

Міністерство освіти і науки України  
Харківський національний університет радіоелектроніки

Факультет Комп'ютерних наук  
(повна назва)

Кафедра Інформаційних управляючих систем  
(повна назва)

## КВАЛІФІКАЦІЙНА РОБОТА Пояснювальна записка

рівень вищої освіти другий (магістерський)

Дослідження моделей соціальних мереж в задачах виявлення уподобань користувачів  
(тема)

Виконала:  
студентка 2 курсу, групи ІУСТМ-22-1

Олена КУЛИКОВА

(Власне ім'я ПРІЗВИЩЕ)

Спеціальність 122 Комп'ютерні науки

(код і повна назва спеціальності)

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

Освітня програма Інформаційні управляючі системи та технології

( повна назва освітньої програми)

Керівник Тетяна БІЛОВА  
(Власне ім'я ПРІЗВИЩЕ)

Допускається до захисту

Зав. кафедри

  
(підпис)


Костянтин ПЕТРОВ  
(Власне ім'я ПРІЗВИЩЕ)

2024

## Харківський національний університет радіоелектроніки

Факультет Комп'ютерних наук  
 Кафедра Інформаційних управляючих систем  
 Рівень вищої освіти другий (магістерський)  
 Спеціальність 122 Комп'ютерні науки  
(код і повна назва)  
 Тип програми освітньо-професійна  
(освітньо-професійна або освітньо-наукова)  
 Освітня програма Інформаційні управляючі системи та технології  
(повна назва)

ЗАТВЕРДЖУЮ:

Зав. кафедри   
(підпис)  
 « 20 » листопада 2023 р.

## ЗАВДАННЯ НА КВАЛІФІКАЦІЙНУ РОБОТУ

студентці КУЛИКОВІЙ Олені Володимирівні  
(прізвище, ім'я, по батькові)

1. Тема роботи Дослідження моделей соціальних мереж в задачах виявлення уподобань користувачів

затверджена наказом університету від 16 листопада 2023 р. № 1359Ст

2. Термін подання студентом роботи до екзаменаційної комісії 17 січня 2024 р.

3. Вихідні дані до роботи: Науково-технічна література, публікації, інформація з інтернет-ресурсів

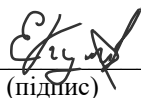
4. Перелік питань, що потрібно опрацювати в роботі: Аналіз вподобань користувачів, аналіз моделей соціальних мереж, огляд існуючих інформаційних систем аналізу соціальних мереж, удосконалення моделей аналізу соціальних мереж, розроблення моделей додатку для аналізу соціальних мереж, реалізація розробленої моделі, експериментальна перевірка отриманих результатів, рекомендації щодо використання системи

## КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Терміни виконання етапів роботи	Примітка
1	Аналіз матеріалів з теми роботи	20.11.2023 – 28.11.2023	Виконано
2	Аналіз вподобань користувачів в соціальних мережах	28.11.2023 – 03.12.2023	Виконано
3	Аналіз моделей соціальних мереж	03.12.2023 – 06.12.2023	Виконано
4	Огляд існуючих інформаційних систем аналізу соціальних мереж в задачах виявлення	06.12.2023 – 09.12.2023	Виконано
5	Огляд моделей аналізу соціальних мереж та їх особливостей	09.12.2023 – 12.12.2023	Виконано
6	Удосконалення моделей аналізу соціальних мереж	12.12.2023 – 18.12.2023	Виконано
7	Розроблення моделі даних додатку аналізатору	12.12.2023 – 18.12.2023	Виконано
8	Діаграми класів, компонентів та структурно функціональна модель додатку	18.12.2023 – 20.12.2023	Виконано
9	Модель варіантів використання додатку та загальна архітектура додатку	20.12.2023 – 22.12.2023	Виконано
10	Реалізація розробленої моделі аналізу соціальних	22.12.2023 – 27.12.2023	Виконано
11	Експериментальна перевірка отриманих	27.12.2023 – 31.12.2023	Виконано
12	Рекомендації щодо використання системи	31.12.2023 – 02.01.2023	Виконано
13	Підготовка пояснювальної записки та графічного матеріалу	02.01.2023 – 07.01.2023	Виконано
14	Захист перед екзаменаційною комісією	18.01.2023	Виконано

Дата видачі завдання 20 листопада 2023 р.

Студент \_\_\_\_\_

  
(підпис)

Керівник роботи \_\_\_\_\_



(підпис)

доц. каф. ІУС Тетяна БІЛОВА

(посада, власне ім'я, прізвище)

## РЕФЕРАТ

Пояснювальна записка містить: 90 с., 1 табл., 42 рис., 1 дод., 21 джерело.

### АНАЛІЗ СОЦІАЛЬНИХ МЕРЕЖ, РЕКОМЕНДАЦІЇ, РЕКОМЕНДАЦІЙНА СИСТЕМА , СОЦІАЛЬНА МЕРЕЖА.

Об'єктом дослідження є процес визначення вподобань користувачів у соціальних мережах.

Предметом дослідження є методи та моделі визначення вподобань користувачів соціальних мереж на основі рекомендаційних систем.

Мета роботи – дослідження методів та моделей для ефективного визначення вподобань користувачів у соціальних мережах. Дослідження зосереджено на використанні моделей рекомендаційних систем як основного інструменту для аналізу та прогнозування індивідуальних інтересів користувачів.

Методи дослідження: модель Neu-CF, багатошаровий перцептрон та нейронні колаборативні шари для моделювання взаємодій користувачів та об'єктів, поєднання матричної факторизації та архітектури MLP для ефективного прогнозування рекомендацій.

## ABSTRACT

The explanatory note contains: 90 p., 1 tabl., 42 pic., 1 add., 21 sources.

SOCIAL NETWORK, SOCIAL NETWORK ANALYSIS,  
RECOMMENDATIONS, RECOMMENDATION SYSTEM.

The object of the study is the process of determining user preferences in social networks.

The subject of the study is methods and models for determining the preferences of social network users based on recommendation systems.

The purpose of this work is to research methods and models for effectively determining user preferences in social networks. The research focuses on the use of recommendation system models as the main tool for analyzing and predicting individual user interests.

Research methods: Neu-CF model, multilayer perceptron and neural collaborative layers to model user-object interactions, combining matrix factorization and MLP architecture for efficient recommendation prediction.

## ЗМІСТ

Скорочення та умовні позначки.....	7
Вступ.....	8
1 Аналіз задач виявлення вподобань користувачів в соціальних мережах.....	9
1.1 Аналіз вподобань користувачів в соціальних мережах.....	9
1.2 Аналіз моделей соціальних мереж.....	12
1.3 Огляд існуючих інформаційних систем аналізу соціальних мереж в задачах виявлення вподобань для отримання рекомендацій.....	18
1.4 Постановка задачі дослідження.....	24
2 Дослідження моделей соціальних мереж.....	26
2.1 Огляд моделей аналізу соціальних мереж та їх особливостей.....	26
2.2 Удосконалення моделей аналізу соціальних мереж.....	35
3 Розроблення моделей додатку для аналізу соціальних мереж.....	42
3.1 Розроблення моделі даних додатку аналізатору.....	42
3.2 Діаграми класів та структурно функціональна модель додатку.....	45
3.3 Модель варіантів використання додатку та загальна архітектура додатку.....	48
3.4 Проектування системи додатку.....	50
4 Реалізація та перевірка роботи додатку аналізу соціальних мереж .....	54
4.1 Реалізація розробленої моделі аналізу соціальних мереж.....	54
4.2 Експериментальна перевірка отриманих результатів.....	56
4.3 Рекомендації щодо використання системи.....	60
Висновки.....	63
Перелік джерел посилання.....	64
Додаток А Графічний матеріал.....	67

## СКОРОЧЕННЯ ТА УМОВНІ ПОЗНАКИ

БД – база даних

ІС – інформаційна система

ОС – операційна система

ПЗ – програмне забезпечення

ПК – персональний комп'ютер

LSTM - Long Short-Term Memory

Neu-CF - Neural Collaborative Filtering

MLP - Multilayer Perceptron

## ВСТУП

У сучасному світі, насиченому соціальними мережами та медіа, спосіб взаємодії та спілкування значно трансформувалася завдяки розповсюдженню Інтернету. Соціальні платформи вплинули на спілкування, соціалізацію та інтенсивність взаємодії між людьми. З цим з'явилися нові можливості для маркетологів та підприємців, які тепер можуть краще представити свої продукти широкому загалу. Соціальні медіа надають фірмам різні способи співпраці, такі як підвищення бренду, підтримка клієнтського спілкування, збільшення обсягів продажів та обмін інформацією в контексті бізнесу.

Останнє десятиліття принесло з собою збільшення ролі онлайн-соціальних мереж, які функціонують за допомогою веб-технологій і захоплюють понад 70% користувачів Інтернету. Вони не лише впливають на поведінку у мережі, але також стають ключовим чинником в управлінні соціально-економічними системами. Це відкриває нові можливості для ефективного аналізу соціальних мереж з метою ідентифікації багатофакторних залежностей, прогнозування поширення і сприйняття інформації, а також розробки рекомендацій.

Метою даної роботи є проведення аналізу моделей соціальних мереж з метою подальшого вдосконалення та підвищення ефективності систем рекомендацій, заснованих на користувацьких вподобаннях. Ця робота спрямована на розуміння та вдосконалення динаміки взаємодії в соціальних мережах, зокрема в контексті оптимізації рекомендаційних систем.

Результати дослідження було апробовано на п'ятій міжнародній студентській науковій конференції «Цифрофізація науки та сучасні тренди її розвитку» [18].

# 1 АНАЛІЗ ЗАДАЧ ВИЯВЛЕННЯ ВПОДОБАНЬ КОРИСТУВАЧІВ В СОЦІАЛЬНИХ МЕРЕЖАХ

## 1.1 Аналіз вподобань користувачів в соціальних мережах

Поведінка на ринках товарів і послуг, як для індивідуумів, так і для організацій, залежить від процесів, чий параметри пов'язані з особливостями комунікативного середовища, в яке вбудовані ринкові учасники. Формування споживчих вподобань є ключовим етапом перед покупкою і сильно залежить від психічних процесів сприйняття дійсності, що виявляються у вербальних та невербальних реакціях особи. Вивчення цих реакцій відкриває можливості для прогнозування поведінки покупців.

Споживчі вподобання формуються на основі інформації, отриманої в процесі спілкування з соціальним оточенням. У сучасних умовах соціальні мережі, зокрема Інтернет і платформи соціальних мереж, стали ключовими інструментами соціалізації. Це особливо актуально в умовах високого попиту на соціальні мережі, які стали важливим елементом соціальної інтеграції та значно збільшили свою популярність, особливо в Україні.

Проте, не всі аспекти взаємодії між вмістом соціальних мереж та формуванням інтересу до товарів на ринку були ретельно вивчені. Метою цього дослідження є обґрунтування використання семантичних елементів особистих сторінок користувачів соціальних мереж для аналізу процесу формування поведінки покупців [1].

Розглянемо ключові компоненти семантичної структури особистих сторінок користувачів (рисунок 1.1), які загалом присутні у більшості соціальних мереж. Ці елементи можуть вигравати значущу роль у визначенні вподобань користувачів та відображенні їх особистих особливостей, емоцій і уподобань.

№ п/п	Семантичний елемент	Тип	Примітка
1	Ідентифікатор користувача	Текстовий	Повне, скорочене або видозмінене ім'я користувача, або обраний ім псевдонім
2	Показник присутності в мережі	Текстовий	В деяких соціальних мережах цей елемент може бути вираженим за допомогою графічного об'єкту
3	Текстовий статус (закріплений)	Текстовий	Слово, фраза або кілька фраз, розташованих вгорі сторінки, які можуть відображати поточний настрій, інтереси або життєву позицію користувача
4	Аудіостатус	Звуковий	Посилання на аудіозапис, який користувач прослуховує в даний час
5	Зображення користувача	Графічний або змішаний	Як правило, це деякий графічний або графіко-текстовий об'єкт
6	Елементи особистої інформації	Текстовий або графічний	Інформація про користувача, яка організована у встановленому форматі
7	Дані про «друзів»	Текстовий та графічний	Загальна кількість «друзів» користувача та зображення тих з них, з якими користувач нещодавно спілкувався в мережі
8	Елементи фотоінформації	Графічний та текстовий	Розміщені користувачем фотографії, дані про їх кількість, коментарі до них
9	Елементи вибраних відеозаписів	Відео і текстовий	Розміщені користувачем посилання на відеозаписи, дані про їх кількість, коментарі до них
10	Елементи вибраних аудіозаписів	Звуковий та текстовий	Розміщені користувачем посилання на аудіозаписи, дані про їх кількість, коментарі до них
11	Дані про групи	Текстовий та графічний	Назви та логотипи груп користувачів (мережових спільнот), в яких користувач є зареєстрованим
13	Елементи повідомлень «стрічки»	Різні типи	Різноманітна інформація, яку користувач розміщує в спеціальному розділі сторінки (так звана «стрічка»)
14	Уподобання (likes)	Текстовий та графічний	Відмітки, які вказують на те, що розміщена інформація подобалася користувачам
15	Знаки емоцій	Графічний	Спеціальні символи, що відображують емоції
16	Повідомлення про поширення інформації	Текстовий та графічний	Відмітки, які вказують на те, що користувачі поділилися інформацією з іншими користувачами

Рисунок 1.1 - Основні семантичні елементи особистих сторінок користувачів в соціальних мережах

Вивчення вподобань користувачів за допомогою аналізу особистих сторінок у соціальних мережах (рисунок 1.2) відкриває нові можливості у сфері отримання персоналізованих рекомендацій. Запропонована психосемантична модель купівельної поведінки, яка базується на ідеї сприйняття об'єктивних факторів через комунікативне середовище, стає ключовим інструментом для надання рекомендацій у різних аспектах життя користувача.

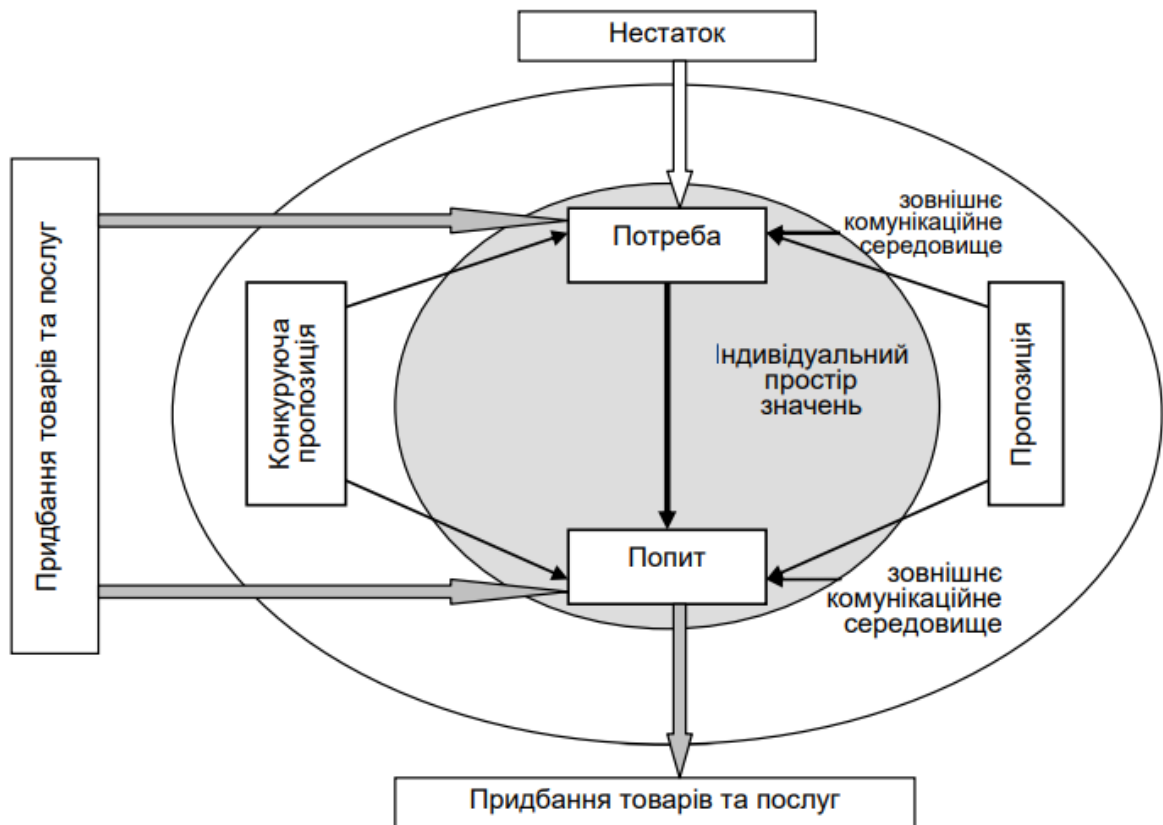


Рисунок 1.2 - Психосемантична модель купівельної поведінки

Таким чином, згідно з цією моделлю, виявлені вподобання під час аналізу особистих сторінок визначають індивідуальний психосемантичний простір. Цей простір не лише відображає купівельні уподобання, але і включає інтереси, смаки та попередній досвід користувача. Аналіз цієї інформації може бути корисним для надання рекомендацій користувачам у різних контекстах.

