, що з використанні відносної тимчасової шкали і з урахуванням використання F-правил можна виділяти інтервали за кількістю покупок товарів чи послуг. У ряді робіт була показана можливість формувати опис процесу шляхом упорядкування подій на часовій шкалі «раніше-пізніше» з урахуванням їх атрибутів, що дозволяє обґрунтувати дане припущення.

Експеримент включав дві фази.

На першій фазі виконано детектування атак на вхідному наборі даних і аналіз способів приховування атаки. На другій фазі виконано порівняння ефективності запропонованого методу з існуючими методами які, аналогічно запропонованому методу, реалізують розбиття набору вихідних даних за часовими інтервалами.

Під час проведення експерименту використали набір даних із інформацією про читанні і рейтингах кількох мільйонів книг.

Записи про читання та рейтинги впорядковані за часом, проте мітки з абсолютними значеннями часу у вихідних даних відсутні.

Факти та формувалися на підмножинах вхідних даних із фіксованої кількості записів.

На першій фазі розглядалися можливості виявлення атак у межах тривалих періодів часу, представлених великою кількістю покупок. При побудові фактів використовувалися підмножини з 100 000 послідовних елементів (покупок, рейтингів). Такі підмножини відповідають вихідним інтервалам методу. Т

Результати ключових етапів для однієї з книг представлені в таблиці 4.1.

Таблиця 4.1 – Результати експерименту

Номер етапу,	Результати					
	Δ_2	Δ_3	Δ_4	Δ_5	Δ_6	Δ_{10}
4.1	0,19	0,19	-0,41	0,04	-0,63	-0,59
4.2	-0,10	-0,10	-0,30	**-0,30**	**-0,30**	1,00
5.1	0,28	0,28	0,00	0,33	0,00	1,59
5.2	-1	-1	0	0	0	1

Показник відображає загальне падіння продажів на Δ4 порівняно з Δ1, Δ2 та Δ3. Ознака має негативне значення і тому вказує на можливу атаку, спрямовану зниження продажів.

відхилення і має менше, ніж , то ймовірна атака відбулася на інтервалі Δ2.

Розбіжність вказує на можливу атаку на інтервалах від Δ2 до Δ10. Оскільки це значення значно більше, ніж і , то ймовірна атака відбулася на інтервалі Δ10. Значення свідчить про можливу шиллінг-атаку на підвищення рейтингу.

Було створено прототип інформаційної системи, в якій було реалізовано метод виявлення шиллінг-атак. На рисунку 4.1 та 4.2 зображений прототип сторінки музичного релізу.

Рисунок 4.1 – Сторінка музичного релізу, постраждалого від шиллінг-атаки

Рисунок 4.2 – Сторінка музичного релізу після видалення фейкових оглядів, виявлених за допомогою алгоритму виявлення шиллінг-атак

Сторінка містить назву та тип релізу, посилання на виконавця, оцінку від 0 до 100, дані про жанри та певні музичні характеристики релізу, а також мову виконання.

На рисунку 4.3 зображена форма, в якій користувач може поставити оцінку та написати більш детальний огляд релізу.

