

Міністерство освіти і науки України
Харківський національний університет радіоелектроніки

Факультет Комп'ютерних наук
(повна назва)
Кафедра Інформаційних управляючих систем
(повна назва)

КВАЛІФІКАЦІЙНА РОБОТА
Пояснювальна записка

рівень вищої освіти другий (магістерський)

Дослідження методів формування рекомендацій вакансій
для працівників фріланс-платформи
(тема)

Виконав:
здобувач 2 року навчання, групи ІУСТм-24-1
Артем ЄМЕЛЬЯНОВ
(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 зав. каф. ІУС Костянтин ПЕТРОВ
(посада, власне ім'я, ПРІЗВИЩЕ)

Допускається до захисту
Зав. кафедри ІУС


(підпис) Костянтин ПЕТРОВ
(власне ім'я, ПРІЗВИЩЕ)

2025 р.

Харківський національний університет радіоелектроніки

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

ЗАТВЕРДЖУЮ:

Зав. кафедри _____
(підпис)

“ 24 ” листопада 2025 р.

ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ

здобувачеві _____ Ємельянову Артему Віталійовичу _____
(прізвище, ім'я, по батькові)

1. Тема роботи Дослідження методів формування рекомендацій вакансій для працівників фріланс-платформи
затверджена наказом університету від «24» листопада 2025 р. № 1055Ст
2. Термін подання студентом роботи до екзаменаційної комісії «19» грудня 2025 р.
3. Вихідні дані до роботи наукова література з тематики кваліфікаційної роботи, матеріали передатестаційної практики, матеріали про методи рекомендації, джерела інформації про рекомендаційні системи
4. Перелік питань, що потрібно опрацювати в роботі аналіз предметної області та опис рекомендаційних систем; аналіз існуючих інформаційних систем і методів надання рекомендацій щодо пошуку вакансій та фріланс-проектів, їх переваг та недоліків; аналіз науково-технічної літератури у сфері побудови персоналізованих рекомендаційних систем; постановка задачі дослідження; дослідження методів формування рекомендацій вакансій; розробка гібридного методу побудови рекомендацій у інформаційній системі фріланс-платформи; практична реалізація гібридного методу; експериментальна оцінка гібридного методу.

КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Терміни виконання етапів роботи	Примітка
1	Аналіз предметної області	24.11.2025–25.11.2025	Виконано
2	Опис існуючих методів формування рекомендацій вакансій	26.11.2025–27.11.2025	Виконано
3	Постановка задачі дослідження	28.11.2025–29.11.2025	Виконано
4	Дослідження існуючих методів для формування рекомендацій	30.11.2025–02.12.2025	Виконано
5	Розробка гібридного методу побудови рекомендацій вакансій	03.12.2025–04.12.2025	Виконано
6	Практична реалізація гібридного методу	05.12.2025–07.12.2025	Виконано
7	Експериментальна оцінка гібридного методу	08.12.2025–09.12.2025	Виконано
8	Оформлення пояснювальної записки	10.12.2025	Виконано
9	Оформлення графічних матеріалів	11.12.2025	Виконано
10	Захист кваліфікаційної роботи	19.12.2025	Виконано

Дата видачі завдання «24» листопада 2025 р.

Здобувач


(підпис)

Керівник роботи



(підпис)

зав. каф. ІУС Костянтин ПЕТРОВ
(посада, власне ім'я, ПРІЗВИЩЕ)

РЕФЕРАТ

Пояснювальна записка кваліфікаційної роботи: 113 с., 7 рис., 2 табл., 2 дод., 30 джерел.

BERT, TF-IDF, ФРІЛАНС-ПЛАТФОРМА, РЕКОМЕНДАЦІЙНА СИСТЕМА, ГІБРИДНИЙ МЕТОД, КОНТЕНТНО-ОРІЄНТОВАНИЙ ПІДХІД, КОЛАБОРАТИВНА ФІЛЬТРАЦІЯ, СЕМАНТИЧНА МОДЕЛЬ.

Об'єктом дослідження є процес формування персоналізованих рекомендацій вакансій.

Предметом дослідження є методи формування рекомендацій вакансій.

Метою дослідження є розробка гібридного методу формування рекомендацій вакансій на фріланс-платформах для підвищення релевантності та точності підбору вакансій.

У ході роботи проаналізовано особливості ринку фріланс-послуг. Розглянуті основи методи формування рекомендацій. Виявлено основні проблеми рекомендаційних систем.

Розроблено гібридний метод формування рекомендацій вакансій для фріланс-платформи.

Створено програмний прототип рекомендаційної підсистеми із використанням Java, Python, Spring Boot, MySQL та зовнішнього сервісу семантичної обробки текстів. Експериментальна перевірка на основі стандартних метрик показала, що запропонований гібридний підхід забезпечує кращі показники якості порівняно з окремим застосуванням контентного, колаборативного та семантичного методів.

ABSTRACT

Explanatory note of the qualification work: 113 p., 7 fig., 2 table, 2 appendix, 30 sources.

BERT, TF-IDF, FREELANCE PLATFORM, RECOMMENDATION SYSTEM, HYBRID METHOD, CONTENT-ORIENTED APPROACH, COLLABORATIVE FILTERING, SEMANTIC MODEL.

The object of the study is the process of forming personalized job recommendations.

The subject of the study is the methods of forming job recommendations.

The purpose of the work is to analyze modern approaches to building recommendation systems for the freelance services market, identify their limitations and develop a hybrid method for forming job recommendations.

In the course of the work, the features of the freelance services market were analyzed. The basics of the recommendation formation method were discussed. The main problems of the recommendation system were identified.

A hybrid method for generating job recommendations for a freelance platform has been developed.

A software prototype of the recommendation subsystem has been created using Java, Python, Spring Boot, MySQL, and an external semantic text processing service. Experimental evaluation based on standard metrics has shown that the proposed hybrid approach provides better quality indicators compared to the separate use of content, collaborative, and semantic methods.

ЗМІСТ

Скорочення та умовні позначки	8
Вступ.....	9
1 Аналіз предметної області та постановка задачі дослідження.....	10
1.1 Аналіз предметної області та опис рекомендаційних систем	10
1.1.1 Стан та специфіка ринку фріланс-послуг в Україні та світі.....	10
1.1.2 Сутність та функціональні особливості рекомендаційних систем .	11
1.1.3 Аналіз існуючих методів надання рекомендацій, що використовуються в ІС	13
1.2 Аналіз існуючих ІС та методів надання рекомендацій щодо пошуку вакансій, їх переваги та недоліки	14
1.2.1 Огляд функціональних можливостей провідних платформ пошуку роботи.....	14
1.2.2 Порівняльний аналіз існуючих рішень	16
1.3 Аналіз існуючих методів та моделей для вирішення задач рекомендації вакансій	18
1.3.1 Метод контентної фільтрації на основі векторної моделі	18
1.3.2 Метод колаборативної фільтрації на основі матричної факторизації	19
1.3.3. Моделі на основі семантичних векторів та нейронних мереж.....	20
1.4 Постановка задачі дослідження.....	21
2 Розробка гібридного методу формування рекомендацій вакансій для фріланс платформи.....	23
2.1 Дослідження контентно-орієнтованої моделі	23
2.2 Дослідження методу колаборативної фільтрації	25
2.3 Дослідження семантичних методів	27
2.4 Розробка гібридного методу для формування рекомендацій вакансій .	29

3 Інформаційна технологія формування рекомендацій вакансій на фріланс-платформі	33
3.1 Контекстна діаграма процесу формування рекомендацій вакансій на фріланс-платформі	33
3.2 Опис діаграми декомпозиції	35
4 Практична реалізація гібридного методу	38
4.1 Програмна реалізація методів рекомендаційної системи	38
4.1.1 Обґрунтування технологічних засобів	38
4.1.2 Опис реалізації контентно-орієнтованої моделі	39
4.1.3 Опис реалізації колаборативної фільтрації	40
4.1.4 Опис реалізації семантичної моделі та BERT	41
4.1.5 Опис реалізації гібридного методу	42
4.1.5 Демонстрація працездатності розробленого гібридного методу	43
4.2 Експериментальна перевірка працездатності гібридного методу	47
4.2.1 Опис використаних метрик оцінювання.....	47
4.2.2 Результати експерименту та висновки.....	50
Висновки	53
Перелік джерел посилання	55
Додаток А Фрагменти коду	59
А.1 – Реалізація контентно-орієнтованої моделі	59
А.2 – Реалізація колаборативної фільтрації	65
А.3 – Реалізація семантичної моделі	74
А.4 – Реалізація BERT	80
А.5 – Реалізація гібридної моделі	82
Додаток Б Графічний матеріал кваліфікаційної роботи	100

СКОРОЧЕННЯ ТА УМОВНІ ПОЗНАКИ

ІС – інформаційна система

ІУС – інформаційні управляючі системи

КФ – колаборативна фільтрація

AP – average precision

BERT – Bidirectional Encoder Representations from Transformers

DCG – discounted cumulative gain

IDCG – ideal discounted cumulative gain

MAP – mean average precision

NDCG – normalized discounted cumulative gain

NLP – Natural Language Processing

NPS – net promoter score

TF-IDF – term frequency – inverse document frequency

ВСТУП

Розвиток цифрових технологій призвів до зростання популярності фріланс-платформ, які поєднують замовників і виконавців на основі проєктів. Зі збільшенням кількості користувачів і вакансій зростає й інформаційне навантаження: фахівцям стає складно знаходити релевантні пропозиції, а замовникам – швидко відбирати відповідних кандидатів. Це зумовлює потребу у використанні рекомендаційних систем, які автоматизують підбір вакансій.

У сучасних рекомендаційних системах використовують різні підходи до побудови рекомендацій: аналіз вмісту об'єктів, узагальнення поведінки користувачів, а також методи, що враховують семантичний зміст текстів на основі моделей глибокого навчання. Разом із тим кожен із цих підходів окремо має обмеження – розрідженість даних, ефект «холодного старту», неповне розуміння контексту текстових описів. Тому актуальним є пошук таких рішень, які дозволяють поєднувати сильні сторони різних методів у межах єдиного рекомендаційного механізму.

Об'єктом дослідження є процес формування персоналізованих рекомендацій вакансій.

Предметом дослідження є методи формування рекомендацій вакансій.

Метою дослідження є розробка гібридного методу формування рекомендацій вакансій на фріланс-платформах для підвищення релевантності та точності підбору вакансій.

Для досягнення мети проаналізовано існуючі методи рекомендацій, розроблено каскадний гібридний метод надання рекомендацій, реалізовано програмний прототип рекомендаційної підсистеми та проведено експериментальну оцінку його ефективності.

1 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ ТА ПОСТАНОВКА ЗАДАЧІ ДОСЛІДЖЕННЯ

1.1 Аналіз предметної області та опис рекомендаційних систем

1.1.1 Стан та специфіка ринку фріланс-послуг в Україні та світі

Сучасна світова економіка все більше відзначається поширенням гнучких форм зайнятості, що відобразилося у так званій "гіг-економіці". Ця економічна модель, що будується на короткострокових контрактах та проєктній роботі замість звичної довгострокової роботи, показує швидкі темпи зростання [1]. Основою та рушієм цього процесу стали цифрові фріланс-платформи. Такі світові гравці, як Upwork, Freelancer та Toptal, працюють як складні двосторонні ринки, виступаючи посередниками, що з'єднують мільйони замовників із виконавцями по всьому світу. Стан цього ринку характеризується високою динамікою та постійним зростанням обсягів [2].

Водночас, стрімке зростання кількості учасників та проєктів на цих платформах створює ключову проблему – інформаційне перевантаження. Ця проблема стосується обох сторін – виконавці не в змозі вручну обробити тисячі нових вакансій, що з'являються щодня, а замовники стикаються зі складним завданням вибору кваліфікованого кандидата з величезної кількості претендентів [3]. Така ситуація призводить до значних втрат часу та зусиль, знижує загальну ефективність ринку, що в кінцевому підсумку веде до втрати потенційних контрактів та зниження задоволеності користувачів.