Наприклад, аналіз вподобань може допомогти визначити основні теми публікацій, які цікавлять користувача, і забезпечити персоналізовані рекомендації щодо подальших публікацій. Також, цей аналіз може служити основою для рекомендацій нових знайомств у соціальних мережах, спільнот або подій, які відповідають профілю індивіда [1].

У сфері рекламних товарів та послуг, вивчення вподобань дозволяє створювати персоналізовані рекламні кампанії, спрямовані на конкретні інтереси та потреби користувача. Це може включати рекомендації товарів, подій чи послуг, які найбільш відповідають його особистим уподобанням, підсилюючи ефективність рекламної взаємодії.

## 1.2 Аналіз моделей соціальних мереж

Перший шлях аналізу соціальних мереж включав в себе використання візуалізаційних форм, зокрема соціограм Я. Морено. Він першим введенням невеликих груп у двовимірний простір за допомогою точок та відрізків для відображення осіб та їх зв'язків. Соціограми застосовувалися для ретельного розгляду структури малих груп, таких як елітні спільноти та внутрішньокорпоративні зв'язки.

Другий напрямок, а саме структурний баланс, проводив дослідження позитивних та негативних взаємозв'язків у групі, зосереджуючись на транзитивності мережі. Це сприяло перевірці гіпотези "ворог мого ворога – мій друг" і сприяло аналізу триплетів – тризначних відносин між учасниками.

Третій напрямок базувався на застосуванні математичної статистики. Деякі дослідники висували ймовірнісні розподіли для випадкових направлених графів, щоб вивчати структурні тенденції у соціальних мережах. Це включало визначення різних типів моделей для аналізу бінарних, тризначних та багатовідносинних відносин [2].

Ці напрямки аналізу структури соціальних мереж сприяли розробці різних характеристик, таких як центральність актора, централізація групи, взаємодія,

структурний баланс, транзитивність, роль, соціальний статус та інші. Це дало можливість глибше розуміти взаємозв'язки та структуру соціальних мереж у різних сферах.

Існує кілька моделей та методів аналізу виявлення вподобань в соціальних мережах.

По-перше, теорія графів створила точний набір термінів для математичного аналізу властивостей соціальних структур. По-друге, вона надає потужні математичні інструменти для кількісної оцінки параметрів соціальних мереж. По-третє, ця теорія та її математичні методи забезпечують здатність формалізовано доводити теореми щодо властивостей соціальних структур. Узагальнено, вершини графа відображають акторів соціальної мережі, а ребра графа визначають комунікації між цими акторами (рис. 3.3). Розглянемо граф  $G$  певної соціальної мережі, де вершин  $N = \{n_1, n_2, \dots, n_g\}$  та ребра  $E = \{l_1, l_2, \dots, l_L\}$ . У проведенні аналізу соціальних мереж використовуються різні характеристики графів. Ступінь вершини в ненаправленому графі визначається як кількість ребер, які безпосередньо пов'язані з даною вершиною. У випадку направленого графа ступінь розраховується окремо для вхідних і вихідних дуг, що вказує на активність конкретного учасника у взаємодії [2]. Середній ступінь вершини слугує узагальненим показником для всієї мережі, відображаючи середню активність її учасників і надаючи більш глибокий вигляд на їхню взаємодію. Середній ступінь вершини є узагальненим показником для мережі:

$$\bar{d} = \sum_{i=1}^g d(n_i) = \frac{2L}{g}, \quad (1.1)$$

$$\bar{d}_1 = \bar{d}_0 = L/g, \quad (1.2)$$

де  $d(n_i)$  – ступінь вершини;

$L$  – кількість ребер/дуг в мережі;

$g$  – кількість вершин в мережі;

$\bar{d}_1$  та  $\bar{d}_0$  – середній ступінь вершини розраховується окремо для вхідних та вихідних дуг, що надає інформацію про середню кількість вхідних та вихідних зв'язків для кожної вершини відповідно [2].

Щільність графа - це показник зв'язаності графа, який визначається як відношення реальної кількості ребер до їх максимально можливої кількості:

$$\Delta = g \frac{L}{(g-1)/2} = \frac{2L}{g(g-1)}, \quad (1.3)$$

$$\frac{\bar{d}}{(g-1)}. \quad (1.4)$$

Центральність – це характеристика активності актора, що визначається як відношення ступеня відповідної вершини графа до загальної кількості вершин, з якими може бути встановлено зв'язок:

$$C_D(n_i) = \frac{d(n_i)}{g-1}. \quad (1.5)$$

Централізованість групи можна визначити таким чином:

$$s_C^2 = \frac{\sum_{i=1}^g (C_D(n_i) - \bar{C}_D)^2}{g}, \quad (1.6)$$

де  $C_D(n_i)$  - дисперсія центральності акторів.

У регулярного графа централізованість дорівнює нулю і досягає максимуму в разі з'єднання, подібного до «зірки». Існують різні методи

визначення центральності актора; наприклад, Л. С. Freeman пропонує узагальнений індекс центральності та його похідні, такі як центральна близькість, центральність посередника та інформаційна центральність. Престиж актора визначається співвідношенням між кількістю вхідних дуг, які спрямовані до відповідної вершини графа, та загальною кількістю акторів у мережі:

$$P_D(n_i) = \frac{d_1(n_i)}{g-1}. \quad (1.7)$$

Іноді акторів оцінюють за допомогою PageRank-алгоритму Google. Теорія графів застосовується для кластеризації соціальних мереж, виявлення щільних внутрішніх груп та розріджених зв'язків між ними. Важливість малозв'язних груп полягає в їх здатності об'єднувати підгрупи. Інформація в щільних групах менше цінна, ніж та, що проходить через слабкі зв'язки з іншими підгрупами. Л. С. Freeman визначав ідеальну підгрупу як повнозв'язну компоненту графа. Розроблено моделі, такі як *n*-clique — максимальний підграф, де відстань між будь-якими парами вершин не перевищує заданого порогу.

Алгебраїчні моделі набули широкого використання для групування акторів. Однією з таких задач є ідентифікація осіб з подібними взаємозв'язками та їх об'єднання в групи. Були розроблені методи для виявлення структурної еквівалентності, зокрема вчений Н. С. White, який упорядкував поняття структурної еквівалентності. Для вимірювання цього аспекту R. S. Burt висунув ідею використання Евклідової відстані:

$$d_{ij} = \sqrt{\sum_{k=1}^g (X_{ik} - X_{jk})^2 + (X_{ki} - X_{kj})^2}, i \neq k, j \neq k, \quad (1.8)$$

де  $x_{ik}$  суміжність вершин( $i, k$ );

$g$  - кількість вершин в графі.

Ще одною популярною мірою можемо виділити коефіцієнт кореляції:

$$r_{ij} = \frac{\sum (x_{ki} - \bar{x}_{row i})(x_{kj} - \bar{x}_{row j}) + \sum (x_{ik} - \bar{x}_{col i})(x_{jk} - \bar{x}_{col j})}{\sqrt{\sum (x_{ki} - \bar{x}_{row i})^2 + \sum (x_{ik} - \bar{x}_{col i})^2} \sqrt{\sum (x_{kj} - \bar{x}_{row j})^2 + \sum (x_{jk} - \bar{x}_{col j})^2}}, \quad (1.9)$$

$$i \neq k, \quad j \neq k$$

де  $x_{ik}$  - суміжність вершин  $(i, k)$ ;

$\bar{x}_{row i}$  та  $\bar{x}_{col i}$  - відповідно, середнє значення  $i$ -го рядка та стовпчика в соціоматриці;  $g$  - кількість вершин в графі.

Обидві міри стають однаковими лише для абсолютно структурно еквівалентних акторів. Проте у інших випадках їх значення можуть значно відрізнятись. Для визначення соціальних позицій було запропоновано кілька методів кластеризації, базованих на вищезазначених мірах, серед яких CONCOR та ієрархічна кластеризація.

Метод ієрархічної кластеризації є жадібним алгоритмом, що використовує критерій схожості. На кожній ітерації алгоритм об'єднує дві підгрупи. У відміну від CONCOR, кількість утворених соціальних позицій залежить від умов зупинки ітераційного процесу та обраного критерію схожості акторів. Схожість акторів можна визначити як за кореляційною матрицею, так і за матрицею евклідових відстаней.

Стохастичні моделі соціальних мереж використовуються для аналізу великих графів акторів. Ці моделі можна поділити на статичні та динамічні. Статичні моделі базуються на теорії випадкових графів і застосовуються для великих мереж акторів. Вони дозволяють визначити долю ключових акторів в мережі, а не знаходження одного центрального актора. Задача полягає у

виявленні долі ключових акторів, вилучення яких може розіб'є мережу або збільшить середню відстань між акторами. Динамічні моделі розглядають мережу як багатоагентну систему, де соціальні актори приймають рішення, змінюючи структуру мережі з часом. Моделі статички розглядають зв'язки як ймовірнісні події, враховуючи неповноту інформації та суб'єктивні фактори. Ці моделі використовують граф, в якому кількість вершин і ребер визначається випадковим розподілом. Різні закони випадкового розподілу зв'язків вивчаються в літературі, від рівномірного до експоненціального розподілу ймовірностей. Використовуючи ці припущення ймовірнісний розподіл ступеня вершини можемо визначити так:

$$P(d(n_i) = k) = \frac{(g-1)!}{k!(g-1-k)!} p^k (1-p)^{g-1-k}, \quad (1.10)$$

де  $p$  - ймовірність існування зв'язку між вершинами.

Розподіл (1.10) наближається до Пуасонівського коли  $g \rightarrow \infty$ . Якщо для всіх зв'язків  $p = \frac{1}{2}$ , отримаємо вираз:

$$P(d(n_i) = k) = \frac{(g-1)!}{k!(g-1-k)!} p^{g-1}. \quad (1.11)$$

Модель (1.11) дозволяє створити випадкову досліджувану мережу з наперед заданими характеристиками, такими як кількість ребер, і навіть оцінити розмір найбільшої компоненти графа. Емпіричні дослідження свідчать про те, що реальні мережі, створені людиною або природою, мають безмасштабний характер, тобто розподіл ступенів вершин є показовим. Для відтворення цієї безмасштабності була розроблена модель Барабаші–Альберта, яка відноситься до еволюційних динамічних мереж.

Згідно з цією моделлю, на кожному кроці до початкової мережі додається вершина з кількома зв'язками. При цьому існує більше шансів утворити новий зв'язок для вершини, яка вже має найбільше зв'язків. Такий принцип росту мережі інтерпретується як правило "багатий стає ще багатшим". В результаті цього зростання формується мережа, яка має виділені потужні вузли та розріджену периферію, що часто спостерігається в реальних системах [2]. Вірогідність додавання зв'язку  $X_{ti}$  на кроці  $t$  в моделі Барабаші–Альберта можна визначити через відношення ступенів вершини:

$$D(X_{ti} = 1) = \frac{d(n_i)}{\sum d(n_j)}. \quad (1.12)$$

1.3 Огляд існуючих інформаційних систем аналізу соціальних мереж в задачах виявлення вподобань для отримання рекомендацій

Інструменти аналізу соціальних мереж використовуються для ідентифікації, аналізу, візуалізації та моделювання вузлів і ребер на основі різних типів даних, включаючи математичні моделі. Ці інструменти можуть бути реалізовані у вигляді графічних інтерфейсів або пакетів/бібліотек для програмування. Міжнародна мережа аналізу соціальних мереж (ММАСМ) підтримує різні програмні пакети та бібліотеки для вибору, залежно від потреб користувача.

Інструменти з графічним інтерфейсом спрощують вивчення та використання, тоді як інструменти для сценаріїв надають потужні можливості для аналізу та розширюють можливості користувача.

Програмне забезпечення Pajek (рисунк 1.3) – це широко використовуване програмне забезпечення для Windows, яке призначене для створення та аналізу великих мереж. Воно доступне безкоштовно для некомерційного використання і спроектоване для обробки різноманітних типів великих мереж з різних джерел.

Pajek володіє аналітичними можливостями і може використовуватися для розрахунку різноманітних показників центральності, ідентифікації структурних отворів, побудови блокових моделей та інших завдань аналізу соціальних мереж. Також важливо відзначити, що Pajek підтримує загальноповживану технологію .NET, що дозволяє використовувати його в різних середовищах.

Це потужний інструментарій для дослідження та візуалізації мереж, особливо у галузі аналізу соціальних мереж та мережі зв'язків [3].

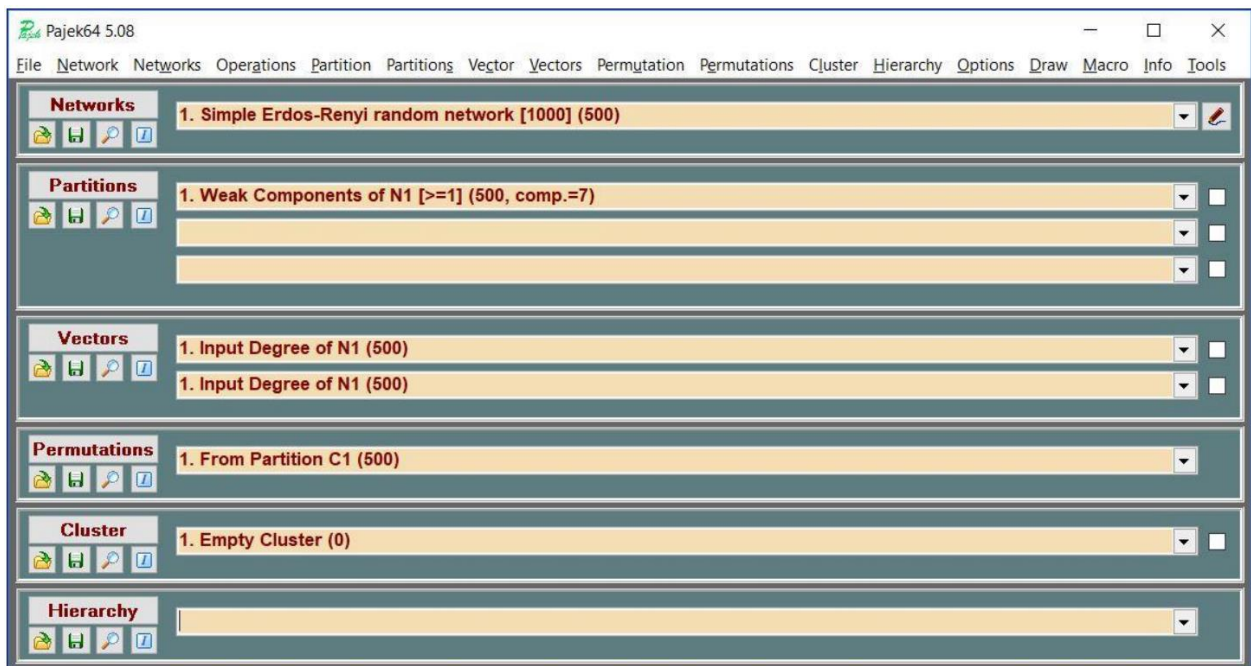


Рисунок 1.3 – Вигляд програми Pajek

Програмне забезпечення Gephi (рисунк 1.4) – є інтерактивним інструментом для візуалізації та дослідження мереж і складних систем. Він

підтримує аналіз як статичних, так і динамічних графів, а також ієрархічних структур. Gephi дозволяє користувачам взаємодіяти з графічним представленням, виявляти приховані властивості та маніпулювати формами, кольорами та структурами графів.