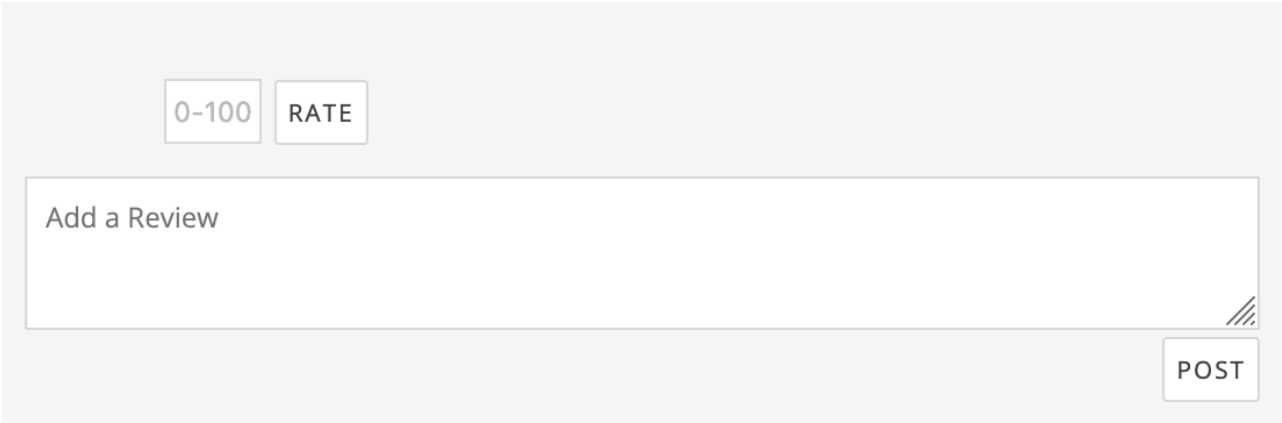

Рисунок 4.3 – Форма надання рейтингу користувача

На рисунку 4.4 зображена форма авторизації користувача.

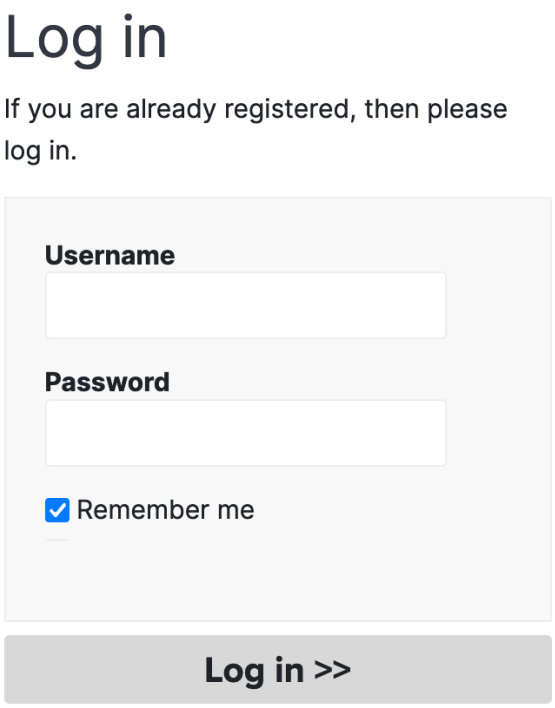

Рисунок 4.4 – Форма авторизації користувача

На рисунку 4.5 зображена форма реєстрації користувача.

