Міністерство освіти і науки України
Харківський національний університет радіоелектроніки

Факультет Комп'ютерних наук
(повна назва)
Кафедра Штучного інтелекту
(повна назва)

КВАЛІФІКАЦІЙНА РОБОТА
Пояснювальна записка

рівень вищої освіти другий (магістерський)

Дослідження моделей темпоральних знань для виявлення шиллінг-атак користувачів в рекомендаційних системах
(тема)

Виконав:
студент 2 курсу, групи СШІм-22-3
Гончаренко Д.О.
(прізвище, ініціали)

Спеціальність 122 Комп'ютерні науки
(код і повна назва спеціальності)
Тип програми освітньо-наукова
(освітньо-професійна або освітньо-наукова)
Освітня програма Системи штучного інтелекту
(повна назва спеціалізації)
Керівник проф. Чалий С.Ф.
(посада, прізвище, ініціали)

Допускається до захисту

Зав. кафедри _____ В.О. Філатов
(підпис) (прізвище, ініціали)

2024 р.

Харківський національний університет радіоелектроніки

Факультет Комп'ютерних наук
(повна назва)

Кафедра Штучного інтелекту
(повна назва)

Рівень вищої освіти другий (магістерський)

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

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

Освітня програма Системи штучного інтелекту
(повна назва)

ЗАТВЕРДЖУЮ:
Зав. кафедри _____
(підпис)
«_____» _____ 20 ___ р.

ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ

студентові Гончаренку Данилу Олександровичу
(прізвище, ім'я, по батькові)

1. Тема роботи Дослідження моделей темпоральних знань для виявлення шиллінг-атак користувачів в рекомендаційних системах

затверджена наказом університету від 1 квітня 20 24 р. № 260Ст
2. Термін подання студентом роботи до екзаменаційної комісії 14 червня 20 24 р.
3. Вихідні дані до роботи документація до мови програмування Python, пакетів NumPy, Pandas, mathplotlib, датасет з ресурсу, документація та наукові матеріали за темою

4. Перелік питань, що потрібно опрацювати в роботі
1) Проаналізувати існуючі методи виявлення шиллінг-атак на рекомендаційних системах та їх обмеження
2) Дослідити можливості використання темпоральних моделей для виявлення шиллінг-атак
3) Розробити модель виявлення шиллінг-атак, засновану на аналізі темпоральних даних користувача
4) Протестувати розроблену модель на реальних даних, проаналізувати результати та порівняти їх із результатами існуючих методів

КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Терміни виконання етапів роботи	Примітка
1	Отримання завдання на кваліфікаційну роботу	01.04.2024	виконано
2	Аналіз методів щодо вирішення поставленої	02.04.2024–08.04.2024	виконано
	задачі		
3	Постановка задачі	08.04.2024–10.04.2024	виконано
4	Експериментальне моделювання та навчання	10.04.2024–18.04.2024	виконано
5	Розробка моделей	19.04.2024–25.04.2024	виконано
6	Написання пояснювальної записки	26.04.2024–19.05.2024	виконано
7	Попередній захист	24.05.2024	виконано
8	Захист перед ЕК	14.06.2024	виконано

Дата видачі завдання 1 квітня 20 24 р.

Студент _____
(підпис)

Керівник роботи _____ проф. Чалий С.Ф.
(підпис) (посада, прізвище, ініціали)

РЕФЕРАТ

Пояснювальна записка: 70 с., 9 рис., 7 табл., 2 дод., 28 джерел.

ДАТАСЕТ, КЛАСИФІКАЦІЯ, КОРИСТУВАЧІ, МОДЕЛЬ, РЕКОМЕНДАЦІЙНІ СИСТЕМИ, ТЕМПОРАЛЬНІ ЗНАННЯ, ШИЛЛІНГ-АТАКИ.

Об'єкт дослідження – рекомендаційні системи, специфічні моделі темпоральних знань та їх використання в детекції шиллінг-атак.

Предмет дослідження – розробка, адаптація або вдосконалення підходів і алгоритмів для розпізнавання шиллінг-атак в структурі даних, які включають у себе часові аспекти, а також оцінка їх ефективності в реальних умовах рекомендаційних систем.

Мета роботи – вивчення можливості використання моделей темпоральних знань для ефективного виявлення шиллінг-атак в рекомендаційних системах. Це передбачає розробку або вдосконалення алгоритмів розпізнавання шиллінг-атак на основі аналізу темпоральних даних, а також подальше тестування і оцінка ефективності цих методів в контексті реальних рекомендаційних систем.

Методи дослідження: дослідження існуючих праць та документів стосовно шиллінг-атак, вивчення можливості використання моделей темпоральних знань для ефективного виявлення шиллінг-атак в рекомендаційних системах, розробка або вдосконалення алгоритмів розпізнавання шиллінг-атак на основі аналізу темпоральних даних.

У результаті цього дослідження буде розробленим інструмент для виявлення шиллінг-атак, який буде ефективно знаходити потенційні атаки, шляхом використання датасетів, маючи необхідну вибірку даних.

ABSTRACT

Master’s thesis contains: 70 pp., 9 fig., 7 tabl., 2 ann., 28 references.

CLASSIFICATION, DATASET, MODEL, RECOMMENDER SYSTEMS, SCHILLING ATTACKS, TEMPORAL KNOWLEDGE, USERS.

The object of the study is recommender systems, specific models of temporal knowledge and their use in the detection of shilling attacks.

The subject of research is the development, adaptation or improvement of approaches and algorithms for recognizing shilling attacks in the data structure, which include time aspects, as well as evaluating their effectiveness in real conditions of recommender systems.

The purpose of the work is to study the possibility of using temporal knowledge models for effective detection of shilling attacks in recommender systems. This involves the development or improvement of algorithms for the recognition of shilling attacks based on the analysis of temporal data, as well as further testing and evaluation of the effectiveness of these methods in the context of real recommender systems.

Research methods: research of existing works and documents related to shilling attacks, study of the possibility of using temporal knowledge models for effective detection of shilling attacks in recommender systems, development or improvement of shilling attack recognition algorithms based on temporal data analysis.

As a result, a tool for detecting shilling attacks will be developed, which will effectively find potential attacks by using datasets with the necessary data sampling.

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ, СКОРОЧЕНЬ І ТЕРМІНІВ

ТД – так далі;

ТП – тому подібне;

AI – Artificial Intelligence – штучний інтелект;

API – Application Programming Interface – інтерфейс програмування додатків;

AWS – Amazon Web Services – платформа хмарних обчислень;

CSV – Comma-Separated Values – значення, розділені комами;

GCP – Google Cloud Platform – платформа хмарних обчислень;

JSON – JavaScript Object Notation – текстовий формат обміну даними.

ВСТУП

У сучасному світі рекомендаційні системи стали невід'ємною частиною багатьох онлайн-платформ. Вони допомагають користувачам знайти цікаві продукти, музику, фільми та інший контент, ґрунтуючись на їхніх попередніх взаємодіях та уподобаннях. Основна мета рекомендаційних систем – покращити користувацький досвід, зробивши його більш персоналізованим та ефективним. Рекомендаційні системи використовуються в різних галузях, включаючи електронну комерцію, стримінгові сервіси, соціальні мережі та інші.

Однак, разом з поширенням рекомендаційних систем виникає загроза шилінг-атак. Шиллінг-атаки – це навмисні дії, спрямовані на маніпулювання рейтингами або відгуками для підвищення або зниження популярності певних продуктів або послуг. Такі атаки можуть здійснюватися окремими користувачами або групами зловмисників з метою вплинути на рекомендаційні алгоритми та спотворити результати.

Існуючі підходи до виявлення шиллінг-атак мають певні недоліки. Традиційні методи, такі як колаборативне та контентне фільтрування, часто не враховують динаміку змін у поведінці користувачів. Вони можуть бути недостатньо ефективними у виявленні аномальних патернів, що робить їх вразливими до маніпуляцій. Колаборативне фільтрування, наприклад, може не розпізнати зміни в поведінці зловмисника, який залишає позитивні відгуки після тривалого періоду нейтральних або негативних відгуків.

Для подолання цих недоліків доцільно використовувати темпоральні знання. Темпоральні знання дозволяють аналізувати зміни в поведінці користувачів протягом часу. Вони включають інформацію про часові аспекти взаємодії користувачів з системою, такі як частота та послідовність дій, тривалість сеансів роботи та зміни у поведінці протягом часу.

Темпоральні знання дозволяють враховувати динаміку поведінки користувачів та виявляти незвичайні патерни, які можуть свідчити про

маніпуляції. Наприклад, аналіз часових рядів може допомогти виявити раптові зміни в оцінках або відгуках, що можуть бути індикаторами шиллінг-атак. Використання темпоральних моделей також дозволяє підвищити точність рекомендацій, враховуючи зміни в інтересах користувачів протягом часу.

Об’єктом дослідження є процеси атак користувачів на рейтинги в рекомендаційній системі.

Предметом дослідження є моделі та методи виявлення шиллінг-атак з використанням темпоральних знань.

Мета дослідження зводиться до підвищення ефективності виявлення атак зміни вподобань користувачів в рекомендаційній системі шляхом використання темпоральних знань.

1 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ ТА ПОСТАНОВКА ЗАДАЧІ

1.1 Дослідження характеристик рекомендаційних систем

Рекомендаційні системи використовуються для виявлення шаблонів у великих масивах даних, щоб надавати індивідуальні пропозиції користувачам. Вони стали ключовим елементом багатьох онлайн-сервісів, зокрема електронної комерції, стримінгових платформ музики та відео, соціальних медіа, ігрових платформ та багатьох інших. Загальну структуру рекомендаційної системи можна побачити на рисунку 1.1, взаємодія між користувачем і веб-сторінкою, а також між користувачем і його профілем, забезпечує ефективність рекомендаційної системи.

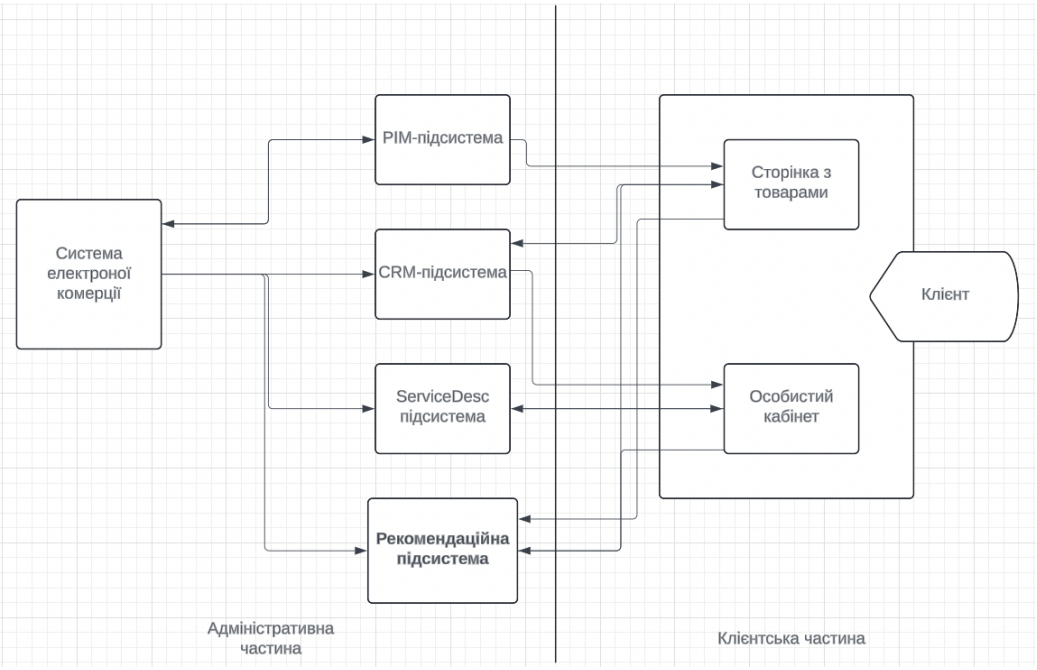

Рисунок 1.1 – Загальна структура рекомендаційної системи

Рекомендаційні системи зазвичай базуються на одному з двох основних підходів: колаборативній фільтрації та контентному фільтруванні. Колаборативна фільтрація робить висновки на основі поведінки

користувачів, наприклад, на основі оцінок, які вони надають певним продуктам. Контентне фільтрування, у свою чергу, зосереджується на об'єктивних характеристиках продуктів.

Деякі рекомендаційні системи використовують гібридний підхід, який комбінує обидва методи. І хоча такі системи взагалі добре працюють, вони не виключають можливості шиллінг-атак, коли користувачі штучно змінюють рейтинги продуктів за допомогою фальшивих відгуків. Це створює виклики для підтримки надійності та точності рекомендацій.

Також у сфері рекомендаційних систем можна виділити декілька підходів, які представлені у таблиці 1.1.

Таблиця 1.1 – Підходи, що використовуються в рекомендаційних системах

Підхід	Опис
Колаборативна фільтрація	Основа на поведінкових даних користувачів з використанням рейтингів та ін., проте є слабкість до шиллінг-атак через можливість маніпуляцій з поведінковими даними
Контентне фільтрування	Основа на характеристиках контенту, менш схильне до шиллінг-атак, обмежене у врахуванні поведінки користувачів.
Гібридні методи	Комбінація колаборативного та контентного фільтрування, покращенна точність рекомендацій, необхідність ефективних механізмів задля запобігання шиллінг-атак

Продовження таблиці 1.1

Підхід	Опис
Виявлення шаблонів	Аналіз поведінки користувачів для виявлення стандартних та аномальних шаблонів, що свідчать про шиллінг-атаки
Аналіз аномалій	Виявлення відхилень в поведінці користувачів, використання машинного навчання для виявлення аномалій
Захист від шиллінг-атак	Розробка моделей та алгоритмів для запобігання фальшивих рейтингів, включення метрик що враховують темпоральні аспекти даних
Моделювання темпоральних знань	Використання темпоральних моделей для точнішого розуміння динаміки поведінки, виявлення довго- та короткострокових тенденцій у даних користувачів

Рекомендаційні системи також можуть включати методи, що використовують машинне навчання, такі як нейронні мережі, для аналізу складних даних і створення персоналізованих рекомендацій. Ці методи дозволяють виявляти приховані залежності та тренди в даних, що значно підвищує точність рекомендацій.