Україна є важливою частиною світового ринку фрілансу та посідає на ньому помітні позиції, особливо у сегменті інформаційних технологій. Український ІТ-сектор є однією з ключових експортних галузей, причому значна частка фахівців працює саме у форматі фрілансу на світових платформах [4]. Окрім цього, в Україні працюють і власні місцеві платформи та ІТ-агрегатори, що мають свою специфіку та орієнтовані на місцевий ринок праці. Висока конкуренція серед виконавців, з одного боку, та постійний попит на

кваліфіковані кадри, з іншого, роблять задачу точного та своєчасного зіставлення вакансій та кандидатів особливо важливою для українського сегменту [5]. Таким чином, загальний стан ринку характеризується гострою потребою в автоматизованих інструментах для роботи з великими та різномірними потоками даних.

1.1.2 Сутність та функціональні особливості рекомендаційних систем

В умовах інформаційного перевантаження, головним інструментом для підвищення ефективності цифрових платформ стають рекомендаційні системи. Рекомендаційну систему можна визначити як клас програм, що автоматично прогнозує ступінь зацікавленості користувача (у вигляді рейтингу або уподобання) до певних об'єктів (наприклад, вакансій) та надає йому персоналізований список найбільш відповідних пропозицій [6]. На відміну від класичних пошукових систем, де користувач повинен сам ввести, що він шукає, рекомендаційні системи намагаються передбачити його справжні потреби. Вони роблять це, базуючись на аналізі профілю користувача, його попередньої поведінки та поведінки інших користувачів. Сутність рекомендаційних систем полягає у переході від загального формулювання "один для всіх" до особистого підходу до кожного користувача, що стало стандартом для успішних веб-платформ.

Основна мета впровадження рекомендаційної системи має дві сторони:

- підвищення користі сервісу для кінцевого користувача шляхом спрощення пошуку;
- досягнення конкретних бізнес-цілей платформи.

Залежно від сфери, це може бути збільшення обсягів продажів, підвищення тривалості перебування на сайті, або, як у випадку фріланс-платформ, збільшення кількості успішних взаємодій (поданих заявок, укладених

контрактів).

Функціонування рекомендаційної системи є постійним процесом і складаються з трьох основних етапів [6].

Перший етап – збір даних. Це базовий крок, що визначає якість майбутніх рекомендацій. Система збирає інформацію про те, як користувачі взаємодіють з об'єктами.

Збір інформації може здійснюватися двома способами:

- явний – коли користувач свідомо надає оцінку;
- неявний – збирається автоматично, коли система стежить за поведінкою.

До таких даних належать перегляди сторінок вакансій, час, проведений на сторінці, кліки, додавання до обраного, подання заявки або, навпаки, ігнорування пропозиції.

Наступний етап – аналіз та навчання. На цьому етапі зібрані дані обробляються з використанням алгоритмів машинного навчання. Мета полягає у пошуку прихованих закономірностей та аналізу уподобання користувачів. Саме тут застосовуються різні підходи (контент-базовані, колаборативні, гібридні тощо), які аналізують або схожість об'єктів між собою, або схожість у поведінці користувачів.

Заключний етап – формування рекомендації. Це результат, який безпосередньо бачить користувач. Формування рекомендацій починається з того моменту, коли система обчислює числове значення для кожного об'єкта-кандидата. Далі відбувається ранжування, де на основі цих прогнозів, система формує впорядкований список, сортуючи об'єкти від найбільш до найменш цікавих.

1.1.3 Аналіз існуючих методів надання рекомендацій, що використовуються в ІС

Аналіз практичних реалізацій рекомендаційних систем у провідних технологічних компаніях дозволяє прослідкувати еволюцію методів та оцінити їхню ефективність у різних предметних областях. Одним із фундаментальних підходів, що довів свою комерційну ефективність, є система, що впроваджена Amazon. У класичній науковій праці описано їх підхід, що базується на методі колаборативної фільтрації "елемент-до-елементу". На відміну від підходу "користувач-до-користувача", який є обчислювально складним на великих наборах даних через необхідність пошуку "сусідів" у динамічному просторі користувачів, метод Amazon полягає у розрахунку статичної матриці схожості між елементами на основі аналізу спільних покупок. Такий підхід характеризується високою масштабованістю та здатністю генерувати релевантні рекомендації в реальному часі, що стало індустріальним стандартом для сфери електронної комерції [7].

Сфера медіа-стрімінгу, представлена Netflix та Spotify, стала каталізатором для розробки більш складних моделей. Сучасні системи Netflix є складними багаторівневими гібридними архітектурами, що інтегрують колаборативні дані, контентні ознаки та контекстну інформацію. Особливістю є глибока персоналізація, що виходить за межі простого ранжування контенту і включає динамічну адаптацію користувацького інтерфейсу, наприклад, персоналізацію постерів до кінофільмів [6].

Платформа Spotify також використовує гібридний підхід, комбінуючи колаборативний аналіз плейлистів користувачів з глибоким контент-базованим аналізом аудіосигналів для виявлення схожості треків за музичними характеристиками.

У контексті даного дослідження, особливий інтерес представляє рекомендаційна система професійної мережі LinkedIn, оскільки вона вирішує

схожу задачу відповідності на ринку праці. Система LinkedIn є яскравим прикладом гібридної моделі, адаптованої до специфіки задачі "job matching" [8]. Вона використовує комбінацію контент-базованого аналізу та потужних колаборативних сигналів. Ефективність системи досягається шляхом врахування багатовимірних даних профілю, що дозволяє рекомендувати не лише очевидні, але й потенційно корисні кар'єрні можливості. Таким чином, аналіз промислових рекомендаційних систем доводить, що провідні платформи відійшли від простих алгоритмів на користь складних гібридних моделей, що глибоко інтегровані у бізнес-логіку платформи.

1.2 Аналіз існуючих ІС та методів надання рекомендацій щодо пошуку вакансій, їх переваги та недоліки

1.2.1 Огляд функціональних можливостей провідних платформ пошуку роботи

У рамках дослідження предметної області доцільно провести детальний аналіз існуючих ІС, що займають провідні позиції на світовому ринку праці та фрілансу.

Об'єктами аналізу обрано популярні платформи, що репрезентують різні концептуальні підходи до організації взаємодії між замовником та виконавцем:

- Upwork;
- LinkedIn;
- Fiverr;
- Freelancer.

Upwork є найбільшою глобальною платформою для фрілансу. Її підсистема рекомендацій відіграє ключову роль у розподілі трудових ресурсів. Функціонал платформи базується на використанні алгоритму оптимального підбору, який аналізує семантичну відповідність між описом проєкту та

профілем виконавця [9]. Окрім прямого збігу професійних навичок, система враховує приховані фактори, такі як історія успішного виконання проєктів у конкретних категоріях та зворотний зв'язок від попередніх замовників. Особливістю алгоритму є інтеграція прогнозової аналітики – система оцінює не лише релевантність навичок, але й ймовірність успішного завершення контракту на основі аналізу поведінкових патернів користувача.

LinkedIn використовує відмінний підхід, зумовлений наявністю специфічного типу даних – соціального графа [8]. Рекомендаційний механізм платформи виходить за межі стандартного аналізу тексту резюме, інтегруючи інформацію про професійні зв'язки користувача. Алгоритм ранжування надає пріоритет вакансіям у компаніях, де вже працюють особи з мережі контактів користувача або випускники того ж навчального закладу. Такий підхід базується на гіпотезі про те, що наявність соціального капіталу позитивно корелює з ймовірністю працевлаштування. Додатково застосовуються методи колаборативної фільтрації, що аналізують поведінку користувачів зі схожими професійними профілями для генерації релевантних пропозицій.

Fiverr реалізує модель "послуга як товар", де об'єктом рекомендації виступають стандартизовані картки послуг, а не вакансії. Архітектура рекомендаційної системи даної платформи наближена до систем електронної комерції. Механізм базується на аналізі історії переглядів, пошукових запитів та транзакцій [10]. Система формує персоналізовану стрічку пропозицій, використовуючи методи пошуку схожих об'єктів. Для виконавців реалізовано функцію аналітичної рекомендації, що на основі аналізу ринкового попиту пропонує створення нових типів послуг, які можуть бути затребувані в поточному періоді.

Freelancer представляє одну з найстаріших моделей на ринку, що поєднує класичний фріланс із конкурсною механікою. Рекомендаційна система платформи значною мірою є контент-базованою і спирається на чітке співставлення заявлених навичок із вимогами проєкту [11]. Відмінною рисою є інтеграція елементів гейміфікації: рейтинг та видимість пропозицій користувача

залежать від набраних балів досвіду та внутрішніх кваліфікаційних іспитів. Система характеризується високою інтенсивністю сповіщень, використовуючи механізми миттєвого інформування про нові проєкти, що відповідають ключовим словам профілю, часто нехтуючи глибоким контекстним аналізом на користь швидкості доставки пропозиції.

Проведений аналіз дозволяє стверджувати, що сучасні системи не обмежуються єдиним методом фільтрації, а використовують комплексний підхід, в якому поєднуються аналіз тексту, поведінкова історія та соціальні зв'язки.

1.2.2 Порівняльний аналіз існуючих рішень

На основі огляду функціональних можливостей провідних платформ проведено систематизацію ключових характеристик, що визначають ефективність рекомендаційних систем у даній предметній області. Метою порівняльного аналізу є виявлення переваг та недоліків існуючих реалізацій для врахування їх при розробці власної моделі.

Основними критеріями порівняння визначено: тип алгоритмічної моделі, джерела вхідних даних для аналізу та специфічні функціональні обмеження. Результати порівняльного аналізу наведено в табл. 1.1.

Таблиця 1.1. Порівняльний аналіз рекомендаційних систем фріланс-платформ

Платформа	Тип моделі	Ключові дані для аналізу	Особливості реалізації	Переваги
Upwork	Гібридна	Історія пропозицій, теги навичок, семантика опису вакансії, рейтинг успішності.	Маркування вакансій за ступенем відповідності; персоналізована стрічка.	Висока точність для користувачів з історією; врахування економічних факторів.

Продовження таблиці 1.1.

Платформа	Тип моделі	Ключові дані для аналізу	Особливості реалізації	Переваги
LinkedIn	Гібридна	Профіль користувача, соціальний граф, дані про компанію, навички.	Окремий розділ рекомендованих вакансій; сповіщення про наявність контактів у компанії.	Використання унікальних соціальних даних; високий рівень довіри до рекомендацій.
Fiverr	Об'єктно-орієнтована	Історія переглядів, транзакцій, пошукові запити.	Стрічки рекомендацій на основі історії переглядів.	Ефективна робота зі структурованим каталогом послуг; адаптивність до зміни інтересів.
Freelancer	Контент-базована	Заявлені навички, результати кваліфікаційних тестів, бали досвіду.	Інтенсивні поштові розсилки; фільтрація проєктів за тегами; конкурси.	Прозорість роботи; швидкість інформування; гейміфікація процесу.

Результати аналізу, наведені в табл. 1.1, свідчать про те, що найбільш перспективною є гібридна модель, реалізована у системах Upwork та LinkedIn. Такий підхід дозволяє компенсувати обмеження окремих методів фільтрації. Водночас, спільним недоліком розглянутих систем є низький рівень інтерпретованості результатів для кінцевого користувача та високий бар'єр входження для нових учасників через відсутність накопиченої історії взаємодій.

Враховуючи виявлені особливості, при розробці власної системи доцільно орієнтуватися на гібридну архітектуру з посиленням блоком семантичного аналізу текстових даних для вирішення проблеми нових вакансій та користувачів.

1.3 Аналіз існуючих методів та моделей для вирішення задач рекомендації вакансій

1.3.1 Метод контентної фільтрації на основі векторної моделі

Векторна модель є фундаментальним підходом в теорії інформаційного пошуку, який застосовується для формалізації текстових даних у задачах рекомендації вакансій. Основна ідея методу полягає у представленні текстових документів як наборів числових характеристик у багатовимірному просторі, де кожен вимір відповідає певному слову зі словника.

Ключовим етапом даного підходу є визначення важливості кожного терміна в документі, для чого традиційно використовується статистична міра TF-IDF. Ця метрика дозволяє оцінити вагу слова, базуючись на двох протилежних факторах:

- частота терміна – чим частіше слово зустрічається в конкретному документі (наприклад, вакансії), тим важливішим воно вважається для опису змісту цього документа;

- обернена частота документа – цей компонент знижує вагу слів, які зустрічаються занадто часто в усій базі даних, якщо слово зустрічається скрізь, воно не є унікальним і не допомагає відрізнити одну вакансію від іншої.