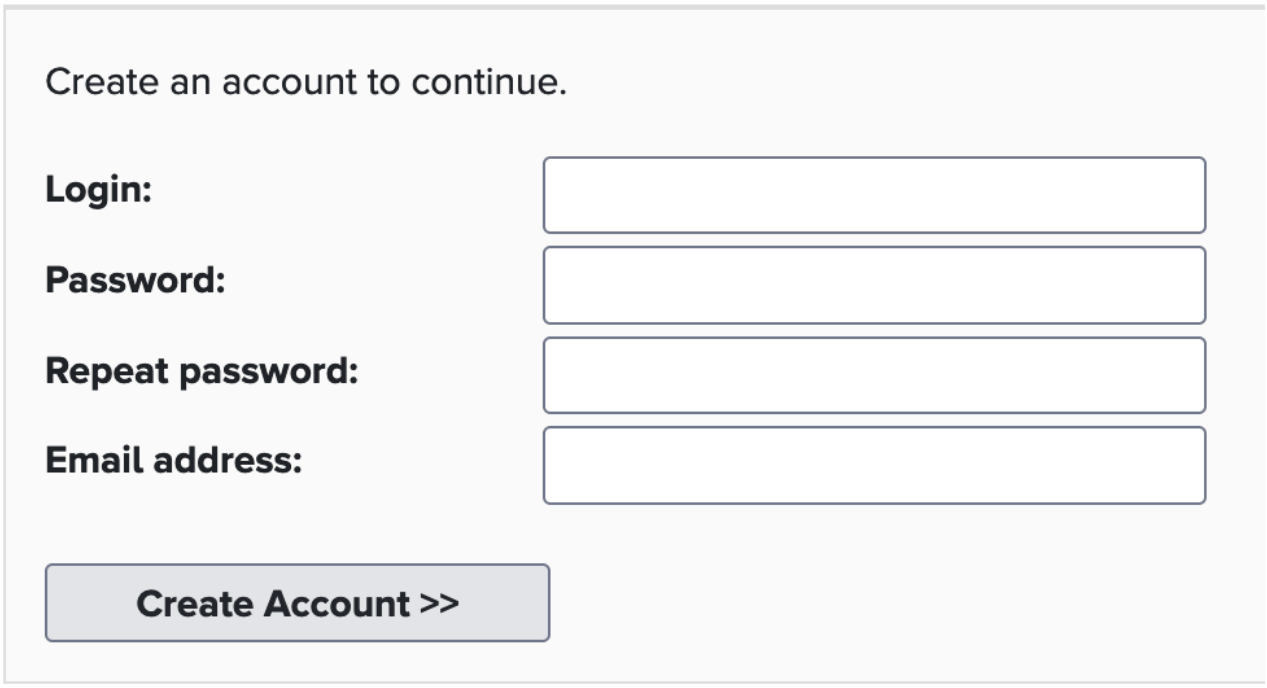

Рисунок 4.5 – Форма реєстрації користувача

На рисунку 4.6 зображено приклад меню вибору музичних релізів за жанром.