Ще один важливий аспект рекомендаційних систем – це обробка великих обсягів даних у реальному часі. Сучасні системи повинні бути здатні швидко обробляти дані від великої кількості користувачів, враховувати їхню поточну активність та швидко генерувати рекомендації. Для цього використовуються високопродуктивні обчислювальні платформи та алгоритми оптимізації, які дозволяють зменшити час обробки даних і підвищити ефективність системи.

Крім того, рекомендаційні системи можуть використовувати інформацію з різних джерел, таких як соціальні мережі, щоб забезпечити більш повний аналіз поведінки користувачів. Це дозволяє системам

враховувати не тільки історію взаємодії користувачів з продуктами, але й їхню соціальну активність та інші зовнішні фактори, які можуть впливати на їхні уподобання.

Важливою характеристикою рекомендаційних систем є їх здатність адаптуватися до змін у поведінці користувачів. Системи повинні постійно оновлювати свої моделі на основі нових даних, щоб забезпечити актуальні та точні рекомендації. Це вимагає використання методів онлайн-навчання та регулярного оновлення моделей машинного навчання.

1.2 Аналіз відмінностей шиллінг-атак в рекомендаційних системах

Протягом останніх років, з поширенням рекомендаційних систем, шиллінг-атаки стали актуальною проблемою. Ці атаки відбуваються, коли особи або групи осіб навмисно маніпулюють рейтингами окремих товарів або послуг, з метою збільшення або зменшення їхнього рейтингу. Такі дії не тільки спотворюють рейтинг, але й погіршують якість рекомендацій, що призводить до зниження довіри користувачів до рекомендаційних систем.

Шиллінг-атаки можуть бути надзвичайно різноманітними за своїм характером і методами виконання. Вони можуть включати автоматизовані боти, що масово створюють фальшиві акаунти, або реальних користувачів, які діють з певною метою. Важливо зазначити, що такі атаки не тільки шкодять репутації продуктів, але й ускладнюють процес аналізу даних, адже істинні рекомендації переплітаються з фальшивими.

Рекомендаційні системи повинні мати механізми для виявлення та запобігання таким атакам, інакше їхня ефективність буде значно знижена. Використання методів машинного навчання та аналізу великих даних дозволяє створювати алгоритми, що можуть виявляти аномалії у поведінці користувачів та виявляти можливі шиллінг-атаки.

Прикладом шиллінг-атаки може бути ситуація, коли група користувачів навмисно залишає позитивні відгуки та оцінки для певного

товару, щоб підвищити його рейтинг і популярність. Подібні маніпуляції можуть призвести до того, що товар, який насправді може бути низької якості, стає більш привабливим для інших користувачів через високий рейтинг. Іншим прикладом може бути випадок, коли конкуренти навмисно залишають негативні відгуки про продукт, щоб знизити його рейтинг та пригнічити продажі.

Аналіз темпоральних даних може допомогти знайти нестандартні шаблони в поведінці користувачів, що може вказувати на шиллінг-атаку. Детектування таких атак може перешкоджати спотворенню рекомендацій та допомогти у покращенні якості рекомендаційних систем.

Шиллінг-атаки є маніпулятивними діями, спрямованими на зміну рекомендаційних результатів шляхом внесення фальшивих даних. Існують різні типи шиллінг-атак, кожна з яких має свої характеристики та мету. Аналіз відмінностей між цими атаками є важливим для розробки ефективних методів захисту. Було виділено декілька типів шиллінг-атак та їх властивостей, які представлені в таблиці 1.2.

Таблиця 1.2 – Типи шиллінг-атак та їх властивості

Тип шиллінг-атаки	Методи виявлення атак	Переваги	Недоліки методу
Атака підвищення рейтингу (Push Attack)	Колаборативне фільтрування, контентне фільтрування	Висока точність для виявлення аномалій у рейтингах	Зловмисники можуть використовувати різні стратегії для підвищення рейтингу, що ускладнює визначення чітких патернів. Також, можливі маскування атак зі сторони зловмисників.

Продовження таблиці 1.2

Тип шиллінг-атаки	Методи виявлення атак	Переваги	Недоліки методу
Атака зниження рейтингу (Nuke Attack)	Контентне фільтрування, аналіз відгуків	Простота реалізації, можливість швидкого виявлення негативних відгуків	Природні зміни у вподобаннях користувачів або негативні відгуки можуть бути сплутані з атакою. Зловмисники можуть швидко адаптувати свої стратегії відповідно до виявлених вразливостей у системі
Атака ботів (Bot Attack)	Аналіз поведінки користувачів, кластеризація	Можливість виявлення великих груп фальшивих акаунтів	Зловмисники можуть розробляти ботів, які імітують дії реальних користувачів. Для ефективного виявлення ботів необхідні великі обсяги історичних даних про поведінку користувачів
Атака на основі схожих відгуків	Аналіз тексту відгуків, NLP	Виявлення шаблонів у текстах відгуків, які свідчать про фальшиві відгуки	Схожі відгуки можуть бути природними. Для ефективного виявлення атак необхідні великі обсяги історичних даних про відгуки користувачів

Продовження таблиці 1.2

Тип шиллінг-атаки	Методи виявлення атак	Переваги	Недоліки методу
Атака зміни уподобань (Preference Attack)	Аналіз змін уподобань користувачів, колаборативне фільтрування	Висока точність у виявленні аномалій в уподобаннях користувачів	Реалізація складних моделей для виявлення змін уподобань користувачів може вимагати значних обчислювальних ресурсів і часу. Зловмисники можуть адаптувати свої методи, щоб уникнути виявлення зміни уподобань.
Атака на нові продукти (Cold Start Attack)	Аналіз поведінки нових користувачів, виявлення аномалій	Ефективність у виявленні аномалій для нових або рідко оцінюваних продуктів	Відсутність даних для порівняння з іншими продуктами ускладнює виявлення аномальної поведінки. Аналіз нових продуктів потребує великих обчислювальних ресурсів для моніторингу і виявлення аномалій

Шиллінг-атаки становлять значну загрозу для ефективності та надійності рекомендаційних систем. Вони можуть істотно спотворити результати рекомендацій, що може призвести до втрати довіри користувачів

і зниження доходів компаній, що використовують такі системи. Оскільки атаки можуть бути дуже різноманітними, важливо використовувати комбінацію різних методів для їх виявлення та запобігання.

Темпоральні дані відіграють ключову роль у покращенні методів виявлення шиллінг-атак. Вони дозволяють виявляти не тільки очевидні маніпуляції, але й більш складні аномалії, які можуть проявлятися лише у певні періоди часу. Врахування темпоральних аспектів дозволяє створювати більш стійкі та надійні рекомендаційні системи. Це досягається шляхом аналізу часових рядів та використання темпоральних правил для виявлення незвичайних патернів у поведінці користувачів.

Крім того, для ефективного виявлення шиллінг-атак необхідно використовувати підхід машинного навчання, що дозволяє системам адаптуватися до нових патернів атак. Регулярне оновлення моделей на основі нових даних забезпечує актуальність та точність виявлення. Це особливо важливо в умовах швидко змінюваного середовища, де нові методи атак можуть з'являтися дуже швидко. Наприклад, використання рекурентних нейронних мереж (RNN) або глибоких навчальних алгоритмів дозволяє обробляти великі обсяги даних і враховувати темпоральні залежності.

Крім того, важливо звертати увагу на багатоканальний підхід до збору даних, який включає аналіз соціальних мереж, поведінкових патернів на різних платформах, а також взаємодії користувачів з контентом. Це дозволяє отримати більш повну картину активності користувачів і виявити навіть найменші аномалії.

Таким чином, комплексний підхід, що включає аналіз темпоральних даних, машинне навчання та багатоканальний збір даних, є ключем до створення ефективних рекомендаційних систем, стійких до шиллінг-атак. Врахування часу та поведінкових патернів користувачів допомагає забезпечити високу точність і надійність рекомендацій, що підвищує довіру користувачів і покращує загальний досвід використання систем.

1.3 Роль темпоральних моделей

У контексті рекомендаційних систем, темпоральні моделі дозволяють аналізувати зміни в поведінці користувачів протягом часу. Користувачі можуть змінювати свої інтереси, вибір продуктів над часом, і ці зміни можуть бути дуже інформативними для прогнозування майбутніх дій.

Для прикладу, якщо користувач-атакувальник в основному залишає нейтральні або негативні відгуки, а потім раптово починає залишати багато позитивних відгуків для певного продукту, це може бути підозрілим поведінкою і може вказувати на шиллінг-атаку. Темпоральні моделі дозволяють виявляти такі аномалії, аналізуючи часові ряди даних.

Більше того, аналіз часових масивів даних може допомогти визначити «нормальну» поведінку для кожного користувача або групи користувачів, що може бути використано як базис для виявлення аномалій. Наприклад, якщо користувач зазвичай купує продукти певного типу, а потім раптово змінює свої уподобання, це може свідчити про зміну поведінки або можливу атаку. Темпоральні моделі дозволяють враховувати ці зміни та адаптувати рекомендації відповідно до нових даних.

Поряд з цим, темпоральні моделі можуть мати прямий вплив на якість рекомендацій. Рекомендаційні системи, які враховують зміну інтересів користувачів впродовж часу, можуть надавати більш точні та особисті рекомендації. Наприклад, користувач, який зазвичай купує спортивний інвентар у літні місяці, може отримувати відповідні рекомендації саме у цей період. Це підвищує задоволеність користувачів та ефективність рекомендаційної системи.

Темпоральні моделі також дозволяють враховувати сезонні та трендові зміни в уподобаннях користувачів. Наприклад, аналіз покупок перед новорічними святами може показати, що певні продукти мають високу популярність у цей період. Використання темпоральних моделей

дозволяє системі адаптувати свої рекомендації, враховуючи ці сезонні тренди та змінюючи пріоритети продуктів залежно від поточного часу.

Таким чином, темпоральні моделі мають важливу роль в розробці рекомендаційних систем, які можуть протистояти шиллінг-атакам та надавати корисні рекомендації. Вони дозволяють не тільки виявляти аномалії та запобігати маніпуляціям, але й покращувати точність та релевантність рекомендацій, враховуючи динамічні зміни в поведінці користувачів.

Темпоральні моделі також забезпечують можливість прогнозування майбутньої поведінки користувачів на основі їхніх попередніх дій. Це дозволяє не тільки виявляти атаки, але й передбачати можливі дії користувачів, що може бути корисним для маркетингових кампаній та управління запасами.

Застосування методів машинного навчання, таких як рекурентні нейронні мережі (RNN), для аналізу темпоральних даних дозволяє моделювати складні залежності між подіями та часом. Це забезпечує високу точність прогнозування та виявлення аномалій, що робить рекомендаційні системи більш надійними та ефективними.

Таким чином, впровадження темпоральних моделей у рекомендаційні системи є ключовим фактором для забезпечення їхньої ефективності та надійності в умовах активних маніпуляцій та швидких змін у поведінці користувачів.

1.4 Сфера застосування темпоральних моделей

Темпоральні моделі можуть бути використані у численних областях, де використовуються рекомендаційні системи та де є можливість виконання шиллінг-атак.

– електронна комерція: у сфері електронної комерції темпоральні моделі можуть бути використані для виявлення підозрілих візерунків

поведінки користувачів, наприклад раптових змін у відгуках або незвично великої кількості відгуків у короткий проміжок часу;

– стримінгові платформи: на стримінгових платформах, таких як Netflix або Spotify, темпоральні моделі можуть бути використані для відстеження змін у переглянутих або прослуханих фільмах або піснях користувачів протягом часу;

– соціальні медіа: соціальні медіа платформи, такі як Facebook або Twitter, можуть використовувати темпоральні моделі для виявлення шиллінг-атак, коли користувачі здійснюють атаки, спрямовані на надування або знищення популярності певного посту, продукту або особи;

– відгуки про продукти або послуги: сайти відгуків, які надають рекомендації на основі відгуків користувачів, також можуть використовувати темпоральні моделі для виявлення шиллінг-атак. Всі ці області можуть вигідно використати від переваг темпоральних моделей для виявлення шиллінг-атак і покращення якості своїх рекомендаційних систем.

1.4.1 Електронна комерція

Електронна комерція є однією з найпоширеніших областей, де використовуються рекомендаційні системи. Завдяки величезному обсягу інформації, яка збирається про користувачів (історія покупок, перегляди товарів, відгуки та оцінки тощо), це також область, де використання темпоральних моделей стає особливо корисним.

Основні напрями застосування темпоральних моделей в електронній комерції включають:

– виявлення шиллінг-атак: аналіз змін метрик, таких як рейтинги та відгуки користувачів, може допомогти виявити незвичні закономірності. Наприклад, якщо певний товар отримує велику кількість позитивних відгуків у короткий проміжок часу, це може бути індикатором шиллінг-атаки;

– обробка та сегментація клієнтів: зчитування та перетворення темпоральних даних, таких як частота покупок, час між покупками та історія купівлі, може допомогти виявити важливі шаблони поведінки споживачів і групувати клієнтів за схожими шаблонами;

– прогнозування преференцій: на основі шаблонів у темпоральних даних можна побудувати моделі для прогнозування майбутніх покупок або переваг користувачів;

– особисті рекомендації: враховуючи інформацію, втягнену з темпоральних даних, системи можуть надавати більш особисті рекомендації, які відповідають не лише поточним інтересам користувачів, але і їх змінюючимся трендам.