Таким чином, найбільшу вагу отримують слова, які часто вживаються в конкретній вакансії, але рідко зустрічаються в інших.

Для оцінки ступеня схожості між профілем користувача та вакансією використовується косинусна міра подібності. Вона дозволяє визначити, наскільки близькими є два документи за своїм змістом, вимірюючи кут між їхніми векторами. Якщо вектори знаходяться в одному напрямку, документи вважаються дуже схожими. Перевагою цього методу є те, що він коректно порівнює тексти різної довжини. Однак суттєвим обмеженням є ігнорування змісту слів [12].

1.3.2 Метод колаборативної фільтрації на основі матричної факторизації

Методи матричної факторизації належать до класу моделей з латентними факторами і є найбільш ефективним підходом у рамках колаборативної фільтрації. Задача рекомендації в цьому контексті розглядається як спроба передбачити, як користувач оцінив би вакансію, з якою він ще не взаємодіяв, на основі його минулих дій.

Суть методу полягає у розкладанні величезної таблиці взаємодій (де рядки – це користувачі, а стовпці – вакансії) на дві менші, компактні матриці:

- перша матриця описує кожного користувача через набір прихованих характеристик;

- друга матриця описує кожну вакансію через той самий набір характеристик.

Ці "приховані характеристики" не задаються вручну, а виявляються алгоритмом автоматично під час навчання на основі аналізу закономірностей у діях усіх користувачів системи. Прогнозована оцінка для конкретної пари "користувач-вакансія" розраховується шляхом поєднання їхніх векторів прихованих характеристик.

Головна перевага цього методу – здатність знаходити неочевидні зв'язки та рекомендувати вакансії, які користувач міг би пропустити при звичайному пошуку. Головним недоліком є проблема "холодного старту": метод не може сформувати профіль прихованих факторів для нового користувача або нової вакансії, доки вони не накопичать історію взаємодій [13].

1.3.3. Моделі на основі семантичних векторів та нейронних мереж

Для подолання обмежень класичних статистичних методів, які ігнорують змістовий зв'язок між словами, у сучасних рекомендаційних системах застосовуються моделі глибокого навчання для генерації щільних векторних представлень. Фундаментальною відмінністю цього підходу є перехід від простого підрахунку слів до розуміння їхнього контексту.

В основі даного класу моделей лежить гіпотеза, що слова, які зустрічаються у схожих контекстах, мають близькі значення. Реалізація цього принципу відбулася у два етапи:

- статичні моделі – архітектури типу Word2Vec та GloVe використовують нейронні мережі для навчання на великих обсягах тексту. У результаті кожному слову зі словника ставиться у відповідність фіксований набір чисел. Схожість слів визначається відстанню між ними у математичному просторі. Проте недоліком є те, що вектор для слова залишається незмінним незалежно від контексту;

- контекстуальні моделі – сучасний етап розвитку базується на архітектурі трансформерів. Ключовим елементом тут є механізм уваги, який дозволяє моделі аналізувати зв'язки між усіма словами в реченні одночасно. На відміну від статичних підходів, BERT генерує унікальне представлення для слова в кожному конкретному випадку його використання, динамічно змінюючи його значення залежно від оточення.

Для задач рекомендації вакансій це дозволяє системі аналізувати та розуміти зміст цілих абзаців тексту. Отримані семантичні вектори дозволяють знаходити релевантні вакансії навіть за повної відсутності спільних ключових слів, спираючись виключно на змістову близькість понять.

Встановлено, що класичні методи мають критичні обмеження для сфери фрілансу. Прості статистичні методи не враховують семантику, а методи факторизації не працюють з новими об'єктами без історії.

Після аналізу існуючих систем-аналогів, методів та моделей та їх переваг і недоліків, можна зробити висновок, що для якісного вирішення задачі рекомендацій вакансій на фріланс-платформі необхідно розробити гібридний метод формування рекомендацій [14].

1.4 Постановка задачі дослідження

В ході аналізу існуючих систем-аналогів та методів формування рекомендацій вакансій виявлено низку суттєвих недоліків, що обмежують їх ефективність у динамічних умовах фріланс-платформ:

- неспроможність класичних текстових методів враховувати семантичний зв'язок слів, що призводить до втрати релевантних вакансій, якщо вони описані іншими термінами;
- проблема "холодного старту" в методах колаборативної фільтрації, що унеможливорює надання точних рекомендацій для нових користувачів та нових вакансій, які ще не мають історії взаємодій;
- проблема розрідженості даних, характерна для великих платформ, де один користувач взаємодіє лише з мізерною часткою доступних проєктів, що знижує точність прогнозів;
- складність інтерпретації результатів роботи складних нейромережових моделей, що знижує довіру користувачів до системи.

Таким чином, враховуючи вищезазначені недоліки, можна сформулювати мету, об'єкт та предмет дослідження.

Метою дослідження є розробка гібридного методу формування рекомендацій вакансій на фріланс-платформах для підвищення релевантності та точності підбору вакансій.

Для досягнення поставленої мети необхідно вирішити такі основні завдання:

- провести детальний аналіз існуючих підходів до побудови рекомендаційних систем та визначити їх слабкі сторони в контексті ринку праці;
- дослідити методи векторизації тексту на основі глибокого навчання для покращення розуміння змісту вакансій;
- розробити власний гібридний метод формування рекомендацій вакансій на фріланс-платформах;
- розробити програмний прототип рекомендаційної системи для експериментальної перевірки працездатності запропонованого методу.
- оцінити ефективність розробленого методу за допомогою стандартних метрик якості та порівняти його з базовими підходами.

Об'єктом дослідження є процес формування персоналізованих рекомендацій вакансій.

Предметом дослідження є методи формування рекомендацій вакансій.

Основним результатом роботи є розробка гібридного методу формування рекомендацій та реалізація, на його основі, програмного модуля рекомендаційної системи. Цей модуль повинен забезпечити можливість автоматичного ранжування вакансій з урахуванням як явних вимог, так і прихованого змісту описів, що дозволить значно скоротити час пошуку роботи для фрілансерів.

Таким чином, основна задача полягає у створенні гібридного методу, який нівелює недоліки окремих методів фільтрації, забезпечуючи високу точність рекомендацій навіть за умов неповноти даних або відсутності історії дій, що сприятиме підвищенню ефективності взаємодії між замовниками та виконавцями.

2 РОЗРОБКА ГІБРИДНОГО МЕТОДУ ФОРМУВАННЯ РЕКОМЕНДАЦІЙ ВАКАНСІЙ ДЛЯ ФРІЛАНС ПЛАТФОРМИ

2.1 Дослідження контентно-орієнтованої моделі

Контентно-орієнтована модель для рекомендації вакансій спирається на перетворення текстів профілів користувачів і вакансій у вектори, потім порівнює їхню схожість. На фріланс-платформах це можуть бути заголовки замовлень, деталі роботи, технічні умови, необхідна кваліфікація чи знання мов – окремо взяті або разом з даними про працівника.

Для реалізації цього підходу текстові дані подаються у вигляді векторів ознак у просторі термінів. Однією з базових і найбільш поширених схем зважування термів є TF-IDF, що використовується в інформаційному пошуку та текстових рекомендаційних системах [15].

Нехай D – корпус документів (описів вакансій), $N = |D|$ – кількість документів у корпусі, $d \in D$ – окремий документ, t – термін.

Термінна частота у документі d визначається за наступною формулою [15]:

$$\text{tf}(t, d) = \frac{f_{t,d}}{\sum_{t' \in d} f_{t',d}}, \quad (2.1)$$

де $f_{t,d}$ – кількість входжень терміна t у документі d ;

$\sum_{t' \in d} f_{t',d}$ – загальна кількість термінів у документі.

Обернена частота документа, яка відображає рідкісність терміна в корпусі, визначається за наступною формулою [15]:

$$\text{idf}(t, D) = \log \frac{N}{n_t}, \quad (2.2)$$

де N – кількість документів у корпусі;

n_t – кількість документів, що містять термін t .

Відповідно, вага терміна за схемою TF-IDF записується у вигляді [15]:

$$\text{tfidf}(t, d, D) = \text{tf}(t, d) \cdot \text{idf}(t, D), \quad (2.3)$$

де $\text{tf}(t, d)$ – нормована частота терміна в документі;

$\text{idf}(t, D)$ – обернена частота документа;

$\text{tfidf}(t, d, D)$ – підсумкова вага терміна t у документі d відносно корпусу D .

За допомогою такої схеми зважування зменшується вплив поширених, малозмістовних слів і підсилює терміни, характерні для певних доменів чи технологій, що є важливим для рекомендацій технічних вакансій.

Кожен опис вакансії d подається як вектор:

$$\mathbf{v}_d = (\text{tfidf}(t_1, d, D), \dots, \text{tfidf}(t_m, d, D)) \quad (2.4)$$

де t_1, \dots, t_m – терміни словника;

\mathbf{v}_d – вектор ознак документа у просторі розмірності m .

Аналогічним чином створюється вектор профілю користувача, якщо проаналізувати його профіль із усіма описовими характеристиками.

Після процесу побудови векторів, наступним етапом є визначення міри подібності між профілем користувача та вакансією. Для знаходження подібності використовується косинусна міра схожості. Вона визначається за наступною формулою [16]:

$$\text{sim}_{\cos}(\mathbf{v}_u, \mathbf{v}_i) = \frac{\mathbf{v}_u \cdot \mathbf{v}_i}{\|\mathbf{v}_u\| \|\mathbf{v}_i\|} = \frac{\sum_{k=1}^m v_{u,k} v_{i,k}}{\sqrt{\sum_{k=1}^m v_{u,k}^2} \sqrt{\sum_{k=1}^m v_{i,k}^2}} \quad (2.5)$$

де \mathbf{v}_u – векторне подання профілю користувача;

\mathbf{v}_i – векторне подання опису вакансії;

$v_{u,k}$ та $v_{i,k}$ – ваги k -го терміна у векторах відповідно;

m – розмірність простору ознак;

$\|v_u\|$ та $\|v_i\|$ – евклідові норми векторів.

Значення цієї міри інтерпретується як косинус кута між двома векторами: чим ближчим до одиниці є значення, тим більш схожими за текстовим змістом є профіль користувача та вакансія.

Поєднуючи TF-IDF з косинусною мірою подібності для обробки текстових описів дозволяє описати як зміст вакансій, так і характеристики профілів фрілансерів і використовувати цю інформацію для побудови контентно-орієнтованих рекомендацій.

2.2 Дослідження методу колаборативної фільтрації

Колаборативна фільтрація виконує задачу рекомендацій з точки зору аналізу історії взаємодій користувачів із вакансіями. Алгоритм колаборативної фільтрації полягає в тому, що вона підбирає користувачів з подібними взаємодіями щодо однакових об'єктів та мають схожі вподобання і в майбутньому може будуть зацікавлені в подібних вакансіях. У рамках фріланс-платформи взаємодіями можуть бути перегляди, відгуки, прийняття пропозицій, успішне завершення контрактів, збереження та інші.

Нехай матриця взаємодій позначається як:

$$R = (r_{u,i}) \quad (2.6)$$

де u – індекс користувача;

i – індекс вакансії;

$r_{u,i}$ – показник взаємодії користувача u із вакансією i .

У варіанті КФ на основі користувачів для кожного користувача шукають множину користувачів з подібною поведінкою – і на основі їхніх взаємодій прогнозують корисність вакансій для цільового користувача. Для оцінки

подібності профілів користувачів часто застосовується коефіцієнт лінійної кореляції Пірсона.