Цей інструмент призначений для аналізу та маніпулювання графами з використанням інтерактивних функцій в реальному часі. Gephi підтримує різні формати графів, має простий у використанні інтерфейс, а також багатофункціональність. Особливість включає динамічний аналіз мережі з фільтрацією в реальному часі, поєднання статичних і динамічних показників, різноманітні макети та компоненти часової шкали для генерації різноманітних звітів. Gephi є потужним інструментом для дослідження та візуалізації комплексних взаємозв'язків у мережах [3].

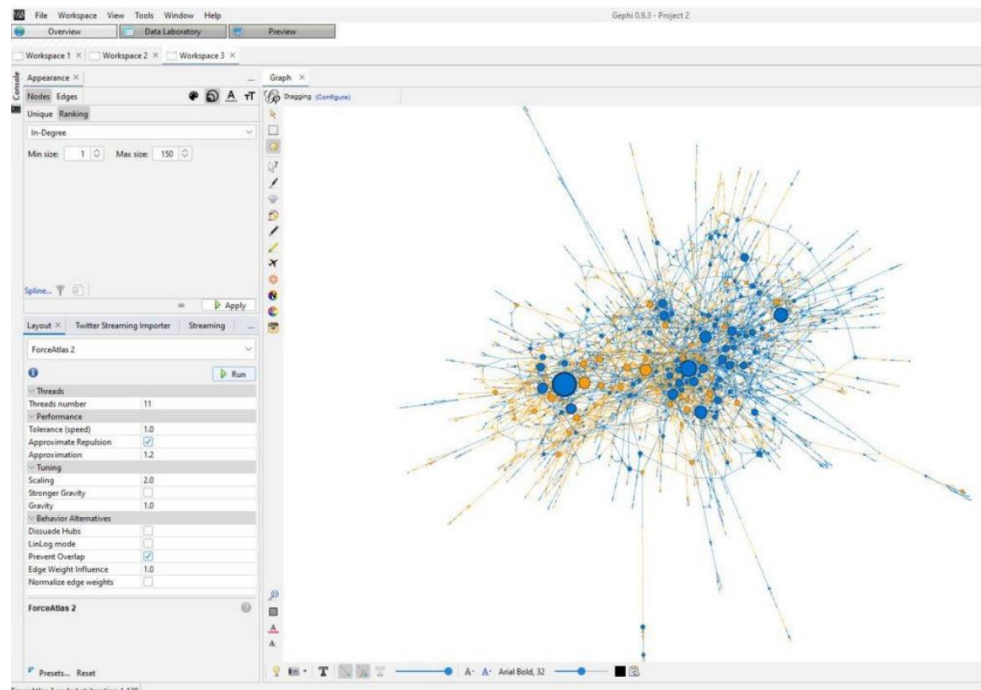


Рисунок 1.4 - Вигляд програми Gephi

Програмне забезпечення Netdraw (рисунок 1.5) – це інструмент для візуалізації соціальних мереж, яке здатне читати дані в різних форматах, таких як VNA, Pajek, UCINET і UCINET DL. Це програмне забезпечення пропонує різноманітні можливості, такі як редагування розмірів та кольорів вузлів, редагування тексту, параметри зовнішнього вигляду та макету. Крім того, інструмент дозволяє візуалізувати міцність зв'язків, фільтрувати дані за різними критеріями і зберігати зображення у різних форматах. Програма має аналітичні можливості, що схожі на UCINET [3].

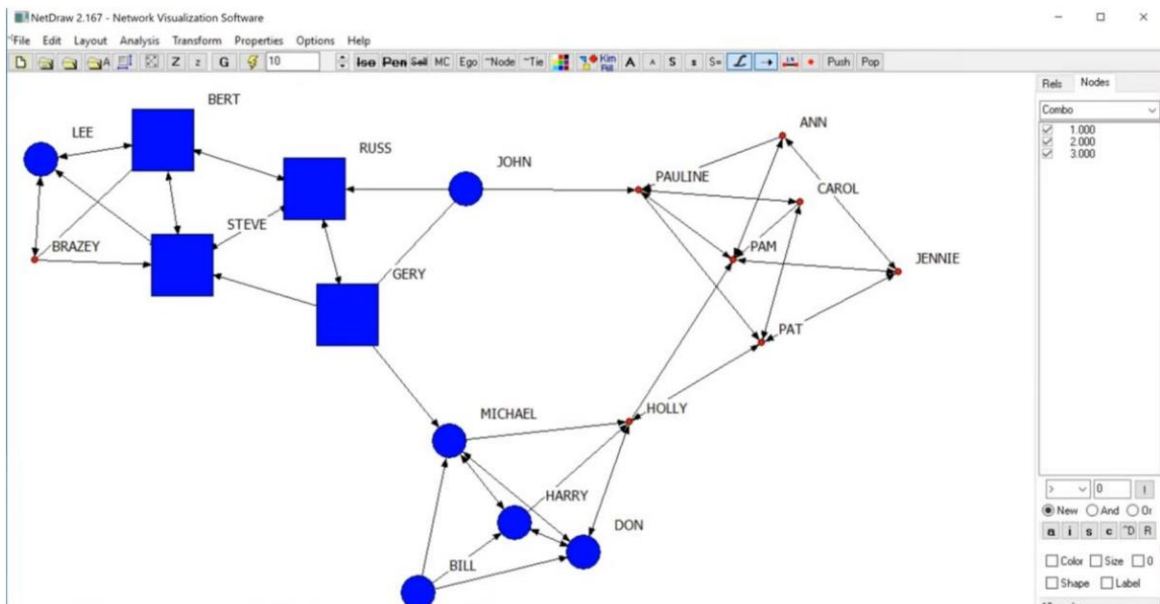


Рисунок 1.5 - Вигляд програми Netdraw

Програмне забезпечення NetworkX (рисунок 1.6) – це Python-пакет, спрямований на створення, управління та аналіз структури складних мереж. Воно дозволяє працювати з різними форматами даних, забезпечує генерацію випадкових та класичних мереж, а також проведення аналізу структури, моделювання мережі та малювання графіків.

NetworkX володіє рядом функцій, таких як різноманітні структури даних для графів, підтримка орієнтованих та графів з багатьма ребрами, а також можливість роботи з різними типами вузлів та ребер. Цей пакет є потужним інструментом для аналізу та взаємодії з мережами в середовищі програмування Python.

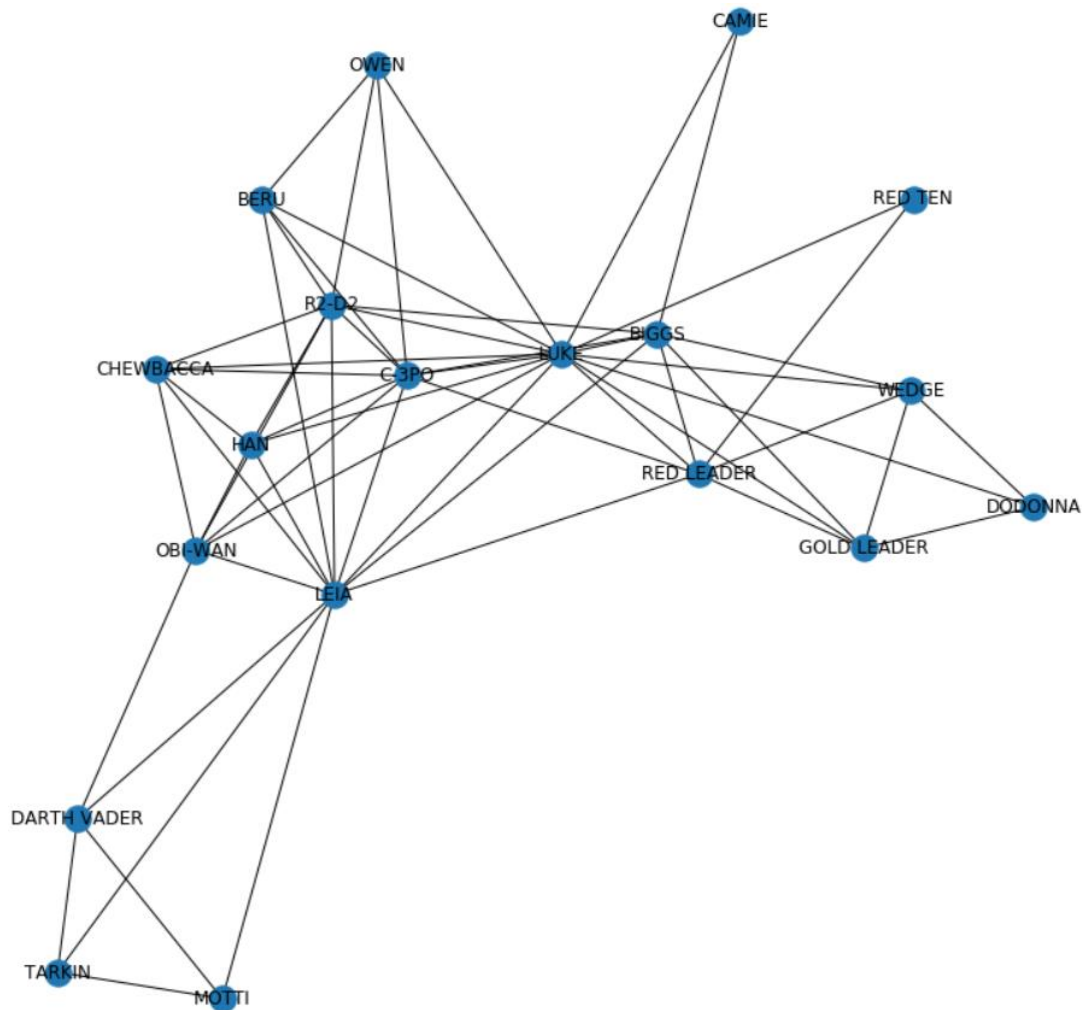


Рисунок 1.6 – Вигляд графу в програмі NetworkX

Програмне забезпечення IGraph – є безкоштовний інструментом для створення та обробки графів, який охоплює класичні завдання теорії графів, такі

як побудова мінімальних охоплюючих дерев та обчислення мережевого потоку. Крім того, воно реалізує алгоритми, такі як виявлення структур спільнот. Ефективна реалізація дозволяє ефективно опрацьовувати граfi навіть з мільйонами вузлів і ребер. IGraph можна використовувати як бібліотеку для мов програмування C, R, Python і Ruby [3].

Для задач виявлення вподобань та отримання рекомендацій інструменти аналізу соціальних мереж мають свої особливості. Розглянемо кожен програму окремо в таблиці 1.1.

Таблиця 1.1 – Переваги й недоліки інформаційних систем аналізу соціальних мереж

№	Інструмент	Переваги	Недоліки
1	Rajek	Широкі аналітичні можливості, підтримка .NET	Засноване на Windows, можливі обмеження для інших ОС
2	Gephi	Інтерактивне програмне забезпечення, динамічний аналіз	Може виявитися складним для роботи з великими мережами
3	Netdraw	Простий у використанні, підтримка різних форматів даних	Може бути менш потужним для складних аналітичних завдань
4	NetworkX	Менш інтуїтивний інтерфейс для не-програмістів	Менш інтуїтивний інтерфейс для не-програмістів
5	IGraph	Безкоштовний, реалізація класичних алгоритмів	Робота з командним рядком може бути складною

Таблиця надає узагальнену інформацію про переваги та слабкі сторони кожного інструменту, що може бути корисним при виборі для конкретних завдань виявлення вподобань та рекомендацій [18].

#### 1.4 Постановка задачі дослідження

Під час дослідження соціальних мереж вивчалися різні методи виявлення вподобань користувачів та отримання рекомендацій. Проведено аналіз різноманітних моделей, включаючи статичні та динамічні, ієрархічні та стохастичні підходи. Огляд наявних інформаційних систем для аналізу соціальних мереж вказав на унікальні характеристики таких інструментів, як Pajek, Gephi, Netdraw, NetworkX та IGraph. Pajek визначається своїми аналітичними можливостями, Gephi - динамічністю та інтерактивністю, Netdraw - ефективністю візуалізації, NetworkX - гнучкістю в роботі з Python, а IGraph - ефективністю обробки графів.

Метою даного дослідження є оцінка потенціалу використання зазначених моделей та інструментів для удосконалення аналізу соціальних мереж з акцентом на виявленні вподобань користувачів і забезпеченні персоналізованих рекомендацій. Проведення аналізу може визначити оптимальні методи та сприяти створенню інформаційної системи, яка в змозі аналізувати моделі соціальних мереж та розробляти ефективні стратегії виявлення вподобань користувачів. Основна мета полягає в створенні засобів для підбору рекомендаційних варіантів, таких як публікації, користувачі, товари та інше, з максимальною відповідністю індивідуальним уподобанням користувачів.

Об'єктом дослідження є процес визначення вподобань користувачів у соціальних мережах.

Предметом дослідження є методи та моделі визначення вподобань користувачів соціальних мереж на основі рекомендаційних систем.

Це програмне забезпечення може бути використане різними категоріями користувачів з різноманітними цілями. Наприклад, для бізнес-користувачів, які прагнуть знайти свою цільову аудиторію серед користувачів соціальних мереж. Також, воно може бути корисним для дослідників, які проводять аналіз утворення вподобань користувачів, базуючись на їх взаємодії та зв'язках з іншими учасниками соціальних мереж.

Крім того, це програмне забезпечення може слугувати інструментом для проведення тренувань та соціальних опитувань, спрямованих на вивчення вподобань великої групи користувачів соціальних мереж. Його універсальність робить його корисним для широкого спектру завдань, пов'язаних із вивченням та аналізом поведінки користувачів в цьому контексті.

## 2 ДОСЛІДЖЕННЯ МОДЕЛЕЙ СОЦІАЛЬНИХ МЕРЕЖ

### 2.1 Огляд моделей аналізу соціальних мереж та їх особливостей

Графові моделі. Узагальнено, вузли графа представляють учасників соціальної мережі, а ребра графа відображають взаємодії чи комунікації між цими учасниками. Проналізуємо граф соціальної мережі, фокусуючись на конкретному вузлі  $g \in \{n_1, n_2, \dots, n_g\}$  і ребрами  $E = \{1, 1, 1, 2, L\}, \{1, 1, 1, 2, L\}$ . Для комп'ютерного аналізу графа соціальної мережі використовується матриця суміжності, що отримала назву соціоматриця. На прикладі, на рисунку 2.1 представлена соціоматриця даного графа:

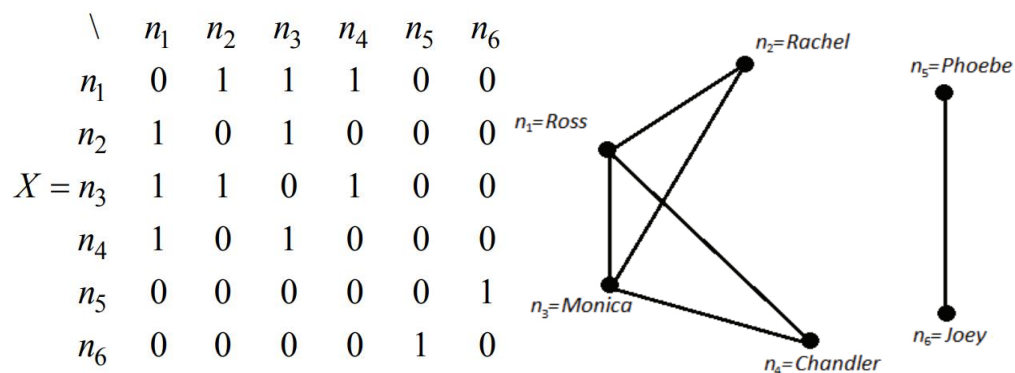


Рисунок 2.1 - Граф соціальної мережі

Однією з ключових мет наукового використання теорії графів є проведення кластеризації соціальних мереж, тобто формального визначення внутрішніх груп. Тут "група" означає частину графа, де щільність взаємозв'язків всередині переважає над щільністю зовнішніх зв'язків. Група з низькою щільністю зв'язків має обмежені внутрішні зв'язки, але поєднує декілька важливих підгруп. Інформація всередині щільних груп виявляється схожою та надто повторюваною,

що зменшує її цінність для групи. В той час як інформація, яка проникає через різні підгрупи з слабкими зв'язками, вважається більш цінною.