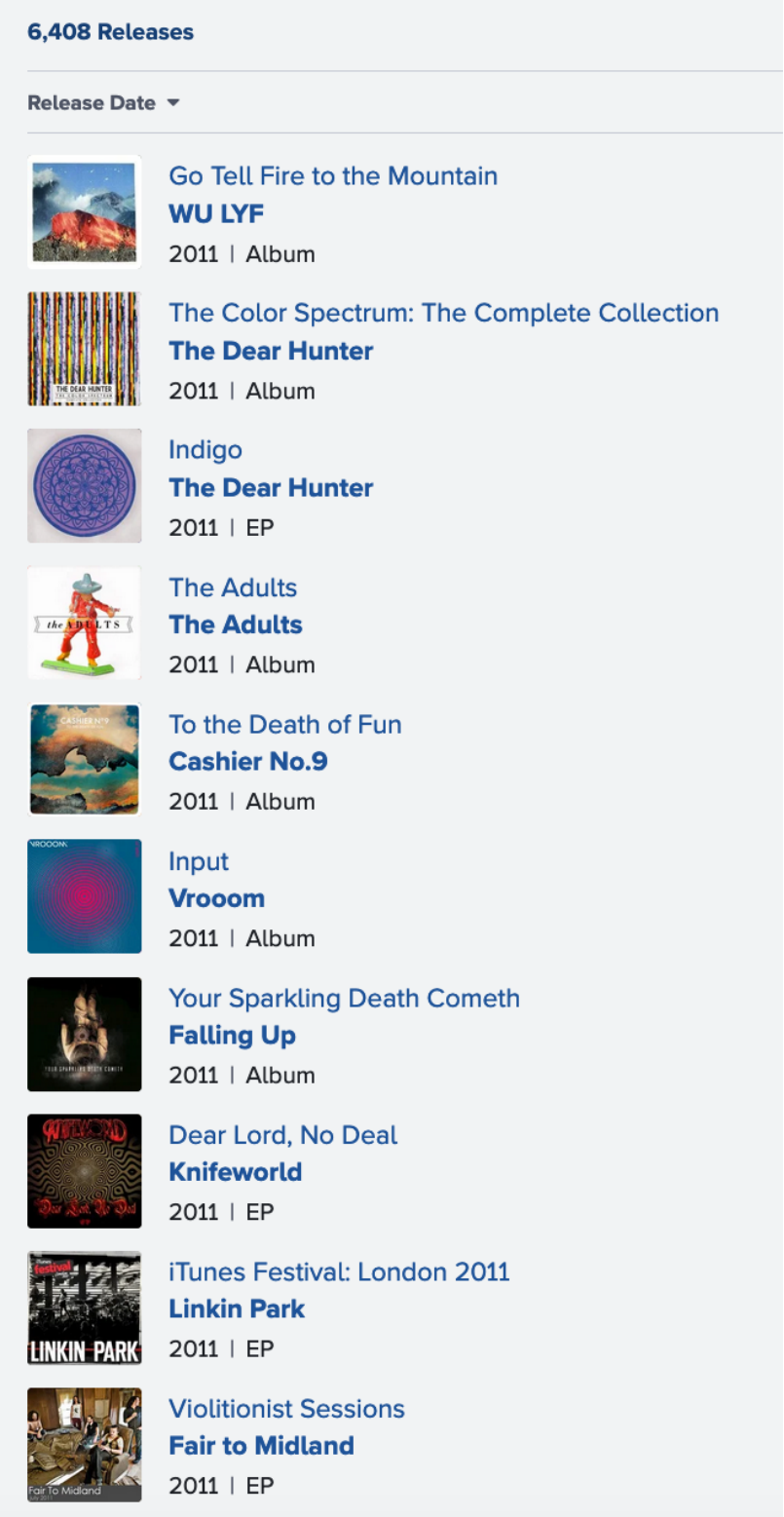

Рисунок 4.6 – Меню вибору музичних релізів за жанром

На рисунку 4.7 зображена сторінка-рекомендатор, яка надає користувачеві актуальні для користувача рекомендації певних музичних релізів.