Таким чином, в електронній комерції темпоральні моделі можуть значно покращити якість рекомендаційних систем, а також допомогти у боротьбі з шиллінг-атаками.

1.4.2 Стримінгові платформи

Стримінгові платформи, такі як Netflix, Amazon Prime Video або Spotify, стали невід'ємною частиною життя сучасних користувачів інтернету. Вони залежать від високоякісних рекомендаційних систем для надання персоналізованих пропозицій користувачам. Це може включати фільми, серіали, музику, подкасти та багато іншого в залежності від платформи.

Основні напрямки застосування темпоральних моделей на стримінгових платформах включають:

– прогнозування поведінки користувачів: Темпоральні моделі можуть допомогти виявити шаблони в поведінці користувачів над часом - наприклад, коли вони найбільше ймовірно слухатимуть музику або дивитимуться фільми, які жанри вони найбільше переглядають у певний час дня або року та інше;

– виявлення шиллінг-атак: Якщо певний фільм, серіал або пісня раптово стає надзвичайно популярним без очевидної причини, це може бути підозріло і вказувати на можливу шиллінг-атаку;

– розширений аналіз вмісту: Платформи можуть використовувати темпоральні моделі для відстеження популярності певного вмісту протягом часу, що може допомогти визначити тривалі тренди або уподобання в аудиторії.

Таким чином, в стримінгових платформах, темпоральні моделі можуть допомогти у виявленні шиллінг-атак, а також пропонувати більш персоналізовані і актуальні рекомендації користувачам.

1.4.3 Соціальні медіа

Соціальні медіа, такі як Facebook, Twitter або Instagram, є потужним інструментом для комунікації та взаємодії в сучасному цифровому світі. Також їх використовують для дослідження продуктів та послуг, і вони надають рекомендації на основі інтересів та взаємодії користувачів. Однак, це також область, де шиллінг-атаки можуть відбутися, особливо у вигляді фейкових користувачів або ботів, що спрямовують аудиторію на певний контент.

Основні напрями застосування темпоральних моделей у соціальних медіа включають:

– виявлення шиллінг-атак: Виявлення незвичайних шаблонів активності, таких як раптове збільшення позитивних або негативних коментарів на певному пості, може бути індикатором шиллінг-атаки;

– виявлення ботів або фейкових облікових записів: Темпоральні моделі можуть допомогти виявити нестандартні шаблони активності, які можуть вказувати на автоматичну активність, таку як боти;

– аналіз сезонності: Соціальні медіа мають змінні тренди над часом. Темпоральні моделі можуть допомогти розібратися в цих змінах та прогнозувати майбутні тренди;

– персоналізовані рекомендації: Аналіз темпоральних даних дозволяє системам краще розуміти змінний інтерес користувачів і надавати більш актуальні рекомендації.

Отже, темпоральні моделі в соціальних медіа можуть допомогти у боротьбі з шиллінг-атаками, виявленні ботів та фейкових облікових записів, розумінні сезонних змін та наданні персоналізованих рекоментацій.

1.4.4 Відгуки про продукти або послуги

Веб-сайти відгуків про продукти та послуги, такі як Yelp, TripAdvisor, або Amazon, зібрали величезну кількість важливих даних про перегляди, вподобання і поведінку користувачів. Ці додатки використовують рекомендаційні системи для надання персоналізованих пропозицій, що ґрунтуються на цих даних. Але вони також стають потенційним полем для шиллінг-атак.

Основні напрями застосування темпоральних моделей в отзивах про продукти або послуги включають:

– виявлення шиллінг-атак: Темпоральні моделі можуть допомогти виявити раптові та незвичайні зміни в оцінках і відгуках користувачів, що можуть бути показником шиллінг-атаки;

– покращення рекомендацій: Складні моделі, що враховують час, можуть допомогти виявити тенденції та зміни вподобань користувачів протягом часу, що дає можливість надавати актуальніші рекомендації;

– аналіз відгуків: Темпоральні моделі можуть допомогти аналізувати, як відгуки та оцінки змінюються протягом часу, що може допомогти бізнесам виявити проблеми або покращити свої продукти або послуги.

Таким чином, у контексті веб-сайтів з відгуками, темпоральні моделі можуть допомогти у виявленні шиллінг-атак, покращенні якості рекомендацій та аналізі відгуків та оцінок користувачів.

1.5 Дослідження особливостей представлення та моделювання темпоральних знань

Представлення темпоральних знань є ключовим аспектом у дослідженні та розробці рекомендаційних систем. Цей етап включає в себе визначення формату даних, використання структур для представлення взаємодій між об'єктами з плином часу, а також вибір методів обробки даних для ефективного моделювання темпоральних знань.

Темпоральні знання можна представити за допомогою різних методів, що дозволяють моделювати часові залежності та враховувати зміни в поведінці користувачів протягом часу. Основні методи представлення темпоральних знань включають часові ряди, графові моделі, матриці залежностей та темпоральні правила, які представлені у таблиці 1.3.

Таблиця 1.3 – Методи представлення темпоральних знань

Метод представлення темпоральних знань	Опис
Графові структури	Об'єкти представляються у вигляді вузлів, а зв'язки між ними - у вигляді ребер. Можливість виявлення патернів та класифікації темпоральних знань.
Часові ряди	Використання часових рядів для моделювання темпоральних знань, де аналізується зміна показників часу (вісь X) відповідно до параметрів (вісь Y).
Матриці залежностей	Представлення темпоральних знань у вигляді матриць, де кожен елемент вказує ступінь зв'язку між сутностями з плином часу.

Продовження таблиці 1.3

Метод представлення темпоральних знань	Опис
Моделі з розподіленими параметрами	Використання моделей, що дозволяють моделювати змінні параметри в часі, такі як моделі авторегресії або нейронні мережі з рекурентними шарами.
Використання контекстуальної інформації	Включення додаткових факторів, таких як геолокація або соціальні взаємодії, які можуть впливати на темпоральні знання та їх моделювання.

Дослідження особливостей даних та їх аналіз також є важливим етапом у розробці та застосуванні рекомендаційних систем. Цей пункт включає аналіз доступних даних, їх типи та об'єм, а також методи аналізу, які можуть бути застосовані для вивчення та розуміння цих даних.

Аналіз темпоральних знань потребує глибокого розуміння різних типів даних, їх структури та обсягу. Важливо розуміти, які дані є релевантними для аналізу та як їх правильно використовувати. Для цього використовуються різні методи аналізу, що дозволяють виявляти тренди, сезонні варіації та аномалії у поведінці користувачів, деякі з них представлені на таблиці 1.4.

Таблиця 1.4 – Особливості даних та їх аналіз

Особливість даних або метод аналізу	Опис
Типи даних	Включає аналіз різних типів даних, таких як структуровані, напівструктуровані та неструктуровані дані.

Продовження таблиці 1.4

Особливість даних або метод аналізу	Опис
Об’єм даних	Оцінка обсягу доступних даних для дослідження та визначення їхньої достатності для побудови моделей та аналізу.
Аналіз часових залежностей	Використання методів для аналізу та розуміння часових залежностей у даних, включаючи виявлення трендів, сезонних варіацій та аномалій.
Виявлення аномалій	Розробка та застосування методів для виявлення аномалій у даних, які можуть бути індикаторами шиллінг-атак або інших неправильних поведінок користувачів.

Використання темпоральних моделей має численні переваги для рекомендаційних систем. Вони дозволяють враховувати динамічні зміни в поведінці користувачів, що підвищує точність та релевантність рекомендацій. Темпоральні моделі також допомагають виявляти складні взаємозв'язки між подіями, що можуть бути важливими для прогнозування майбутніх дій користувачів.

Однак, використання темпоральних моделей також має свої виклики. По-перше, вони вимагають значних обчислювальних ресурсів для обробки великих обсягів даних у реальному часі. По-друге, темпоральні моделі можуть бути складними у налаштуванні та оптимізації, що вимагає високого рівня експертизи у галузі машинного навчання та аналізу даних.

Подальший розвиток темпоральних моделей у рекомендаційних системах відкриває нові можливості для їх використання у різних галузях. Наприклад, вони можуть бути використані для прогнозування попиту на товари та послуги, управління запасами, оптимізації маркетингових кампаній та підвищення задоволеності користувачів.

Врахування темпоральних аспектів дозволяє створювати рекомендаційні системи, що краще розуміють контекст та уподобання користувачів. Це забезпечує більш персоналізований підхід до надання рекомендацій, що сприяє підвищенню лояльності користувачів та збільшенню доходів компаній.

Загалом, використання темпоральних моделей є важливим напрямком розвитку рекомендаційних систем, що дозволяє підвищити їх ефективність та надійність в умовах швидких змін та високої конкуренції.

1.6 Постановка задачі дослідження

Рекомендаційні системи використовують інформацію про зроблені користувачем вибір товарів та послуг для того, щоб визначити його вподобання та запропонувати нові товари та послуги згідно з цих вподобань. Одним із видів вхідних даних для таких систем є дані про рейтинги товарів та послуг. Помилки в таких даних або їх навмисне спотворення приводять до некоректних рекомендацій. Спотворення даних відбувається за допомогою шилінг-атак. Ці атаки штучно занижують рейтинг товарів конкурентів або ж завищуються рейтинги цільових товарів. Результати такої атаки можуть виглядати як зміни вподобань користувачів. Існуючі підходи до виявлення шилінг-атак не завжди виявляють таку штучну зміну вподобань, оскільки зловмисники можуть адаптувати свої методи з часом. Тому, для виявлення таких атак доцільно використовувати темпоральні знання, які дають можливість описати зміну вподобань користувачів з часом.

Враховуючи особливості та проблеми рекомендаційних систем, основною задачею цього дослідження є розробка та впровадження методу виявлення шиллінг-атак на основі темпоральних моделей.

Об’єктом дослідження є процеси атак користувачів на рейтинги в рекомендаційній системі.

Предметом дослідження є моделі та методи виявлення шиллінг-атак з використанням темпоральних знань.

Мета дослідження зводиться до підвищення ефективності виявлення атак зміни вподобань користувачів в рекомендаційній системі шляхом використання темпоральних знань.

Задачі дослідження:

– проаналізувати існуючі методи виявлення шиллінг-атак на рекомендаційних системах та їх обмеження;

– дослідити можливості використання темпоральних моделей для виявлення шиллінг-атак;

– розробити модель виявлення шиллінг-атак, засновану на аналізі темпоральних даних користувача;

– протестувати розроблену модель на реальних даних, проаналізувати результати та порівняти їх із результатами існуючих методів.

Очікується, що в результаті цього дослідження буде розробленим інструмент для виявлення шиллінг-атак, який буде ефективно знаходити потенційні атаки, шляхом використання датасетів, маючи необхідну вибірку даних.

2 ТЕОРЕТИЧНЕ ДОСЛІДЖЕННЯ ОБРАНОЇ ПРОБЛЕМИ

2.1 Розробка підходу до виявлення шиллінг-атак з використанням темпоральних знань

Важливість забезпечення безпеки та надійності систем рекомендацій в сучасних умовах важко переоцінити. Однією з серйозних загроз для цих систем є шиллінг-атаки, коли зловмисники маніпулюють рейтингами або іншими видами зворотного зв’язку з метою штучного підвищення або зниження популярності певних продуктів або послуг.

Традиційні методи виявлення шахрайства в основному зосереджуються на аналізі явного зворотного зв’язку, такого як рейтинги або відгуки користувачів. Проте, ці методи часто виявляються недостатньо ефективними, оскільки зловмисники можуть легко адаптувати свої дії, щоб уникнути виявлення. Вони можуть, наприклад, створювати множинні акаунти, імітуючи реальних користувачів, або використовувати автоматизовані скрипти для генерації великої кількості відгуків.

Запропонований підхід базується на використанні темпоральних знань, отриманих з неявного зворотного зв’язку, для виявлення шиллінг-атак. Темпоральні знання включають інформацію про часові аспекти взаємодії користувачів з системою, такі як частота та послідовність їхніх дій, тривалість сеансів роботи та зміни у поведінці протягом часу.

Використання темпоральних знань дозволяє враховувати не лише поточні дії користувачів, але й їхню поведінку в контексті часу. Це може бути особливо корисним для виявлення "сплячих" акаунтів, які активуються лише для проведення атак, або для виявлення патернів, які вказують на координацію між різними акаунтами.

Загальна схема підходу до виявлення шиллінг-атак з використанням темпоральних знань зображена на рисунку 2.1. Вона складається з 5-ти етапів, які мають підпункти. Кожний з яких буде описано нижче.

