

**Київський національний університет імені Тараса Шевченка**

**Економічний факультет**

**Кафедра економічної кібернетики**

**КВАЛІФІКАЦІЙНА РОБОТА МАГІСТРА**

**Розробка системи рекомендацій у рекрутингу на основі**

**методів штучного інтелекту**

студентки 2 курсу магістратури  
спеціальності 051 «Економіка»  
ОНП «Економічна кібернетика»  
денної форми навчання  
Кожановської Ольги Сергіївни

**Науковий керівник:**

д.е.н., проф.

Чорноус Галина Олександрівна

Засвідчую, що в цій роботі немає запозичень із  
праць інших авторів без відповідних посилань

Студент \_\_\_\_\_

Роботу допущено до захисту перед ЕК  
рішенням кафедри економічної кібернетики  
від 7 травня 2025 р., протокол № 12

Завідувач кафедри:

д.е.н., професор

Ляшенко Олена Ігорівна \_\_\_\_\_

КИЇВ – 2025

## РЕФЕРАТ

Кваліфікаційна робота магістра містить: 70 ст., 20 рис., 5 табл., 51 джерел, додатки.

Ключові слова: *рекомендаційна система, ШІ, машинне навчання, рекрутинг, обробка природної мови, TF-IDF, Random Forest, косинусна подібність, персоналізація.*

Об'єкт дослідження: процеси рекрутингу як складова системи управління персоналом.

Мета дослідження: моделювання та розробка ефективної рекомендаційної системи.

Методи дослідження: класифікаційні моделі машинного навчання, методи обробки природної мови (TF-IDF, косинусна подібність), оцінка ефективності алгоритмів.

Наукова новизна, теоретична значимість дослідження: систематизація та удосконалення сучасних підходів до побудови рекомендаційних систем у сфері рекрутингу.

Практична цінність: система може бути інтегрована у існуючі HR-платформи або працювати як автономний інструмент для автоматизації підбору вакансій та підвищення ефективності рекрутингових процесів.

## RESUME

Taras Shevchenko National University of Kyiv,

Faculty of Economics, Department of Economic Cybernetics

Key words: *recommender system, recruitment, AI, machine learning, natural language processing, TF-IDF, Random Forest, cosine similarity, personalization.*

The graduation research of student Olha Kozhanovska deals with the development of a recommendation system for recruitment based on artificial intelligence methods. The work focuses on analyzing user skills and matching them with job vacancies using machine learning and natural language processing techniques.

The work is interesting for HR professionals, recruitment platforms, and researchers in the field of AI applications in labor markets. It provides a practical solution for improving the efficiency and accuracy of job matching processes.

Pages 70, tables 5, bibliog. 51, append. 2.

## ЗМІСТ

ВСТУП.....	4
РОЗДІЛ 1. ВИКОРИСТАННЯ МЕТОДІВ ШТУЧНОГО ІНТЕЛЕКТУ В РЕКРУТИНГУ .7	7
1.1. Вплив цифровізації та штучного інтелекту на ринок праці та рекрутинг .....	7
1.2. Використання методів штучного інтелекту у сучасних рекомендаційних моделях .11	11
1.3. Аналіз наукових і практичних підходів до створення рекомендаційних систем з впровадженням машинного навчання .....	14
Висновки до розділу 1 .....	19
РОЗДІЛ 2. ТЕОРЕТИКО-МЕТОДОЛОГІЧНІ АСПЕКТИ МОДЕЛЮВАННЯ РЕКОМЕНДАЦІЙНОЇ СИСТЕМИ .....	21
2.1. Рекомендаційні системи .....	21
2.2. Методи машинного навчання в ШІ .....	26
2.3. Архітектура рекомендаційної системи з використанням методів машинного навчання .....	32
Висновки до розділу 2 .....	34
РОЗДІЛ 3. МОДЕЛЮВАННЯ РЕКОМЕНДАЦІЙНОЇ СИСТЕМИ З ВИКОРИСТАННЯМ МЕТОДІВ ШІ .....	36
3.1. Опис та аналіз бази даних.....	36
3.2. Розробка моделей машинного навчання та аналіз результатів моделювання .....	43
3.3. Розробка та імплементація рекомендаційної системи.....	46
Висновки до розділу 3 .....	51
ВИСНОВОК .....	52
ДОДАТКИ.....	60

## ВСТУП

Сучасний ринок праці характеризується стрімким зростанням кількості вакансій та складністю їх ефективного підбору для кандидатів. Традиційні методи рекрутингу часто виявляються неефективними через обмежену здатність аналізувати великі обсяги даних та враховувати індивідуальні характеристики кандидатів. У цьому контексті рекомендаційні системи на основі штучного інтелекту (ШІ) та машинного навчання стають ключовим інструментом для оптимізації процесів пошуку роботи. Вони дозволяють не лише оптимізувати підбір кандидатів, а й прогнозувати їхню ефективність, персоналізувати навчання та розвиток, а також забезпечувати стратегічне планування кадрового потенціалу компаній. Розробка таких систем поєднує новітні технології машинного навчання, обробки природної мови та аналітики великих даних, що відкриває нові можливості для HR-аналітики. Попри значні досягнення розробці сучасних рекомендаційних систем, питання комбінування різних алгоритмів для підвищення точності рекомендацій залишається актуальним, особливо щодо аналізу взаємодії між профілями користувачів та описом вакансій.

*Актуальність* дослідження полягає у необхідності подолання ключових проблем сучасного рекрутингу, зокрема обмеженої персоналізації рекомендацій та низької швидкості обробки даних у умовах стрімкого зростання кількості інформації та їхньої складності. Розробка інтелектуальної системи на основі методів машинного навчання дозволяє автоматизувати аналіз навичок користувачів, підвищити точність підбору за рахунок гібридного підходу та забезпечити адаптацію до індивідуальних потреб ринку праці, що є особливо важливим у контексті цифровізації HR-процесів та дефіциту кваліфікованих фахівців у ключових галузях.

*Об'єкт дослідження* даної роботи – процеси рекрутингу як складова системи управління персоналом.

*Предмет дослідження* – методи та алгоритми машинного навчання, що використовуються для підвищення ефективності рекрутингу шляхом впровадження рекомендаційних систем.

*Метою* даної роботи є моделювання та розробка ефективної рекомендаційної системи для підбору вакансій на основі аналізу навичок користувача з використанням методів ШІ, зокрема, методів машинного навчання та обробки природної мови.

Для досягнення вищезазначеної мети було поставлено наступні *завдання*:

- дослідження впливу цифровізації та впровадження технологій ШІ на ринку праці та в рекрутингу;
- аналіз сучасних підходів до побудови рекомендаційних систем у процесах рекрутингу;
- аналіз наукових і практичних підходів до створення рекомендаційних систем з впровадженням машинного навчання;
- дослідження теоретико-методологічних аспектів моделювання рекомендаційних систем;
- дослідження ефективності різних алгоритмів машинного навчання для класифікації;
- розробка архітектури гібридної системи, що поєднує різні методи та інструменти машинного навчання;
- реалізація рекомендаційної системи, оцінка точності та релевантності рекомендацій.

Теоретична значимість та *наукова новизна* дослідження полягає у систематизації та удосконаленні сучасних підходів до побудови рекомендаційних систем у сфері рекрутингу. Було проведено моделювання архітектури гібридної рекомендаційної системи шляхом інтеграції TF-IDF для більш точної векторизації текстових даних, вибору ефективного алгоритму для класифікації вакансій та застосування косинусної подібності для персоналізованого ранжування результатів.