Припустимо, що I_u є множиною вакансій, з якими взаємодіє користувач u , а через $I_{uv} = I_u \cap I_v$ позначимо множину вакансій, спільних для двох користувачів u та v . Середнє значення взаємодій користувача u позначимо як \bar{r}_u . Тоді коефіцієнт Пірсона між користувачами u та v буде обчислюватись за наступною формулою [17]:

$$\text{sim}_P(u, v) = \frac{\sum_{i \in I_{uv}} (r_{u,i} - \bar{r}_u)(r_{v,i} - \bar{r}_v)}{\sqrt{\sum_{i \in I_{uv}} (r_{u,i} - \bar{r}_u)^2} \sqrt{\sum_{i \in I_{uv}} (r_{v,i} - \bar{r}_v)^2}}, \quad (2.7)$$

де $r_{u,i}$ та $r_{v,i}$ – значення взаємодій користувачів u та v з вакансією i ;

\bar{r}_u та \bar{r}_v – середні значення відповідних взаємодій по користувачах;

I_{uv} – множина вакансій, з якими взаємодіяли обидва користувачі.

Значення $\text{sim}_P(u, v)$ змінюється у діапазоні від -1 до 1 і відображає ступінь лінійної залежності між шаблонами оцінювання двох користувачів.

Після процесу обчислення схожості користувачів обирається підмножина найбільш подібних $N_u(i)$ для користувача u при прогнозуванні відповідної вакансії i . Для прогнозу значення $\hat{r}_{u,i}$ використовується класична формула колаборативної фільтрації на основі користувачів, яка виглядає як [17]:

$$\hat{r}_{u,i} = \bar{r}_u + \frac{\sum_{v \in N_u(i)} \text{sim}_P(u, v) (r_{v,i} - \bar{r}_v)}{\sum_{v \in N_u(i)} |\text{sim}_P(u, v)|} \quad (2.8)$$

де $\hat{r}_{u,i}$ – прогнозоване значення взаємодії користувача u з вакансією i ;

\bar{r}_u – середнє значення оцінок користувача u ;

$N_u(i)$ – множина сусідів користувача u , які вже взаємодіяли з вакансією i ;

$\text{sim}_P(u, v)$ – коефіцієнт Пірсона між користувачами u та v ;

$r_{v,i}$ – значення взаємодії сусіда v із вакансією i ;

\bar{r}_v – середнє значення його оцінок.

Така нормалізація дозволяє компенсувати індивідуальні зсуви у шкалах оцінювання різних користувачів.

Підбиваючи підсумки, можна сказати, що КФ дає змогу враховувати досвід фрілансерів та виявляти вакансії, які виявились якісними для користувачів із подібними взаємодіями. Але варто врахувати, що даний підхід має відомі обмеження, у вигляді, розрідженості матриці взаємодій при великій кількості користувачів та проблема «холодного старту», у випадку, коли користувач не має історії взаємодії.

Для розширення колаборативної фільтрації на основі користувачів, зазвичай, використовуються КФ на основі об'єктів, у якій для прогнозування значень використовується подібність між вакансіями. Ідея даного типу фільтрації полягає в тому, що об'єкти, з якими взаємодіяли одні і ті ж самі користувачі, можуть мати спільні характеристики, і тому їхня схожість може бути використана для визначення, наскільки інша вакансія підходить для користувача [18].

Подібність між вакансіями обчислюється аналогічним чином, як і для користувачів, але з транспонованою матрицею взаємодій.

КФ на основі об'єктів вважається більш стійким у випадках, коли кількість вакансій суттєво менша за кількість користувачів, оскільки поведінкові профілі вакансій формуються щільніше, ніж профілі користувачів.

У фріланс-платформі цей підхід дозволяє ефективно визначати вакансії, які підходять користувачеві, навіть при відносно невеликій кількості взаємодій, оскільки вагому роль відіграє структура схожості самих вакансій.

2.3 Дослідження семантичних методів

При дослідженні контентно-орієнтованої моделі, можна прийти до висновку, що TD-IDF трактує документи як набір незалежних термінів і не може

якісно враховувати контекстні зв'язки між словами. Дане обмеження є проблемним для фріланс-платформ, так як у багатьох випадках, зміст вакансій може бути сформульований різними словами та конструкціями.

Для вирішення цієї проблеми існують сучасні підходи, які орієнтуються на контекстні векторні представлення, що будуються за допомогою трансформерних моделей. Однією із найбільш популярних та впливових моделей такого типу є BERT (Bidirectional Encoder Representations from Transformers).

BERT складається з послідовності енкодерних шарів. Кожен такий шар містить механізм лінійного перетворення, завдяки чому текстова послідовність на вході переводиться у послідовність векторів прихованих станів. Сукупність шарів утворює енкодер, тобто відображення від простору текстів до простору векторних представлень фіксованого розміру. Якщо казати узагальнено, енкодер можна розглядати як функцію, що отримує токенизований текст і повертає для нього числове подання, яке в подальшому буде використовуватись для порівняння та класифікації [19].

Для практичних застосувань, створено декілька варіантів архітектури, зокрема BERT-Base (12 енкодерних шарів, розмір прихованого стану 768) та BERT-Large (24 шари, прихований розмір 1024), які відрізняються якістю представлень та обчислювальною вартістю [19].

Нехай $\text{enc}(\cdot)$ позначає енкодер, побудований на основі BERT. Тоді опис вакансії i і текстовий опис профілю користувача u можуть бути перетворені у вектори фіксованої розмірності за формулами:

$$e_i = \text{enc}(\text{опис вакансії } i); e_u = \text{enc}(\text{профіль користувача } u) \quad (2.9)$$

де e_i – векторне представлення текстового опису вакансії i ;

e_u – векторне представлення тексту профілю користувача u ;

$\text{enc}(\cdot)$ – відображення, що реалізується послідовністю енкодерних шарів BERT і перетворює вхідну послідовність токенів на один вектор фіксованої

розмірності.

Наступним етапом, після перетворення описів профіля користувача та вакансії, є обчислення семантичної близькості між ними. Доцільно це робити також за допомогою косинусної міри подібності [20]:

$$\text{sim}_{\text{BERT}}(u, i) = \frac{e_u \cdot e_i}{\|e_u\| \|e_i\|}, \quad (2.10)$$

де e_u та e_i – вектори, отримані з моделі BERT для профілю користувача та вакансії;

$\|e_u\|$ і $\|e_i\|$ – їх евклідові норми.

Таким чином, високе значення $\text{sim}_{\text{BERT}}(u, i)$ вказує на семантичну близькість двох текстів з урахуванням контексту, а не лише збігу окремих термінів.

2.4 Розробка гібридного методу для формування рекомендацій вакансій

На основі проаналізованих методів розроблено гібридний підхід до формування рекомендацій вакансій для фріланс-платформи. Запропонований метод має каскадну структуру та працює у двох режимах залежно від стану користувача. Для користувачів з наявною історією взаємодій застосовуються колаборативна фільтрація, контентно-орієнтований метод та семантичний підхід. Для нових користувачів використовується модифікований підхід, що спирається на контентний та семантичний методи з додатковим урахуванням популярності та новизни вакансій. Такий підхід дозволяє поєднати сильні сторони колаборативної та контентно-орієнтованої фільтрації, а також сучасних семантичних моделей на основі BERT, мінімізуючи їхні окремі недоліки.

У разі, коли користувач має заповнений профіль із навичками й хоча б одну взаємодію з вакансіями, система працює у каскадному режимі. Розглядається

матриця взаємодій (2.6). Для кожного користувача обчислюється середнє значення його взаємодій. Міра схожості між двома користувачами обчислюється за формолою (2.7).

На основі цього розрахунку, для кожної вакансії, яка ще не отримувала оцінки від поточного користувача, прогнозується очікуване значення взаємодій за схемою КФ на основі взаємодій користувачів (2.8). Таким же чином використовується КФ на основі об'єктів, де аналізуються схожі вакансії.

Для підвищення стійкості оцінки, запропонована комбінована модель, де підсумковий колаборативний бал розраховується за формулою:

$$s_{u,i}^{CF} = 0,5 \hat{r}_{u,i}^{user} + 0,5 \hat{r}_{u,i}^{item}, \quad (2.11)$$

де u – індекс користувача в системі;

i – індекс вакансії;

$s_{u,i}^{CF}$ – узагальнений колаборативний бал вакансії i для користувача u , отриманий у результаті комбінування двох типів колаборативної фільтрації;

$\hat{r}_{u,i}^{user}$ – прогнозований рейтинг вакансії i для користувача u , обчислений колаборативною фільтрацією на основі схожості між користувачами та їхніх історій взаємодій;

$\hat{r}_{u,i}^{item}$ – прогнозований рейтинг вакансії i для користувача u , обчислений колаборативною фільтрацією на основі схожості між вакансіями;

Коефіцієнти відображають рівну вагу обох типів КФ. На основі значень $s_{u,i}^{CF}$ формується впорядкована множина кандидатів.

Другий етап у каскадному режимі реалізує контентно-орієнтовану фільтрацію в межах сформованого списку кандидатів. Для кожного користувача формується текстове представлення профілю, що об'єднує його навички, біографічну інформацію та опис досвіду. Аналогічна процедура відбувається по відношенню до кожної вакансії. Формуються тексти, які містять назву, опис, вимоги до навичок та категоріальну інформацію. На основі цих текстів,

створюються TF-IDF-вектори (2.1) – (2.4), а подібність між профілем користувача та вакансією оцінюється за косинусною мірою (2.5).

Третій етап каскадного режиму полягає у семантичному ранжуванню, попередньо відібраних кандидатів, за допомогою моделі BERT. Для профілю користувача та описів вакансій будуються контекстні векторні представлення, сформовані енкодером BERT (2.9). Перед обчисленням міри подібності, ці вектори нормалізуються до одиничної довжини. Семантичний бал визначається за косинусною мірою (2.10).

Отримані на різних етапах бали мають різні шкали, тому перед об'єднанням вони нормалізуються за схемою min–max для кожного користувача окремо. Нехай $s_{u,i}^{CF}$, $s_{u,i}^{cont}$ та $s_{u,i}^{sem}$ – відповідно колаборативний, контентний і семантичний бали для вакансії i , а i перебирає множину кандидатів для користувача u . Для кожного методу $m \in \{CF, cont, sem\}$ вводяться значення $s_u^{m,min}$ та $s_u^{m,max}$, що відповідають мінімальному та максимальному значенню оцінки серед усіх кандидатів. Нормалізований бал визначається за формулою [21]:

$$\tilde{s}_{u,i}^m = \frac{s_{u,i}^m - s_u^{m,min}}{s_u^{m,max} - s_u^{m,min}}, \quad (2.12)$$

де $\tilde{s}_{u,i}^m$ – нормалізований бал методу m для вакансії i ;

$s_{u,i}^m$ – вихідне значення;

$s_u^{m,min}$ та $s_u^{m,max}$ – мінімальне і максимальне значення серед кандидатів для цього користувача.

Після нормалізації фінальний бал вакансії i для користувача u обчислюється як зважена сума нормалізованих оцінок [22]:

$$s_{u,i} = 0,2 \tilde{s}_{u,i}^{CF} + 0,4 \tilde{s}_{u,i}^{cont} + 0,4 \tilde{s}_{u,i}^{sem}, \quad (2.13)$$

де $s_{u,i}$ – підсумкова оцінка релевантності вакансії,

$\tilde{s}_{u,i}^{CF}$, $\tilde{s}_{u,i}^{cont}$ та $\tilde{s}_{u,i}^{sem}$ – нормалізовані бали колаборативного, контентного та семантичного методів.

Ці коефіцієнти задають баланс між компонентами гібридної моделі. Колаборативній фільтрації надано меншу вагу, оскільки її якість суттєво залежить від наявності достатньої історії взаємодій і вона є більш вразливою до проблеми розрідженості даних та «холодного старту». Контентному та семантичному підходам призначено однакові ваги, що відображає припущення про їх більшу стійкість і універсальність. Сума коефіцієнтів дорівнює 1, що забезпечує інтерпретованість підсумкового бала як комбінації трьох методів рекомендації.

У режимі «холодного старту», коли користувач не має історії взаємодій або профіль недостатньо заповнений для застосування колаборативної фільтрації, використовується спрощений варіант гібридної моделі. За наявності мінімального текстового опису профілю на першому етапі, формуються TF-IDF вектори та обчислюється косинусна подібність із усіма вакансіями, що дозволяє сформувати початкову множину кандидатів. На другому етапі, ця множина сортується за допомогою семантичної моделі, яка уточнює порядок вакансій з урахуванням контекстної близькості текстів. Завершальний етап передбачає врахування показників популярності та новизни вакансій, що компенсує відсутність поведінкових даних користувача.