Л. К. Фрімен запропонував концепцію ідеальної підгрупи, яка є повністю зв'язною компонентою графа, відомою як сильний альянс. З цими передумовами були розроблені різноманітні моделі підгруп:

- $n$ -кліки: це максимальні підграфи, де відстань між будь-якою парою вершин менша або дорівнює  $n$ ;

- $k$ -комплекси: максимальні підграфи, що включають  $g_s$  вершин і мають щонайменше  $(g_s - k)$  сусідніх вершин;

- $k$ -ядро: максимальний підграф, в якому кожна вершина має щонайменше  $k$  сусідніх вершин всередині цього підграфа;

- LS-кластер: підграф, де будь-яка множина вершин має більше внутрішніх ребер, ніж зовнішніх.

- $\lambda$ -множина: підграф, де міра зв'язності будь-якої пари вершин всередині підгрупи переважає міру зв'язності між цими вершинами та іншими частинами графа. Зв'язаність вершин  $\lambda(i, j)$  визначається як мінімальна кількість зв'язків, які потрібно видалити, щоб вершини  $(i, j)$  були недосяжні одна для одної.

Знаходження підгруп є завданням високої складності; наприклад, визначено, що складність пошуку LS-множин та  $\lambda$ -множин дорівнює  $O(n^5)$ . Це означає, що точні алгоритми для виявлення таких підгруп можна успішно застосовувати лише в обмеженому масштабі, зазвичай для невеликих соціальних мереж.

Алгебраїчні моделі. Для прикладу можемо розглянути граф  $G$  з рис. 2.1 з трьома групами акторів, які є структурно еквівалентними. Шляхом формули (1.8) ми обчислимо відстань між вершинами графа:

$$\begin{array}{c|cccccc}
 & n_1 & n_2 & n_3 & n_4 & n_5 & n_6 \\
 n_1 & - & \sqrt{2} & 0 & \sqrt{2} & \sqrt{8} & \sqrt{8} \\
 n_2 & \sqrt{2} & - & \sqrt{2} & 0 & \sqrt{6} & \sqrt{6} \\
 d = n_3 & 0 & \sqrt{2} & - & \sqrt{2} & \sqrt{8} & \sqrt{8} \\
 n_4 & \sqrt{2} & 0 & \sqrt{2} & - & \sqrt{6} & \sqrt{6} \\
 n_5 & \sqrt{8} & \sqrt{6} & \sqrt{8} & \sqrt{6} & - & 0 \\
 n_6 & \sqrt{8} & \sqrt{6} & \sqrt{8} & \sqrt{6} & 0 & -
 \end{array}$$

$$D = \begin{array}{c|ccc}
 & \beta_1 & \beta_2 & \beta_3 \\
 \beta_1 & 1 & 1 & 0 \\
 \beta_2 & 1 & 0 & 0 \\
 \beta_3 & 0 & 0 & 1
 \end{array}$$

(2.1)

Як показано у формулі (2.1) існує висока щільність зв'язків між групами та акторами, свідчачи про активну міжгрупову комунікацію, в контрасті до відокремлених груп  $\beta_1$  та  $\beta_2$ . Точне визначення зв'язків між групами стає складним завданням через різні значення щільності. Для ідентифікації цікавих груп на основі матриць щільності з неперервними значеннями застосовують блокові моделі (запропоновано Н. С. White, S. A. Boorman та R. L. Breiger у 1976 році). Цей метод включає бінаризацію матриці щільності  $D$  за певним порогом загальногрупових зв'язків або середньою щільністю зв'язків між групами [6].

За допомогою блокової моделі маємо можливість систематизувати соціальні ролі як узагальнені патерни взаємозв'язків між позиціями в соціальній мережі. Основний упор робиться на розгляданні взаємодій між учасниками. Наприклад, якщо в соціальній мережі існують відносини  $A$  – Допомога,  $F$  – Дружба, тоді застосування композиції відносин може породжувати нові типи відносин, такі як  $F \circ F$  – "друг одного",  $A \circ A$  – "допомагати тому, хто вже допомагає іншим", і так далі. Важливо відзначити, що соціальні ролі можна сприймати як конфігурації відносин із значущими інтерпретаціями. Таким чином, аналіз

соціальних мереж базується на розгляданні структури вивчених взаємозв'язків, виявленні шаблонів з типовими структурами відносин та їх формалізації.

Стохастичні моделі. Імовірнісні моделі широко використовуються для вивчення великих соціальних мереж, що мають тисячі або навіть мільйони учасників. Під час аналізу таких мереж, основна мета не полягає в пошуку одного ключового актора, а визначенні частки ключових акторів у мережі. Важливим завданням є визначення стійкості мережі шляхом оцінки частки акторів, видалення яких призвело б до розділення мережі на окремі частини або суттєвого збільшення середньої відстані між учасниками.

Імовірнісні моделі розглядають взаємозв'язки як стохастичні події, враховуючи обмежену інформацію про структуру мережі та вплив суб'єктивних факторів. Динамічні моделі аналізують мережу як мультиагентну систему, де соціальні актори (активні та адаптивні агенти) приймають рішення, що призводять до змін у структурі мережі. Соціальні та когнітивні процеси можуть впливати на взаємозв'язки всередині мережі.

Моделі соціальних мереж, що ґрунтуються на статистичних підходах, оперують графами, де випадковий розподіл визначає як кількість вершин і ребер, так і частоту їхнього виникнення. В рамках досліджень вивчається різноманіття законів випадкового розподілу зв'язків, від простих рівномірних до послідовності експоненціальних розподілів. В моделі Ердоша-Реньї кількість вершин є фіксованою, а розподіл зв'язків є рівномірним. Відповідно до цих умов, ймовірність порядку вершин визначається наступним чином:

$$P(d(n_i) = k) = \frac{(g-1)!}{k!(g-1-k)!} p^k (1-p)^{g-1-k}, \quad (2.2)$$

де  $p$ — вірогідність існування зв'язку між вершинами.

Розподіл (2.2) буде наближатися до Пуасонівського, якщо  $g \rightarrow \infty$ . У випадку, коли для всіх зв'язків, отримаємо вираз:

$$P(d(n_i) = k) = \frac{(g-1)!}{k!(g-1-k)!} p^{g-1}. \quad (2.3)$$

Модель (2.3) може служити інструментом для створення випадкових мереж з попередньо визначеними характеристиками, такими як кількість ребер, а також для оцінки розміру найбільших компонентів графа. На основі емпіричних досліджень встановлено, що реальні мережі, які виникають в результаті діяльності людей та природи, мають безмасштабну структуру. Це виявляється в законі розподілу ступенів вершин. Для відтворення безмасштабної природи мереж була розроблена модель Барабаші-Альберта, пов'язана з еволюційними динамічними мережами.

Згідно з цією моделлю, на кожному етапі до початкової мережі додаються вершини з кількома зв'язками. При цьому вершини з найбільшою кількістю зв'язків мають вищу ймовірність утворювати нові зв'язки. Цей принцип росту мережі можна інтерпретувати як "багаті стають багатшими". Результатом такого зростання є мережа з окремими потужними вузлами та розрідженою периферією. У моделі Барабаші-Альберта ймовірність додавання зв'язку  $X_{ti}$  на етапі  $t$  визначається відношенням частот вершин:

$$P(X_{ti} = 1) = \frac{d(n_i)}{\sum d(n_j)}. \quad (2.4)$$

Ще однією характерною особливістю реальних мереж є явище "малого світу", коли діаметр графа є відносно невеликим порівняно з кількістю вершин.

Наприклад, середня кількість вершин, які розділяють двох жителів США, становить лише шість. Для вивчення цього явища була розроблена модель Вотца-Строгаца, яка передбачає будівництво кільцевого графа із  $g$  вершин. Кожна вершина графа може переосмислити свої зв'язки з іншими вершинами з однаковою ймовірністю  $p$ . Ця модель може бути використана для створення мережі з певним діаметром. Експериментально було виявлено, що середня кількість вершин, які розділяють будь-які дві вершини мережі, збігається при значенні ймовірності  $p \rightarrow 1$  [2].

$$L = \frac{\ln g}{\ln d}. \quad (2.5)$$

Статистичний аналіз має велике застосування в аналізі соціальних мереж, особливо для побудови алгебраїчних моделей, вивчення взаємодії та тенденцій вибору, аналізу еволюції мережі в часі та порівняння складних взаємозв'язків з багатьма параметрами. Використання статистичних підходів в аналізі соціальних мереж є одним з найбільш плідних та перспективних напрямків, і вони використовуються для дослідження різних моделей, таких як багатопараметричні моделі, моделі ймовірності двосторонніх графів і моделі латентного простору в маркетингу.

Моделі інших типів. Протягом останнього десятиліття стрімкий розвиток інформаційних технологій відкрив можливості для збору та обробки інформації про топологію реальних мереж, які можуть включати мільйони вузлів. Доступ до великої кількості інформації сприяв міждисциплінарним дослідженням складних соціальних мереж, де застосовуються гібридні моделі.

Гібридний підхід об'єднує теорію графів, що вивчає структуру мереж, і теорію ігор, яка аналізує моделі поведінки, що виникають внаслідок взаємодії

акторів. Графо-ігровий підхід застосовується для аналізу структури мереж ринків та влади, моделювання взаємодії окремих осіб і груп, вивчення динаміки епідемій і моделювання процесів пошуку в інформаційних системах.

У контексті теорії управління соціальні мережі аналізуються з використанням таких понять, як агенція, думка, вплив, довіра, репутація та дифузія. Використовуються методи аналізу, засновані на порогових моделях, моделях незалежних кроків, моделях проникнення і зараження, моделях Ісінга, клітинних автоматах і ланцюгах Маркова.

Останніми роками використання статистично-фізичних моделей для аналізу соціальних мереж набуло значного поширення. У цій статті узагальнено застосування моделей статистичної фізики та динаміки для аналізу соціальних мереж, наприклад, у зв'язку з теорією перколяції та моделлю випадкових графів Ердоса-Реньї. Теорія перколяції описує поведінку зв'язаних множин у випадкових графах. З цієї теорії випливає, що якщо ймовірність  $p$  моделі (2.2) нижче деякого критичного значення  $p_c$ , то мережа складається з ізольованих кластерів, а якщо  $p > p_c$ , то велика складова покриває всю мережу.

У сфері статистичної фізики використовується інша цікава модель, яку відомо як фітнес-модель Альберта-Барабаші. Початкова модель Альберта-Барабаші враховує безмасштабну природу реальних мереж, але не враховує можливість появи нових привабливих вузлів. Для врахування цього аспекту у фітнес-модель додається параметр, який визначає, як обирається значення для нового вузла згідно з певним розподілом ймовірностей. При високому значенні параметра фітнесу новий вузол може активно нарощувати свої зв'язки. Формула для фітнес-моделі має наступний вигляд:

$$P(X_{ii} = 1) = \frac{\eta_i d(n_i)}{\sum \eta_j d(n_j)}. \quad (2.6)$$

Річард Р. Єгер розробив концепцію, яка об'єднує точне математичне моделювання соціальних мереж і лінгвістичну оцінку їхніх характеристик. Замість використання лінгвістичних термінів, таких як "лідер" чи "трендсеттер", Єгер використовує нечіткі множини для висловлення лінгвістичних питань стосовно соціальних мереж. Наприклад, для опису зв'язності підгруп використовуються терміни, такі як "дуже низька", "низька", "середня" і "висока". Автор пропонує концепцію, яка дозволяє використовувати нечіткі множини для лінгвістичного опису характеристик соціальних мереж, зокрема ступеня зв'язності підгруп. Це можна досягти за допомогою побудови нечіткої бази правил "якщо-то", яка враховує різноманітні експертні та емпіричні знання про структурні та композиційні змінні в мережі.

Порівняльна характеристика базових рекомендаційних алгоритмів. У даному розділі проведено порівняльний аналіз базових алгоритмів для систем рекомендацій, і отримані результати представлені на рисунку 2.2.

Альтернативний підхід з використанням архітектури LSTM. Альтернативним підходом до розробки рекомендаційної системи може бути використання рекурентних нейронних мереж, а конкретно — мереж з довгостроковою короткостроковою пам'яттю (LSTM). Такі мережі спроектовані для створення моделі, яка прогнозує наступну подію на основі послідовності попередніх подій. Загальна архітектура відображена на рисунку 2.3.

Цей метод виявляється надзвичайно гнучким і дозволяє використовувати будь-який контекст інформації про взаємодію користувача з об'єктом рекомендації. Додатково, враховується послідовність подій, що дозволяє віддавати перевагу недавнім подіям і, водночас, адекватно реагувати на різку зміну характеру взаємодії користувача з об'єктами рекомендацій. Ще однією перевагою моделей на основі LSTM є їхню здатність забезпечити отримання нових рекомендацій у режимі реального часу.



## 2.2 Удосконалення моделей аналізу соціальних мереж

Соціальні мережі набувають значущого значення як канал для взаємодії зі споживачами. Західні компанії розуміють великий потенціал соціальних мереж в сфері Інтернет-комунікацій і використовують їх для підвищення усвідомленості споживачів щодо своїх продуктів і брендів. Крім того, соціальні медіа стають додатковим елементом основних рекламних стратегій. Наприклад, керівництво компаній визнає, що соціальні мережі можуть підвищити ефективність їхніх маркетингових кампаній. Рекламні заходи в соціальних мережах сприяють зростанню трафіку та інформують споживачів про ключові переваги продуктів і послуг [8].

Для порівняння розглянемо результати опитування керівників українських компаній (рисунок 2.4), оцінюючи їхнє ставлення до використання сторінок компаній у соціальних мережах. За результатами дослідження виявлено, що третина українських компаній має намір просувати свої товари та послуги через соціальні медіа, тоді як лише 20% планують взаємодіяти зі споживачами через ці канали. Інакше кажучи, вітчизняний ринок ще не повністю розуміє потенціал соціальних медіа у сфері онлайн-маркетингових комунікацій. Крім традиційних платформ, таких як Facebook та Twitter, які продовжують збільшувати свою популярність, нові соціальні мережі, зокрема YouTube та Instagram, також набувають все більшого застосування на ринку [9].

Крім того, вплив комунікації через конкретну соціальну мережу залежить не лише від охоплення цільової аудиторії, але й від специфіки цільового ринку. Наприклад, взаємодія в професійній мережі LinkedIn суттєво впливає на промисловий ринок, тоді як Facebook зазвичай активно використовується в

споживчому секторі. З цього можна зробити висновок, що іноземні компанії розглядають соціальні мережі та їх можливості як дуже перспективний фактор для успіху своєї маркетингової стратегії взагалі.