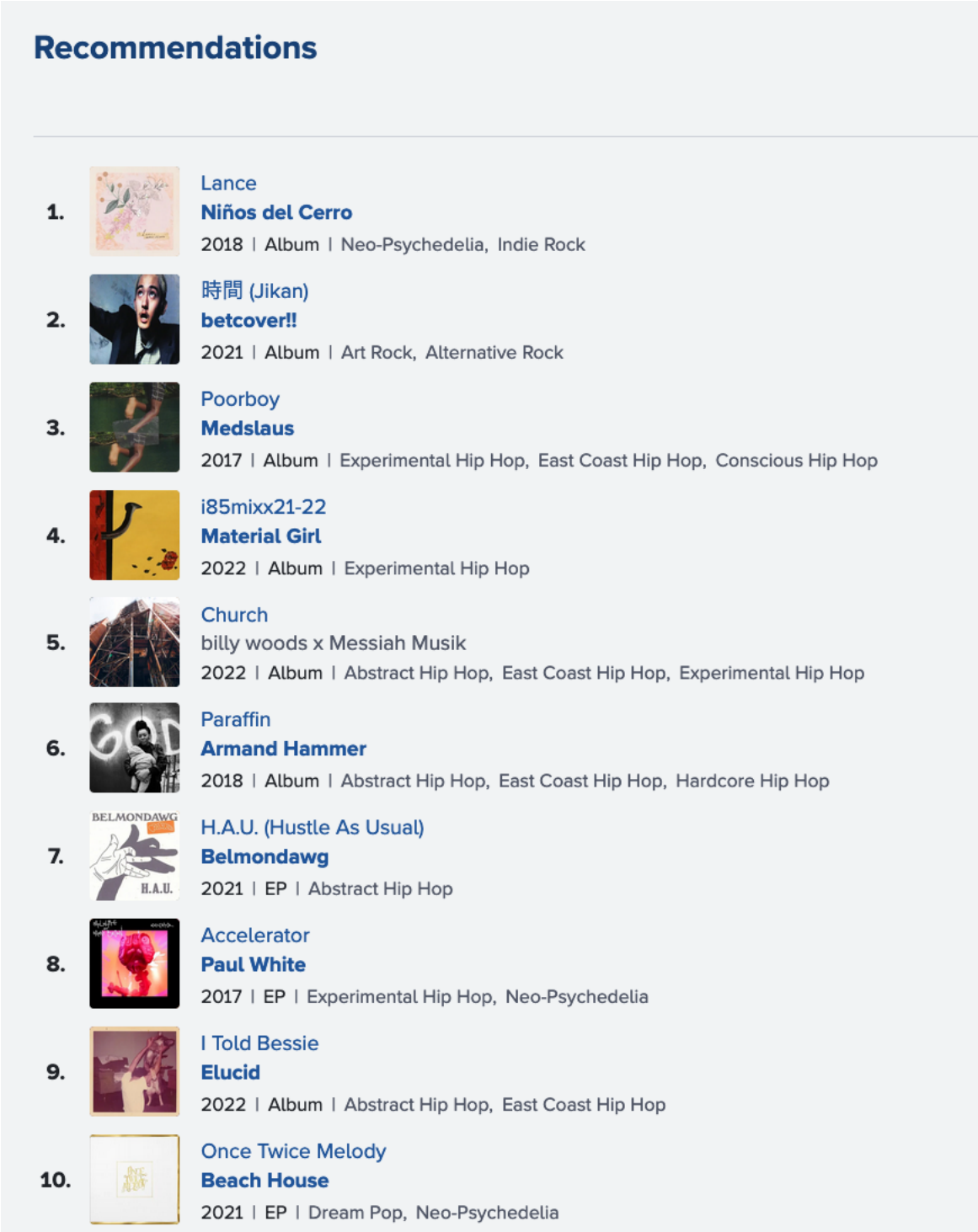

Рисунок 4.7 – Сторінка-рекомендатор

На рисунку 4.8 зображено сторінку вибору популярних на даний момент музичних релізів.