Запропонований гібридний метод формування рекомендацій вакансій об'єднує вищезазначені методи в єдину каскадну структуру, що забезпечує підвищену точність та адаптивність рекомендацій [23].

3 ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ ФОРМУВАННЯ РЕКОМЕНДАЦІЙ ВАКАНСІЙ НА ФРІЛАНС-ПЛАТФОРМІ

3.1 Контекстна діаграма процесу формування рекомендацій вакансій на фріланс-платформі

Для відображення процесу формування рекомендацій вакансій на фріланс-платформі та демонстрації логіки взаємодії основних інформаційних потоків, методів і технічних засобів, що забезпечують роботу рекомендаційної підсистеми, створено контекстну діаграму. Головний блок «Формування рекомендацій вакансій на фріланс-платформі» відображає інтегрований процес опрацювання даних користувача, у межах якого відбувається об'єднання колаборативних, контентних та семантичних підходів до створення рекомендацій вакансій. Процес функціонує на основі гібридного методу.

До системи надходять два базові інформаційні потоки, які визначають зміст подальшого аналізу. Історія взаємодій користувача – включає відомості про перегляди вакансій, їх збереження, подані заявки та інші поведінкові ознаки, що формують основу для застосування колаборативної фільтрації. Другий потік складають описові характеристики профілю користувача, які містять інформацію про його навички, професійний досвід, категорію діяльності та інші текстові атрибути, необхідні для контентно-орієнтованого та семантичного аналізу. Ці дані виступають визначають здатність системи формувати релевантні персоналізовані рекомендації.

Функціонування головного процесу контролюється трьома різними підходами, які накладають обмеження та задають правила обробки даних. Колаборативна фільтрація спрямовує систему на виявлення схожості користувачів та вакансій на основі взаємодії користувачів; контентно-орієнтована модель забезпечує побудову векторних подань текстових описів та оцінювання їхньої подібності за допомогою TF-IDF і косинусної міри. Семантична модель, заснована на контекстних векторних репрезентаціях, в

нашому випадку це BERT, дозволяє системі аналізувати зміст вакансій на рівні контекст. У сукупності ці методи формують керуючий вплив, який визначає послідовність оброблення даних, спосіб вибору релевантних вакансій та правила побудови фінального ранжування.

Для реалізації процесу використовуються технічні засоби нижнього рівня діаграми. Мова програмування Java забезпечує реалізацію серверної логіки платформи. СУБД MySQL відповідає за зберігання даних користувачів, вакансій і взаємодій між ними. Фреймворк Spring Boot виконує роль базового середовища для організації веб-сервісів і викликів, пов'язаних із рекомендаційним модулем. Сукупність цих засобів формує механізм системи, який забезпечує виконання алгоритмів і доступ до даних у режимі реального часу.

Результатом роботи процесу, позначеним на правій стороні діаграми, є сформовані рекомендації – впорядкований список вакансій, який уже пройшов каскадну обробку початкових даних за допомогою всіх зазначених методів. Цей вихідний потік відображає кінцеву мету системи – надання користувачеві максимально релевантних пропозицій, що відповідають його професійному профілю та історії поведінки на платформі.

Таким чином, діаграма показує узагальнену архітектуру рекомендаційної підсистеми, у якій вхідні дані користувача перетворюються на персоналізований набір вакансій завдяки поєднанню методів і технологічних засобів.

Контекстна діаграма зображена на рисунку 3.1

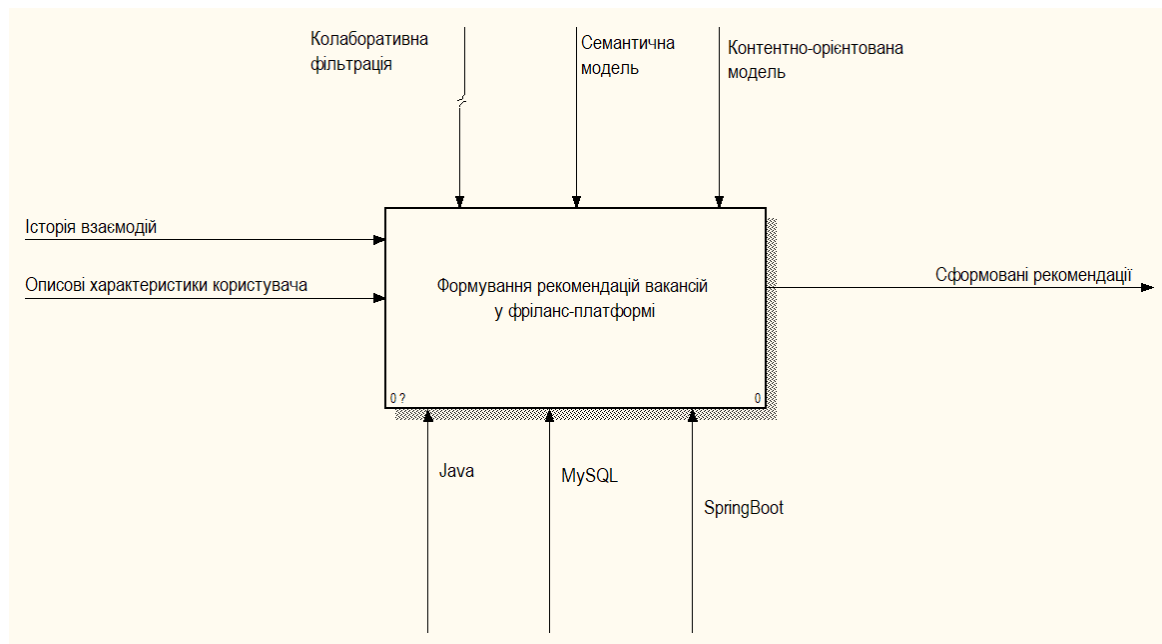


Рисунок 3.1 – Контекстна діаграма

3.2 Опис діаграми декомпозиції

Для відображення деталізованої структури процесу формування рекомендації вакансій прийнято рішення провести декомпозицію контекстної діаграми. Дана діаграма демонструє внутрішню логіку функціонування рекомендаційної підсистеми, фіксуючи послідовність етапів обробки даних та взаємодію між ними. Кожен елемент моделі виконує специфічну функцію, пов'язану з аналізом поведінкових або текстових характеристик користувача, що дозволяє поєднувати колаборативні, контентні та семантичні методи у єдиному алгоритмі.

Перший блок «Отримання та обробка існуючих даних користувача» – відповідає за первинний прийом інформації, яка надходить до системи у вигляді історії взаємодій та описових характеристик профілю. На цьому етапі обробляються дані про перегляди вакансій, заявки, попередні контракти, а також текстова інформація про навички, досвід та інші атрибути профілю. Отримані

відомості структуруються й уніфікуються, після чого передаються до наступного процесу. Саме цей блок вирішує задачу відокремлення поведінкових сигналів від текстової інформації, готуючи їх до подальшого використання у різних компонентах гібридного методу.

Блок «Відбір схожих користувачів та формування початкового списку вакансій» – реалізується колаборативна фільтрація. На основі даних, сформованих на попередньому етапі, здійснюється пошук користувачів зі схожими взаємодіями, що дозволяє побудувати початковий набір вакансій-кандидатів. У цьому ж блоці генерується початкова множина вакансій, яка надалі підлягає уточненню за допомогою текстового та семантичного аналізу.

Блок «Формування векторних представлень та порівняння вакансій із профілем користувача» – відповідає за контентно-орієнтовану обробку інформації. На даному етапі система трансформує текстовий опис профілю користувача та кожної вакансії у векторні подання, побудовані на основі TF-IDF. За допомогою косинусної міри подібності отримуємо оцінку, яка каже про схожість змісту вакансій із атрибутами профілю користувача. Цей процес виконує функцію фільтрації та перерозподілу початкової множини кандидатів, отриманої з попереднього блоку, уникаючи вакансій, які текстово не відповідають профілю користувача.

Блок «Ранжування вакансій та формування кінцевого списку рекомендацій» – реалізує завершальну фазу рекомендаційного процесу. На цьому етапі до системи входять результати попередніх компонентів, які піддаються кінцевій обробці із застосуванням семантичної моделі. Використовуючи контекстні векторні уявлення, що базуються на архітектурі трансформерів, алгоритм виконує високоточне порівняння вакансій з профілем користувача та формує підсумковий ранжований список. Завдяки цьому завершується каскадна інтеграція всіх методів. Колаборативна фільтрація забезпечує пошук релевантних вакансій за взаємодіями, контентно-орієнтована модель відповідає за структурну текстову відповідність, а семантична модель займається глибоким зіставленням змісту.

Таким чином, діаграма декомпозиції зображує послідовність процесів, починаючи із збору та нормалізації даних користувача закінчуючи формуванням кінцевої множини персоналізованих рекомендацій. Вона демонструє логіку гібридної моделі, у якій кожен етап уточнює результати попереднього, забезпечуючи цілісність, адаптивність і високу точність рекомендацій у динамічних умовах фріланс-платформи.

Діаграма декомпозиції зображена на рис. 3.2.

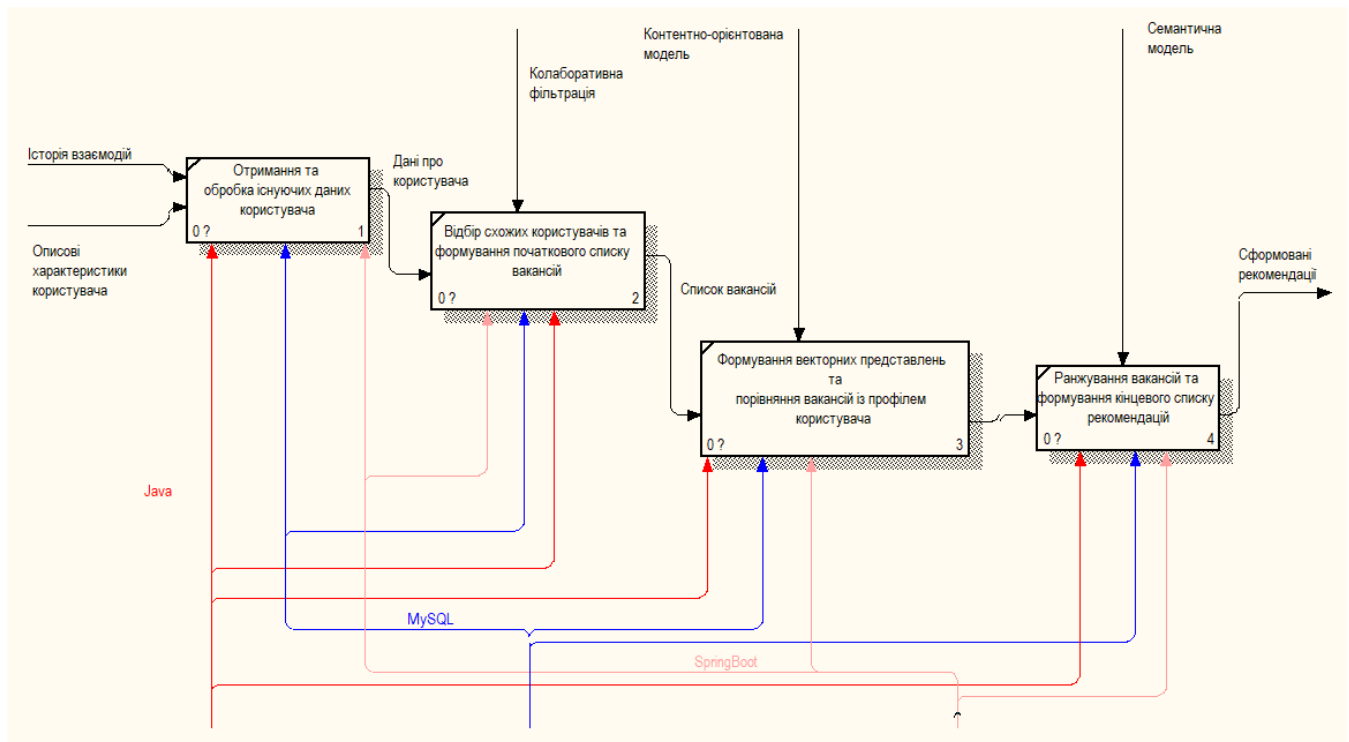


Рисунок 3.2 – Діаграма декомпозиції

4 ПРАКТИЧНА РЕАЛІЗАЦІЯ ГІБРИДНОГО МЕТОДУ

4.1 Програмна реалізація методів рекомендаційної системи

4.1.1 Обґрунтування технологічних засобів

Розробка гібридної рекомендаційної системи для фріланс-платформи повинна використовувати технології, які здатні забезпечити стабільну обробку великих обсягів даних, гнучкість архітектури та високу продуктивність під час формування рекомендацій.