*Методи дослідження* включають класифікаційні моделі на основі навчання з учителем, серед яких Logistic Regression, KNN, Random Forest, SVM, Naive Bayes, а також методи обробки природної мови (NLP) –TF-IDF (Term Frequency-Inverse Document

Frequency) та косинусну подібність (cosine similarity). Моделювання рекомендаційної системи виконане за допомогою мови програмування Python.

Інформаційною та аналітичною базою дослідження моделювання рекомендаційної системи з використанням методів ШІ є наукові праці вітчизняних та зарубіжних вчених, інтернет-ресурси, періодичні видання тощо.

Дипломна робота складається з вступу, першого розділу, в якому розглянуто теоретичні основи застосування ШІ в рекрутингу, сучасні підходи до побудови рекомендаційних систем та аналіз існуючих методів машинного навчання. Другий розділ присвячений дослідженню архітектури гібридних рекомендаційних систем, вибору оптимальних алгоритмів обробки даних та методикам оцінювання їх ефективності. В третьому розділі представлено практичну реалізацію системи, аналіз результатів моделювання та порівняльну характеристику роботи різних алгоритмів. Також робота містить висновки, перелік використаних джерел, що включає наукові праці, статті, електронні ресурси та додатки з програмним кодом.

## РОЗДІЛ 1. ВИКОРИСТАННЯ МЕТОДІВ ШТУЧНОГО ІНТЕЛЕКТУ В РЕКРУТИНГУ

### 1.1. Вплив цифровізації та штучного інтелекту на ринок праці та рекрутинг

На сьогодні, під час активного впровадження цифровізації в усі сфери життя та стрімкого розвитку ШІ світовий ринок праці трансформується надзвичайно швидкими темпами, модифікуючи традиційні підходи до галузі рекрутингу, управління персоналом та професійного розвитку. Автоматизація рутинних процесів, поява нових професій, зростання попиту на висококваліфікованих фахівців у сфері Bid Data та Data Science – все це ставить перед бізнесом та HR-спеціалістами виклики, які потребують впровадження інноваційних рішень. Класичні методи в рекрутингу вже не задовольняють сучасні потреби, що спонукає до пошуку рішень, здатних не лише ефективно аналізувати наявні дані, а й адаптуватись до динамічних змін на ринку.

ШІ є ключовим драйвером технологічного та економічного прогресу, відкриваючи нові можливості для підвищення ефективності та прибутковості підприємств. Його впровадження змінює глобальні ринки, трансформуючи моделі конкуренції та розвитку. Зокрема, провідні компанії активно розробляють стратегії інтеграції ШІ, водночас зосереджуючись на створенні етичних та соціально відповідальних систем, які відповідають сучасним цінностям.

Сучасний рекрутинг все частіше використовує ШІ для оптимізації рутинних процесів, аналізу даних та персоналізації взаємодії з кандидатами. На відміну від звичайної автоматизації, ШІ не просто виконує завдання за шаблоном, а вміє навчатись, аналізувати інформацію та пристосовуватись до змін, що дозволяє вирішувати складні завдання з високою точністю.

Основними напрямками застосування ШІ у галузі рекрутингу та ринку праці є:

- обробка великих даних – модернізована система може швидко аналізувати резюме, оцінювати відповідність кандидатів і прогнозувати успішність найму;
- автоматизація складних етапів – пошук кандидатів, скринінг, запрошення на співбесіди, що значно прискорює процес;

- покращення якості підбору – ШІ допомагає виявляти найбільш релевантні посади і дає рекомендації на основі аналітики.

Важливо також зазначити, що ШІ не витісняє HR-спеціалістів, а лише допомагає їм концентруватись на стратегічних аспектах, таких як комунікація з кандидатами та побудова ефективних команд. Найкращі ШІ-рішення інтегровані в рекрутингові платформи та ATS-системи як «віртуальні асистенти», що спрощують пошук відповідних кандидатів для конкретних потреб певної компанії [1].

Сучасні ШІ-технології змінюють сферу управління персоналом, надаючи HR-фахівцям потужні інструменти для автоматизації рутинних процесів та прийняття обґрунтованих рішень на основі даних. Використання машинного навчання, обробки природної мови та інших аналітичних систем дозволяє оптимізувати такі ключові процеси, як оцінка персоналу, планування кадрового резерву та розвиток співробітників. ШІ-рішення в HR поєднують потужну технічну інфраструктуру (хмарні сховища, сервери для обробки даних), інтелектуальне програмне забезпечення (віртуальні асистенти, чат-боти, системи прогнозової аналітики) та спеціалізовані послуги для інтеграції та підтримки. Важливою перевагою даних технологій є здатність обробляти великі обсяги даних у режимі реального часу, що забезпечує об'єктивність і точність прийняття HR-рішень.

Інновації з використанням ШІ-технологій особливо актуальні для компаній, які прагнуть підвищити ефективність кадрової політики – від стартапів до міжнародних корпорацій у таких сферах, як IT, фінанси, охорона здоров'я, ритейл тощо.

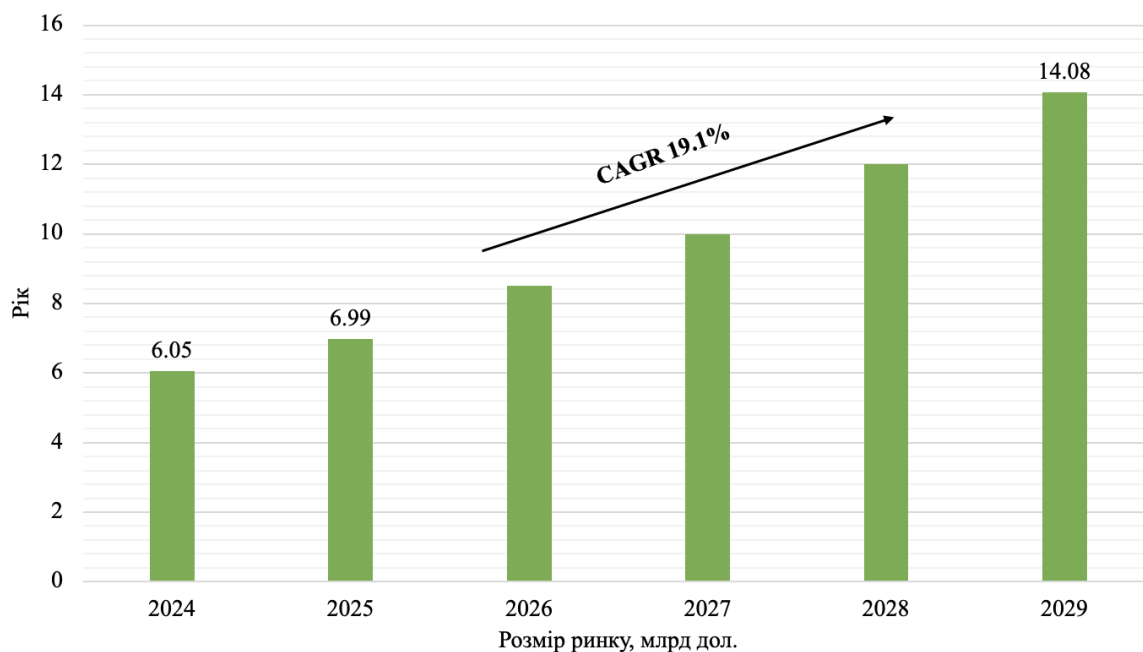


Рис. 1.1. Динаміка світового ринку ШІ у HR-сфері

Джерело: складено автором на основі [2]

Світовий ринок ШІ в HR-сфері демонструє стрімке зростання: у 2024 році його обсяг становив 6,05 млрд доларів, а вже у 2025 - 6,99 млрд доларів із середньорічним темпом приросту (CAGR) 15,6%. Цей значний прогрес пов'язаний із підвищеним попитом на автоматизацію кадрових процесів, необхідністю аналізу великих масивів даних та зростанням витрат на управління талантами. За прогнозами, до 2029 року ринок досягне 14,08 млрд доларів із темпом приросту 19,1%, що зумовлено наступними чинниками: перехід до персоналізації навчання та розвитку персоналу, впровадження етичних стандартів використання ШІ, а також активне застосування передових інструментів – від чат-ботів зі ШІ до аналітики продуктивності. Ключові тренди найближчих років включатимуть розширення можливостей анонімізації даних, створення інноваційних систем менторства та активізацію партнерств між технологічними компаніями, що сформує новий стандарт управління персоналом [3].