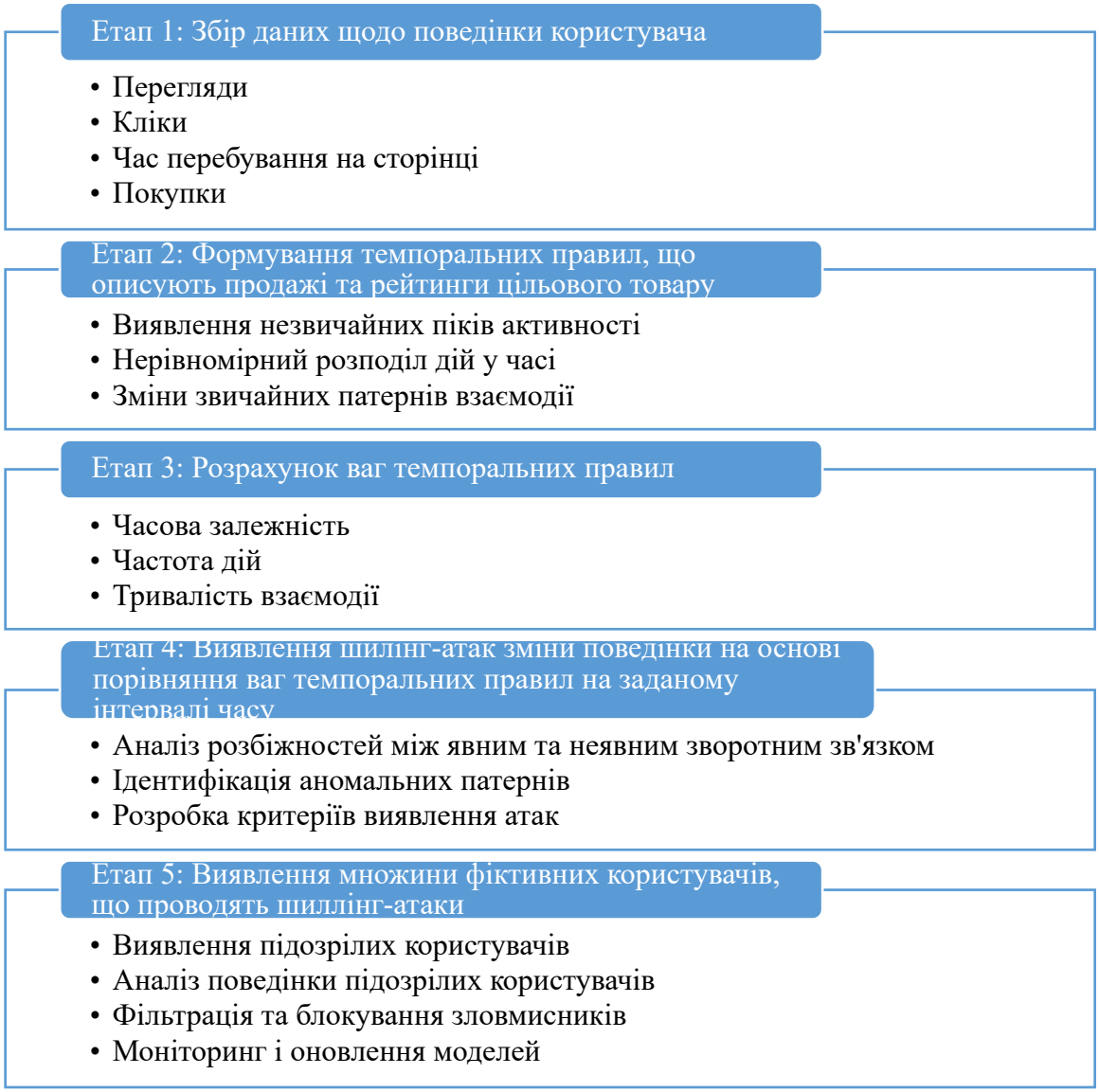

Рисунок 2.1 – Підхід до виявлення шиллінг-атак з використанням темпоральних знань

Почнемо з першого етапу: збір даних щодо поведінки користувача. По-перше, перегляди – аналіз часу і частоти переглядів користувачами сторінок з товарами або послугами, що дозволяє визначити, які продукти найбільше приваблюють увагу користувачів. Другим йдуть кліки, тобто фіксація часу і послідовності кліків на елементи інтерфейсу, що дозволяє зрозуміти, які елементи інтерфейсу найбільш впливають на поведінку користувачів та якими саме ці послідовності є. Час перебування на сторінці

передбачає вимірювання тривалості перебування користувача на сторінці, що допомагає визначити, які сторінки є більш привабливими для користувачів або який товар чи продукт викликає більший інтерес у користувачів. Та врешті – покупки. Аналіз часу здійснення покупок та частоти покупок певних товарів. Виявлення піків активності під час акцій та знижок, а також аналіз послідовностей покупок для виявлення аномалій. Це дозволяє визначити, які фактори впливають на рішення користувачів про покупку та виявити можливі аномалії у поведінці користувачів.

Для другого етапу виявляємо незвичні піки активності, тобто аналіз статистичних аномалій в активності користувачів. Наприклад, виявлення раптових сплесків активності, що можуть свідчити про масові шиллінг-атаки. Нерівномірний розподіл дій у часі – виявлення нерегулярних інтервалів між діями, визначення періодів високої та низької активності впродовж дня, наприклад, що може допомогти з визначенням патернів, які можуть бути пов'язані з шиллінг-атаками. Та визначаємо зміни звичайних патернів взаємодії користувачів з системою чи ресурсом, тобто зміни у їх поведінці, що може бути використано для прогнозування майбутньої поведінки користувачів на основі історичних даних, за допомогою часових рядів.

Наступним етапом буде розрахунок ваг темпоральних правил, в якому будемо визначати значимість кожної дії на основі часових інтервалів між діями, наприклад, дії, що виконуються з короткими інтервалами між ними, можуть мати більшу вагу. Формула вагового коефіцієнта може виглядати так:

$$w_{time} = \frac{1}{1+e^{-\lambda(t_{current} - t_{action})}}, \quad (2.1)$$

де λ – параметр затухання;

$t_{current}$ – поточний час;

t_{action} – час дії.

Частота дій відповідає за врахування частоти виконання дій для визначення їх ваги. Чим частіше користувач виконує певну дію, тим більше її вага. Формула може виглядати так:

$$w_{freq} = \log(1 + n_{actions}), \tag{2.2}$$

де $n_{actions}$ – кількість виконаних дій.

Тривалість взаємодії, у свою чергу, вбирає в себе врахування часу, який користувач проводить на виконанні певної дії, наприклад, тривалість перегляду може бути індикатором зацікавленості. Формула ваги може виглядати так:

$$w_{duration} = \frac{t_{view}}{t_{total}}, \tag{2.3}$$

де t_{view} – час перегляду сторінки;

t_{total} – загальний час перегляду.

До четвертого етапу входить аналіз розбіжностей між явним та неявним зворотним зв'язком, де визначаємо розбіжності між середніми значеннями рейтингу та кількістю покупок та використовуємо метрики, такі як середнє квадратичне відхилення, для визначення розбіжностей. Ідентифікація аномальних патернів відбувається завдяки використанню методів статистичного аналізу, наприклад, використання Z-оцінки для визначення аномально високих або низьких значень параметрів. Розробка критеріїв виявлення атак визначенням порогових значень для розбіжностей, які можуть свідчити про можливі атаки.

Фінальним етапом є знаходження користувачів шиллінг-атак, який можливо реалізувати за допомогою виявлення підозрілих користувачів, тобто ідентифікації користувачів, чиї дії відповідають виявленим аномальним патернам. Наприклад, визначення користувачів, які

здійснюють велику кількість однотипних дій за короткий період часу. Аналізом поведінки підозрілих користувачів, а саме детальним аналізом дій користувачів для підтвердження їх участі в атаках. Їх фільтрацію та подальше блокування та постійний моніторинг поведінки користувачів і оновлення моделей виявлення атак для підвищення ефективності.

2.2 Удосконалення моделей темпоральних правил для задачі виявлення шиллінг-атак

Моделі темпоральних правил є потужним інструментом для аналізу послідовностей подій у часі. Вони використовуються для виявлення закономірностей і залежностей між подіями, які відбуваються в різні моменти часу. Це дозволяє створювати прогнози, виявляти аномалії та аналізувати поведінку користувачів у різних контекстах. Темпоральні моделі широко застосовуються у різних галузях, включаючи системи рекомендацій, фінансовий аналіз, медицину та інші сфери.

Факти є ключовими елементами в моделях темпоральних правил, оскільки вони представляють конкретні події або стани, які можуть відбуватися в певні моменти часу. У контексті моделей темпоральних правил, факти використовуються для визначення темпоральних залежностей та взаємозв'язків між подіями. Факти можуть відображати різні аспекти поведінки користувачів, такі як вибір продукту, залишення відгуку або будь-які інші дії, які мають часову прив'язку.

Темпоральні залежності описують взаємозв'язки між подіями у часі, що дозволяє розуміти, як одні події впливають на інші або слідують за ними. Темпоральні залежності можуть бути представлені різними способами, зокрема за допомогою темпоральних правил, часових рядів, графових структур тощо. Використання часових рядів дозволяє моделювати темпоральні залежності, аналізуючи зміни показників часу (вісь X) відповідно до параметрів (вісь Y). Часові ряди можуть відображати тренди,

сезонні зміни та інші закономірності в даних. Наприклад, аналіз продажів продукту протягом року, де можна відстежувати сезонні зміни та тренди.

Темпоральні правила взагалі представляють собою загальні принципи та залежності між подіями у часі. Вони використовуються для моделювання темпоральних залежностей і можуть застосовуватися у різних контекстах, таких як аналіз даних, рекомендаційні системи, виявлення аномалій та прогнозування поведінки. Темпоральні правила можуть бути представлені у вигляді простих формул, що описують відносини між подіями.

Для вдосконалення моделей темпоральних правил у задачі виявлення шиллінг-атак використовується підхід, що дозволяє враховувати динаміку змін інтересів користувачів та їхню поведінку. Цей підхід базується на формуванні та використанні вагових темпоральних правил, які описують відношення між часовими інтервалами вибору продуктів та їх оцінюванням. Важливим аспектом є можливість виявлення розбіжностей між процесами вибору та оцінювання продуктів, що дозволяє ідентифікувати шиллінг-атаки навіть при відсутності оцінок у певні часові інтервали.

Темпоральні правила використовуються для опису послідовності подій у часі. Вони представлені у вигляді правил типу «Next» та «Future». Кожне з цих правил визначає часове відношення між парою фактів S_t та S_p, що відображають вибір (купівлю) продукту або встановлення його рейтингу.

Темпоральні правила визначають відносний часовий порядок типу «раніше-пізніше»:

$$r_{t,p}^{j} = \begin{cases} X(S_t, S_p) \\ F(S_t, S_p) \end{cases}, \tag{2.4}$$

де X та F - часові оператори, що визначають порядок подій у часі.

Оператор X з'єднує дві послідовні події, тоді як оператор F з'єднує непослідовні події з принаймні однією проміжною подією.

Вага правила $w_{t,p}^{j}$ визначається нормалізованою різницею між кількістю покупок або середнім значенням оцінок продукту ij на інтервалах $\Delta\tau_t$ та $\Delta\tau_p$:

$$w_{t,p}^{j} = n_p - n_t, \tag{2.5}$$

де n_p та n_t – кількість покупок або середнє значення оцінок на інтервалах $\Delta\tau_p$ та $\Delta\tau_t$ відповідно.

Це дозволяє враховувати динаміку змін інтересів користувачів.

Формування моделей процесу зміни вподобань користувачів для продукту ij представляється послідовністю оцінок G_p^j, упорядкованих за інтервалами $\Delta\tau_p$:

$$M(j) = \left(G_1^j, G_2^j, .., G_p^j\right), \tag{2.6}$$

де G_p^j – сума ваг правил між j та p інтервалами.

Ця модель описує зміну вподобань користувачів для кожного інтервалу $\Delta\tau_p$, що дозволяє виявити фальсифікацію оцінок шляхом порівняння відповідних оцінок G_p^j.

Удосконалення моделей темпоральних правил для задачі виявлення шиллінг-атак дозволяє значно підвищити точність та ефективність виявлення маніпуляцій у рейтингах. Використання вагових коефіцієнтів, нормалізованих відхилень та темпоральних операторів дозволяє враховувати динаміку змін у поведінці користувачів, що є ключовим аспектом у боротьбі з шиллінг-атаками. Подальші дослідження у цій галузі можуть зосередитися на інтеграції з іншими методами машинного навчання та аналізу великих даних для подальшого підвищення ефективності та точності алгоритмів.

2.3 Формат вхідних даних

Формат вхідних даних є ключовим аспектом у розробці системи виявлення шиллінг-атак з використанням темпоральних знань. Правильна організація та структуризація даних дозволяє забезпечити ефективне їхнє оброблення та аналіз, що є необхідним для виявлення аномалій у поведінці користувачів. У цьому розділі буде розглянуто формат вхідних даних на прикладі датасету «All Beauty» з Amazon Review Data, який використовується для експериментів.

Датасет «All Beauty» містить інформацію про відгуки на продукти з категорії краси, включаючи рейтинги, тексти відгуків, голоси за корисність відгуків, перевірені покупки, метадані продуктів, описання, категорії, ціну, бренд та зображення продуктів. Ці дані зібрані у форматі JSON, де кожен запис представляє один відгук, структура представлена у таблиці 2.2.

Таблиця 2.2 – Структура вхідних даних

Назва стовпця	Опис
reviewerID	Ідентифікатор рецензента, який залишив відгук
asin	Ідентифікатор продукту, на який залишено відгук
reviewerName	Ім'я рецензента
helpful	Кількість голосів «корисно/некорисно» за відгук
reviewText	Текст відгуку
overall	Рейтинг, який рецензент поставив продукту

Продовження таблиці 2.2

Назва стовпця	Опис
Summary	Короткий опис або заголовок відгуку
unixReviewTime	Час відгуку в форматі Unix timestamp
reviewTime	Час відгуку у форматі дати
verified	Інформація про те, чи була покупка перевіреною
style	Додаткові характеристики продукту
vote	Кількість голосів за корисність відгуку
image	URL зображення продукту

Структура даних визначає, як інформація організована та представлена для обробки та аналізу. Використання структурованих даних дозволяє застосовувати різні методи та алгоритми для виявлення аномалій та маніпуляцій. Вхідні дані мають бути достатньо детальними, щоб забезпечити глибокий аналіз, але водночас добре організованими для ефективного доступу та обробки.