Порівняльна характеристика	Західна Європа та США	Україна
Рівень проникнення Інтернету	Високий. Покриває усі вікові та соціальні групи.	Середній. Переважає молода аудиторія та міське населення.
Цілі використання Інтернету споживачами	Пошук контенту, здійснення покупок у рівній пропорції	Пошук контенту, тенденція до збільшення покупок онлайн
Точка виходу в Інтернет	Персональні комп'ютери, смартфони та планшети. Використовують декілька пристроїв одночасно	Переважають стаціонарні комп'ютери
Рівень проникнення соціальних мереж	Високий. Переважають мережі Facebook, Twitter	Середній. Переважають локальні мережі Odnoklassniki, VKontakte
Динаміка ринку Інтернет-реклами	Позитивна. Збільшується за рахунок соціальних мереж	Позитивна. Збільшується за рахунок всіх Інтернет- маркетингових інструментів
Причини зростання ринку	Розуміння потреб споживачів, зміни їх купівельної поведінки	Відносна дешевизна Інтернету порівняно з традиційними каналами комунікацій
Активність компаній в мережі Інтернет	Висока. Використовують значну кількість інструментів роботи зі споживачами	Середня. Використовують Інтернет переважно для внутрішньо-фірмової комунікації, пошуку інформації, однак не спілкуються зі споживачами
Контент у соціальних мережах	Висока якість контенту, діалог зі споживачем. Реалізують професійні підрядники	Переважно низька якість контенту через непрофесійність підрядників. Ведуть роботу у соціальних мережах самостійно

Рисунок 2.4 – Порівняння використання Інтернет-технологій у країнах Західної Європи, США та України

У той час як вітчизняні компанії ще не повністю розуміють зручність використання соціальних мереж та сучасних інтернет-технологій у своїй маркетинговій комунікаційній діяльності та не враховують можливості, які вони можуть надати для їхнього бізнесу. Таким чином, рівень використання Інтернет-маркетингу як споживачами, так і бізнесом, в Україні та Західній Європі, значно відрізняється [10].

Українському ринку надається обмежена увага у сфері реклами та просування через соціальні мережі, незважаючи на те, що багато компаній мають свої сторінки у Facebook.

Соціальні мережі визначаються як онлайн-платформи, які дозволяють користувачам створювати свої профілі та взаємодіяти з іншими у віртуальному просторі. Існують два основних типи соціальних мереж: непрофесійні та професійні. Непрофесійні, або загальні, мережі охоплюють віртуальні спільноти без чіткої тематичної спеціалізації. Професійні спільноти є віртуальними об'єднаннями, спрямованими на обмін та розвиток знань у конкретній галузі [22].

Французький аналітик Ф. Кавацца, проводячи щорічний аналіз соціальних медіа, класифікує їх за сімма функціональними категоріями (див. рисунок 2.5).

Група	Підгрупа	Приклад
Сервіси для публікації текстів	мікроблоги	Twitter
	стрічки новин	FriendFeed, Tumblr, Posterous
Сервіси для розповсюдження контенту	фото	Instagram, Flickr, Picasa
	відео	YouTube, DailyMotion, Vimeo
	музика	Last.fm, Spotify
	документи	SlideShare
Сервіси, пов'язанні з комерційною діяльністю		Blippy, Groupon, LivingSocial
Геолокаційні сервіси		Foursquare, Gowalla, Facebook Places, Google Places
Соціальні мережі	професійних контактів	LinkedIn
	для знайомств	Badoo
	«традиційні мережі»	Facebook, MySpace, Вконтакте <small>на Windows</small>

Рисунок 2.5 – Класифікація соціальних медіа за Ф. Кавацца

Збільшення ефективності рекомендаційних систем можливо через розширення обсягу використовуваної інформації, зокрема, шляхом урахування контексту. Контекст включає атрибути, які описують ситуацію, в якій користувач розглядає об'єкт або отримує рекомендації. Таким чином, можна виділити два основних типи контексту у системах рекомендацій:

Контекст при фіксації інтересів користувача та вподобань:

– фізичний контекст, який включає час, місцеположення, вид діяльності, погодні умови, освітленість і т.д.;

- соціальний контекст, що визначається наявністю та роллю оточуючих людей навколо користувача;
- пристрійовий контекст, який описує вид і характеристики пристрою, через який користувач отримує доступ до інформації;
- модальний контекст, що включає настрій користувача, його мету, досвід та когнітивні особливості.

Контекст при формуванні рекомендацій:

- використання інформації про фізичний, соціальний, пристрійовий та модальний контекст для забезпечення більш точних рекомендацій.

Цей підхід розв'язує ряд проблем, які часто виникають у більшості рекомендаційних систем. Зокрема, він допомагає подолати "холодний старт" для нових користувачів, надає можливість отримання рекомендацій для нетипових аудиторій, уникає видачі тривіальних порад та враховує контекст для уникнення обмежень, пов'язаних з попередніми точками зору користувачів [12].

Впровадження індивідуалізації в рекомендації за допомогою врахування контексту може бути реалізоване шляхом застосування нейромережових архітектур у поєднанні з традиційними методами.

Було обрано модель Neu-CF. Структуру загальної моделі для нейронної колаборативної фільтрації представлено на рисунку 2.6:

Для моделювання взаємодії користувача з об'єктом рекомендації використовується структура з багатьох рівнів, де вихід одного рівня служить входом для наступного. Нижній рівень включає вектори, які можуть відображати контекстні, контентозалежні або схожість користувачів та їхньої поведінки. Ідентифікатор користувача та об'єкта піддаються кодуванню і використовуються як вхідні функції.

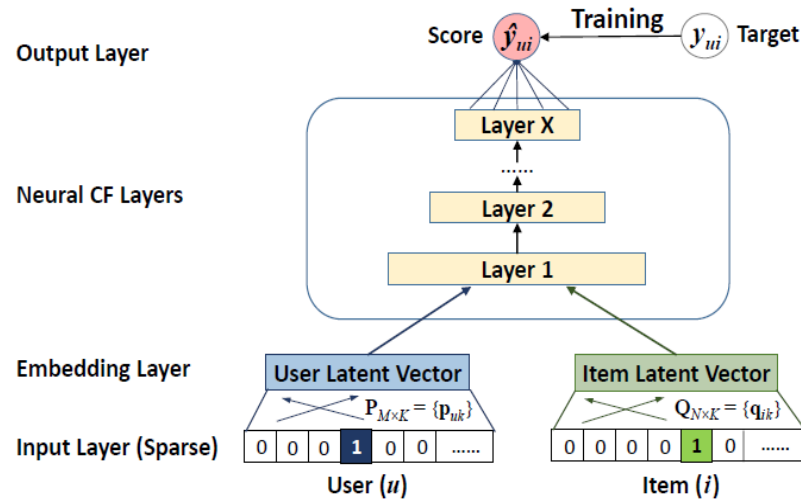


Рисунок 2.6 - Загальна схема нейронної колаборативної фільтрації

Вище рівнем знаходиться шар ембедінгів, який перетворює розріджене представлення об'єкта на щільний вектор. Отриманий ембедінг розглядається як прихований вектор користувача чи об'єкта в рамках моделі прихованого фактору. Ембедінги вводяться в багатошарову нейронну архітектуру, яка містить нейронні колаборативні шари для фільтрації та порівняння прихованих векторів з прогнозованими показниками.

Кожен шар нейронних колаборативних шарів використовується для визначення прихованих взаємодій між користувачем та об'єктом.

Модель для прогнозування методом Neu-CF можна задати як:

$$\hat{y}_{ui} = f(\mathbf{P}^T \mathbf{v}_u^U, \mathbf{Q}^T \mathbf{v}_i^I | \mathbf{P}, \mathbf{Q}, \Theta_f), \quad (2.7)$$

де  $\mathbf{P}$ ,  $\mathbf{Q}$  - це матриці прихованих факторів для користувача та об'єкта відповідно. представляє собою багатошарову нейронну мережу, яка використовується для

моделювання взаємодії між користувачем та об'єктом у контексті рекомендацій, таким чином задається взаємозв'язок між вихідним шаром та NCF-шаром з певним порядковим номером.

Впровадження концепції багатошарового перцептрон у класичні методи. Дана ідея рекомендацій полягає в тому, щоб об'єднати вектори, які представляють користувачів та об'єкти, шляхом конкатенації, щоб врахувати обидві аспекти в нейронній колаборативній фільтрації. Однак звичайна конкатенація не враховує прихованих взаємодій між користувачем та об'єктами, що може бути недостатнім для ефективного моделювання ефекту спільної фільтрації.

Для вирішення цього виклику запропоновано додати приховані шари до об'єданого вектора, використовуючи стандартну архітектуру багатошарового перцептрон (див. рисунок 2.7). Це дозволяє вивчати приховані функції взаємодій між користувачем та об'єктами, забезпечуючи моделі високий рівень гнучкості та більшу чутливість до нелінійності при моделюванні взаємодії між користувачами та об'єктами.

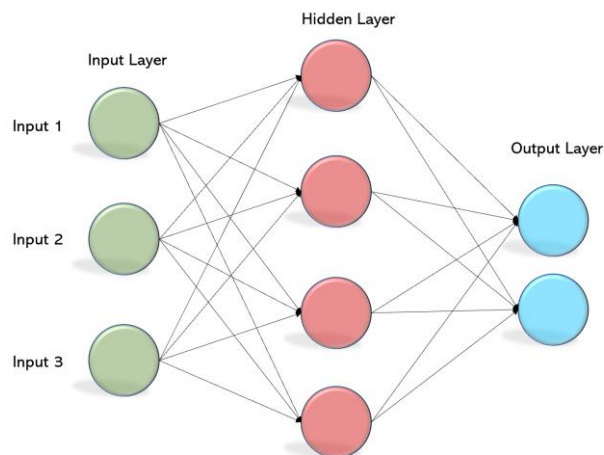


Рисунок 2.7 – Загальний вигляд багатошарового перцептрон

Модель MLP в рамках нейронної колаборативної фільтрації визначатимемо як:

$$\begin{aligned} \mathbf{z}_1 &= \phi_1(\mathbf{p}_u, \mathbf{q}_i) = \begin{bmatrix} \mathbf{p}_u \\ \mathbf{q}_i \end{bmatrix}, \\ \phi_2(\mathbf{z}_1) &= a_2(\mathbf{W}_2^T \mathbf{z}_1 + \mathbf{b}_2), \\ &\dots\dots \\ \phi_L(\mathbf{z}_{L-1}) &= a_L(\mathbf{W}_L^T \mathbf{z}_{L-1} + \mathbf{b}_L), \\ \hat{y}_{ui} &= \sigma(\mathbf{h}^T \phi_L(\mathbf{z}_{L-1})), \end{aligned}$$

В цьому випадку, ці вектори відповідатимуть користувачеві та об'єкту відповідно, а також векторам відхилення та функціям активації на відповідних рівнях.

Об'єднання методів матричної факторизації та MLP-архітектури представлено на рисунку 2.8 у загальній структурі моделі.

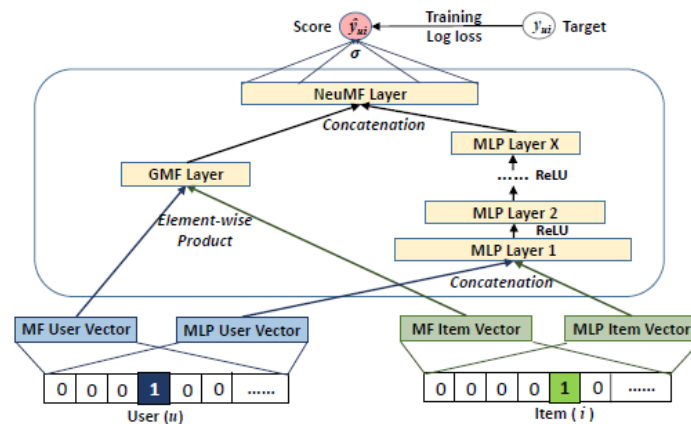


Рисунок 2.8 – Модель з поєднанням методів матричної факторизації та MLP-архітектури

В розробленій моделі, що ґрунтується на подібній архітектурі, забезпечено значну гнучкість та чутливість до нелінійних взаємодій. Цього досягнуто через те, що окремі шари, що відповідають за матричну факторизацію та стандартні нейронні шари, незалежно навчають ембедінги, що представляють приховані взаємодії. Після цього обидві моделі об'єднуються за допомогою останнього прихованого шару [13].

## **3 РОЗРОБЛЕННЯ МОДЕЛЕЙ ДОДАТКУ ДЛЯ АНАЛІЗУ СОЦІАЛЬНИХ МЕРЕЖ**

### **3.1 Розроблення моделі даних додатку аналізатору**

Була використана відкрита бібліотека нейромереж Keras, розроблена на мові програмування Python, як основний фреймворк для розробки програмного продукту. Ця бібліотека спеціалізується на роботі з глибоким навчанням та відрізняється компактністю, модульністю та розширюваністю. Keras пропонує високорівневий та інтуїтивно зрозумілий інтерфейс, що полегшує створення нейронних мереж. Фреймворк підтримує різноманітні шари нейронних мереж, такі як згорткові, рекурентні та щільні. Деякі переваги Keras включають швидке та просте прототипування, повністю налаштовані модулі, інтуїтивно зрозумілий інтерфейс, підтримку навчання на декількох типах GPU (включаючи NVIDIA та TPU від Google), а також сумісність з різними бекендами, такими як Tensorflow, CNTK та Theano.

Для обробки даних використана бібліотека Pandas, яка надає спеціальні структури даних та функції для роботи з числовими таблицями та часовими рядами. Для візуалізації процесу навчання мереж використана бібліотека Matplotlib.

Представлено модель, побудовану за допомогою гібридного методу, який комбінує нейронну колаборативну фільтрацію та матричну факторизацію. Зображення цієї моделі за допомогою методу Neu-CF-MF подано на рисунку 3.1.

Подання моделі, створеної за допомогою методу матричної факторизації із реальними значеннями. Модель, побудовану за методом Neu-MF, представлено на рисунку 3.2.