Рис. 1.2. Використання штучного інтелекту в рекрутингу та HR

Джерело: складено автором на основі [3]

Варто зазначити, HR-фахівці активно інтегрують ШІ у процеси підбору персоналу. Наприклад, дослідження показали, що 44% керівників відділів кадрів вже використовують ШІ-технології для рекрутингу та найму. При цьому 27% спеціалістів поки що не застосовують ШІ у своїй роботі, а 29% повністю автоматизували один або кілька етапів підбору за допомогою ШІ.

Що стосується обліку виплат та управління пільгами, ситуація виглядає так:

- 44% HR-директорів повністю перевели ці процеси на ШІ-платформи;
- 33% лише починають тестувати такі рішення;
- 23% досі не використовують технології ШІ в цій сфері.

Ці дані демонструють, що ШІ-технології поступово стають невід'ємною частиною сучасного HR-менеджменту, хоча рівень їх адаптації залишається нерівномірним у різних компаніях [3].

## **1.2. Використання методів штучного інтелекту у сучасних рекомендаційних моделях**

Сучасний рекрутинг стикається з низкою викликів: велика кількість кандидатів, суб'єктивність оцінювання, трудомісткість пошуку та аналізу даних. У цих умовах ШІ стає ключовим інструментом для оптимізації процесів підбору персоналу. Одним із найперспективніших напрямів імплементації новітніх технологій в галузі ринку праці є розробка рекомендаційної системи з використанням методів ШІ, що дозволяють автоматизувати процеси та приймати рішення на основі data-driven підходу.

Такі системи можуть застосовуватися як компаніями, що здійснюють пошук та відбір персоналу (у тому числі рекрутинговими агентствами), так і самими кандидатами для пошуку найбільш релевантних вакансій, що сприяє автоматизації, персоналізації та підвищенню точності прийняття рішень у процесі рекрутингу.

В умовах стрімкого розвитку технологій обробки даних інформаційний простір набуває надзвичайної складності. В цьому контексті рекомендаційні системи виступають ключовим інструментом подолання інформаційного перевантаження, знаходячи широке застосування в різних сферах цифрових сервісів. Зокрема, у сфері рекрутингу такі системи базуються на двох основних аспектах: по-перше, аналіз відповідності кваліфікації та досвіду кандидата вимогам вакансії, по-друге – врахування його суб'єктивних переваг, таких як бажаний рівень заробітної плати, форма зайнятості чи віддаленість робочого місця. Таким чином, для побудови ефективної системи рекомендацій у сфері зайнятості необхідно інтегрувати як об'єктивні критерії оцінки компетенцій, так і суб'єктивні фактори вибору [5].

Сучасні рекомендаційні системи, що використовують передові технології ШІ, демонструють значну перевагу над класичними методами. Впровадження генеративних моделей та векторних баз даних відкриває нові можливості для персоналізації та ефективності. Нижче наведено ключові аспекти, які підтверджують їхню перевагу:

- Інтерактивна адаптація у реальному часі: на відміну від традиційних алгоритмів, які базуються на статичних правилах, сучасні чат-боти здатні аналізувати запити

користувачів у формі діалогу. Це дозволяє системі динамічно корегувати рекомендації, враховуючи не лише явні вподобання, а й контекст спілкування. В результаті користувач отримує більш точні та релевантні пропозиції;

- Обробка мультимодальних даних: сучасні системи використовують різноманітні типи даних, включаючи текст, зображення та соціальні взаємодії. Інтеграція великих мовних моделей (LLM) з векторними базами даних дозволяє ефективно керувати масштабними каталогами продуктів. Такі технології забезпечують швидкий пошук схожих елементів на основі семантичної близькості, що значно підвищує якість рекомендацій;

- Контекстно-орієнтований аналіз: генеративні моделі здатні інтерпретувати складні запити, враховуючи поточні потреби користувача. Наприклад, система може адаптувати рекомендації на основі часу доби, місця знаходження або попередніх дій. Це відкриває можливості для створення інтелектуальних асистентів, які не лише пропонують товари чи послуги, але й надають корисні поради [6].

Сучасний рекрутинг проходить значну трансформацію завдяки інтеграції ШІ (значною мірою з використанням генеративного ШІ), який дозволяє автоматизувати рутинні процеси, підвищити точність відбору та зменшити упередженість. Використання ШІ охоплює всі етапи підбору персоналу — від аналізу вакансій до онбордингу нових співробітників [7].

У таблиці нижче систематизовано основні методи та інструменти ШІ, що застосовуються на кожному етапі, а також наведено конкретні технології, які використовують провідні компанії для оптимізації рекрутингових процесів.

## Застосування ШІ в процесах рекрутингу на кожному з етапів

Етап	Основні кроки етапу	Методи та інструменти ШІ
Аналіз вакансії	Генерація опису вакансії, визначення ключових навичок, прогнозування успішності	Текстова генерація (NLP, GPT-4, ChatGPT), аналіз історичних даних (методи класифікації та кластеризації), прогнозування (Random Forest, Gradient Boosting)
Пошук кандидатів	Скрапінг профілів, рекомендації	Обробка неструктурованих даних (NLP, spaCy, BERT), системи рекомендацій, чат-боти
Скринінг резюме	Парсинг резюме, ранжування кандидатів, виявлення упередженості	Виділення ключових збігів (TF-IDF, BERT), класифікація (XGBoost, LightGBM), Bias detection algorithms (AI Fairness 360)
Оцінка кандидатів	Аналіз відеоінтерв'ю, перевірка технічних завдань, гейміфіковані тестові завдання	Емоційний аналіз (Computer Vision (OpenCV), Sentiment Analysis), автоматична оцінка коду (Codex, Kattis), поведінковий аналіз (використання нейронних мереж)
Інтерв'ю	Проведення співбесіди, аналіз відповідей	Аудіоаналіз (Whisper, Google Speech-to-Text), контекстне розуміння (LSTM, GPT)
Онбординг	Персоналізація адаптації	Machine Learning, Survival Analysis (Kaplan-Meier)

Джерело: складено автором на основі [8-10]

Отже, використання NLP, машинного навчання, комп'ютерного зору та рекомендаційних систем дозволяє не лише прискорити процес найму, але й підвищити його якість за рахунок об'єктивних даних. Однак, незважаючи на переваги ШІ, важливо враховувати потенційні ризики, такі як, наприклад, втрата персоналізованого підходу. Тому оптимальна стратегія передбачає поєднання автоматизації з експертною оцінкою HR-фахівців. Майбутнє рекрутингу лежить саме в такій синергії — де ШІ допомагає приймати обґрунтовані рішення, а людина залишається ключовим учасником процесу.