Мовою програмування для серверної частини обрано Java. Ця мова є однією з найпоширеніших мов у розробці високонавантажених систем, що робить її оптимальною для проєктів, де важлива продуктивність, масштабованість і безпека. Вона забезпечує сувору типобезпеку, що зменшує кількість помилок на етапі компіляції та підвищує надійність розробки. Java має величезну екосистему бібліотек, що спрощує інтеграцію з базами даних, системами безпеки, сервісами машинного навчання та інструментами документування API.

У процесі формування семантичних ознак система прийнято рішення використовувати додатковий сервіс, реалізований мовою Python, який відповідає за обробку текстових даних за допомогою моделей сімейства BERT.

Python-сервіс функціонує як окремий компонент, який отримує текстові дані, генерує високорозмірний embedding і повертає його у вигляді масиву чисел для подальшого використання на Java-бекенді [24].

У ролі бекенд-фреймворку використано Spring Boot. Його обрано завдяки своїй модульності, широкій екосистемі та підтримці автоматичної конфігурації, яка значно скорочує час розробки. Spring Boot надає велику кількість готових рішень, що дозволяє зосередитися на реалізації бізнес-логіки рекомендаційної системи.

Для збереження даних обрано MySQL 8.0+, що є надійною реляційною системою з оптимізованими механізмами індексації.

Сукупність обраних технологій - Java, Python, Spring Boot, MySQL - формує надійну основу для гібридної рекомендаційної системи. Така архітектура забезпечує масштабованість, продуктивність, гнучкість і легку підтримку. Вона дозволяє безпечно працювати з даними користувачів, оптимізує процес формування рекомендацій та забезпечує швидку роботу сервісу в умовах зростаючого навантаження.

4.1.2 Опис реалізації контентно-орієнтованої моделі

Контентно-орієнтована модель реалізована через сервіс `ContentBasedFilteringService`. Фрагмент коду A.1 з додатку A, відображає реалізацію даного сервісу, який оцінює релевантність вакансій на основі інформації з профілю користувача та текстового опису вакансій. Реалізація поєднує TF-IDF векторизацію та точне зіставлення навичок.

На першому кроці формується текстове представлення профілю користувача та вакансії.

Побудовані текстові об'єкти перетворюються на числові вектори методом TF-IDF. У модулі використовується фіксований словник технічних термінів. На основі отриманих векторів, обчислюється косинусна схожість між профілем користувача та вакансією, що дає оцінку текстової схожості.

Паралельно виконується зіставлення навичок. Навички нормалізуються, приводяться до однакового регістру і розбиваються на множини. Далі визначається кількість збігів між множинами навичок користувача та вакансії. Формується окрема оцінка навичок, що є важливим доповненням до TF-IDF.

Фінальна оцінка контентно-орієнтованої моделі визначається як комбінація TF-IDF та оцінки навичок. Якщо збіг навичок є суттєвим, частка навичок у комбінованій оцінці збільшується, що дозволяє надавати пріоритет вакансіям, які користувач потенційно може виконати професійно.

4.1.3 Опис реалізації колаборативної фільтрації

Колаборативна фільтрація реалізована у вигляді сервісу CollaborativeFilteringService, який працює з історією взаємодій користувачів. Реалізація поєднує два підходи –КФ на основі взаємодій користувачів та КФ на основі об'єктів, що дозволяє формувати більш узгоджені прогнози. Реалізація даного сервісу наведено у фрагменті коду A.2 з додатку A.

На першому етапі для кожного користувача формується вектор його взаємодій із вакансіями. Усі події переводяться у числову оцінку, а взаємодії з однією вакансією сумуються. У такий спосіб формується набір рейтингів, необхідний для подальших обчислень.

КФ на основі взаємодії користувачів реалізовано через обчислення кореляції Пірсона між користувачами. Для кожної пари користувачів враховуються лише спільні вакансії, а для коректної оцінки необхідна наявність мінімум двох спільних елементів у векторах. Кореляція використовується як коефіцієнт схожості. Значення, які менше або дорівнюють 0 відкидаються, а позитивні коефіцієнти слугують вагами під час формування прогнозів. Для кожної вакансії, з якою взаємодіяли схожі користувачі, обчислюється зважена оцінка, що є прогнозом для цільового користувача.

КФ на основі об'єктів реалізовано аналогічним чином. Замість порівняння користувачів порівнюються вакансії. Для кожної нової вакансії визначаються інші, з якими вона має значущу схожість за історією оцінок. Після обчислення кореляції вакансії формуються прогнози на основі рейтингів, поставлених користувачем схожим вакансіям.

Фінальна оцінка методу є комбінацією двох видів КФ. Для об'єднання використовується середнє з фіксованими вагами, що дозволяє збалансувати внесок кожного підходу в загальний результат.

4.1.4 Опис реалізації семантичної моделі та BERT

Семантична фільтрація реалізована в модулі `SemanticFilteringService` ґрунтується на використанні векторних представлень тексту, сформованих за допомогою моделі `Sentence-BERT`. Даний підхід дозволяє аналізувати вакансії не лише на рівні ключових слів, а й на рівні сенсових та контекстуальних зв'язків між профілем користувача та вакансіями. Реалізація даного сервісу наведено у фрагменті коду А.3 з додатку А. Завдяки цьому сервісу виявляється концептуальна схожість між запитам, які можуть мати суттєві лексичні розбіжності.

Генерація семантичних векторів винесена у зовнішній Python-сервіс, оскільки сучасні трансформерні моделі значно ефективніше виконуються у Python-екосистемі, що забезпечує кращу продуктивність та стабільність. Реалізація BERT-сервісу наведено у фрагменті коду А.4 з додатку А. Сервіс працює на основі моделі `all-MiniLM-L6-v2`, яка створює `embeddings` розмірністю 384 компоненти, оптимізовані для семантичного порівняння коротких та середніх текстів. Java передає на сервіс сформований текст, отримує `embedding` у форматі JSON та перетворює його у масив чисел типу `float`. Той самий механізм використовується як для профілів користувачів, так і для вакансій.

Семантична схожість між профілем і вакансіями визначається за допомогою косинусної міри, яка є стандартом для порівняння багатовимірних векторів. Косинусна схожість дозволяє ефективно вимірювати схожість між `embeddings`, незалежно від їх абсолютних значень, що робить метод стійким до варіацій текстових описів. На основі цього для кожної вакансії формується числовий показник, який відображає ступінь концептуальної відповідності між профілем користувача та вакансією.

Завдяки застосуванню BERT, сервіс здатний розпізнавати синонімію, парафрази, контекстно близькі формулювання та семантичні зв'язки. Це підвищує якість рекомендацій, особливо у випадках, коли описи вакансій або

профілі користувачів недостатньо структуровані або містять різні варіанти термінології.

Узагальнюючи, семантична фільтрація забезпечує найвищий рівень змістовної точності в системі, оскільки оперує повноцінним семантичним простором. Вона підсилює обидва попередні методи формування рекомендацій і дозволяє формувати фінальні рекомендації, які найбільш повно відображають реальні компетенції та професійні інтереси користувача.

4.1.5 Опис реалізації гібридного методу

Гібридний метод у системі виконує роль інтеграційного механізму, який поєднує результати трьох незалежних підходів – колаборативного, контентно-орієнтованого та семантичного – у єдину модель формування рекомендацій. Реалізацію гібридного методу наведено у фрагменті коду А.5 з додатку А.

На початку гібридний модуль формує оцінки від колаборативної фільтрації. Цей етап виконує звуження списку вакансій, відокремлюючи ті, які мають поведінкову схожість. КФ працює з великими масивами даних взаємодій, тому вона здатна виявити закономірності, що будуються на реальному досвіді користувачів платформи. Результатом першого етапу є сформований набір вакансій, відсортованих за взаємодіями та придатних для подальшого етапу аналізу.

Після визначення цього набору, алгоритм переходить до контентно-орієнтованого аналізу. Контентна модель переглядає кожну із відібраних вакансій та порівнює їх із профілем користувача за допомогою TF-IDF представлень та оцінки відповідності навичок. На цьому етапі система поступово переходить від поведінкових сигналів до змістовних характеристик, аналізуючи, наскільки текстові описи вакансій та вказані навички відповідають даним користувача. Результати контентно-орієнтованого аналізу повторно

впорядковуюють вакансії, підвищуючи позиції тих, що мають значний змістовний збіг. Таким чином формується більш точний список вакансій, який переходить на семантичний метод рекомендації.

Вакансії, відібрані по змісту, проходять семантичну обробку. На цьому етапі для кожної вакансії і профілю користувача, за допомогою Sentence-BERT, генеруються векторні представлення, які відображають внутрішню семантику текстів. Косинусна міра подібності між цими векторами дозволяє визначити концептуальну близькість між професійним описом користувача та описом вакансій, навіть якщо текстові формулювання відрізняються по сенсу. На основі цієї обробки формується найточніша оцінка відповідності, здатна враховувати контекстні зв'язки, синонімію та відтінки значення.

Після отримання результатів трьох методів виконується їхнє об'єднання. Усі оцінки нормалізуються до єдиного масштабу, після чого беруть участь у фінальній ваговій комбінації. Комбінація цих методів будує збалансовану кінцеву оцінку. Після впорядкування за підсумковим значенням, формується остаточний список рекомендацій.

У результаті гібридний метод забезпечує найвищий рівень точності рекомендації. Він компенсує слабкі сторони кожного окремого підходу. Така архітектура дозволяє системі стабільно працювати як для нових користувачів з невеликим профілем, так і для досвідчених фрілансерів із великою історією взаємодій, забезпечуючи універсальність та надійність рекомендацій.

4.1.5 Демонстрація працездатності розробленого гібридного методу

Для того, щоб продемонструвати роботу запропонованого гібридного методу була розроблена підсистема надання рекомендацій вакансій на фріланс-платформі.

Розглянемо її роботу детальніше.

На першому етапі необхідно авторизуватись або зареєструватись у системі. Так, як у нас є список існуючих користувачів, авторизуємось у профіль одного із них. Даний користувач має інформацію у профілі та взаємодії по проектах, тому гібридний метод працюватиме успішно. Профіль користувача та сформовані рекомендації представлені на рис. 4.1-4.2.

Професійні деталі

Біографія

Experienced Java backend developer with strong skills in building stable and clean server-side applications. Working with Spring Boot, REST APIs and MySQL. Actively improving my web development skills and expanding my tech stack to deliver better solutions.

Місцезнаходження

Kharkiv

Номер телефону

+380505654552

Навички (через кому)

Java, Spring Boot, React, REST API, MySQL, Maven, Git

Годинна ставка (\$)

2000

Роки досвіду

5

Рисунок 4.1 – Профіль користувача

☀ Рекомендовані для вас
Вакансії, які найкраще підходять вашому профілю

[Оновити](#)

95% релевантність

Build E-commerce Website

Need an experienced developer to build a modern e-commerce platform with payment integration.

\$5000.00
Фіксована ціна

Web Development

Навички: Spring Boot, React, MySQL +1 більше

[Переглянути](#) [Зберегти](#)

83% релевантність

Розробка REST API на Spring Boot

Потрібен досвідчений Java розробник для створення REST API з використанням Spring Boot, PostgreSQL. Проект включає інтеграцію з платіжними системами.