Layer (type)	Output Shape	Param #	Connected to
Item (InputLayer)	(None, 1)	0	
User (InputLayer)	(None, 1)	0	
course-Embedding-MLP (Embedding)	(None, 1, 10)	97250	Item[0][0]
User-Embedding-MLP (Embedding)	(None, 1, 8)	4888	User[0][0]
Flattencourse-MLP (Flatten)	(None, 10)	0	course-Embedding-MLP[0][0]
FlattenUsers-MLP (Flatten)	(None, 8)	0	User-Embedding-MLP[0][0]
dropout_203 (Dropout)	(None, 10)	0	Flattencourse-MLP[0][0]
dropout_205 (Dropout)	(None, 8)	0	FlattenUsers-MLP[0][0]
Concat (Concatenate)	(None, 18)	0	dropout_203[0][0] dropout_205[0][0]
dropout_207 (Dropout)	(None, 18)	0	Concat[0][0]
FullyConnected (Dense)	(None, 200)	3800	dropout_207[0][0]
Batch (BatchNormalization)	(None, 200)	800	FullyConnected[0][0]
Dropout-1 (Dropout)	(None, 200)	0	Batch[0][0]
FullyConnected-1 (Dense)	(None, 100)	20100	Dropout-1[0][0]
Batch-2 (BatchNormalization)	(None, 100)	400	FullyConnected-1[0][0]
course-Embedding-MF (Embedding)	(None, 1, 3)	29175	Item[0][0]
User-Embedding-MF (Embedding)	(None, 1, 3)	1833	User[0][0]
Dropout-2 (Dropout)	(None, 100)	0	Batch-2[0][0]
Flattencourse-MF (Flatten)	(None, 3)	0	course-Embedding-MF[0][0]
FlattenUsers-MF (Flatten)	(None, 3)	0	User-Embedding-MF[0][0]
FullyConnected-2 (Dense)	(None, 50)	5050	Dropout-2[0][0]
dropout_204 (Dropout)	(None, 3)	0	Flattencourse-MF[0][0]
dropout_206 (Dropout)	(None, 3)	0	FlattenUsers-MF[0][0]
FullyConnected-3 (Dense)	(None, 20)	1020	FullyConnected-2[0][0]
Dot (Concatenate)	(None, 6)	0	dropout_204[0][0] dropout_206[0][0]
Activation (Dense)	(None, 1)	21	FullyConnected-3[0][0]
Dot (Concatenate)	(None, 6)	0	dropout_204[0][0] dropout_206[0][0]
Activation (Dense)	(None, 1)	21	FullyConnected-3[0][0]
Concat-MF-MLP (Concatenate)	(None, 7)	0	Dot[0][0] Activation[0][0]
Combine-MF-MLP (Dense)	(None, 100)	800	Concat-MF-MLP[0][0]
FullyConnected-4 (Dense)	(None, 100)	10100	Combine-MF-MLP[0][0]
Prediction (Dense)	(None, 1)	101	FullyConnected-4[0][0]
Total params: 175,338			
Trainable params: 174,738			
Non-trainable params: 600			

Рисунок 3.1 - Модель з використанням методу Neu-CF-MF

Layer (type)	Output Shape	Param #	Connected to
Item (InputLayer)	(None, 1)	0	
User (InputLayer)	(None, 1)	0	
Course-Embedding (Embedding)	(None, 1, 3)	29175	Item[0][0]
User-Embedding (Embedding)	(None, 1, 3)	1833	User[0][0]
Flattencourse (Flatten)	(None, 3)	0	Course-Embedding[0][0]
FlattenUsers (Flatten)	(None, 3)	0	User-Embedding[0][0]
DotProduct (Dot)	(None, 1)	0	Flattencourse[0][0] FlattenUsers[0][0]

=====  
 Total params: 31,008  
 Trainable params: 31,008  
 Non-trainable params: 0

Рисунок 3.2 Модель з використанням методу Neu-MF

3.2 Діаграми класів, компонентів та структурно функціональна модель додатку

Діаграма класів призначена для відображення статичної структури системи в термінах класів у об'єктно-орієнтованому програмуванні. Це графічне зображення надає уявлення про взаємозв'язки між класами, об'єктами та підсистемами в межах предметної області. Крім того, вона включає опис внутрішньої структури кожного класу, такий як поля та методи, і також вказує на типи відносин, такі як спадкування чи реалізація інтерфейсів.

Важливо відзначити, що діаграма класів не передає інформацію про часові аспекти роботи системи. Основний акцент робиться на структурі та взаємозв'язках між класами під час проектування системи. Використання

об'єктно-орієнтованого підходу та патернів проектування є ключовими аспектами розробки на даному етапі [14].

Діаграма класів додатку-аналізатора представлена на рисунку 3.3.

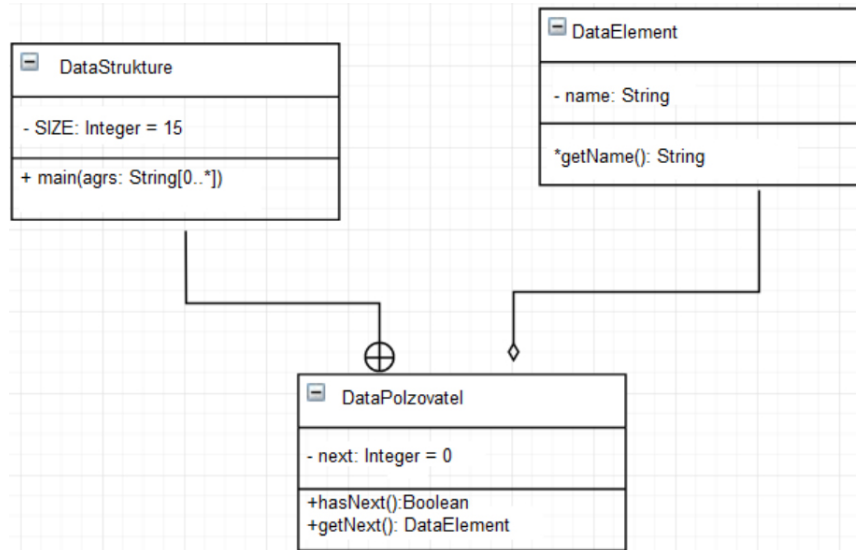


Рисунок 3.3 – Діаграма компонентів додатку бази користувачів

Діаграма компонентів визначає фізичне представлення системи та надає опис її архітектури, встановлюючи взаємозв'язки між програмними компонентами. На цій діаграмі використовуються компоненти, які можуть представляти вихідний, бінарний чи виконуваний код. Кожен компонент може відповідати окремому файлу в багатьох середовищах розробки. Стрілки, що з'єднують модулі, вказують на відношення взаємозалежності, подібні до тих, які виникають під час компіляції вихідного програмного коду. Основні графічні елементи на діаграмі компонентів включають самі компоненти, інтерфейси та залежності між ними.

Діаграма компонентів додатку-аналізатору бази користувачів соціальної мережі представлена на рисунку 3.4.

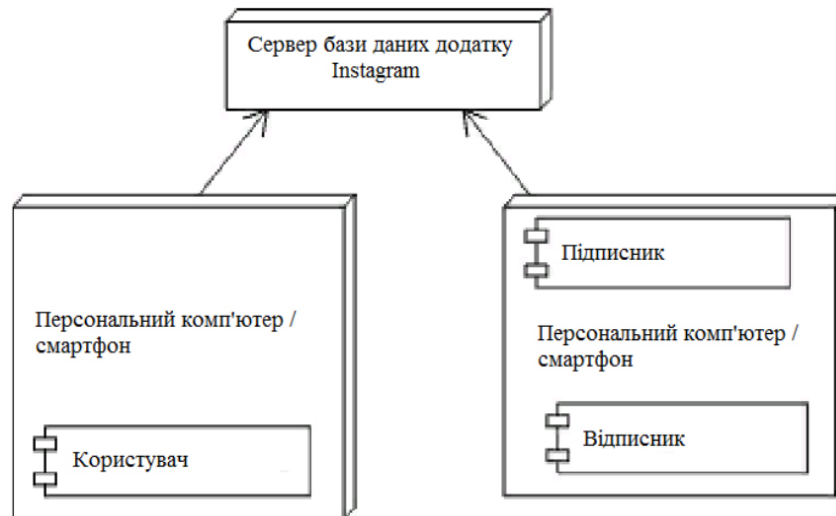


Рисунок 3.4 – Діаграма компонентів додатку бази користувачів

З діаграми видно, що додаток буде обробляти запити користувачів щодо даних підписників або відписників з їхніх персональних комп'ютерів або смартфонів і надсилатиме ці запити до сервера бази даних додатку Instagram.

Далі в моделюванні використовується структурно-функціональна модель. Функціональна модель описує обчислення в системі, вказуючи, як вихідні дані обчислюються за вхідними, не вдаючись у конкретний порядок та спосіб реалізації обчислень. Вона включає діаграми потоку даних, які вказують на потоки значень від вхідних даних через операції та внутрішні сховища даних до вихідних даних.

Метод SADT (IDEF0) використовується для структурування діяльності організації за бізнес-процесами, підкреслюючи перехід від організаційно-штатної структури до процесного підходу. Бізнес-процеси, спрямовані на створення значущих результатів для споживачів, виступають ключовим елементом цього підходу.

Контекстна діаграма структурно-функціональної моделі додатку представлена на рисунку 3.5.

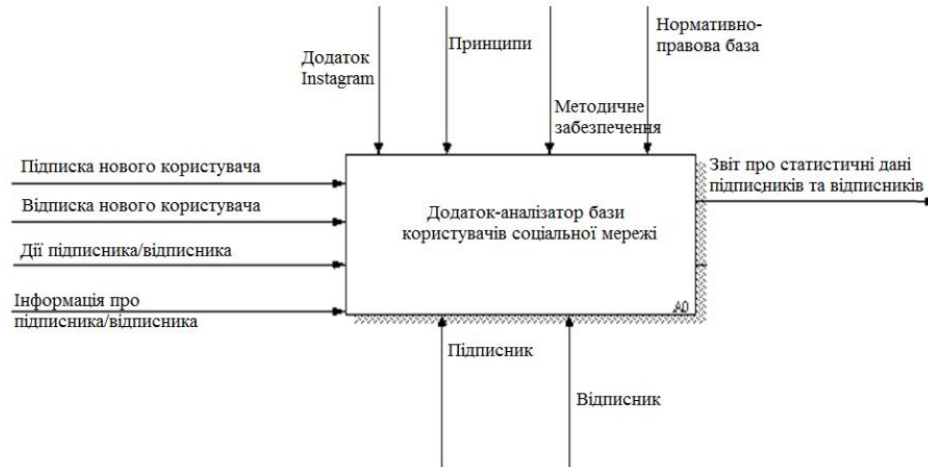


Рисунок 3.5 – Структурно-функціональна модель додатку

### 3.3 Модель варіантів використання додатку та загальна архітектура додатку

Модель, побудовану з використанням методу NN-Neu-MF, зображено на рисунку 3.6, модель MLP-RS – на рисунку 3.7.

Layer (type)	Output Shape	Param #	Connected to
Item (InputLayer)	(None, 1)	0	
User (InputLayer)	(None, 1)	0	
NonNegCourse-Embedding (Embedding)	(None, 1, 3)	29175	Item[0][0]
NonNegUser-Embedding (Embedding)	(None, 1, 3)	1833	User[0][0]
FlattenCourse (Flatten)	(None, 3)	0	NonNegCourse-Embedding[0][0]
FlattenUsers (Flatten)	(None, 3)	0	NonNegUser-Embedding[0][0]
DotProduct (Dot)	(None, 1)	0	FlattenCourse[0][0] FlattenUsers[0][0]

Total params: 31,008  
 Trainable params: 31,008  
 Non-trainable params: 0

Рисунок 3.6 - Модель, побудована з використанням методу NN-Neu-MF

Layer (type)	Output Shape	Param #	Connected to
Item (InputLayer)	(None, 1)	0	
User (InputLayer)	(None, 1)	0	
Course-Embedding (Embedding)	(None, 1, 8)	77800	Item[0][0]
User-Embedding (Embedding)	(None, 1, 5)	3055	User[0][0]
FlattenCourse (Flatten)	(None, 8)	0	Course-Embedding[0][0]
FlattenUsers (Flatten)	(None, 5)	0	User-Embedding[0][0]
dropout_222 (Dropout)	(None, 8)	0	FlattenCourse[0][0]
dropout_223 (Dropout)	(None, 5)	0	FlattenUsers[0][0]
Concat (Concatenate)	(None, 13)	0	dropout_222[0][0] dropout_223[0][0]
FullyConnected-1 (Dense)	(None, 100)	1400	Concat[0][0]
FullyConnected-2 (Dense)	(None, 50)	5050	FullyConnected-1[0][0]
FullyConnected-3 (Dense)	(None, 20)	1020	FullyConnected-2[0][0]
Activation (Dense)	(None, 1)	21	FullyConnected-3[0][0]

Total params: 88,346  
Trainable params: 88,346  
Non-trainable params: 0

Рисунок 3.7 - Модель MLP-RS

Представимо модель, яка ґрунтується на використанні рекурентної нейронної мережі з використанням довгої короткочасної пам'яті (LSTM). Модель, побудована за допомогою цієї архітектури, представлена на рисунку 3.8.

Layer (type)	Output Shape	Param #
embedding_56 (Embedding)	(None, 3, 3)	21
spatial_dropout1d_55 (Spatial	(None, 3, 3)	0
lstm_56 (LSTM)	(None, 3)	84
dense_112 (Dense)	(None, 3)	12
dropout_59 (Dropout)	(None, 3)	0
dense_113 (Dense)	(None, 7)	28

Total params: 145  
Trainable params: 145  
Non-trainable params: 0

Рисунок 3.8 - LSTM-based модель

Отже, з урахуванням результатів роботи програмного продукту можна зробити висновок, що найефективнішими є методи побудови рекомендаційних систем, які комбінуються між собою, а також моделі, що ґрунтуються на аналізі послідовностей. У випадку, якщо бізнес-вимоги цього вимагають, покращити представлені ефективні моделі можна шляхом:

- виділення додаткового часу для тренування моделей;
- додавання агрегованої інформації, що стосується користувацького досвіду взаємодії з МООС.

### 3.4 Проектування системи додатку

Порівняємо предиктивні здатності розроблених рекомендаційних систем. На рисунку 3.9 зображено предиктивну здатність рекомендаційної системи, побудованої з використанням методу Neu-CF-MF.

```
In [620]: from sklearn.metrics import mean_absolute_error
y_hat_2 = np.round(model.predict([test.user_id, test.course_id]),0)
print(mean_absolute_error(y_true, y_hat_2))
print(mean_absolute_error(y_true, model.predict([test.user_id, test.course_id])))
executed in 8.73s, finished 02:58:39 2020-05-30
0.6774097580325268
0.673544501072296
```

Рисунок 3.9 – Якість методу Neu-CF-MF

Рисунок 3.10 відображає прогностні можливості рекомендаційної системи, побудованої за допомогою методу матричної факторизації з використанням реальних значень.

```
In [641]: from sklearn.metrics import mean_absolute_error
mean_absolute_error(y_true, y_hat)
executed in 7ms, finished 03:24:07 2020-05-30
Out[641]: 0.779353431178104
```

Рисунок 3.10 - Якість алгоритму Neu-MF

Отже, після оцінювання предиктивної здатності висновки стосовно якості та ефективності найвище оцінили гібридний метод Neu-CF-MF. Цей метод об'єднує ключові концепції традиційних підходів колаборативної фільтрації та матричної факторизації, переносячи їх на основу нейронних мереж. Крім того, також високе значення ефективності було зафіксовано у випадку моделі LSTM, яка ґрунтується на рекурентних нейронних мережах. Процес навчання моделі Neu-CF-MF наведено на рисунку 3.13 [15].

На рисунку 3.11 зображено предиктивну здатність рекомендаційної системи, що побудована на основі MLP.

```
In [659]: y_hat_2 = np.round(model.predict([test.user_id, test.course_id]),0)
print(mean_absolute_error(y_true, y_hat_2))

print(mean_absolute_error(y_true, model.predict([test.user_id, test.course_id])))
executed in 8.19s, finished 04:07:16 2020-05-30
0.6706664022213408
0.6660313889922154
```

Рисунок 3.11 - Якість моделі MLP-RS

Рисунок 3.12 ілюструє прогностні можливості рекомендаційної системи, створеної за допомогою рекурентної нейронної мережі.

```
Epoch 100/100
1742/1742 [=====] - 1s 515us/step - loss: 1.0968 - accuracy: 0.5482 - val_loss: 0.5990 - val_acc
uracy: 0.8889
```

Рисунок 3.12 - Якість моделі на основі LSTM

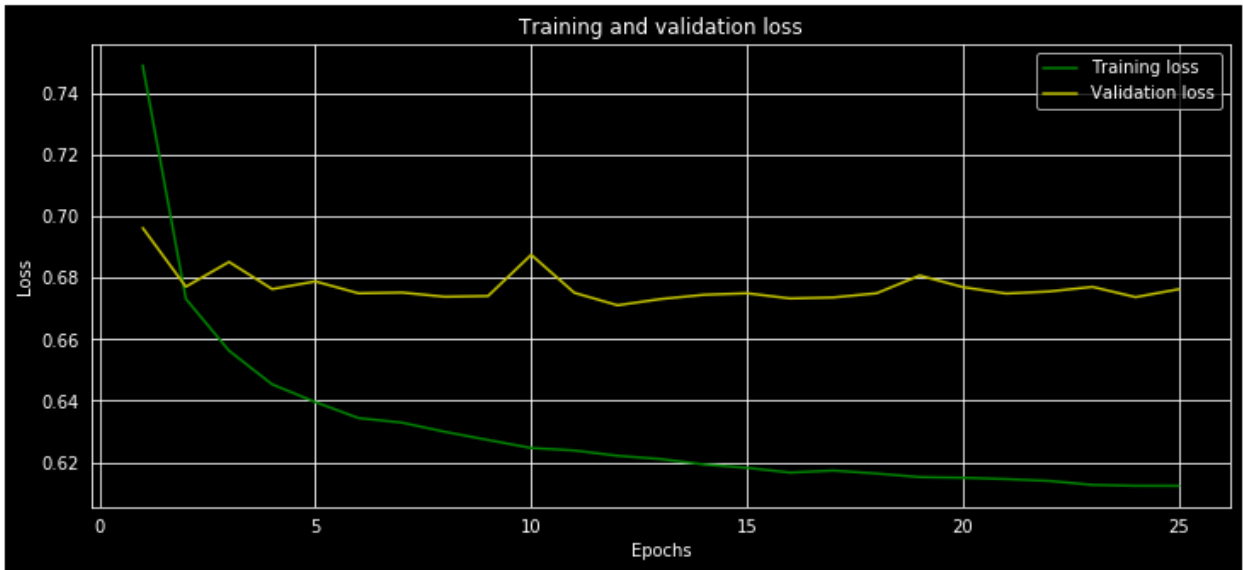


Рисунок 3.13 – Процес навчання Neu-CF-MF

Зобразимо процес навчання LSTM-моделі на рисунках 3.14 та 3.15.