### **1.3. Аналіз наукових і практичних підходів до створення рекомендаційних систем з впровадженням машинного навчання**

Сучасні дослідження в галузі рекомендаційних систем для ринку праці охоплюють широкий спектр підходів та технологічних рішень. Зокрема, при аналізі частоти пошуку ключових слів було виокремлено 75 термінів, що сформували сім кластерів, представлених на рисунку нижче. Слова згруповані навколо таких понять: «recommender systems» (11 слів), «collaborative filtering» (12 слів), «social recommendation» (4 слова), «information» (15 слів), «algorithms» (6 слів), «personalization» (12 слів) та «social networks» (15 слів). Усі кластери відображають ключові напрямки досліджень у сфері рекомендаційних систем: від технологій та моделей персоналізації до алгоритмів, методів фільтрації, інформаційного забезпечення та оцінювання ефективності. Кластер «recommender systems» охоплює поняття, що стосуються принципів реалізації, технічних підходів і сфер застосування рекомендаційних систем, які вдосконалюються завдяки розвитку ШІ і технологій побудови користувачьких профілів [11].

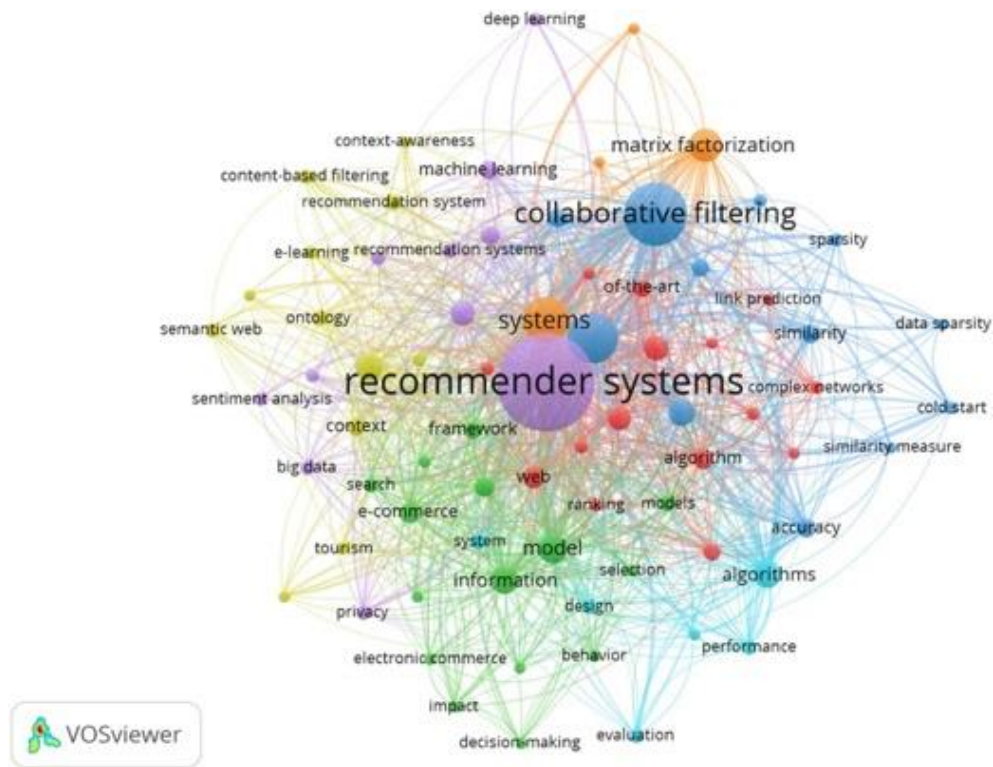


Рис. 1.3. Топ ключових слів для рекомендаційних систем

Джерело: [12]

Сучасний цифровий ринок праці характеризується стрімким зростанням кількості вакансій та їхньої складності, що потребує впровадження інтелектуальних систем для ефективного підбору відповідних пропозицій для користувачів.

Дослідження застосування рекомендаційних систем охоплюють роботи як українських, так і зарубіжних науковців. Серед вітчизняних дослідників варто відзначити праці Скібської К. та Панасюк О. з аналізу автоматизації рекрутингу, Євдокимова В.В. та Морозова А.В. з архітектур рекомендаційних систем, а також Чередніченко О.Ю., який досліджував кластерний аналіз для персоналізації.

Зарубіжні дослідження представлені роботами Widodo R. та Herdiyanto R. (система SWR), Charan S. та співавторів (модифікований алгоритм ROCK), Maria-Iuliana Dascălu (гібридні системи на базі SVD), Mulay A., Sutar Sh., Yifan Hu, Yehuda K. тощо.

Зокрема, в своєму дослідженні Скібська К. та Панасюк О. [13] аналізують сучасні тенденції автоматизації процесів підбору персоналу. Автори висвітлюють потенціал ШІ в рекрутингу, відзначаючи при цьому проблеми адаптації цих технологій у бізнес-

середовищі. Робота містить цінні висновки щодо перспектив впровадження інтелектуальних систем для оптимізації пошуку та найму кваліфікованих фахівців.

У роботі В. В. Євдокимова В.В., Морозова А.В. [14] проведено комплексний аналіз архітектурних рішень для рекомендаційних систем на основі ШІ. Автори систематизують знання про основні типи таких систем, детально описуючи етапи їх роботи - від збору даних до фільтрації результатів. Особливу увагу приділено факторам, що впливають на ефективність рекомендаційних алгоритмів у практичній реалізації.

У своєму дослідженні Чередніченко О.Ю. [15] розглянув вдосконалення рекомендаційних механізмів у сфері електронної торгівлі. Також було обґрунтовано використання кластерного аналізу з застосуванням евклідової відстані для групування як користувачів за їхніми вподобаннями, так і товарів за їхніми атрибутами. Цей підхід дозволив суттєво підвищити точність рекомендацій у системах електронної комерції.

У науковій літературі широко досліджуються гібридні підходи, які поєднують методи колаборативної фільтрації та фільтрації на основі вмісту. Такі системи дозволяють підвищити точність рекомендацій шляхом аналізу як історії взаємодій користувачів, так і змісту вакансій та профілів кандидатів [16].

Одним із перспективних рішень в роботі авторів Widodo R. та Herdiyanto R. є Smart Work Recommendation System (SWR), яка інтегрує передові технології обробки природної мови (NLP) для аналізу навичок користувачів та порівняння їх із вимогами вакансій. Ця система також враховує поведінкові фактори, такі як уподобання щодо корпоративної культури, та надає прогнози щодо заробітної плати, що значно підвищує якість персоналізованих рекомендацій [17].

В роботі Charan S., Suhas G., Yathisha L варто зазначити велику цінність прикладу застосування методу колаборативної фільтрації, використовуючи модифікований алгоритм кластеризації ROCK (A Robust Clustering Algorithm for Categorical Attributes), який вирішує проблему утворення хибних кластерів на завершальних етапах обробки даних. Особливу цінність становить розроблений гібридний метод, що поєднує числові

рейтинги та демографічні характеристики для більш точного визначення груп користувачів [18].

При дослідженні проблеми інформаційного перевантаження в онлайн-рекрутингу запропоновано гібридну систему рекомендацій вакансій. Модель вдало поєднала контентно-орієнтований підхід (аналіз навичок кандидатів з історії поданих заявок) з колаборативною фільтрацією (врахування поведінки схожих користувачів). Для обробки текстів застосовано модель Word2Vec, а для визначення схожості - метод k-найближчих сусідів і коефіцієнт кореляції Пірсона [19].