\$2000.00
Фіксована ціна

Web Development

Навички: Java, Spring Boot, REST API +1 більше

[Переглянути](#) [Зберегти](#)

63% релевантність

Розробка корпоративного порталу на Spring Boot

Створення корпоративного порталу з інтеграцією LDAP, системою ролей та правами доступу. Потрібен досвідчений Java розробник.

\$4500.00
Фіксована ціна

Web Development

Навички: Java, Spring Boot, Spring Security +3 більше

[Переглянути](#) [Зберегти](#)

61% релевантність

Java Backend з MongoDB

Розробка backend на Java з використанням MongoDB та Spring Boot. Створення API для мобільного додатку.

\$1800.00
Фіксована ціна

Web Development

Навички: Java, Spring Boot, MongoDB

[Переглянути](#) [Зберегти](#)

44% релевантність

Microservices на Spring Cloud

Розробка мікросервісної архітектури з використанням Spring Cloud, Eureka, Zuul, Config Server.

\$5000.00
Фіксована ціна

Web Development

Навички: Spring Cloud, Microservices, Eureka +2 більше

[Переглянути](#) [Зберегти](#)

38% релевантність

Мікросервісна архітектура на Java

Розробка мікросервісів з використанням Spring Cloud, Docker, Kubernetes. Потрібен досвідчений розробник з архітектурними навичками.

\$3000.00
Фіксована ціна

Web Development

Навички: Java, Spring Cloud, Docker +2 більше

[Переглянути](#) [Зберегти](#)

Рисунок 4.2 – Сформовані за допомогою гібридного методу рекомендації

Як ми можемо побачити, рекомендації формуються успішно та підходять для нашого користувача. Також відображається відсоток релевантності вакансії для зручності вибору.

Тепер переглянемо випадок, коли у систему приходить новий користувач і виникає проблема «холодного старту» із-за відсутності взаємодій. Це означає, що при відсутній історії взаємодій, колаборативна фільтрація не буде працювати взагалі. У системі дану проблему автоматично вирішують контентно-орієнтована та семантична моделі. Так як, при реєстрації, користувач повинен внести про себе інформацію, ці два методи спрацьовують одразу при створенні його профіля і формують рекомендації. Спробуємо створити користувача та показати, чи коректно працює гібридний метод. Процес реєстрації користувача у системі та формування рекомендацій при ситуації «холодного старту» наведено на рис. 4.3 – 4.4.

Ваші навички (через кому)

Design, User Flows, Usability Testing, Design Systems, Visual Design

Коротко про себе

use. Constantly learning new design approaches and refining my skills in Figma, prototyping and user-centered design.

Роки досвіду

5

Годинна ставка (USD)

50

Рисунок 4.3 – Реєстрація профіля користувача

☀ Рекомендовані для вас
Вакансії, які найкраще підходять вашому профілю Оновити

97% релевантність

UI/UX дизайн фінтех додатку

Створення повного UI/UX дизайну для фінтех мобільного додатку з фокусом на безпеку та зручність.

\$1800.00

Фіксована ціна

Навички: Figma UI/UX Design Mobile Design +1 більше

Design

Переглянути Зберегти

59% релевантність

Mobile App UI/UX Design

Looking for a creative designer to design mobile app interfaces.

\$2000.00

Фіксована ціна

Навички: Figma Adobe XD Mobile Design

Design

Переглянути Зберегти

28% релевантність

Веб-дизайн для SaaS платформи

Дизайн сучасного веб-інтерфейсу для SaaS платформи з акцентом на конверсію та користувацький досвід.

\$2200.00

Фіксована ціна

Навички: Figma Web Design SaaS +2 більше

Design

Переглянути Зберегти

Рисунок 4.4 – Сформовані рекомендації для нового користувача

Як бачимо, рекомендації сформовані успішно та відповідають профілю створеного користувача. Єдиний недолік даного підходу, для вирішення «холодного старту», полягає у швидкості формування рекомендацій. Так як відсутня КФ, яка звужує набір релевантних рекомендацій, то контентно-

орієнтованій та семантичній моделям необхідно проходити по всьому списку вакансій, тим самим уповільнюючи час формування пропозицій. Для прискорення роботи підходу, користувачу достатньо почати взаємодіяти із будь-якою вакансією.

4.2 Експериментальна перевірка працездатності гібридного методу

4.2.1 Опис використаних метрик оцінювання

Метрика $Precision@K$ характеризує частку релевантних об'єктів серед перших K елементів списку рекомендацій. Вона відображає “чистоту” верхньої частини списку: чим вищою є точність, тим менше нерелевантних вакансій система подає користувачу серед перших K позицій.

Формула $Precision@K$ виглядає як [25]:

$$Precision@K(u) = \frac{\text{кількість релевантних елементів у топ-}K}{K} \quad (4.1)$$

де K – розмір списку;

u – індекс користувача.

Метрика $Recall@K$ відображає, яку частку від усіх релевантних вакансій система змогла виявити в межах перших K елементів списку. Таким чином, $Recall@K$ характеризує повноту рекомендацій.

Формула $Recall@K$ виглядає як [25]:

$$Recall@K(u) = \frac{\text{кількість релевантних елементів у топ-}K}{\text{загальна кількість релевантних елементів}} \quad (4.2)$$

де K – розмір списку;

u – індекс користувача.

Average Precision@K для користувача u обчислює скільки релевантних вакансій знайдено в топ-K, і на яких позиціях вони знаходяться. На кожній позиції k , де рекомендований елемент релевантний, враховується Precision@k, після чого середнє цих значень нормується на кількість релевантних елементів.

Формула для користувача u [26]:

$$AP@K(u) = \frac{1}{r_u} \sum_{k=1}^K * Precision@k(u) \cdot rel_u(k) \quad (4.3)$$

де K – розмір списку;

k – позиція в ранжованому списку;

$Precision@k(u)$ – точність на префіксі довжини k для користувача u ;

$rel_u(k)$ – індикатор релевантності елемента на позиції k для користувача u ;

u – індекс користувача;

r_u – загальна кількість релевантних елементів для користувача u .

Mean Average Precision@K є середнім значенням AP@K по всіх користувачах. Ця метрика узагальнює якість ранжування в системі. Оцінка метрики висока тільки тоді, коли для більшості користувачів вакансії зосереджені у верхніх позиціях списку.

Формула MAP@K виглядає як [26]:

$$MAP@K = \frac{1}{N} \sum_{u=1}^N * AP@K(u) \quad (4.4)$$

де N – кількість користувачів у тестовій вибірці;

$AP@K(u)$ – середня точність для користувача u ;

u – індекс користувача;

Метрика NDCG@K використовується для оцінювання якості ранжування в рекомендаційних системах. Вона показує, наскільки якісно система розміщує релевантні елементи у верхніх позиціях списку рекомендацій, порівнюючи отриманий порядок із ідеальним, у якому всі релевантні елементи стоять

максимально високо.

$CG@K$ – це метрика, що просто підсумовує релевантність елементів у топ- K , без урахування їх позицій.

Формула $CG@K$ виглядає як [27]:

$$CG@K = \sum_{i=1}^K rel_i \quad (4.5)$$

де $CG@K$ – сумарний здобуток релевантності у межах перших K позицій;

K – розмір списку;

rel_i – релевантність елемента на позиції i .

DCG вводить знижувальний коефіцієнт, через який релевантні елементи на нижчих позиціях роблять менший внесок.

Формула $DCG@K$ виглядає як [27]:

$$DCG@K = rel_1 + \sum_{i=2}^K \frac{rel_i}{\log_2(i+1)} \quad (4.6)$$

де $DCG@K$ – знижена кумулятивна релевантність у топ- K ;

rel_1 – релевантність першого елемента: перша позиція не дисконтується;

rel_i – релевантність елемента на позиції i ;

$\log_2(i+1)$ – дисконт, що зменшує внесок елементів на нижчих позиціях;

i – позиція у списку рекомендацій.

IDCG@K – це DCG, обчислений для ідеального впорядкування, де всі релевантні елементи стоять на максимально високих позиціях.

Формула IDCG@K виглядає як [27]:

$$IDCG@K = \sum_{i=1}^K \frac{rel_i^*}{\log_2(i+1)} \quad (4.7)$$

де $IDCG@K$ – максимально можлива величина DCG для тих самих релевантностей;

rel_i^* – релевантність елемента на позиції i у ідеальному ранжуванні;

$\log_2(i + 1)$ – дисконт, що зменшує внесок елементів на нижчих позиціях;

K – розмір списку.

Розкривши значення та формули попередніх метрик, з яких складається $NDCG@K$, можемо сформулювати його формулу, яка виглядає як [27]:

$$NDCG@K = \frac{DCG@K}{IDCG@K} \quad (4.8)$$

де $NDCG@K$ – нормалізована оцінка якості ранжування;

$DCG@K$ – фактичний показник ранжування моделі;

$IDCG@K$ – максимальний можливий DCG;

4.2.2 Результати експерименту та висновки

В ході дослідження було проведено експеримент та отримані результати експериментальної оцінки розробленого гібридного методу формування рекомендацій вакансій та здійснено їх порівняння з базовими підходами.

У табл. 4.1 наведено узагальнені результати експериментальної оцінки якості роботи рекомендаційної системи.

Таблиця 4.1 – Узагальнені результати експериментальної оцінки якості роботи рекомендаційної системи

Метод	K	Precision@K	Recall@K	MAP@K	NDCG@K
Гібридний метод	5	0.2556	0.8056	0.5907	0.6220
	10	0.1333	0.8333	0.5709	0.6334
	15	0.1074	0.9722	0.5551	0.6770
	20	0.0833	1.0000	0.5561	0.6850

Продовження таблиці 4.1.

Метод	K	Precision@K	Recall@K	MAP@K	NDCG@K
Контентно-орієнтований метод	5	0.2000	0.5833	0.4833	0.4921
	10	0.1278	0.7500	0.4969	0.5566
	15	0.0926	0.8333	0.5018	0.5804
	20	0.0778	0.9167	0.4949	0.6047
Семантичний метод	5	0.2222	0.6667	0.5250	0.5515
	10	0.1167	0.6944	0.5343	0.5636
	15	0.0926	0.8611	0.5471	0.6111
	20	0.0750	0.9167	0.5381	0.6269
Колаборативна фільтрація	5	0.2500	0.5556	0.3954	0.4205
	10	0.2242	0.6389	0.4042	0.4539
	15	0.2205	0.6667	0.3954	0.4634
	20	0.2205	0.6667	0.3954	0.4634

Аналізуючі значення метрик з табл 4.1, можна зробити висновок, що гібридний метод демонструє себе як найбільш збалансований і якісний підхід для формування рекомендацій вакансій. За інтегральними метриками MAP@K та NDCG@K саме гібридний підхід стабільно випереджає всі інші методи для всіх значень K. Це означає, що релевантні вакансії не просто присутні у списку, а й займають відносно високі позиції, тобто користувач бачить якісні вакансії у верхній частині списку рекомендацій. Варто звернути увагу, що Recall для гібридного підходу є дуже високим, що свідчить про здатність системи охопити практично всі релевантні вакансії для користувача в межах сформованого списку. Контентно-орієнтований підхід на основі TF-IDF та семантичний метод з BERT виглядають логічною конструкцією для формування рекомендацій, але окремо вони поступаються гібридному підходу, тоді як поєднуючи їх з колаборативною фільтрацією вдається досягти помітного приросту якості пропозицій.

Не дивлячись на гарні результати у більшості метрик, виявились також і

недоліки. Система все ще має простір для покращення. Хоча гібридний метод і випереджає інші підходи за MAP та NDCG, його Precision@K при $K = 10, 15, 20$ помітно зменшується, що означає збільшення кількості нерелевантних вакансій у списку. Гібридний підхід поступається колаборативній фільтрації, бо вона, попри нижчі інтегральні метрики, іноді демонструє вищу точність, особливо на великих значеннях K , що вказує на потенціал для більш агресивного використання її сигналів у верхній частині списку.

Таким чином, отримані метрики одночасно підтверджують коректність і ефективність обраного гібридного підходу, але й підкреслюють, що подальше налаштування ваг та логіки переранжування може суттєво покращити якість системи з точки зору користувацького досвіду.

Графіки результатів експериментальної оцінки рекомендаційної системи наведено на рис. 4.5.