Для забезпечення ефективного зберігання та оброблення даних рекомендується використовувати формати, що підтримують великі обсяги даних та швидкий доступ до них. Найбільш підходящими форматами є:

– формат JSON (JavaScript Object Notation): Використовується для зберігання структурованих даних. Забезпечує гнучкість у представленні складних об'єктів та масивів;

– формат CSV (Comma-Separated Values): Простий текстовий формат, який використовується для зберігання табличних даних. Легко обробляється та підтримується багатьма інструментами для аналізу даних;

– бази даних SQL (Structured Query Language): Використовується для зберігання даних у реляційних базах даних. Забезпечує потужні можливості для запиту та маніпуляції даними;

– бази даних NoSQL: Альтернативні бази даних, такі як MongoDB, які дозволяють зберігати неструктуровані або напівструктуровані дані. Забезпечують гнучкість та масштабованість.

Для виявлення шиллінг-атак дані повинні відповідати певним вимогам:

– повнота даних: Дані повинні бути повними та містити всі необхідні атрибути для аналізу. Відсутність важливої інформації може призвести до помилкових результатів або невиявлення аномалій;

– точність та актуальність: Дані повинні бути точними та актуальними. Використання застарілих або неточних даних може знизити ефективність алгоритму;

– структурованість: Дані повинні бути структурованими у зручному форматі для аналізу. Це забезпечує можливість застосування різних алгоритмів машинного навчання та статистичних методів;

– можливість масштабування: Формат даних повинен підтримувати можливість масштабування для обробки великих обсягів інформації. Це особливо важливо для реальних застосувань, де кількість даних може бути дуже великою.

2.4 Порядок роботи розробленого алгоритму

Теоретичний аналіз у контексті виявлення шиллінг-атак з використанням темпоральних знань передбачає розуміння різних типів

шиллінг-атак, існуючих методів їх виявлення та ключових концепцій, що застосовуються у цій галузі.

Шиллінг-атаки можна класифікувати на три основні типи: Push-атаки, Nuke-атаки та Hybrid-атаки.

Push-атаки спрямовані на штучне підвищення рейтингу певного продукту. Зловмисники створюють фальшиві акаунти та залишають позитивні відгуки і високі рейтинги для підвищення популярності продукту. Такі атаки можуть значно спотворити рейтингову систему, роблячи неякісні продукти більш популярними.

Nuke-атаки мають протилежну мету - знизити рейтинг конкурентних продуктів. У цьому випадку зловмисники залишають негативні відгуки та низькі рейтинги, щоб знизити популярність конкурента. Це може призвести до того, що якісні продукти будуть менш помітними або взагалі втрачають довіру користувачів.

Hybrid-атаки поєднують елементи push- та nuke-атак. Зловмисники одночасно підвищують рейтинг свого продукту та знижують рейтинг конкурентних. Це створює двосторонній ефект, де їхні продукти стають більш популярними, а продукти конкурентів менш помітними.

Для виявлення шиллінг-атак використовуються різні методи, що можна розділити на кілька категорій: фільтраційні методи, моделі машинного навчання та гібридні методи, які представлені у таблиці 2.1.

Таблиця 2.1 – Методи виявлення шиллінг-атак

Метод виявлення шиллінг-атак	Опис
Фільтраційні методи	Включають контентну фільтрацію та колаборативне фільтрування

Продовження таблиці 2.1

Метод виявлення шиллінг-атак	Опис
Моделі машинного навчання	Застосовують класифікацію та кластеризацію для виявлення шиллінг-атак. Класифікація використовує такі алгоритми, як Decision Trees та Random Forests, для виявлення атак на основі навчання на мітках з виявленими атаками. Кластеризація, такі як K-means та DBSCAN, групує схожі патерни поведінки для виявлення аномалій
Гібридні методи	Поєднують елементи фільтраційних методів та моделей машинного навчання для підвищення точності виявлення шиллінг-атак. Використовують переваги обох підходів для більш ефективного виявлення атак

Темпоральні моделі враховують часові аспекти взаємодії користувачів з системою. Використання часових рядів та марковських ланцюгів для аналізу змін у поведінці користувачів дозволяє враховувати динаміку змін інтересів користувачів. Це забезпечує більш точне виявлення аномалій у поведінці, які можуть свідчити про шиллінг-атаки.

Вагові темпоральні правила включають використання ваг для визначення значущості кожної дії користувача на основі часових інтервалів, частоти дій та тривалості взаємодії. Цей підхід дозволяє визначити, які дії є найбільш значущими у контексті виявлення атак.

Аналіз розбіжностей включає порівняння моделей процесів зміни вподобань на основі імпліцитного та експліцитного зворотного зв'язку для виявлення розбіжностей, що можуть свідчити про шиллінг-атаки. Це дозволяє виявити аномалії у процесах вибору та оцінювання продуктів, що є індикатором наявності атак.

Теоретичний аналіз показує, що для ефективного виявлення шиллінг-атак необхідно використовувати комплексний підхід, який включає аналіз темпоральних аспектів поведінки користувачів, використання вагових темпоральних правил та моделей машинного навчання. Такий підхід дозволяє враховувати динаміку змін інтересів користувачів та забезпечує високу точність і надійність виявлення атак.

– збір даних користувачів. Проте, необхідно врахувати обмеження API і правила конфіденційності;

– відкриті набори даних: Існують відкриті набори даних, такі як MovieLens, Jester, Netflix Prize, які містять дані про рейтинги, відгуки та активність користувачів, що можуть бути використані для першого тренування моделі.

Попередня обробка даних: Після збору даних, є велика ймовірність, що доведеться провести попередній аналіз даних. Це може включати видалення або заповнення відсутніх значень, перекодування категоріальних змінних, нормалізацію чисельних змінних тощо.

Додавання додаткових функцій: Можна створювати нові функції на основі даних, які допоможуть моделі краще визначати шиллінг-атаки. Наприклад, можна врахувати кількість оцінок, які користувач віддає в одиницю часу, середній рейтинг, який користувач дає товару тощо.

Розроблений алгоритм для виявлення шиллінг-атак з використанням темпоральних знань базується на аналізі поведінкових патернів користувачів. Він складається з кількох послідовних етапів, кожен з яких має свою специфічну задачу, що дозволяє забезпечити точне та надійне виявлення атак. Цей підхід забезпечує збір, обробку та аналіз даних для

виявлення аномалій у поведінці користувачів. Він складається з загальних 6-ти етапів, кожен з яких представлений на рисунку 2.2 та буде описаний нижче.

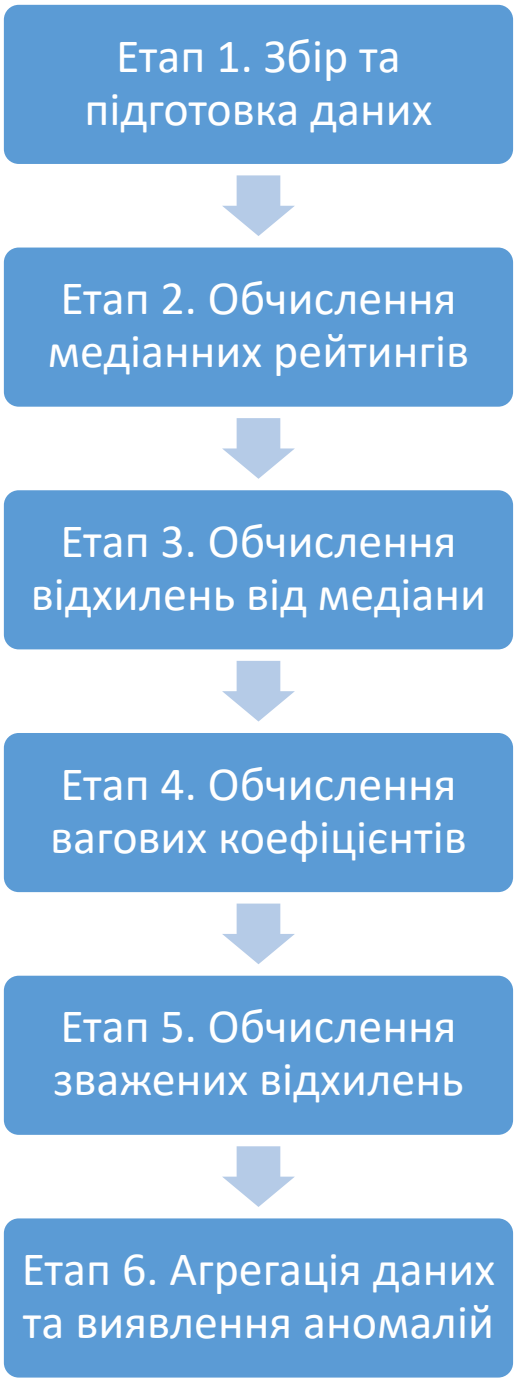

Рисунок 2.2 – Етапи роботи розробленого алгоритму

Робота алгоритму починається зі збору та підготовки даних. Використовується датасет «All Beauty» з Amazon Review Data, який містить інформацію про відгуки на продукти з категорії краси. Дані завантажуються та перетворюються у формат, придатний для подальшого аналізу. Часові позначки перетворюються у формат Unix timestamp, що дозволяє більш ефективно працювати з темпоральними даними.

Далі обчислюються медіанні рейтинги для кожного продукту. Це дозволяє визначити базові рівні оцінок для подальшого порівняння. Медіанні рейтинги використовуються як опорні точки, що допомагають виявити аномалії у поведінці користувачів.

Наступним кроком є обчислення відхилень від медіани для кожного відгуку. Це відхилення дозволяє визначити, наскільки дана оцінка відрізняється від типової. Відхилення, що значно перевищують або нижчі за медіану, можуть свідчити про аномальні дії, які потрібно дослідити.

Важливим етапом є обчислення вагових коефіцієнтів для кожної дії користувача з урахуванням темпоральних аспектів. Ваги визначаються на основі таких факторів, як частота дій, час перегляду сторінок, час залишення відгуків тощо. Врахування цих факторів дозволяє оцінити важливість кожної дії у контексті часу.

Далі вагові коефіцієнти використовуються для обчислення зважених відхилень. Це дозволяє враховувати важливість кожної оцінки у контексті часу, забезпечуючи більш точний аналіз даних та виявлення аномалій.

На останньому етапі зважені відхилення агрегуються для кожного продукту, що дозволяє виявити аномалії, які можуть свідчити про шиллінг-атаки. Порівнюються сумарні вагові відхилення з середніми значеннями, що дозволяє визначити підозрілі патерни у поведінці користувачів.

Розроблений алгоритм для виявлення шиллінг-атак включає всі ці етапи, що забезпечують ефективний збір, обробку та аналіз даних. Використання темпоральних знань та вагових коефіцієнтів дозволяє враховувати динаміку змін інтересів користувачів і забезпечує високу

точність та надійність виявлення атак. Такий підхід дозволяє ефективно ідентифікувати аномалії у поведінці користувачів та запобігати маніпуляціям з рейтингами продуктів, що забезпечує надійність і справедливість систем рекомендацій.

3 ПРОГРАМНА РЕАЛІЗАЦІЯ ТА ЕКСПЕРИМЕНТАЛЬНА ПЕРЕВІРКА

3.1 Потенційні інструменти реалізації

Розробка та впровадження алгоритму виявлення шиллінг-атак з використанням темпоральних знань потребує вибору відповідних інструментів та технологій. Мовою програмування був Python, який є одним з найпопулярніших засобів для обробки даних та машинного навчання. Його широкі бібліотеки, такі як Pandas(використовується для обробки та аналізу табличних даних), NumPy(забезпечує підтримку роботи з багатовимірними масивами та матрицями) та SciPy(доповнює можливості NumPy, надаючи функції для інтеграції, оптимізації та інших наукових обчислень, необхідних для реалізації алгоритму), дозволяють ефективно обробляти великі обсяги даних, виконувати складні обчислення та аналіз. Крім того, Python має потужні інструменти для візуалізації даних, такі як Matplotlib та Seaborn, що дозволяє створювати наочні графіки та діаграми. Крім того, була використана scikit-learn, яка є основною бібліотекою машинного навчання, що була використана для побудови моделей класифікації та кластеризації. Вона забезпечує інструменти для навчання моделей, їхньої валідації та оцінки ефективності.

В майбутньому для реалізації повноцінного застосунку для знаходження, виявлення та подальшої обробки користувачів, які є джерелом шиллінг-атак, можуть бути використані додаткові платформи для обчислень, такі як хмарні платформи, наприклад: Amazon Web Services (AWS), Google Cloud Platform (GCP) або Microsoft Azure, які мають вбудовані ресурси для розробки та тренування моделей, доступ до зберігання великого об’єму даних та подальшої роботи.

3.2 Підготовка вхідних даних для виявлення шиллінг-атак

Правильна обробка та структуризація даних забезпечують ефективність та точність алгоритмів машинного навчання, що використовуються для виявлення аномалій у поведінці користувачів. Розглянемо процес підготовки вхідних даних з використанням датасету «All Beauty» з Amazon Review Data[12].

Використані бібліотеки:

– pandas – бібліотека для обробки та аналізу табличних даних. Вона забезпечує потужні засоби для маніпуляції даними, включаючи їх зчитування, фільтрацію, агрегацію та перетворення. Pandas використовує ефективні алгоритми для роботи з великими обсягами даних. Наприклад, метод `read_json` дозволяє швидко завантажити JSON-файл і перетворити його в DataFrame, який є основною структурою даних в pandas. Функція `groupby` використовується для групування даних за заданим ключем, що дозволяє виконувати агрегаційні операції на групах даних. Метод `join` дозволяє об'єднувати DataFrame на основі спільних колонок, що забезпечує інтеграцію різних наборів даних. Функція `dropna` видаляє рядки з відсутніми значеннями, що забезпечує цілісність даних для аналізу;

– matplotlib – бібліотека для візуалізації даних у вигляді графіків та діаграм. Вона надає широкий спектр інструментів для створення візуалізацій, які допомагають виявляти тенденції та аномалії в даних. Використання matplotlib дозволяє створювати різноманітні типи графіків, включаючи лінійні графіки, гістограми, діаграми розсіювання та інші типи візуалізацій;

– numpy – бібліотека для роботи з багатовимірними масивами та матрицями, а також математичними функціями високого рівня. Numpy забезпечує ефективні методи для обробки числових даних. Наприклад, функція `astype` дозволяє змінювати тип даних масиву, а функція `exp`

обчислює експоненту для масиву, що корисно для обчислення вагових коефіцієнтів.