В своїй роботі Maria-Iuliana Dascălu [20] побудувала інноваційну систему рекомендацій вакансій на основі гібридної фільтрації орієнтована на покращення якості підбору як для початківців, так і для досвідчених фахівців. Особливістю системи є використання NLP для аналізу зворотного зв'язку користувачів, що дозволяє постійно вдосконалювати модель. Зокрема, пропонується реалізація у вигляді мобільного додатку на Flutter, з подальшим вдосконаленням через інтеграцію з глибоким навчанням. Для memory-based колаборативної фільтрації застосовано метод сингулярного розкладу (SVD), що дозволяє точно прогнозувати рейтинги неоцінених вакансій.

Важливим аспектом у працях зарубіжних науковців в дослідженнях рекомендаційних систем є робота з неявними даними, такими як історія переглядів або дії користувачів на платформі. На відміну від явних оцінок, такі дані потребують спеціальних методів обробки, зокрема факторних моделей, які враховують різні рівні довіри до отриманої інформації. Ці підходи демонструють високу ефективність у застосуванні, наприклад, для рекомендації телевізійного контенту, і можуть бути адаптовані для сфери зайнятості [21].

Розвиток електронної комерції та рекомендаційних систем набуває особливої актуальності в умовах цифровізації економіки та зростання конкуренції на онлайн-ринках. Дослідження науковців, що є представниками економічного факультету Київського національного університету імені Тараса Шевченка відіграють ключову роль

у розробці інноваційних підходів до вдосконалення рекомендаційних систем, зокрема для українського та міжнародного бізнес-середовища.

Білорус Т.В. у навчальному посібнику «Управління персоналом: 800+ запитань та відповідей» систематизує теоретичні та практичні аспекти управління персоналом, що є важливим фундаментом для дослідження рекомендаційних систем у рекрутингу. Особливу цінність для дипломної роботи становлять розділи, присвячені методології управління персоналом, оскільки вони допомагають зрозуміти ключові вимоги до підбору кадрів та обґрунтувати необхідність автоматизації цього процесу за допомогою рекомендаційних алгоритмів [22].

Зокрема, у своїй статті Білорус Т.В. акцентує увагу на складності сучасного відбору персоналу та його впливі на конкурентоспроможність компанії. Автор підкреслює, що традиційні методи пошуку та оцінки кандидатів потребують удосконалення через зростання вимог до якості кадрів. В статті обґрунтовується потреба впровадження алгоритму для оптимізації процесу рекрутингу, зокрема для швидкого й ефективного підбору кандидатів з урахуванням професійних та особистісних характеристик [23].

Робота Черноус Г.О. та Лем Т. присвячена гібридним рекомендаційним системам для українського ринку, де рівень їх впровадження залишається низьким. Автори пропонують оригінальну методику, яка включає чотири типи рекомендацій (персоналізовані, "найкращі покупки", новинки, на основі опитувань) та альтернативну систему оцінки товарів (Wilson, Bayes, Hacker). Підхід враховує такі фактори, як час публікації, відгуки та уподобання користувачів, що підвищує релевантність рекомендацій. Дослідження підкреслює потенціал таких систем для української е-комерції [24].

Дослідження Черноус Г.О. та Харламової Г.О. розглядає гібридну модель рекомендаційної системи (РС) для електронної комерції, яка поєднує «user-based» та «item-based» колаборативну фільтрацію. Автори пропонують метод bagging для підвищення точності рекомендацій без використання глибокого навчання. Запропонована модель демонструє нижчий показник RMSE порівняно з традиційними

підходами, що робить її ефективним рішенням для малих та середніх е-комерційних платформ. Зокрема, ця робота є важливим внеском у розвиток методології рекомендаційних систем, зокрема для компаній з обмеженими ресурсами [25].

Також у статті Черноус Г.О. аналізується вплив регіональних та демографічних факторів на поведінку користувачів, зокрема в умовах війни в Україні. В роботі розглядається гібридна рекомендаційна система, яка поєднує контентно-орієнтовані, колаборативні та демографічні фільтри. Система використовує вагові коефіцієнти для адаптації рекомендацій до соціально-економічного контексту, що дозволяє підтримувати українських виробників. Реалізація на мові програмування Python підкреслює практичну придатність розробки для бізнесу [26].

Отже, було проведено комплексний аналіз сучасних наукових праць вітчизняних та зарубіжних дослідників, що висвітлюють теоретичні й практичні аспекти застосування рекомендаційних систем у сфері ринку праці. Особливу увагу приділено дослідженням, які розглядають інноваційні гібридні підходи, що поєднують методи колаборативної фільтрації, контентного аналізу та машинного навчання для підвищення точності рекомендацій. Аналіз показав, що сучасні системи все частіше використовують NLP для обробки текстів резюме та вакансій, глибоке навчання для виявлення складних залежностей, а також методи поведінкового аналізу для врахування кар'єрних уподобань. Водночас виявлено ключові проблемні аспекти, такі як "холодний старт" для нових користувачів, обмеженість даних про взаємодії та складність оцінки якості рекомендацій, які потребують подальших досліджень.

## **Висновки до розділу 1**

У даному розділі досліджуються ключові тенденції впливу цифрових технологій та ШІ на ринок праці, аналізуються переваги та виклики впровадження ШІ-рішень у рекрутингу, а також обґрунтовується необхідність створення інтелектуальних систем рекомендацій для підвищення ефективності залучення та управління талантами.

Зокрема, було досліджено наукову літературу для систематизації основних методологічних підходів до побудови рекомендаційних систем у сфері зайнятості. У ході дослідження було проаналізовано еволюцію методів від традиційних алгоритмів до сучасних гібридних моделей, що поєднують техніки обробки природної мови та глибокого навчання. Особливу увагу приділено вивченню ключових технологічних рішень, таких як семантичний аналіз резюме, векторні представлення навичок та прогнозування відповідності кандидата. Результати аналізу дозволяють визначити перспективні напрями подальших досліджень у цій галузі.

Отримані висновки є теоретичною основою для подальшої розробки власної системи рекомендацій, спрямованої на вдосконалення процесів підбору персоналу за допомогою використання сучасних алгоритмів ШІ.

## РОЗДІЛ 2. ТЕОРЕТИКО-МЕТОДОЛОГІЧНІ АСПЕКТИ МОДЕЛЮВАННЯ РЕКОМЕНДАЦІЙНОЇ СИСТЕМИ

### 2.1. Рекомендаційні системи

Рекомендаційна система є спеціалізованим підвидом машинного навчання, який аналізує великі масиви даних для передбачення, фільтрації та вибору найбільш релевантних варіантів серед зростаючої кількості альтернатив.

Рекомендаційна система представляє собою алгоритм штучного інтелекту [27], що застосовує методи машинного навчання для обробки великих даних (Big Data) з метою пропозиції споживачам додаткових товарів чи послуг. Враховуючи такі критерії, як історія покупок, пошукові запити, демографічні дані та інші параметри, такі системи значно спрощують процес вибору, допомагаючи користувачам виявляти нові для них продукти .

Функціонування рекомендаційних систем ґрунтується на аналізі взаємодій користувачів з контентом, включаючи перегляди, кліки, вподобання та фактичні покупки. Накопичуючи та обробляючи ці дані, система формує індивідуальні моделі для визначення користувацької поведінки [27].

Сучасні рекомендаційні системи класифікуються за типом даних, що використовуються для прогнозування уподобань користувачів, на такі основні категорії:

- фільтрація вмісту (Content-Based Filtering);
- системи колаборативної фільтрації (Collaborative Filtering);
- системи гібридної фільтрації (Hybrid Filtering) [28].

Метод фільтрації на основі вмісту є одним з підходів у рекомендаційних системах, який пропонує користувачам елементи, схожі за характеристиками на ті, що їх вони вже переглядали або оцінювали. Ця технологія застосовує алгоритми машинного навчання для ідентифікації подібних об'єктів на основі їхніх властивостей та описових параметрів [29].