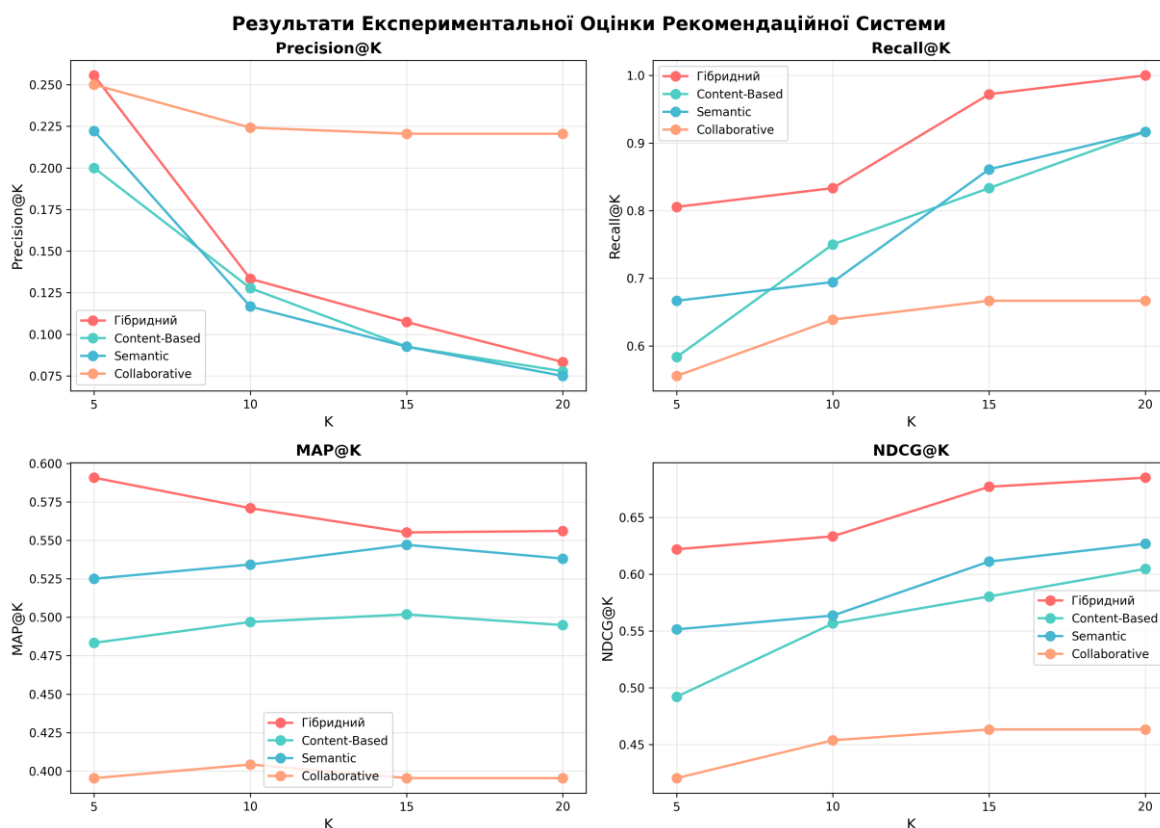


Рисунок 4.5 – Результати експериментальної оцінки рекомендаційної системи

ВИСНОВКИ

У кваліфікаційній роботі вирішено задачу підвищення якості формування рекомендацій вакансій у інформаційній системі фріланс-платформи шляхом розробки гібридного методу, що поєднує контентно-орієнтовану, колаборативну та семантичну фільтрацію. Показано, що ринок фріланс-послуг характеризується великою кількістю користувачів і проєктів, інформаційним перевантаженням та наявністю сценаріїв «холодного старту», де класичні методи працюють обмежено.

У ході дослідження проаналізовано переваги й недоліки контентно-орієнтованої моделі, колаборативної фільтрації та семантичного підходу на основі моделей типу BERT. На основі цього запропоновано каскадний гібридний метод: для активних користувачів спочатку генерується множина кандидатів за допомогою КФ, потім вони ранжуються контентно-орієнтованим методом і, на завершальному етапі, формується кінцевий список семантичним модулем. Для нових користувачів застосовується спрощена схема, що поєднує контентно-орієнтований та семантичний аналіз із урахуванням популярності та новизни вакансій.

Розроблено інформаційну технологію і програмний прототип рекомендаційної підсистеми на базі Java, Spring Boot, MySQL і зовнішнього сервісу семантичної обробки текстів.

Експериментальна перевірка показала, що запропонований гібридний метод демонструє кращі значення метрик якості у порівнянні з окремим застосуванням проаналізованих методів по окремої. Таким чином, експерименти підтвердили успішність розробленого підходу і довели, що поєднання кількох методів у гібридний метод дозволяє підвищити релевантність і стабільність рекомендацій для користувачів фріланс-платформи.

Роботу виконано відповідно до вимог методичних вказівок щодо розробки та оформлення кваліфікаційних робіт [26], а також із дотриманням положень державних стандартів ДСТУ 8302:2015 [27] і ДСТУ 3008:2015 [28].

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Freelance Forward 2022. *Upwork*.
URL: <https://www.upwork.com/research/freelance-forward-2022> (date of access: 01.11.2025).
2. The Gig Economy and the Future of Work. *The World Bank*.
URL: <https://documents1.worldbank.org/curated/en/099060524074041161/pdf/P179647-e104d722-f2f4-4ec8-9397-d1ead6456d2e.pdf> (date of access: 01.11.2025).
3. Jiang K. Reducing Information Overload in Recommender Systems. 2023. 60 p. URL: https://www.politesi.polimi.it/retrieve/43862559-79b4-491d-a7d2-a11d247693b4/Dissertation_Kexin%20Jiang.pdf (date of access: 01.11.2025).
4. Digital Tiger: the Power of Ukrainian IT – 2023 - IT Ukraine Association. *IT Ukraine Association*. URL: <https://itukraine.org.ua/digital-tiger-the-power-of-ukrainian-it-2023/> (date of access: 02.11.2025).
5. Статистика зарплат. *Djinni*. URL: <https://djinni.co/salaries/> (дата звернення: 02.11.2025).
6. Recommendations: what and why?. *Google for Developers*.
URL: <https://developers.google.com/machine-learning/recommendation/overview> (date of access: 06.11.2025).
7. Hardesty L. The history of Amazon's recommendation algorithm. *Amazon Science*. URL: <https://www.amazon.science/the-history-of-amazons-recommendation-algorithm> (date of access: 05.11.2025).
8. Engineering Blog. *LinkedIn Engineering*.
URL: <https://engineering.linkedin.com/blog> (date of access: 05.11.2025).
9. How Upwork improves site performance. *Upwork*.
URL: <https://www.upwork.com/careers/engineering-blog/upwork-improves-site-performance> (date of access: 05.11.2025).
10. Hustler T. What is Fiverr and How does it work- The Complete Beginner's Guide. *Medium*. URL: <https://medium.com/@TheHustler/what-is-fiverr-and-how->

does-it-work-the-complete-beginners-guide-c95ae449f8f5 (date of access: 05.11.2025).

11. Welcome to the Freelancer.com API Docs. *Freelancer developers*. URL: <https://developers.freelancer.com/docs> (date of access: 05.11.2025).

12. Research of Information Filtering Based on Vector Space Model. *IEEE Xplore*. URL: <https://ieeexplore.ieee.org/document/5403435> (date of access: 15.11.2025).

13. THANDAPANI S. P. Recommendation Systems: Collaborative Filtering using Matrix Factorization—Simplified. *Medium*. URL: <https://medium.com/sfu-csmpmp/recommendation-systems-collaborative-filtering-using-matrix-factorization-simplified-2118f4ef2cd3> (date of access: 15.11.2025).

14. Zuidema W. Vector-Based and Neural Models of Semantics. *ResearchGate*. URL: https://www.researchgate.net/publication/364526728_Vector-Based_and_Neural_Models_of_Semantics (date of access: 15.11.2025).

15. Contributors to Wikimedia projects. tf-idf - Wikipedia. *Wikipedia, the free encyclopedia*. URL: <https://en.wikipedia.org/wiki/Tf-idf> (date of access: 30.11.2025).

16. Douglas C. Finding Word Similarity using TF-IDF and Cosine in a Term-Context Matrix from Scratch in Python. <https://towardsdatascience.com/finding-word-similarity-using-tf-idf-in-a-term-context-matrix-from-scratch-in-python-e423533a407/>. URL: <https://towardsdatascience.com/finding-word-similarity-using-tf-idf-in-a-term-context-matrix-from-scratch-in-python-e423533a407/> (date of access: 30.11.2025).

17. Collaborative Filtering. *Statistics & Data Science*. URL: <https://wikistat.fr/pdf/st-m-datSc3-colFil.pdf> (date of access: 30.11.2025).

18. Item-item collaborative filtering. *Grokopedia*. URL: https://grokopedia.com/page/Item-item_collaborative_filtering (date of access: 30.11.2025).

19. Devlin J., Chang M.-W. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv.org e-Print archive*. URL: <https://arxiv.org/pdf/1810.04805> (date of access: 01.12.2025).

20. Kaushicbaravind. Similarity of Word Embeddings with BERT: A Comprehensive Discussion. *Medium*. URL: <https://medium.com/@kaushicbaravind/similarity-of-word-embeddings-with-bert-a-comprehensive-discussion-a641a5710325> (date of access: 02.12.2025).

21. Normalization: Complete Guide to Feature Scaling with Min-Max Implementation - Interactive | Michael Brenndoerfer. *Michael Brenndoerfer | Data & AI, Private Equity, Technology*. URL: <https://mbrenndoerfer.com/writing/normalization-feature-scaling-min-max-machine-learning-guide> (date of access: 02.12.2025).

22. Weighted Sum Method. *ScienceDirect*. URL: <https://www.sciencedirect.com/topics/computer-science/weighted-sum-method> (date of access: 02.12.2025).

23. Ємельянов А., Петров К. Гібридний метод формування рекомендацій вакансій для працівників фріланс-платформи [Електронний ресурс] // *Evolving Science: Theories, Discoveries and Practical Outcomes*, December 15-17, 2025. – Zurich, Switzerland. – Режим доступу: https://www.eoss-conf.com/wp-content/uploads/2025/12/Zurich_Switzerland_15.12.25.pdf (дата звернення: 12.12.2025). – С. 229-232.

24. Step-by-Step Guide: How to Use BERT Word Embeddings in Python. *Medium*. URL: <https://medium.com/@Roy.Wong/step-by-step-guide-how-to-use-bert-word-embeddings-in-python-ac7b621771d8> (date of access: 04.12.2025).

25. Precision and recall at K in ranking and recommendations. *Evidently AI - AI Evaluation & LLM Observability Platform*. URL: <https://www.evidentlyai.com/ranking-metrics/precision-recall-at-k> (date of access: 07.12.2025).

26. Rink K. Mean Average Precision at K (MAP@K) clearly explained. *Toward Data Science*. URL: <https://towardsdatascience.com/mean-average-precision-at-k-map-k-clearly-explained-538d8e032d2/> (date of access: 07.12.2025).

27. Normalized Discounted Cumulative Gain (NDCG) explained. *Evidently AI - AI Evaluation & LLM Observability Platform*.

URL: <https://www.evidentlyai.com/ranking-metrics/ndcg-metric> (date of access: 07.12.2025).

28. Методичні вказівки щодо розробки та оформлення кваліфікаційної роботи другого (магістерського) рівня вищої освіти за освітньо-професійною програмою «Інформаційні управляючі системи та технології» спеціальності 122 Комп'ютерні науки / Упоряд.: К.Е. Петров, В.М. Левикін, С.Ф. Чалий, М.В. Євланов, В.І. Саєнко, Д.К. Міхнов, А.В. Міхнова, О.В. Чала. ХНУРЕ: Харків, 2024. 24 с.

29. ДСТУ 8302:2015. Бібліографічне посилання. Загальні положення та правила складання. К.: ДП «УкрНДНЦ», 2016. 20 с.

30. ДСТУ 3008:2015. Державний стандарт України. Інформація та документація. Звіти у сфері науки і техніки. Структура та правила оформлювання. К.: ДП «УкрНДНЦ», 2016. 31 с.