Для початку необхідно завантажити датасет «All Beauty», який містить інформацію про відгуки користувачів на продукти з категорії краси. Датасет містить такі поля, як `reviewerID`, `asin`, `reviewerName`, `helpful`, `reviewText`, `overall`, `summary`, `unixReviewTime`, `reviewTime`, `verified`, `style`, `vote`, `image`

```
data = pd.read_json('/Users/pogremuhalike/Downloads/All_Beauty_5.json', lines=True)
```

Наступним кроком є перетворення та очищення даних для забезпечення їхньої придатності для аналізу. Час відгуків (`reviewTime`) перетворюється у формат Unix timestamp (`unixReviewTime`), що полегшує подальший аналіз темпоральних патернів.

Час відгуків (`reviewTime`) перетворюється у формат Unix timestamp (`unixReviewTime`), що полегшує подальший аналіз темпоральних патернів.

```
data['reviewTime'] = pd.to_datetime(data['reviewTime'])
data['unixReviewTime'] = data['reviewTime'].astype(np.int64) // 10 ** 9
```

Під капотом `pd.to_datetime` використовує ефективні алгоритми для перетворення текстових дат у формат datetime. Це дозволяє легко виконувати операції порівняння та обчислення з датами. Перетворення у формат Unix timestamp забезпечує уніфіковане представлення часу як числового значення, що спрощує подальші обчислення.

Далі відсутні значення у ключових полях, таких як `overall` (рейтинг) та `unixReviewTime`, повинні бути заповнені або видалені для забезпечення цілісності даних.

```
data.dropna(subset=['overall','unixReviewTime'], inplace=True)
```

Функція `dropna` видаляє всі рядки, які містять відсутні значення у вказаних колонках. Це забезпечує, що всі дані, які використовуються для аналізу, є повними і цілісними.

Також необхідно відфільтрувати дані, залишивши тільки перевірені відгуки (`verified`), що забезпечує більш надійні дані для аналізу.

```
data = data[data['verified'] == True]
```

Фільтрація даних за допомогою умовних виразів дозволяє вибирати тільки ті записи, які відповідають певним критеріям, таким як перевірені відгуки. Це підвищує якість даних для подальшого аналізу.

Для більш детального аналізу необхідно обчислити додаткові показники, які можуть свідчити про аномальну поведінку користувачів. Тож, обчислюються медіанні рейтинги для кожного продукту, що дозволяє визначити базові рівні оцінок для подальшого порівняння.

```
medians = data.groupby('asin')['overall'].median()
data = data.join(medians, on='asin', rsuffix='_median')
```

Метод `groupby` дозволяє групувати дані за унікальними значеннями в колонці `asin`, після чого `median` обчислює медіанне значення для кожної групи. Метод `join` об'єднує отримані медіанні значення з початковим DataFrame, додаючи нову колонку з медіанними значеннями.

Далі для кожного відгуку обчислюється відхилення оцінки від медіанної оцінки продукту, що дозволяє виявити аномальні оцінки.

```
data['delta_R'] = data['overall'] - data['overall_median']
```

Також обчислимо вагові коефіцієнти, вони обчислюються для кожної дії користувача з урахуванням темпоральних аспектів, таких як час перегляду сторінок та частота дій.

```
lambda_decay = 0.01
current_time = data['unixReviewTime'].max()
data['weight'] = np.exp(-lambda_decay * (current_time - data['unixReviewTime']))
```

Функція `np.exp` використовується для обчислення експоненти. Вагові коефіцієнти зменшуються з часом, що дозволяє враховувати важливість кожної оцінки у контексті часу. Це забезпечує, що більш нові оцінки мають більшу вагу, ніж старі.

Після цього обчислюємо зважених відхилень за допомогою вагових коефіцієнтів, що дозволяє врахувати важливість кожної оцінки в контексті часу.

```
data['weighted_delta_R'] = data['weight'] * data['delta_R']
```

Останнім етапом є агрегація зважених відхилень для кожного продукту та виявлення аномалій, які можуть свідчити про шиллінг-атаки. Використовуємо порівняння сумарних вагових відхилень з середніми значеннями.

```
def weighted_rating(group):
    total_weight = group['weight'].sum()
    if total_weight == 0:
        return np.nan

    return (group['weighted_delta_R'].sum()) / total_weight

data_reset = data.reset_index()

# Перетворення індексу назад у стовпець
weighted_ratings = data_reset.groupby('asin').apply(
weighted_rating).reset_index(name='weighted_rating')
anomalies = weighted_ratings[weighted_ratings['weighted_rating'] .abs() > threshold]
```

Функція `weighted_rating` обчислює зважений рейтинг для кожної групи продуктів. Після цього результати групуються за колонкою `asin`, а результат зберігається у новій колонці `weighted_rating`. Аномальні відхилення виявляються шляхом порівняння абсолютних значень зважених рейтингів з заданим порогом (threshold)

Після цього будуємо графіки, щоб показати результат роботи експериментальної перевірки, приклад яких може бути представлений на рисунку 3.1.

```python
# Спрощення графіка
plt.figure(figsize=(14, 7))
plt.plot(interval_data['month'].astype(str),
         interval_data['sales'], 'b--o', label='Sales', markersize=5)
plt.plot(interval_data['month'].astype(str),
         interval_data['avg_rating'], 'r--o', label='Ratings', markersize=5)
plt.plot(interval_discrepancy['month'].astype(str),
         interval_discrepancy['discrepancy'], 'k--o', label='Divergence', markersize=5)

# Визначення можливих атак
threshold = interval_discrepancy['discrepancy'].quantile(0.95)
possible_attack = interval_discrepancy['discrepancy'].apply(lambda x: x if x > threshold else 0)
plt.plot(interval_discrepancy['month'].astype(str),
         possible_attack, 'g-o', label='Possible attack', markersize=5)
```

Рисунок 3.1 – Приклад побудови графіків

3.3 Результати експериментальної перевірки

В рамках реалізації програмного прототипу було створено процес обробки вхідних даних та реалізація алгоритму знаходження шиллінг-атак. Для експериментальної перевірки було обрано декілька датасетів з Amazon Review Data. А також реалізоване формування графіків, на яких можна побачити можливі шиллінг-атаки – як саме результат роботи цієї програми. Приклад вхідних даних у форматі JSON та CSV представлені на рисунках 3.2 та 3.3 відповідно. Дані містять інформацію про відгуки користувачів, включаючи рейтинг, текст відгуку, час відгуку, перевірену покупку та інші метадані.


```
{"overall": 5.0, "verified": true, "reviewTime": "09 1, 2016", "reviewerID": "A3CIUOJXQ5VDQ2", "asin": "B0000530HU", "style": {"Size:": " 7.0 oz", "Flavor:": " Clas
{"overall": 5.0, "verified": true, "reviewTime": "11 14, 2013", "reviewerID": "A3H7T87S984REU", "asin": "B0000530HU", "style": {"Size:": " 7.0 oz", "Flavor:": " Cla
{"overall": 1.0, "verified": true, "reviewTime": "08 18, 2013", "reviewerID": "A3J034YH7UG4KT", "asin": "B0000530HU", "style": {"Size:": " 7.0 oz", "Flavor:": " Cla
{"overall": 5.0, "vote": "25", "verified": false, "reviewTime": "05 3, 2011", "reviewerID": "A2UEO5XR3598GI", "asin": "B0000530HU", "style": {"Size:": " 7.0 oz", "F
{"overall": 5.0, "vote": "3", "verified": true, "reviewTime": "05 6, 2011", "reviewerID": "A3SFRT223XXWF7", "asin": "B00006L9LC", "style": {"Size:": " 200ml/6.7oz"}
{"overall": 5.0, "verified": false, "reviewTime": "05 16, 2010", "reviewerID": "A24HQ2N733ZW7W", "asin": "B00006L9LC", "style": {"Size:": " 366"}, "reviewerName": "
{"overall": 5.0, "verified": false, "reviewTime": "05 7, 2018", "reviewerID": "A2G90R2ZU6KU5D", "asin": "B00006L9LC", "style": {"Size:": " Small"}, "reviewerName":
{"overall": 2.0, "verified": true, "reviewTime": "05 7, 2018", "reviewerID": "A24W4W9E62FZP2", "asin": "B00006L9LC", "style": {"Size:": " Small"}, "reviewerName": "
{"overall": 1.0, "verified": true, "reviewTime": "05 6, 2018", "reviewerID": "A7ID5H7FWLJHC", "asin": "B00006L9LC", "style": {"Size:": " Small"}, "reviewerName": "U
{"overall": 1.0, "verified": true, "reviewTime": "05 6, 2018", "reviewerID": "AYKOSAJTP5AVS", "asin": "B00006L9LC", "style": {"Size:": " Small"}, "reviewerName": "S
{"overall": 5.0, "verified": false, "reviewTime": "05 6, 2018", "reviewerID": "A3HIEBXDI9EQA6", "asin": "B00006L9LC", "style": {"Size:": " Small"}, "reviewerName":
{"overall": 1.0, "verified": true, "reviewTime": "05 5, 2018", "reviewerID": "A2YDF506DAA5W4", "asin": "B00006L9LC", "style": {"Size:": " Small"}, "reviewerName": "
{"overall": 5.0, "verified": false, "reviewTime": "05 5, 2018", "reviewerID": "AKFUG2XFPPWWM", "asin": "B00006L9LC", "style": {"Size:": " Small"}, "reviewerName": "
{"overall": 5.0, "verified": false, "reviewTime": "05 5, 2018", "reviewerID": "A1T7DRMZ9A9KEA", "asin": "B00006L9LC", "style": {"Size:": " Small"}, "reviewerName":
{"overall": 1.0, "verified": true, "reviewTime": "05 3, 2018", "reviewerID": "AEL1DK2OJ41ZZ", "asin": "B00006L9LC", "style": {"Size:": " Small"}, "reviewerName": "A
{"overall": 1.0, "verified": true, "reviewTime": "04 30, 2018", "reviewerID": "A2UM2UI2KVHG64", "asin": "B00006L9LC", "style": {"Size:": " Small"}, "reviewerName":
{"overall": 1.0, "verified": true, "reviewTime": "04 28, 2018", "reviewerID": "ADUKTDKBY4CNP", "asin": "B00006L9LC", "style": {"Size:": " Small"}, "reviewerName": "
{"overall": 5.0, "verified": true, "reviewTime": "04 28, 2018", "reviewerID": "ADW00ZRLWD3KG", "asin": "B00006L9LC", "style": {"Size:": " Small"}, "reviewerName": "
{"overall": 5.0, "verified": true, "reviewTime": "04 25, 2018", "reviewerID": "A1118RD3AJD5KH", "asin": "B00006L9LC", "style": {"Size:": " 511"}, "reviewerName": "D
```


Рисунок 3.2 – Приклад вхідних даних у форматі json

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	overall	verified	reviewTime	reviewerID	asin	style/Size:	style/Fla	reviewerName	reviewText	summary	unixReviewTim	vote	image/0	image/1
2	5,0	true	09 1, 2016	A3CIUOJXQ5VDQ2	B0000530HU	7.0 oz	Classic I	Shelly F	As advertised. Reasonably pri	Five Stars	1472688000			
3	5,0	true	11 14, 2013	A3H7T87S984REU	B0000530HU	7.0 oz	Classic I	houserules18	Like the oder and the feel whe	Good for the face	1384387200			
4	1,0	true	08 18, 2013	A3J034YH7UG4KT	B0000530HU	7.0 oz	Classic I	Adam	I bought this to smell nice afte	Smells awful	1376784000			
5	5,0	false	05 3, 2011	A2UEO5XR3598GI	B0000530HU	7.0 oz	Classic I	Rich K	HEY!! I am an Aqua Velva Man					
6	5,0	true	05 6, 2011	A3SFRT223XXWF7	B00006L9LC	200ml/6.7oz		C. C. Christian	If you ever want to feel pampe	Bvlgari Shampoo	1304640000	3		
7	5,0	false	05 16, 2010	A24HQ2N7332W7W	B00006L9LC	366		Kindle Customer Jo	If you know the scent of Diva,	Diva is Heavenly	1273968000			
8	5,0	false	05 7, 2018	A2G90R2ZU6KU5D	B00006L9LC	Small		Mike	Got this shampoo as a solutio	Outstanding, no complai	1525651200			
9	2,0	true	05 7, 2018	A24W4W9E62FZP2	B00006L9LC	Small		Reb	No change my scalp still itche	No change my scalp still i	1525651200			
10	1,0	true	05 6, 2018	A7ID5H7FWLJHC	B00006L9LC	Small		U. V.	Too expensive for such poor q	Too expensive for such pc	1525564800			
11	1,0	true	05 6, 2018	AYKOSAJTP5AVS	B00006L9LC	Small		Senthil Kumar M	It dries my hair, doesnt help t	Dries my hair, doesnt hel	1525564800			
12	5,0	false	05 6, 2018	A3HIEBXDI9EQA6	B00006L9LC	Small		john robbins	Outstanding! Top organic sha	Five Stars	1525564800			
13	1,0	true	05 5, 2018	A2YDF506DAA5W4	B00006L9LC	Small		justdancing	So watered down, I didn't feel	I didn't feel like it was act	1525478400			
14	5,0	false	05 5, 2018	AKFUG2XFPPWWM	B00006L9LC	Small		Donovan	10 stars right here. This produ	Five Stars	1525478400			
15	5,0	false	05 5, 2018	A1T7DRMZ9A9KEA	B00006L9LC	Small		Robert Roehlk	First hair care product I've de	First hair care product I'v	1525478400			
16	1,0	true	05 3, 2018	AEL1DK2OJ41ZZ	B00006L9LC	Small		Amazon Customer	Made dandruff worse and irrit	One Star	1525305600			
17	1,0	true	04 30, 2018	A2UM2UI2KVHG64	B00006L9LC	Small		Lorie B.	Worst shampoo Ive ever used	Worst shampoo Ive ever	1525046400			
18	1,0	true	04 28, 2018	ADUKTDKBY4CNP	B00006L9LC	Small		Andelika	Made my hair brittle and dull	Disappointment...	1524873600			

Рисунок 3.3 – Приклад вхідних даних у форматі csv

Після обробки даних та обчислення вагових коефіцієнтів було проведено агрегацію зважених відхилень для кожного продукту. На основі цього аналізу було виявлено аномалії, які можуть свідчити про шиллінг-атаки. Для візуалізації результатів було створено три графіки, що показують можливі шиллінг-атаки, їх можна побачити на рисунку 3.4, 3.5 та 3.6.