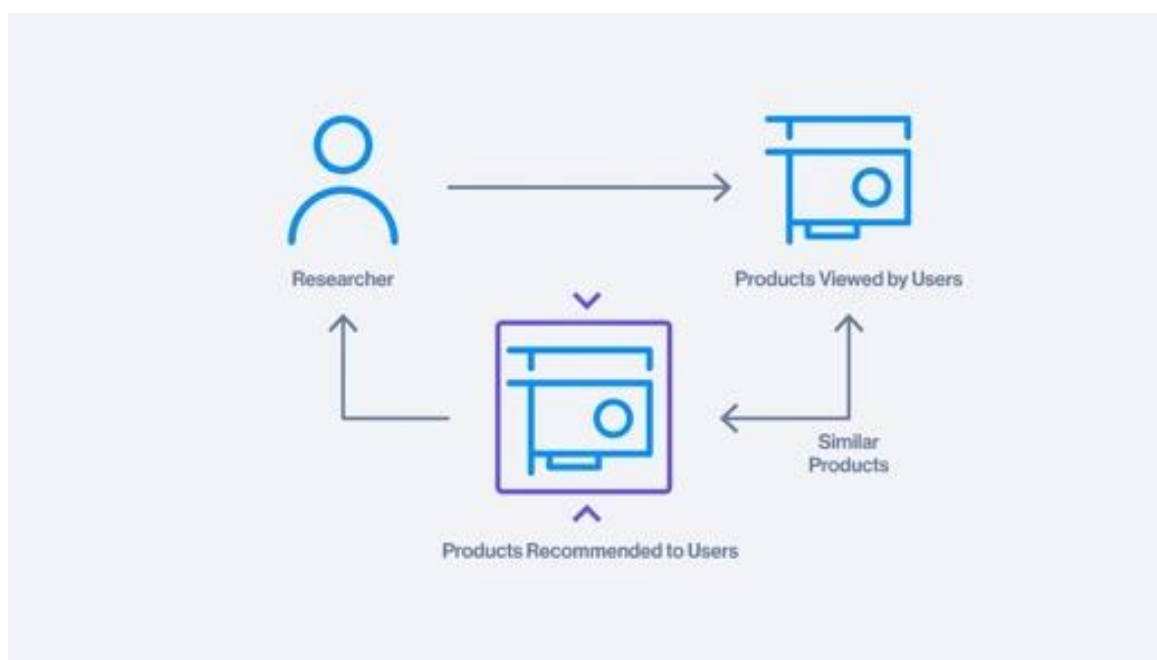


Рис. 2.1. Схема рекомендаційної системи на основі фільтрації вмісту

Джерело: [28]

Системи рекомендацій на основі фільтрації вмісту використовують класифікаційні або регресійні моделі, орієнтовані на інтереси конкретного користувача. На основі цих даних модель формує персоналізовані прогнози щодо нових рекомендацій. Ефективність таких систем може бути значно підвищена за допомогою тегів, створених методами обробки природної мови. Проте процес розмітки великих масивів даних може вимагати значних обчислювальних ресурсів та часу.

Важливою перевагою такого підходу є менша залежність від проблеми "холодного старту" порівняно з методами колаборативної фільтрації, оскільки аналіз базується на метаданих, а не на історії взаємодії користувачів. Однак цей метод має обмеження щодо розширення асортименту рекомендацій, оскільки здебільшого пропонує об'єкти, які схожі на вже відомі користувачеві [30].

Колаборативна фільтрація є одним з найпоширеніших методів побудови рекомендаційних систем, який базується на аналізі взаємодій між користувачами та об'єктами.

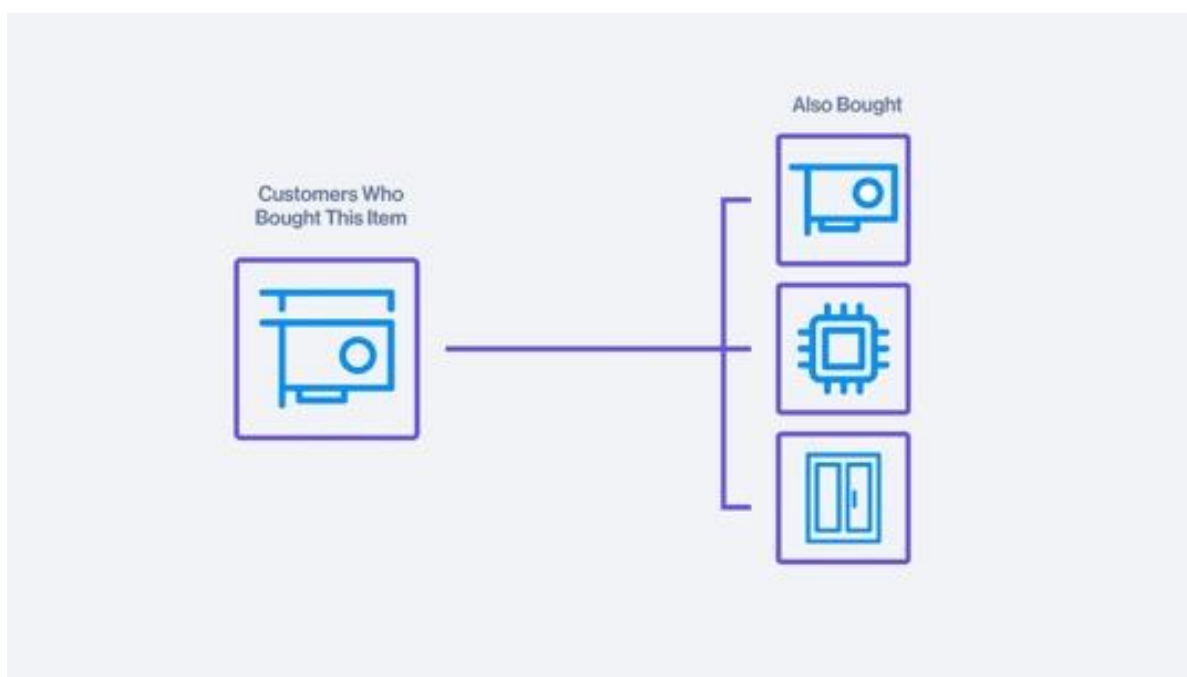


Рис. 2.2. Схема рекомендаційної системи на основі колаборативної фільтрації

Джерело: [28]

Сучасні системи колаборативної фільтрації поділяються на дві категорії: системи, що базуються на пам'яті (Memory-based) та системи, що базуються на моделях (Model-based). Кожна з цих категорій має свої переваги: системи на основі пам'яті простіші у підтримці, тоді як Model-based системи краще масштабуються для великих каталогів і демонструють вищу точність рекомендацій. Вибір конкретного типу залежить від розмірів бази даних, вимог до продуктивності та наявних обчислювальних ресурсів [31].

Рекомендаційні системи на основі пам'яті діляться на два типи:

- колаборативна фільтрація за типом користувачів: даний метод працює за принципом виявлення користувачів зі схожими вподобаннями. Система аналізує історію переглядів або оцінок та знаходить "сусідів" – користувачів, чий інтереси збігаються з цільовим. Наприклад, якщо двоє користувачів високо оцінили одні й ті самі продукти, а один з них додатково переглянув іншу стрічку, цей продукт буде запропоновано другому користувачеві як потенційно цікавий;

- колаборативна фільтрація за типом об'єктів: у цьому випадку система зосереджується на пошуку схожих між собою предметів або контенту. Аналізуючи

поведінку користувачів, алгоритм визначає, які об'єкти часто співвідносяться разом. Наприклад, якщо багато користувачів одночасно оцінюють дві конкретні книги, система вважатиме їх схожими, і при виборі однієї з них запропонує другу як рекомендацію [32].