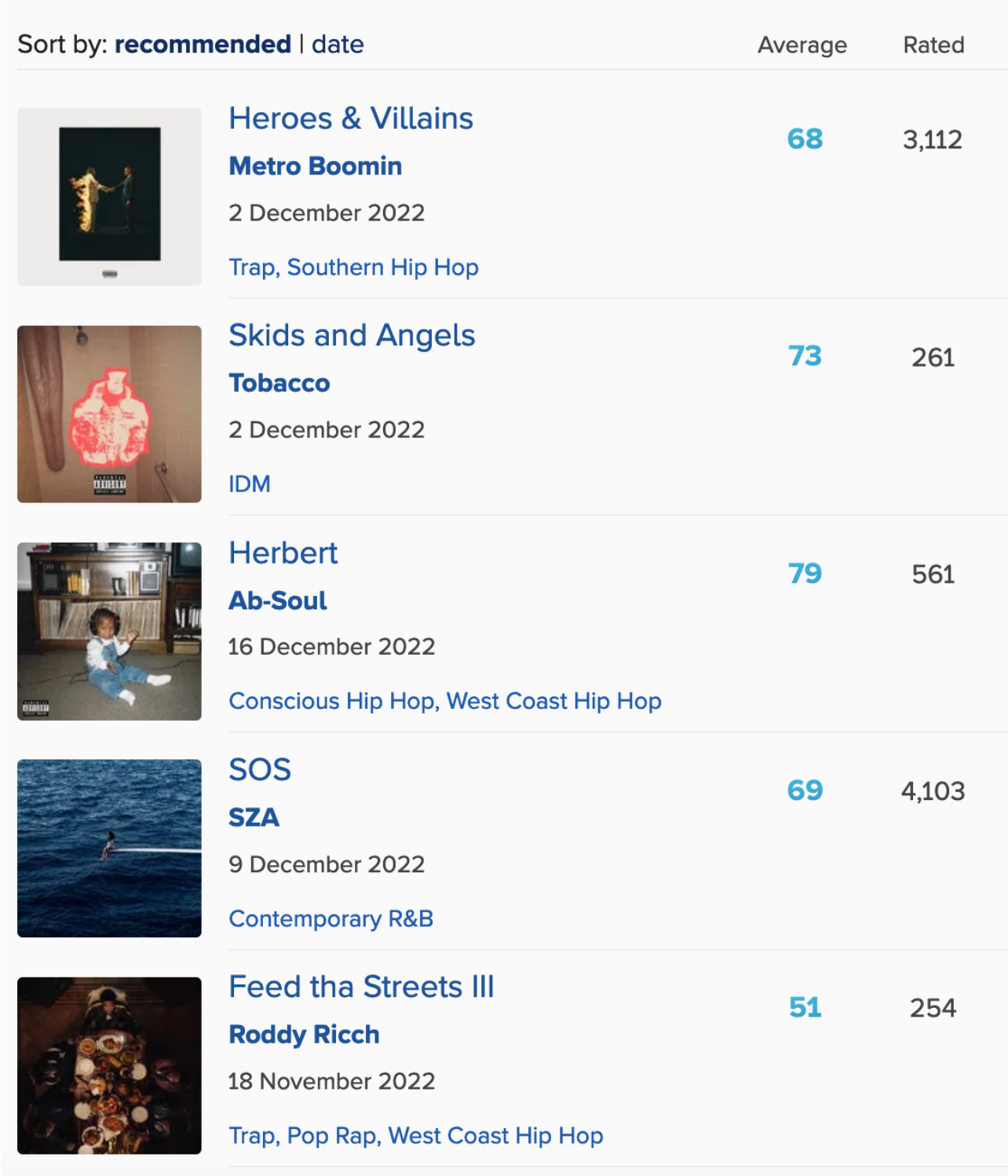

Рисунок 4.8 – Сторінка вибору популярних музичних релізів

ВИСНОВКИ

Була розглянута проблема виявлення шиллінг-атак з урахуванням як явного, так і неявного зворотного зв’язку.

Удосконалено метод виявлення шиллінг-атак шляхом порівняння змін у рейтингах та часу прослуховування на послідовних інтервалах часу.

Обґрунтовано сферу застосування методу для виявлення атак таких типів.

Виконано експериментальну перевірку методу та практичу сферу застосування удосконаленого методу виявлення шиллінг-атак в рекомендаційних системах музичних релізів.

СПИСОК ДЖЕРЕЛ ПОСИЛАННЯ

1. Sundar A. P., Li F., Zou X., Gao T., Russomanno E. D., Understanding Shilling Attacks and Their Detection Traits: A Comprehensive Survey. IEEE Access, 8, 171703-171715, 2020, doi: https://doi.org/10.1109/ACCESS.2020.3022962

2. Gao, M., Yuan, Q., Ling, B., Xiong, Q. (2014). Detection of Abnormal Item Based on Time Intervals for Recommender Systems. The Scientific World Journal, 2014, 1–8. doi: https://doi.org/10.1155/2014/845897

3. Gao, M., Tian, R., Wen, J., Xiong, Q., Ling, B., Yang, L. (2015). Item Anomaly Detection Based on Dynamic Partition for Time Series in Recommender Systems. PLOS ONE, 10 (8), e0135155. doi: https://doi.org/10.1371/journal.pone.0135155

4. Chala, O., Novikova, L., Chernyshova, L. (2019). Method for detecting shilling attacks in e-commerce systems using weighted temporal rules. EUREKA: Physics and Engineering, 5, 29–36. doi: https://doi.org/10.21303/2461-4262.2019.00983

5. Levykin, V., Chala, O. (2018). Method of determining weights of temporal rules in Markov logic network for building knowledge base in information control systems. EUREKA: Physics and Engineering, 5, 3–10. doi: https://doi.org/10.21303/2461-4262.2018.00713

6. Chalyi, S., Leshchynskyi, V., Leshchynska, I. (2019). Method of forming recommendations using temporal constraints in a situation of cyclic cold start of the recommender system. EUREKA: Physics and Engineering, 4, 34–40. doi: https://doi.org/10.21303/2461-4262.2019.00952

7. Chalyi, S., Pribylnova, I. (2019). The method of constructing recommendations online on the temporal dynamics of user interests using multilayer graph. EUREKA: Physics and Engineering, 3, 13–19. doi: https://doi.org/10.21303/2461-4262.2019.00894

8. Чалий С.Ф., Лещинський В.О., Лещинська І.О. Моделювання контексту в рекомендаційних системах. Науковий журнал «Проблеми інформаційних технологій», 2018, №. 1(023). С. 21-26.