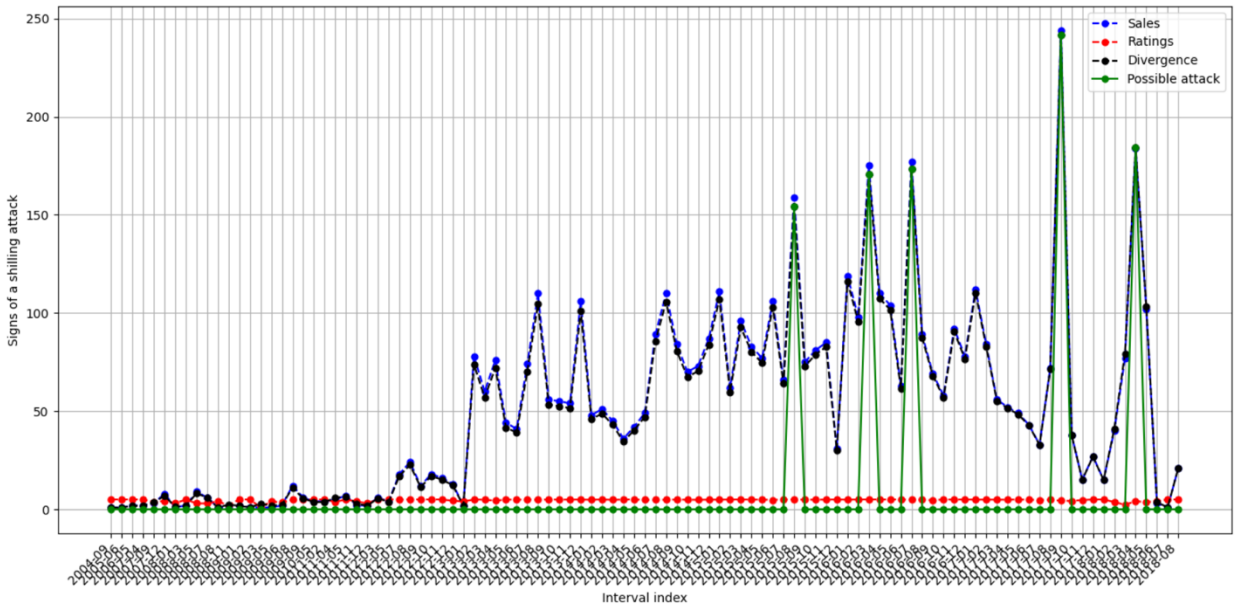

Рисунок 3.4 – Результат роботи програми для All_Beauty

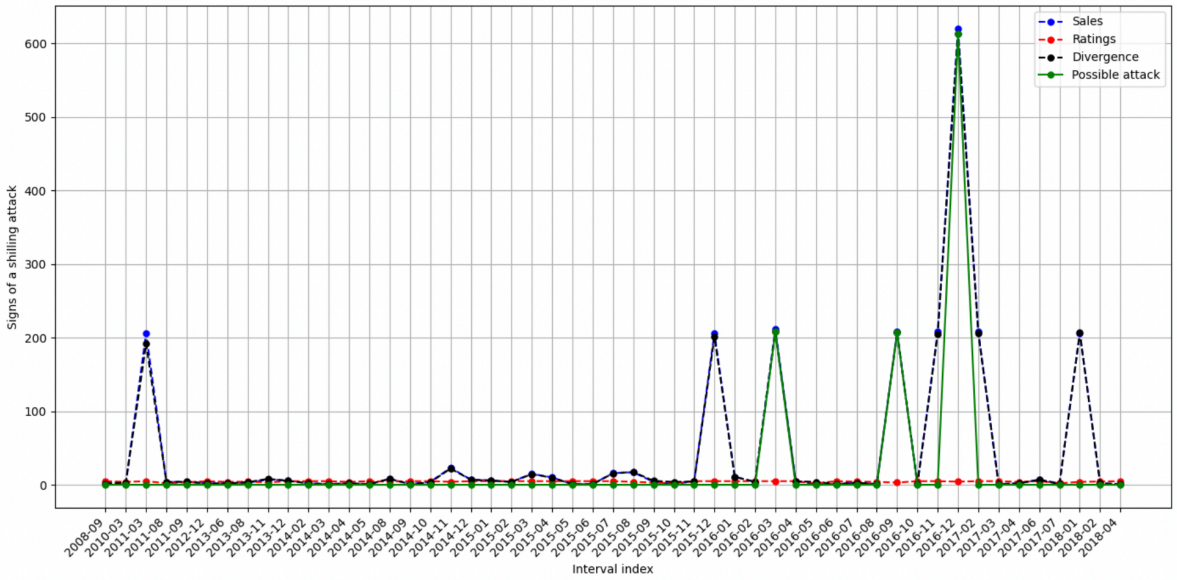

Рисунок 3.5 – Результат роботи програми для Appliances

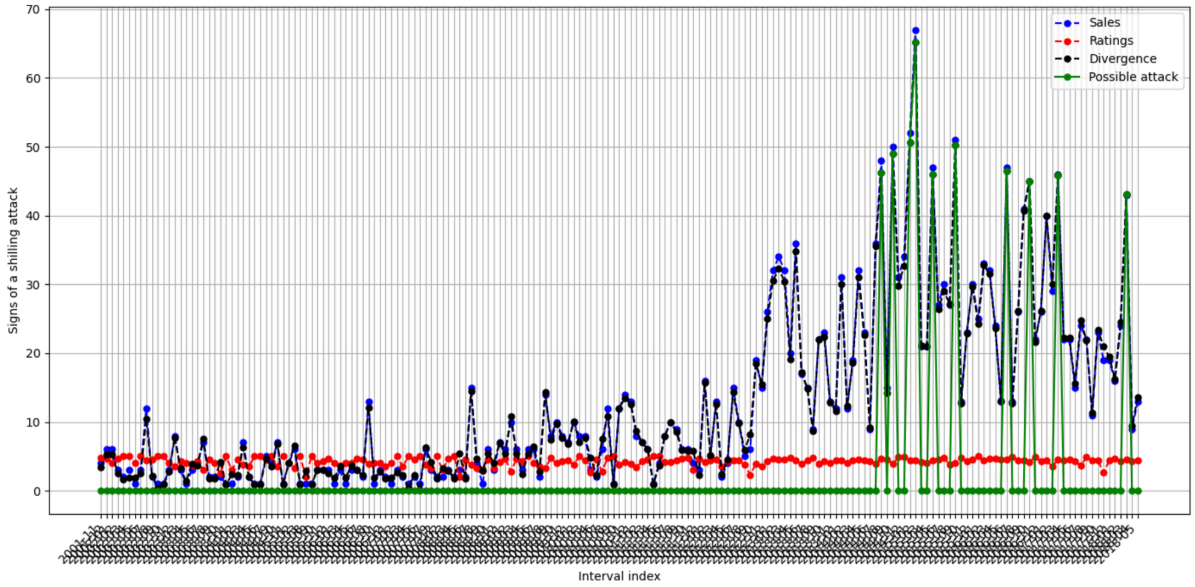

Рисунок 3.5 – Результат роботи програми для Magazine_Subscriptions

На рисунках 3.4, 3.5 та 3.6 представлені наступні дані:

– синя лінія(Sales) - частота продажів у певні часові інтервали;

– червона лінія(Ratings) - оцінки, залишені користувачами у той самий період;

– чорна лінія(Divergence) - відхилення по темпоральним правилам;

– зелена лінія(Possible attack) - моменти часу, коли було виявлено можливі шиллінг-атаки на основі аналізу аномалій.

Тож, можна зробити висновок, що експериментальна перевірка алгоритму для виявлення шиллінг-атак продемонструвала його ефективність у виявленні аномалій у поведінці користувачів. Завдяки використанню темпоральних знань та обчисленню вагових коефіцієнтів вдалося ідентифікувати підозрілі патерни у відгуках, що можуть свідчити про маніпуляції з рейтингами продуктів. Візуалізація результатів аналізу допомагає наочно представити виявлені аномалії та забезпечити розуміння процесу виявлення шиллінг-атак.

3.3.1 Порівняння результатів з існуючими рішеннями

У методі на основі контентного фільтрування використовується інформація про характеристики продукту для рекомендацій. Воно аналізує опис продукту, ключові слова, категорії та інші атрибути для визначення подібних продуктів. Цей метод зазвичай не враховує темпоральні аспекти і може бути менш точним у виявленні аномальних патернів у поведінці користувачів, що є його недоліком. Проте його простота реалізації, висока швидкість обробки даних, можливість працювати з новими або рідко оцінюваними продуктами, робить з цього підходу конкурента.

Метод на основі колаборативного фільтрування базується на історії поведінки користувачів та їхніх уподобаннях. Існують кілька підходів до виявлення шиллінг-атак у рамках цього методу, включаючи аналіз кореляцій між користувачами та обчислення аномалій у рейтингах, також, недоліком є те, що методи на основі колаборативного фільтрування можуть не враховувати час, що знижує їхню точність у виявленні аномалій,

особливо якщо аномалії мають темпоральний характер. Можливі проблеми з новими або рідко оцінюваними продуктами (проблема холодного старту).

Також, є методи на основі гібридного фільтрування, які поєднують елементи контентного та колаборативного. Цей підхід використовує як інформація про продукт, так і поведінкові дані про користувачів. З переваг можна виділити високу точність рекомендацій, здатність працювати з різними типами даних та можливість врахування різних аспектів інформації. Проте, ці методи можуть бути більш складними для реалізації та обробки, так як вони вимагають значних обчислювальних ресурсів.

Досліджуваний підхід до виявлення шиллінг-атак базується на використанні темпоральних знань та вагових коефіцієнтів, що дозволяє враховувати часові аспекти поведінки користувачів. У таблиці 3.1 представлено порівняння методів.

Таблиця 3.1 – Порівняння методів виявлення шиллінг-атак

Параметр	Контентне фільтрування	Колаборативне фільтрування	Гібридне фільтрування	Досліджуваний метод
Точність виявлення	Низька	Середня	Висока	Висока
Врахування темпоральних аспектів	Немає	Обмежене	Обмежене	Повне
Швидкість обробки	Висока	Середня	Середня	Висока

Продовження таблиці 3.1

Параметр	Контентне фільтрування	Колаборативне фільтрування	Гібридне фільтрування	Досліджуваний метод
Складність реалізації	Низька	Середня	Висока	Середня
Обчислювальні ресурси	Низькі	Середні	Високі	Середні

Контентне фільтрування відрізняється простотою реалізації та високою швидкістю обробки даних. Цей метод добре підходить для роботи з новими або рідко оцінюваними продуктами, що є його головною перевагою. Проте, основним недоліком є неврахування темпоральних аспектів, що знижує точність у виявленні шиллінг-атак. Методи на основі контентного фільтрування можуть пропустити аномальні патерни у поведінці користувачів, які мають часовий характер.

Колаборативне фільтрування забезпечує середню точність виявлення шиллінг-атак та враховує поведінкові дані користувачів. Недоліком є можливість неврахування часових аспектів, що також впливає на точність виявлення аномалій. Проблема холодного старту може виникати при роботі з новими або рідко оцінюваними продуктами, що потребує додаткових обчислювальних ресурсів для її вирішення.

Гібридне фільтрування поєднує переваги обох підходів, забезпечуючи високу точність рекомендацій та здатність працювати з різними типами даних. Цей метод враховує різні аспекти інформації про продукти та поведінку користувачів, що робить його потужним інструментом для виявлення шиллінг-атак. Проте, висока складність реалізації та потреба у значних обчислювальних ресурсах можуть бути серйозними перешкодами для його впровадження.

Досліджуваний метод, що базується на темпоральних знаннях, забезпечує високу точність виявлення аномалій у поведінці користувачів. Використання темпоральних аспектів дозволяє виявляти складні патерни та зміни у поведінці користувачів, які можуть вказувати на шиллінг-атаки. Цей підхід також забезпечує високу швидкість обробки даних та середню складність реалізації, що робить його ефективним рішенням для виявлення аномалій у рекомендаційних системах.

Порівняння досліджуваного підходу з існуючими методами шиллінг-атак демонструє, що використання темпоральних знань дозволяє значно підвищити точність виявлення аномалій у поведінці користувачів. Використання ефективних інструментів для обробки даних забезпечує високу швидкість обробки, а також спрощує процес реалізації алгоритму. Хоча деякі існуючі методи можуть бути простішими у реалізації, наш підхід забезпечує кращі результати завдяки врахуванню часових аспектів та динаміки змін у поведінці користувачів.