Model-based системи колаборативної фільтрації передбачають створення прогнозної моделі машинного навчання. В якості навчального набору даних використовується матриця «користувач-елемент», де наявні значення слугують для побудови моделі. Для формування рекомендацій застосовуються різні методи аналізу даних та алгоритми машинного навчання, включаючи дерева рішень, байєсові класифікатори та нейронні мережі [33].

Гібридна рекомендаційна система інтегрує переваги фільтрації на основі вмісту та колаборативної фільтрації, усуваючи недоліки, притаманні кожному з цих методів окремо.

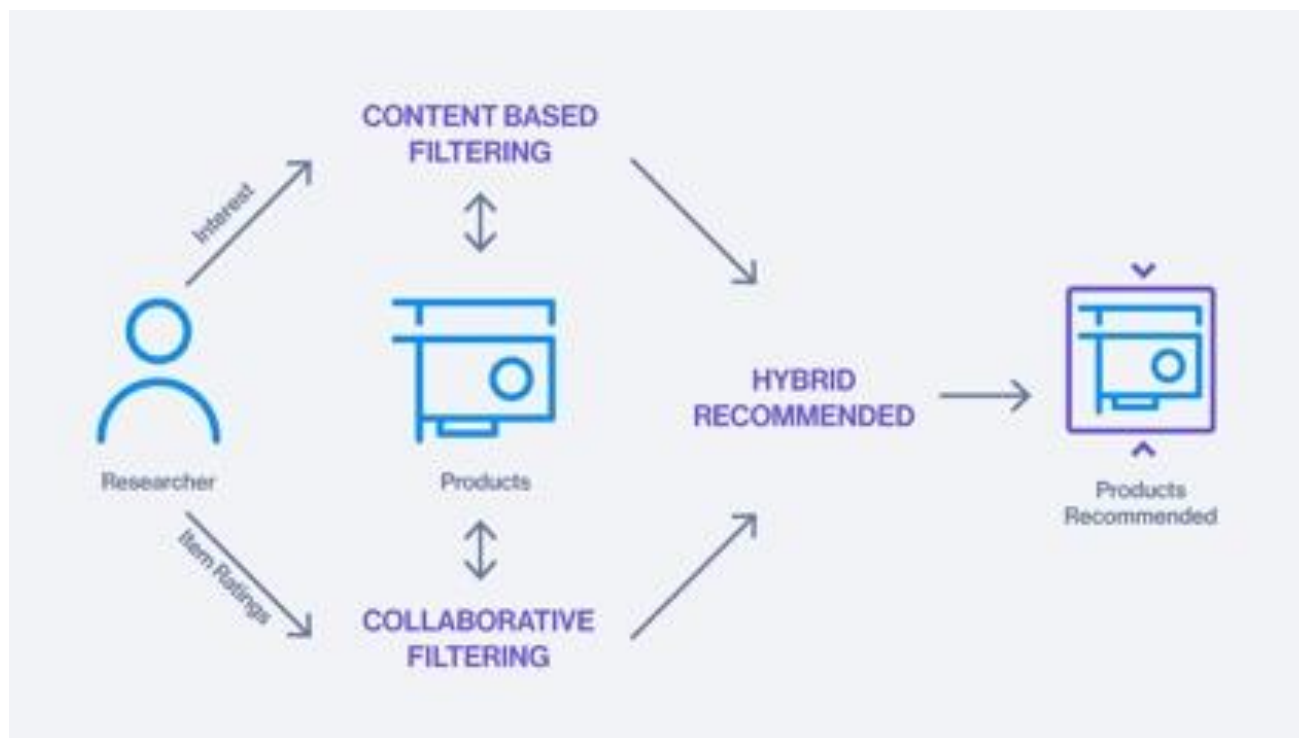


Рис. 2.3. Схема рекомендаційної системи на основі гібридної фільтрації

Джерело: [28]

Гібридні системи можуть реалізовуватись кількома способами:

- паралельна комбінація методів: окремі компоненти системи розробляються незалежно, а потім об'єднуються для отримання остаточних рекомендацій. Це дозволяє зберегти переваги кожного підходу та компенсувати їхні обмеження;

- ієрархічна реалізація: гібридна система адаптує вибір методу в залежності від обсягу доступних даних про користувача. Наприклад, для нових користувачів із обмеженою історією взаємодій застосовується фільтрація вмісту, що ефективно вирішує проблему «холодного старту». Для користувачів з вищим рівнем взаємодії активується механізм колаборативної фільтрації [34].

Така гнучкість дозволяє гібридним системам забезпечувати якісні рекомендації незалежно від стадії життєвого циклу користувача чи об'єкта в системі, підвищуючи задоволеність користувачів та ефективність платформи в цілому.

Зокрема, у таблиці нижче представлено порівняльну характеристику трьох ключових типів рекомендаційних систем: колаборативної фільтрації, фільтрації вмісту та гібридної фільтрації. Аналіз охоплює такі аспекти, як використовувані інструменти, переваги, недоліки та основні виклики, пов'язані з кожним із підходів.

Таблиця 2.1

Порівняння типів рекомендаційних систем

Характеристики	Колаборативна фільтрація	Фільтрація вмісту	Гібридна фільтрація
Інструменти	Apache Mahout, Hadoop	Python, TensorFlow	Apache Spark, scikit-learn
Переваги	Використовує взаємодію «користувач-об'єкт». Ефективність для соціальних мереж	Рекомендації на основі атрибутів об'єктів Незалежність від даних інших користувачів	Поєднує переваги обох підходів Подолання обмежень окремих методів

Недоліки	Проблема «холодного старту»	Обмеженість відомими об'єктами Суб'єктивні уподобання важко врахувати	Складність налаштування Висока обчислювальна складність
Виклики	Масштабування для великих даних	Обмежений аналіз уподобань Складне визначення рекомендацій для нових користувачів без попередньої історії	Оптимізація балансу між підходами Інтеграція з реальними системами

Джерело: складено автором на основі [35]

## 2.2. Методи машинного навчання в ШІ

Штучний інтелект — широка галузь комп'ютерних наук, яка займається створенням систем, здатних виконувати завдання, що традиційно вимагають людського інтелекту. До таких завдань належать розпізнавання образів, прийняття рішень, обробка природної мови та інше. ШІ охоплює різні підходи, включаючи експертні системи, логічне програмування, нейронні мережі та машинне навчання. Останнє є однією з найважливіших складових сучасного ШІ, оскільки саме воно надає можливість комп'ютерам «навчатися» на основі даних без безпосереднього програмування [36].

Машинне навчання (МН) — один із напрямів ШІ, що дозволяє інформаційним системам самостійно виявляти закономірності у великих обсягах даних і знаходити рішення без безпосереднього втручання людини. Це поняття охоплює низку методів і інструментів, які забезпечують здатність комп'ютера навчатися та пристосовуватися до нових умов [37].

На відміну від класичного програмування, де людина створює програму вручну, використовуючи вхідні дані для отримання результату, у машинному навчанні алгоритм самостійно формує модель на основі наданих вхідних даних та очікуваного результату.

Сучасні алгоритми машинного навчання здатні працювати з будь-якими типами інформації - від текстів і зображень до складних математичних моделей. Головна умова - наявність цифрового представлення даних, які можна використати для навчання моделей. Це відкриває практично необмежені можливості для автоматизації процесів у різних галузях [38].