Рисунок 3.14 – Навчання LSTM

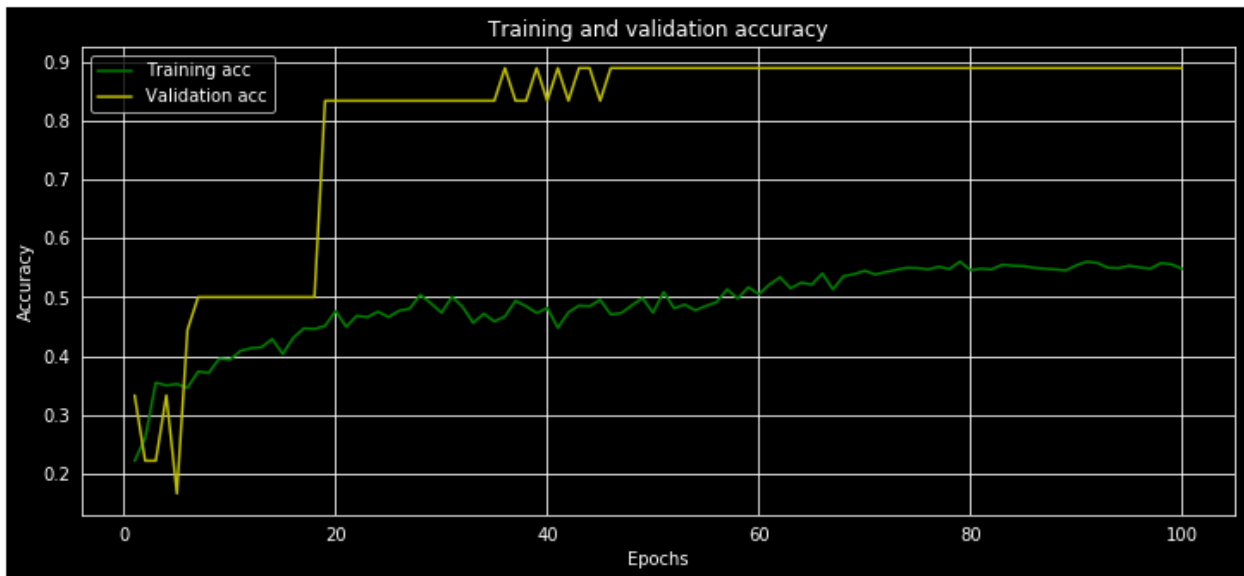


Рисунок 3.15 – Навчання LSTM

Таким чином, з урахуванням результатів функціонування програмного продукту можна зробити висновок, що найбільш ефективними є методи рекомендаційних систем, які комбінуються між собою, а також моделі, що використовують послідовності. З метою вдосконалення цих успішних моделей можна вжити наступні заходи, якщо це відповідає бізнес-вимогам:

- виділити більше часу для тренування моделей, щоб підвищити їхню точність та адаптацію;
- додати агреговані додаткові джерела інформації, що стосуються користувацького досвіду взаємодії з MOOC (масові відкриті онлайн-курси), для більш повного урахування контексту та упереджень користувачів.

## 4 РЕАЛІЗАЦІЯ ТА ПЕРЕВІРКА РОБОТИ ДОДАТКУ АНАЛІЗУ СОЦІАЛЬНИХ МЕРЕЖ

### 4.1 Реалізація розробленої моделі аналізу соціальних мереж

Збір даних та обробка даних. Кожна система рекомендацій у реальному часі передбачає попередню обробку перед тим, як ви передасте її механізму рекомендацій. Обробку даних у реальному часі можна здійснити різними способами. Можна обрати механізм обробки потоків, як-от Apache Flink, для трансформації поточкових даних у русі, або платформу аналізу даних у реальному часі, як-от Tinybird, для обчислення функцій на льоту за допомогою SQL.

Імпортуємо список елементів таблиці з серверної частини обраної соціальної мережі з бази даних SQLite серверної частини в Python (Jupyter Notebook). Вхідні дані користувачів було отримуємо з API у pythoneverywhere за допомогою `urllib.parse` та запитів для отримання запиту списку категорій, який представляє введені користувачами під час пошуку елементів (рис. 4.1) [17].

```
In [12]: df_original = pd.read_sql_query("SELECT * from geofree_api_item", conn)

In [13]: df = df_original.copy()
df.head()
# df.shape
# df.columns

Out[13]:
```

	id	title	description	available	timestamp_creation	latitude	longitude	condition	categories	category_ml	views	likes
0	1	test1	wewerwr	1	2023-04-03 17:01:03.428861	52.448102	13.384797	good	Clothes,Plants,Kitchen	Clothes	0	0
1	2	test1	wewerwr	1	2023-04-03 17:01:21.380178	52.448102	13.384797	good	Clothes,Kitchen,Plants	Clothes	0	0
2	3	test2	wewerwr	1	2023-04-03 17:01:32.759260	52.448102	13.384797	good	Kitchen,Plants	Kitchen	0	0
3	4	test4	wewerwr	1	2023-04-03 17:01:41.241430	52.448102	13.384797	good	Kitchen,Plants,Kids toys	Kitchen	0	0

Рисунок 4.1 – Імпорт даних

Обробка даних. Це можна виконати в наступному порядку: вибір ознак, кодування категоріальної змінної, фільтрація елементів за якими відбувається рекомендація та ранжування за ознаками.

Функції, вибрані з таблиці в базі даних, створеної серверною частиною: id, condition, category, views, likes, reposts.

Елементи були впорядковані за ознаками сортування (рис. 4.2) таким чином: post\_creation (за спаданням), кількість лайків (за зростанням), кількість переглядів (за спаданням) і умова (за зростанням) і обмежено трьома рекомендованими елементами.

```
ranking_df_filtered = ranking_df.copy()
# filter by categories and available
ranking_df_filtered = ranking_df_filtered[(ranking_df_filtered['categories'] == chosen_category) & (ranking_df_filtered['av
# rank these columns
ranking_df_filtered['timedate_creation_rank'] = ranking_df_filtered['timedate_creation'].rank(method='min', ascending=True)
ranking_df_filtered['condition_rank'] = ranking_df_filtered['condition'].rank(method='min', ascending=True)
ranking_df_filtered['views_rank'] = ranking_df_filtered['views'].rank(method='min', ascending=False)
ranking_df_filtered['likes_rank'] = ranking_df_filtered['likes'].rank(method='min', ascending=False)
# ranking_df_filtered['age_rank'] = ranking_df_filtered['age'].rank(method='min', ascending=True)
ranking_df_filtered['overall_rank'] = (ranking_df_filtered['timedate_creation_rank'] + ranking_df_filtered['condition_rank']
# sort by the higher overall_rank
ranking_sorted = ranking_df_filtered.sort_values(by='overall_rank', ascending = False)
# save only the ids
ranked_id = ranking_sorted['id'] # limit to the 1st, 2nd and 3rd results.]
```

### Рисунок 4.2 – Впорядкування елементів

На рисунку 4.3 наведено код який використовувався для отримання категорій (тематик для рекомендацій).

```
1 import urllib.parse
2 import requests
3 import urllib.parse
4 from urllib.parse import urlparse
5
6
7 url_search = 'https://geofree.pythonanywhere.com/api/get-categories/'
8 response2 = requests.get(url_search)
9
10 parsed_url = urlparse(url_search)
11 path = parsed_url.path
12
13 get_categories = response2.json()
14 get_categories
15
16 chosen_category = get_categories[0]['name']
17 chosen_category
```

### Рисунок 4.3 – Отримання категорій

На рисунку 4.4 наведено код з функцією ранжування. На рисунку 4.5 наведено код реалізації бекенду.

```

1 import sqlite3
2 import pandas as pd
3 import urllib.parse
4 import requests
5 from urllib.parse import urlparse
6 import json
7 import random
8
9
10 def create_ranking_df(chosen_category):
11     # connect the db
12     conn = sqlite3.connect('./db.sqlite3')
13     # creates a cursor object that allows Python code to interact with the SQLite database
14     cursor = conn.cursor()
15     # execute the query
16     cursor.execute("SELECT * FROM geofree_api_item;")
17     # the rows from the result set obtained by executing an SQL query
18     rows = cursor.fetchall()
19     # select the table
20     ranking_df = pd.read_sql_query("SELECT * from geofree_api_item", conn)
21     # close sql connection
22     conn.close()
23     # remove the column point
24     ranking_df = ranking_df.drop('point', axis=1)
25     # encode the column condition
26     ranking_df['condition'] = ranking_df['condition'].replace('like_new', 1).replace('good', 2).
27     # convert the column creation_date in datetime
28     ranking_df['timedate_creation'] = pd.to_datetime(ranking_df['creation_date'])
29     ## create a new column with post's ages in hours (it was a duplicity, creation_data is already)
30     # ranking_df['age'] = round((pd.Timestamp.now() - ranking_df['creation_date']) / pd.Timedelta(hours=1))
31     # create a new copy of the df
32     ranking_df_filtered = ranking_df.copy()
33     # filter by categories and available
34     ranking_df_filtered = ranking_df_filtered[(ranking_df_filtered['categories'] == chosen_category)]
35     # rank these columns
36     ranking_df_filtered['timedate_creation_rank'] = ranking_df_filtered['timedate_creation'].rank(ascending=True)
37     ranking_df_filtered['condition_rank'] = ranking_df_filtered['condition'].rank(method='min', ascending=True)
38     ranking_df_filtered['views_rank'] = ranking_df_filtered['views'].rank(method='min', ascending=True)
39     ranking_df_filtered['likes_rank'] = ranking_df_filtered['likes'].rank(method='min', ascending=True)
40     # ranking_df_filtered['age_rank'] = ranking_df_filtered['age'].rank(method='min', ascending=True)

```

Рисунок 4.4 – Функція ранжування

## 4.2 Експериментальна перевірка отриманих результатів

На рисунку 4.6 наведено результати синхронізації та обробки даних з Facebook.

```

1 # For the implementation, it was used Python decorator in Django REST framework to define a vi
2
3 @api_view(['GET'])
4 def getRakingRecomm(request):
5     category_arg = request.query_params.get('chosen_category')
6     # the ML model returns the id of 3 ranked recommended items based on the category selected
7     # recommendation_dir = os.path.join(settings.BASE_DIR, 'm_1', 'recommendation')
8     # sys.path.append(recommendation_dir)
9     # from recommendation_f import create_ranking_df
10    id1, id2, id3 = create_ranking_df(category_arg)
11    obj1 = Item.objects.get(id=id1)
12    obj2 = Item.objects.get(id=id2)
13    obj3 = Item.objects.get(id=id3)
14    ranked_items = []
15    serialized_obj1 = ItemSerializers(obj1).data
16    ranked_items.append(serialized_obj1)
17    serialized_obj2 = ItemSerializers(obj2).data
18    ranked_items.append(serialized_obj2)
19    serialized_obj3 = ItemSerializers(obj3).data
20    ranked_items.append(serialized_obj3)
21
22    return Response(ranked_items)

```

Рисунок 4.5 – Бекенд

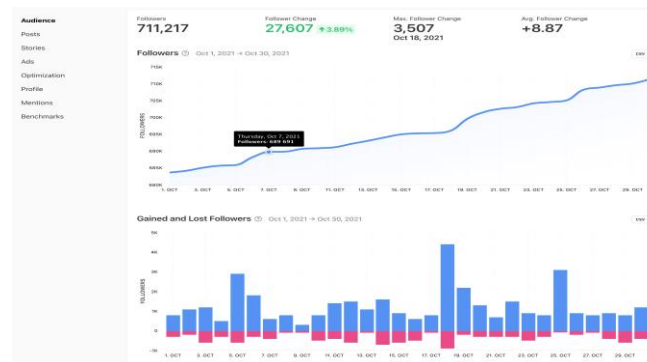


Рисунок 4.6 – Отримані результати синхронізування розроблюваного додатку з Facebook

На рисунках 4.7-4.8 наведено приклад роботи рекомендаційної системи на основі розробленого методу.

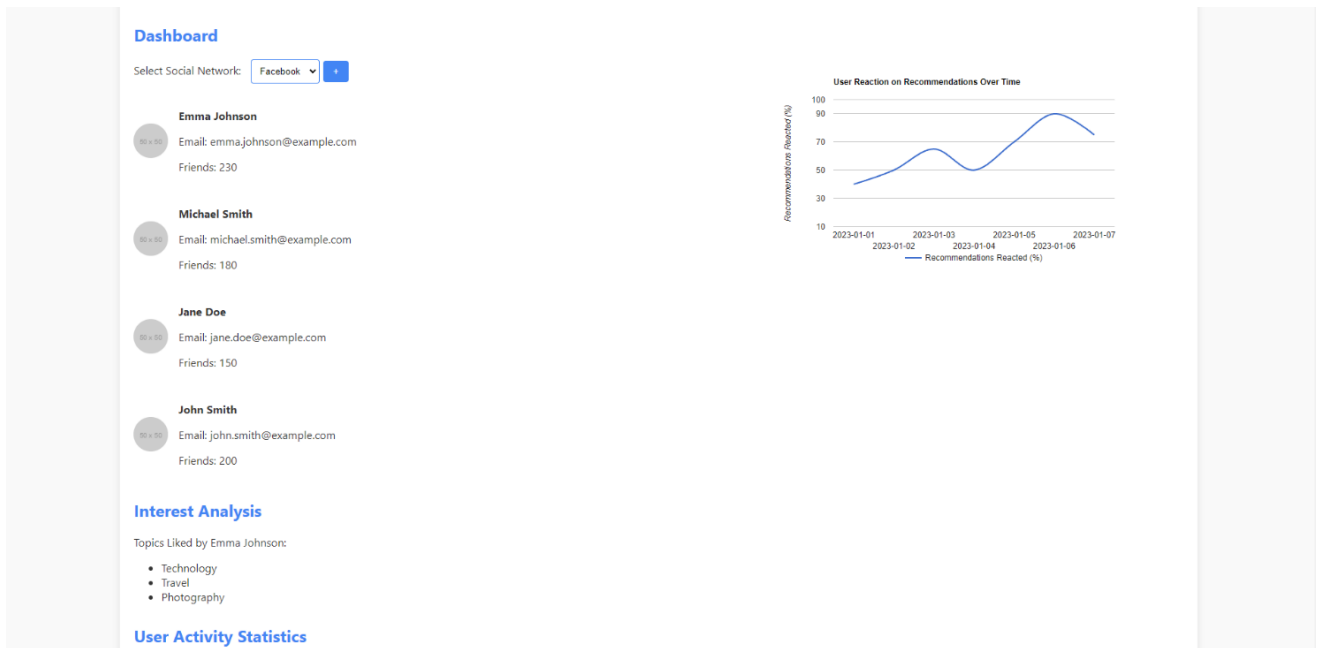


Рисунок 4.7 – Приклад роботи системи



Рисунок 4.8 – Приклад отриманих рекомендацій користувача

Отже можемо побачити, що система аналізує вподобання користувачів за тематиками та надає кращі рекомендації по публікаціям, рекламі та користувачам

зі схожими інтересами. На рисунку 4.9 наведено результати аналізу рекомендаційної системи. Згідно з графіком залежності кількості (відсотку) згенерованих вподобань від часу, можна зробити висновок, що система в більшості випадків надає такі рекомендації, що 50 - 90% з них цікавить користувачів та змушує їх реагувати та взаємодіяти з рекомендованими елементами (публікаціями, рекламою і т.д.).