9. Chalyi S., Bogatov Ie. Method of constructing an attribute description of the business process "as is" in the process approach to enterprise management EUREKA: Physics and Engineering. 2018. Vol. 6. P. 35-40.

10. Chalyi S., Pribylnova I. The method of constructing recommendations online on the temporal dynamics of user interests using multilayer graph. EUREKA: Physics and Engineering. 2019. Vol. 3. P. 13-19. **(SCOPUS)**

11. Чалий С.Ф., Прібильнова І.Б. Ситуаційна модель користувацького вибору в рекомендаційні системі.// Системи управління, навігації та зв'язку. 2019. Вип. 2(54). С.159-163. doi:https://doi.org/10.26906/SUNZ.2019.2.159.

12. Chalyi, S., Leshchynskyi, V., Leshchynska, I. (2019). Method of forming recommendations using temporal constraints in a situation of cyclic cold start of the recommender system. EUREKA: Physics and Engineering, 4, 34–40. doi: https://doi.org/10.21303/2461-4262.2019.00952

13. Chalyi S., Leshchynskyi V., Leshchynska I. Designing explanations in the recommender systems based on the principle of a black box. Сучасні інформаційні системи. 2019. Т. 3, № 2 С. 47-51.

14. Chalyi S., Leshchynskyi V. Knowledge Representation in the Recommendation System Based on the White Box Principle Сучасні інформаційні системи. 2019. Т. 3, № 3. С 82-86.

15. Чалий С.Ф., Лещинський В.О., Лещинська І.О. Моделювання пояснень щодо рекомендованого переліку об'єктів з урахуванням темпорального аспекту вибору користувача. Системи управління, навігації та зв'язку, 2019. Том 6 № 58. С. 97-101.

16. Чалий С.Ф., Лещинський В.О., Лещинська І.О. Концепція формування пояснень в рекомендаційних системах за принципом білого ящику. Системи

управління, навігації та зв’язку. Збірник наукових праць. Полтава: ПНТУ, 2019. Т. 3 (55). С. 156-160. doi:https://doi.org/10.26906/SUNZ.2019.3.156.

17. Чалий С.Ф., Лещинський В.О., Лещинська І.О. Доповнення вхідних даних рекомендаційної системи в ситуації циклічного холодного старту з використанням темпоральних обмежень типу «NEXT». Системи управління, навігації та зв'язку, 2019. Вип. 4(56). С. 105-109.

18. Chalyi S., Leshchynskyi V., Leshchynska I. Detailing explanations in the recommender system based on matching temporal knowledge. Eastern-European Journal of Eenterprise Technologies , 2020, Vol 4, No 2 (106). P. 6-13. **(SCOPUS)**

19. Shilling attack detection utilizing semi-supervised learning method for collaborative recommender system / J.Cao, Z. Wu, M. Bo, Z. Yanchun. // World Wide Web. – 2013. – №16.

20. Bilge A. A Novel Shilling Attack Detection Method / A. Bilge, Z. Ozdemir, H. Polat. // Procedia Computer Science. – 2014. – №31. – С. 165–174.

21. Методичні вказівки щодо розробки та оформлення кваліфікаційної роботи (для студентів усіх форм навчання другого (магістерського) рівня програми"Інформаційні управляючі системи та технології) / Упоряд.:Петров К.Е., Левикін В.М., Чалий С.Ф., Євланов М.В., Саєнко В.І., Міхнов Д.К., Міхнова А.В., Чала О.В. - Харків: ХНУРЕ,2021.- 30с.