Для подальшого підвищення ефективності досліджуваного методу можна розглянути впровадження додаткових технік машинного навчання, таких як глибокі нейронні мережі та алгоритми підкріплювального навчання. Також перспективним напрямком є використання розподілених обчислень та хмарних платформ для обробки великих обсягів даних у реальному часі. Це дозволить ще більше підвищити швидкість та точність виявлення аномалій, забезпечуючи надійну та ефективну роботу рекомендаційних систем.

3.4 Аналіз перспектив і варіанти застосування

Розглянемо можливе подальше вдосконалення та використання застосованого методу.

Одним з основних застосувань алгоритму є виявлення фальшивих відгуків на платформах електронної комерції. Це дозволяє захистити

покупців від маніпуляцій з рейтингами товарів, підвищити довіру до відгуків та забезпечити чесну конкуренцію між продавцями. Алгоритм допомагає виявляти недобросовісних конкурентів, які намагаються знизити рейтинги продукції конкурентів шляхом залишення негативних відгуків. Використання цього методу може значно покращити загальний досвід користувачів, підвищуючи їх задоволеність і довіру до платформи.

У системах бронювання готелів, ресторанів та інших послуг алгоритм може використовуватися для виявлення маніпуляцій з рейтингами, що забезпечує справедливу оцінку послуг і підвищує довіру користувачів. Виявлення аномалій у відгуках дозволяє покращити якість рекомендацій, забезпечуючи більш точні та надійні результати для користувачів. Це важливо для споживачів, які приймають рішення на основі відгуків та рейтингів, а також для постачальників послуг, які прагнуть підтримувати високий рівень обслуговування та репутації.

У сфері соціальних мереж може використовуватися як боротьба з ботами та фальшивими акаунтами, зокрема виявляючи такі акаунти, які створюють штучні взаємодії, зокрема, лайки, коментарі та перегляди. Що допомагає забезпечити автентичність контенту, підвищуючи загальну якість та надійність платформи. Крім того, це сприяє збереженню довіри користувачів, які можуть бути впевнені в тому, що контент і взаємодії є справжніми.

Алгоритм також має потенціал для використання у фінансовій сфері, зокрема для виявлення шахрайських транзакцій та інших аномалій у фінансових даних. Аналіз темпоральних патернів може допомогти у виявленні підозрілих операцій, які відрізняються від нормальної поведінки користувачів. Це може сприяти підвищенню безпеки фінансових систем і зменшенню ризиків, пов'язаних з шахрайством.

У сфері охорони здоров'я алгоритм може бути використаний для аналізу електронних медичних записів та виявлення аномалій у даних пацієнтів. Це може допомогти у виявленні потенційних проблем зі

здоров'ям на ранніх стадіях та сприяти більш ефективному лікуванню. Аналіз темпоральних даних може також допомогти у відстеженні змін у стані пацієнтів і прогнозуванні можливих ускладнень.

Щодо майбутніх перспектив, можливо подальше вдосконалення алгоритму шляхом інтеграції з іншими методами машинного навчання та штучного інтелекту. Використання глибоких нейронних мереж, методів підкріплювального навчання та розподілених обчислень може значно підвищити ефективність та точність виявлення шиллінг-атак. Залучення розподілених обчислювальних ресурсів, таких як хмарні платформи, дозволить обробляти великі обсяги даних у реальному часі, що є важливим для забезпечення оперативного реагування на виявлені загрози.

Таким чином, розроблений метод має широкий спектр застосувань у різних галузях, включаючи електронну комерцію, соціальні мережі, фінансові системи та охорону здоров'я. Його подальший розвиток та вдосконалення можуть сприяти підвищенню безпеки, надійності та ефективності систем, забезпечуючи кращий захист від маніпуляцій та шахрайства.

ВИСНОВКИ

У ході виконання кваліфікаційної роботи було проведено аналіз існуючих методів та підходів до виявлення шиллінг-атак у рекомендаційних системах. Аналіз показав, що багато сучасних систем не враховують темпоральні аспекти поведінки користувачів, що може призвести до невиявлення атак, які мають часові патерни. Це означає, що зловмисники можуть ефективно обходити традиційні методи виявлення атак, що засновані лише на статичних даних.

Для вирішення цього недоліку було обрано використовувати темпоральні правила, які враховують часові залежності між подіями. Цей підхід дозволяє системі краще розуміти та аналізувати поведінкові патерни користувачів з урахуванням часу, що підвищує ефективність виявлення шиллінг-атак. Темпоральні правила, такі як «Next» та «Future», дозволяють виявляти складніші взаємозв'язки між подіями та краще прогнозувати потенційні атаки.

У рамках даної роботи було розроблено програму для виявлення шиллінг-атак у рейтингах користувачів на основі темпоральних знань. Програма використовує аналіз часових залежностей та обчислює ваги відхилень рейтингів від медіанних значень для кожного продукту. Це дозволяє виявляти аномалії у поведінці користувачів та підвищує точність виявлення потенційних атак.

Основною складовою системи є алгоритм, що використовує темпоральні правила для аналізу поведінки користувачів. Було розроблено прототип методу на мові програмування Python та проведено кілька експериментів з використанням даних з Amazon Review Data. Додатково було порівняно результати з існуючими методами виявлення шиллінг-атак, що дозволило оцінити ефективність запропонованого підходу.

Подальші перспективи охоплюють розширення і вдосконалення системи. Зокрема, планується інтеграція додаткових методів машинного

навчання для покращення точності виявлення атак. Також планується розробка мобільного додатку, що дозволить користувачам легко отримувати рекомендації та повідомлення про потенційні атаки безпосередньо на своєму смартфоні. Вдосконалення темпоральних моделей та їх адаптація до різних типів даних також є важливим напрямком для подальших досліджень.

Практична значущість розробленого методу полягає у його здатності забезпечувати більш точне виявлення шиллінг-атак у рекомендаційних системах. Це має важливе значення для багатьох комерційних платформ, таких як інтернет-магазини, стримінгові сервіси та соціальні мережі, де довіра користувачів до рекомендацій є критично важливою. Запропонований підхід допомагає захистити інтереси споживачів, забезпечуючи чесну конкуренцію та покращуючи загальну якість сервісу.

Розроблена система також може бути використана для підвищення ефективності інших алгоритмів, які враховують поведінкові патерни користувачів у часі. Наприклад, вона може бути інтегрована в системи персоналізованого маркетингу для більш точного визначення цільової аудиторії або в системи безпеки для виявлення підозрілої активності.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Hastie, T., Tibshirani, R., & Friedman, J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer-Verlag New York, Inc., 2009. URL: https://web.stanford.edu/~hastie/ElemStatLearn/ (дата звернення: 15.04.2024).

2. Geron, A. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. O'Reilly Media, Inc., 2019. URL: https://www.oreilly.com/library/view/hands-on-machine-learning/9781492032632/ (дата звернення: 15.04.2024).

3. Zhou, W., Yin, Y., Zhou, Y., & Zhang, J. Shilling attack detection for recommender systems based on credibility of group users and rating time series. Knowledge-Based Systems, 2018. URL: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0196533 (дата звернення: 16.04.2024).

4. Lam, S.K., Frankowski, D., & Riedl, J. «Preventing shilling attacks in online recommender systems.» URL: https://www.researchgate.net/publication/220759092_Preventing_shilling_attacks_in_online_recommender_systems (дата звернення: 17.04.2024).

5. Hendrawan, R., Hansun, S., & Indrawan, P. «Shilling attack detection utilizing semi-supervised learning method for collaborative recommender system.» URL: https://www.researchgate.net/publication/257676543_Shilling_attack_detection_utilizing_semi-supervised_learning_method_for_collaborative_recommender_system (дата звернення: 17.04.2024).

6. Bishop, C.M. Pattern Recognition and Machine Learning. Springer-Verlag New York, Inc., 2006. URL: https://www.academia.edu/17851990 (дата звернення: 20.04.2024).

7. Mehta, B., Nejdl, W., & Bao, S. «Defending Shilling Attacks in Recommender Systems using Soft Co-Clustering.» URL: https://www.researchgate.net/publication/315848844_Defending_Shilling_Attacks_in_Recommender_Systems_using_Soft_Co-Clustering (дата звернення: 18.04.2024).

8. Breiman, L. Random Forests. Machine Learning, 2001. URL: https://link.springer.com/article/10.1023/A:1010933404324 (дата звернення: 19.04.2024).

9. Zhang, M., Zhou, Z., & Chen, X. «Defending Recommender Systems against Shilling Attacks Using Soft Co-Clustering.» In Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence (IJCAI-11), 2011. URL: https://www.ijcai.org/Proceedings/11/Papers/382.pdf (дата звернення: 19.04.2024).

10. Mobasher, B., Burke, R., Williams, C., & Bhaumik, R. «Analysis and Detection of Segment-Focused Attacks Against Collaborative Recommendation.» In Proceedings of the 2005 IEEE/WIC/ACM International Conference on Web Intelligence (WI'05), 2005. URL: https://ieeexplore.ieee.org/document/5365077 (дата звернення: 24.04.2024).

11. Chala, O., Novikova, L., & Chernyshova, L. "Method for detecting shilling attacks based on implicit feedback in recommender systems." URL: https://journal.eu-jr.eu/engineering/article/view/1402/1292 (дата звернення: 24.04.2024)

12. Ni, J. Amazon Review Data (2018). URL: https://nijianmo.github.io/amazon/index.html (дата звернення: 24.04.2024).

13. Jannach, D., Lerche, L., & Jugovac, M. «Adaptation and Evaluation of Recommendations for Short-Term Shopping Goals.» Artificial Intelligence Review, 2012. URL: https://link.springer.com/article/10.1007/s10462-012-9364-9 (дата звернення: 24.04.2024).

14. Adomavicius, G., & Tuzhilin, A. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions.

IEEE Transactions on Knowledge and Data Engineering, 2005. URL: https://ieeexplore.ieee.org/document/1423975 (дата звернення: 23.04.2024).

15. Su, X., & Khoshgoftaar, T. M. A survey of collaborative filtering techniques. Advances in Artificial Intelligence, 2009. URL: https://www.hindawi.com/journals/aai/2009/421425/ (дата звернення: 24.04.2024).

16. Pathak, B., Garfinkel, R., Gopal, R., Venkatesan, R., & Yin, F. «Empirical Analysis of the Impact of Recommender Systems on Sales.» ACM Transactions on Management Information Systems, 2022. URL: https://dl.acm.org/doi/10.2753/MIS0742-1222270205 (дата звернення: 24.04.2024).

17. Aggarwal, C. C. Recommender Systems: The Textbook. Springer International Publishing, 2016. URL: https://link.springer.com/book/10.1007/978-3-319-29659-3 (дата звернення: 21.04.2024).

18. Koren, Y., Bell, R., & Volinsky, C. Matrix Factorization Techniques for Recommender Systems. Computer, 2009. URL: https://dl.acm.org/doi/10.1109/MC.2009.263 (дата звернення: 23.04.2024).

19. Shi, Y., Larson, M., & Hanjalic, A. Collaborative filtering beyond the user-item matrix: A survey of the state of the art and future challenges. ACM Computing Surveys, 2014. URL: https://dl.acm.org/doi/10.1145/2556270 (дата звернення: 18.04.2024).

20. Harper, F. M., & Konstan, J. A. The MovieLens Datasets: History and Context. ACM Transactions on Interactive Intelligent Systems, 2016. URL: https://dl.acm.org/doi/10.1145/2827872 (дата звернення: 22.04.2024).

21. Rendle, S., Freudenthaler, C., Gantner, Z., & Schmidt-Thieme, L. BPR: Bayesian Personalized Ranking from Implicit Feedback. In Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence (UAI), 2009. URL: https://dl.acm.org/doi/10.5555/1795114.1795167 (дата звернення: 24.04.2024).

22. Sharma, M., & Singh, S. «A Comprehensive Survey on Recommender Systems: Taxonomies, Challenges, and Future Directions.» ACM Computing Surveys, 2021. URL: https://dl.acm.org/doi/10.1145/3450003 (дата звернення: 21.04.2024).

23. Adomavicius, G., & Tuzhilin, A. «Context-Aware Recommender Systems.» In Proceedings of the 2008 ACM Conference on Recommender Systems, 2008. URL: https://dl.acm.org/doi/10.1145/1454008.1454068 (дата звернення: 20.04.2024).

24. Guo, G., Zhang, J., & Yorke-Smith, N. «A Novel Recommendation Model Regularized with User Trust and Item Ratings.» IEEE Transactions on Knowledge and Data Engineering, 2015. URL: https://ieeexplore.ieee.org/document/7404241 (дата звернення: 17.04.2024).

25. He, X., Liao, L., Zhang, H., Nie, L., Hu, X., & Chua, T. S. «Neural Collaborative Filtering.» In Proceedings of the 26th International Conference on World Wide Web, 2017. URL: https://dl.acm.org/doi/10.1145/3038912.3052569 (дата звернення: 23.04.2024).

26. Zhang, S., Yao, L., Sun, A., & Tay, Y. «Deep Learning Based Recommender System: A Survey and New Perspectives.» ACM Computing Surveys, 2019. URL: https://dl.acm.org/doi/10.1145/3285029 (дата звернення: 22.04.2024).

27. Li, C., Zhu, H., Luo, X., Cai, Y., & Zhao, X. «Cross-Domain Recommendation: Challenges, State-of-the-Art and Future Directions.» In Proceedings of the 2019 World Wide Web Conference, 2019. URL: https://www.researchgate.net/publication/349727889_Cross-Domain_Recommendation_Challenges_Progress_and_Prospects (дата звернення: 24.04.2024).

28. Li, X., & Zhan, Z. «A Study of Two Different Methods for Solving the Recommender System Problem.» In Proceedings of the 2014 IEEE International Conference on Data Mining (ICDM), 2014. URL: https://ieeexplore.ieee.org/document/6889419 (дата звернення: 20.04.2024).