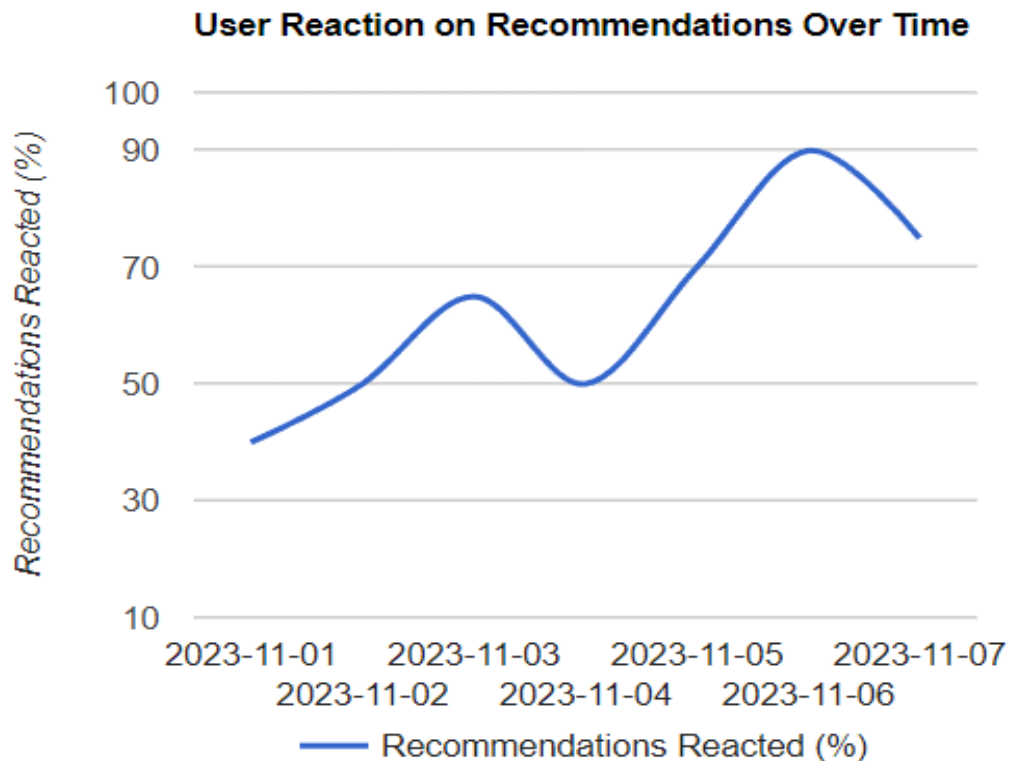


Рисунок 4.9 – Результати аналізу рекомендаційної системи

### 4.3 Рекомендації щодо використання системи

Рекомендаційна система на основі покращеного методі аналізу соціальних мереж може бути реалізована в якості веб-додатку (рисунки 4.11-4.13) та пропонує такі можливості:

- аналіз різноманітних соціальних мереж;
- забезпечення покращених рекомендацій користувачам на основі їхніх вподобань;
- аналіз активності користувачів в реальному часіб що дозволяє генерувати найбільш актуальні рекомендації;
- забезпечення безпеки і приватності користувачів (система не отримує інформацію з особистих переписок та закритих акаунтів).



Рисунок 4.11 – Приклад веб-додатку для аналізу соціальних мереж

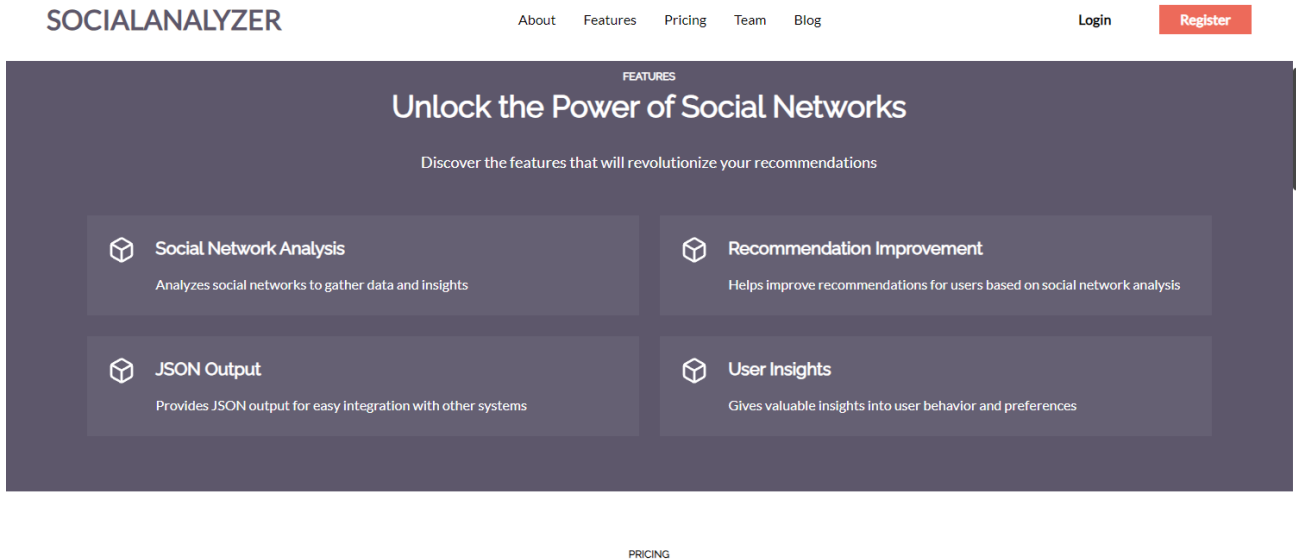


Рисунок 4.12 – Можливості веб-додатку

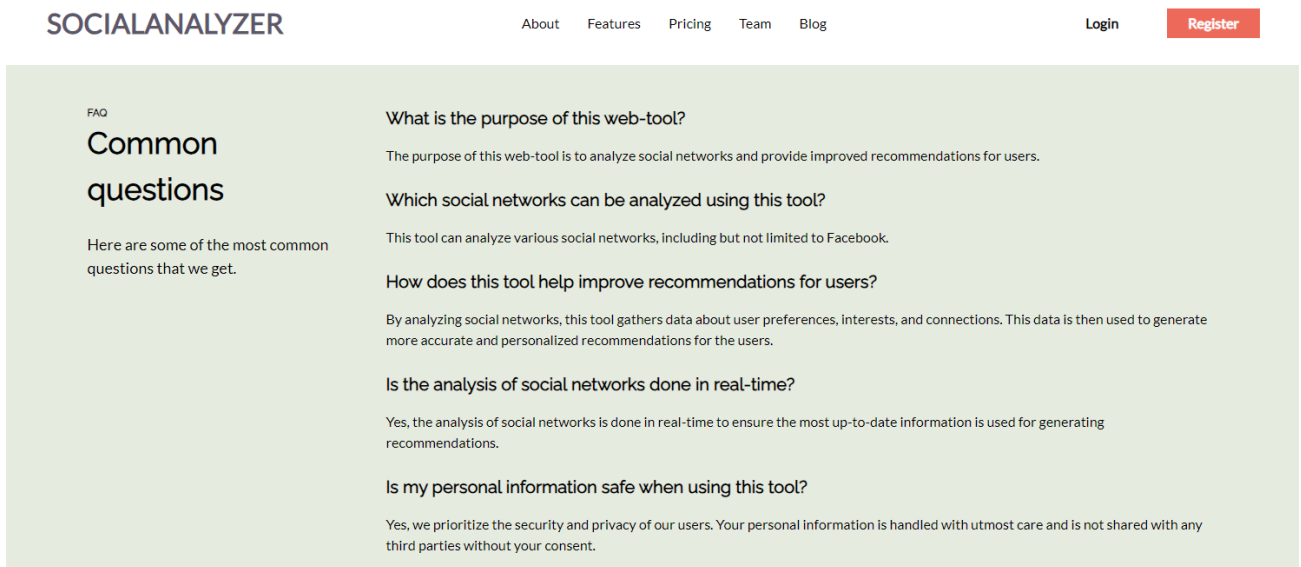


Рисунок 4.13 – Відповіді на поширені питання з приводу роботи додатку

Узагальнюючи результати роботи, можна визначити, що комбінація рекомендаційних систем Neu-CF та MLP, яка базується на аналізі соціальних мереж, виявилася досить ефективною. Однак для подальших досліджень і

вдосконалення системи рекомендацій можна врахувати декілька ключових аспектів.

Перш за все, можна розширити дослідження, вдосконалюючи моделі через виділення додаткового часу для тренування. Це може сприяти підвищенню точності та адаптації системи до змінних умов користувачів. Додатково, дослідження може бути розширене шляхом додавання агрегованої інформації, що стосується користувацького досвіду взаємодії з МООС, для більш повного урахування контексту та упереджень користувачів.

Потенційно цікавим напрямком досліджень є також розгляд розширених методів аналізу послідовностей та їх впливу на ефективність рекомендаційних систем. Додаткові експерименти та аналіз можуть допомогти виявити оптимальні підходи до використання послідовностей для покращення якості рекомендацій.

Отже, подальші дослідження можуть розглядати покращення точності, адаптивності та контексту рекомендаційних систем через розширення моделей та вдосконалення методів аналізу соціальних мереж та послідовностей.

## ВИСНОВКИ

У ході дослідження в магістерській роботі було вивчено вподобання користувачів та розглянуто наявні інформаційні системи для аналізу соціальних мереж. Також проведено аналіз різноманітних моделей соціальних мереж, включаючи графові, алгебраїчні, стохастичні та гібридні підходи.

У процесі вивчення базових рекомендаційних алгоритмів виявлено їх переваги та недоліки. Окрім того, розглянуті можливості удосконалення рекомендацій в соціальних мережах. В результаті аналізу обрано моделі Neu-CF та MLP для подальшого об'єднання, і визначено загальну структуру для нейронної колаборативної фільтрації.

Розроблена модель відрізняється значною гнучкістю та чутливістю до нелінійних взаємодій. Продемонстровано процес реалізації системи, використовуючи розроблений метод. Експериментальне дослідження, проведене в рамках роботи, підтверджує, що рекомендації користувачам на основі розробленої моделі в більшості випадків здійснюються з успіхом, з взаємодією користувачів у межах 50-90% отриманих рекомендацій.

## ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Дослідження моделей соціальних мереж. URL: <https://core.ac.uk/download/pdf/229721929.pdf> (дата звернення 15.12.2023).
2. Огляд моделей аналізу соціальних мереж. URL: <https://shorturl.at/dvES3> (дата звернення 15.12.2023).
3. Graph Models of Social Media Network As Applied to Facebook and Facebook Messenger Groups. URL: [https://www.researchgate.net/publication/368287579\\_Graph\\_Models\\_of\\_Social\\_Media\\_Network\\_As\\_Applied\\_to\\_Facebook\\_and\\_Facebook\\_Messenger\\_Groups](https://www.researchgate.net/publication/368287579_Graph_Models_of_Social_Media_Network_As_Applied_to_Facebook_and_Facebook_Messenger_Groups) (дата звернення 15.12.2023).
4. Використання семантичного простору соціальних мереж для аналізу споживчих уподобань. URL: <https://shorturl.at/ansC0> (дата звернення 15.12.2023).
5. Аналіз вподобань користувачів соціальних мереж на основі графових баз даних. URL: <https://shorturl.at/DMPX5> (дата звернення 15.12.2023).
6. інструменти дослідження. Аналіз соціальних мереж [електронна версія]. URL: <http://www.securityinfowatch.ru/view.php?section=articles&item=3> (дата звернення 18.12.2023).
7. Дослідження типів колаборативної фільтрації для побудови прогнозів у рекомендаційних системах URL: <https://shorturl.at/hjAO2> (дата звернення 20.12.2023).
8. Оцінка ефективності рекомендаційних систем IPTV, що враховують неявні уподобання URL: <http://sap.pstu.edu/article/view/240849> (дата звернення 20.12.2023).
9. Чалий С. Ф., Притула І. Б. Часо-об'єктні моделі подій та процесів в задачах створення програмних продуктів. Темпорально-об'єктні моделі подій і

процесів в задачах моделювання міркувань на основі прецедентів. Прецеденти. Київський науковий вісник. 2015. № 19. С. 71–75.

10. Чалий С.Ф., Притула І.Б., Адаптивні процесні методи Методи управління процесами на основі кейс-стаді підходу. Наукоємні технології. 2019. № 4. С. 410-414.

11. Чалий С.Ф., Лещинський В.О., Лещинська І.О. Моделювання контексту в рекомендаційних системах. Чалий С.Ф., Лещинський В.О., Лещинська І.О. Моделювання контексту в рекомендаційних системах. інформаційних технологій". 2019. № 1. С. 21-26.

12. Гунес І., Калелі К., Більге А., Полат Х. Проти атак Шилінга Атаки Шилінга проти систем номінації: комплексне дослідження Artificial Intelligence Review, 2020. С. 767-799.

13. Розробка та оцінка анотацій для рекомендаційних систем. Річчі, Ф., Рокач, Л., Шапіра, Б. та Кантор, П. (Ред.), Довідник рекомендаційних систем. Springer, Dordrecht. 2020. С. 479-510.

14. Тинтарев Н., Мастхофф Я. Оцінка ефективності анотацій для рекомендаційних систем. Оцінка ефективності анотацій для рекомендаційних систем. User Model User-Adap Inter. 2019. С. 399-439.

15. Браунгофер М. Гібридне вирішення проблеми холодного старту в контекстно-орієнтованих рекомендаційних системах. умар 2020: User Modelling, Adaptation and персоналізація. 2020. С. 484-489.

16. Social Network-Based Recommender Systems URL: [https://pzs.dstu.dp.ua/DataMining/recom/bibl/1schall\\_daniel\\_social\\_network\\_based\\_recommender\\_systems.pdf](https://pzs.dstu.dp.ua/DataMining/recom/bibl/1schall_daniel_social_network_based_recommender_systems.pdf) (дата звернення 27.12.2023).

17. Building a ranking-based recommendations system URL: <https://medium.com/@karina.condeixa/building-a-ranking-based-recommendations-system-e3adeb2c00f7> (дата звернення 27.12.2023).

18. Куликова О.В. Програмні засоби аналізу соціальних мереж. Цифровізація науки та сучасні тренди її розвитку : матеріали V Міжнародної студентської наукової конференції, м. Житомир, 24 листопада, 2023 рік / ГО «Молодіжна наукова ліга». Вінниця: ТОВ «УКРЛОГОС Груп», 2023. С. 315-317.

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

20. ДСТУ 3008:2015. Інформація та документація. Звіти у сфері науки і техніки. Структура і правила оформлювання. Чинний від 22.06.2015. Київ: ДП «УкрНДНЦ», 2016. 31 с.

21. ДСТУ 8302:2015. Інформація та документація. Бібліографічні посилання. Загальні положення та правила складання. Чинний від 04.03.2016. Київ: ДП «УкрНДНЦ», 2016. – 20 с.