Взаємозв'язок між ШІ та МН полягає в тому, що машинне навчання виступає як ключовий механізм реалізації штучного інтелекту. Наприклад, якщо ШІ — це мета (створення розумних систем), то МН — один із найефективніших способів досягнення цієї мети. Без машинного навчання багато сучасних ШІ-систем, таких як чат-боти, системи рекомендацій або автоматичні перекладачі, не могли б функціонувати, оскільки саме МН дозволяє їм аналізувати великі обсяги даних, виявляти закономірності та адаптуватися до нової інформації.

Важливо зауважити, що машинне навчання — не єдиний спосіб реалізації ШІ. Існують системи, які використовують, наприклад, правила, задані людиною (експертні системи), або жорсткі алгоритми обробки даних. Однак саме МН стало основним драйвером прогресу в ШІ завдяки зростанню обчислювальних потужностей і доступності великих даних. Сьогодні такі технології, як глибоке навчання (підвид МН), лежать в основі найскладніших ШІ-систем, таких як ChatGPT або системи комп'ютерного зору.

Таким чином, машинне навчання є ключовим інструментом у розвитку штучного інтелекту, надаючи йому можливість адаптуватися, вдосконалюватися та вирішувати складні завдання. Разом вони формують основу сучасних технологічних інновацій, трансформуючи різні сфери життя — від медицини до фінансів і рекрутингу [39].

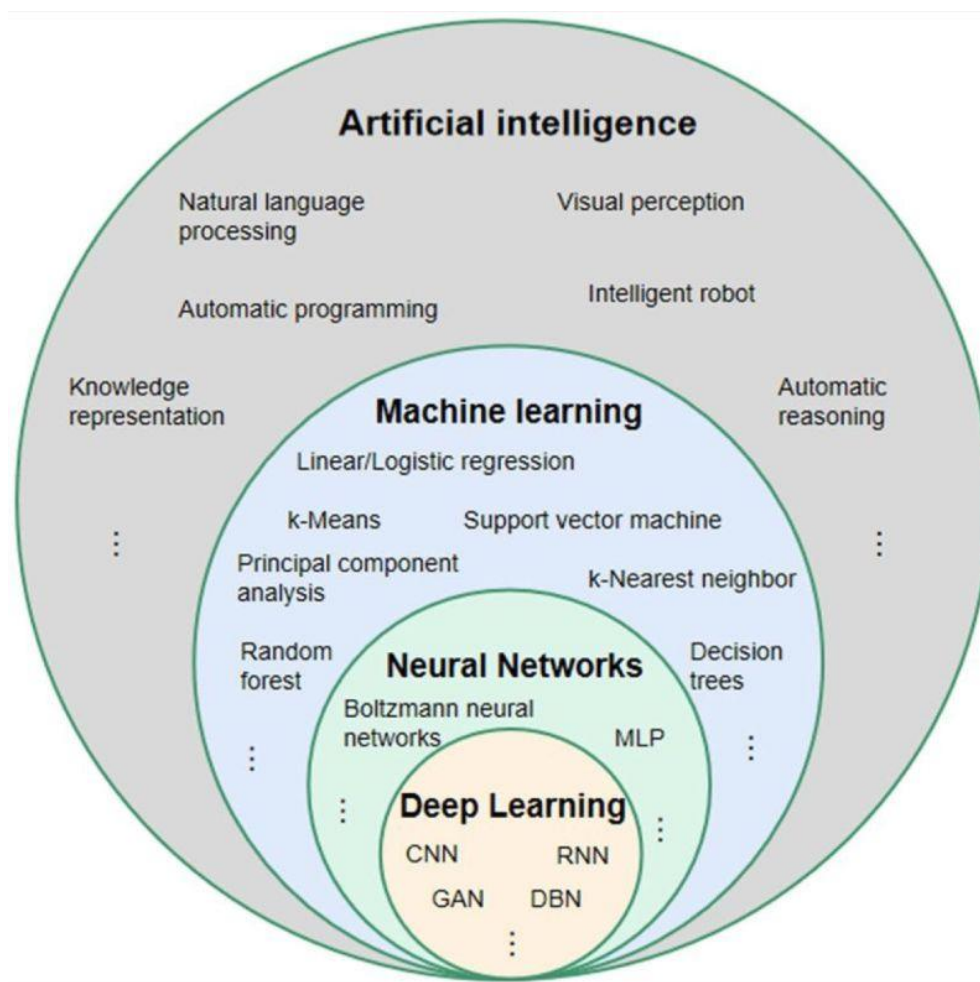


Рис. 2.4. Місце машинного навчання в ІІІ

Джерело: [40]

Алгоритми та методи машинного навчання є основою роботи систем рекомендацій. Завдяки автоматизованому налаштуванню, узгодженню та використанню алгоритмів прогнозування, такі системи можуть обирати оптимальні варіанти для конкретних потреб користувача.

Сучасні рекомендаційні системи активно використовують методи ІІІ та алгоритми машинного навчання для аналізу даних і генерації персоналізованих пропозицій. Одним із найпоширеніших підходів є колаборативна фільтрація, яка базується на аналізі поведінки користувачів. Алгоритми колаборативної фільтрації, такі як SVD (Singular Value Decomposition) або k-NN (k-Nearest Neighbors), дозволяють виявляти приховані закономірності у великих масивах даних. Однак цей метод має обмеження, пов'язані з

проблемою «холодного старту», коли система не може дати рекомендації новим користувачам або для нових товарів через відсутність історичних даних.

Іншим важливим підходом є контентно-орієнтована фільтрація, яка аналізує характеристики об'єктів (наприклад, жанри фільмів, атрибути товарів або ключові навички у вакансіях) для рекомендації схожих елементів. Для обробки текстових даних у таких системах часто використовуються методи TF-IDF (Term Frequency-Inverse Document Frequency) або Word2Vec, які перетворюють текст у числові вектори. Цей підхід особливо ефективний у сфері рекрутингу, де рекомендаційна система може аналізувати резюме та вакансії на основі ключових слів. На відміну від колаборативної фільтрації, контентний метод не залежить від історії взаємодій користувачів, але має інший недолік — обмежену здатність враховувати суб'єктивні уподобання.

Для подолання недоліків окремих методів у сучасних рекомендаційних системах все частіше застосовують гібридні моделі, які поєднують колаборативну та фільтрацію на сонові вмісту. Наприклад, алгоритм DeepFM інтегрує факторизаційні машини (Factorization Machines) з глибоким навчанням для аналізу як структурованих, так і неструктурованих даних. Також популярність набувають трансформерні моделі, такі як BERT або GPT, які можуть обробляти текст і виявляти складні семантичні зв'язки між об'єктами. Ці моделі особливо корисні для рекомендації контенту у соціальних мережах або для підбору персоналізованих навчальних матеріалів [41].

Окремо варто відзначити методи глибокого навчання, зокрема нейронні мережі з згортками (CNN) для обробки зображень або рекурентні нейронні мережі (RNN) для аналізу послідовних даних, таких як історія переглядів. Такі моделі дозволяють враховувати складні нелінійні залежності та контекстні фактори, такі як час доби або місцезнаходження користувача.

Таким чином, вибір конкретних методів ШІ і машинного навчання для побудови рекомендаційної системи залежить від типу даних, специфіки завдання та доступних обчислювальних ресурсів. Найефективніші рішення зазвичай поєднують кілька підходів, таких як колаборативна фільтрація, контентний аналіз і глибоке навчання, щоб

забезпечити точність, персоналізацію та масштабованість рекомендацій. Майбутній розвиток цієї галузі пов'язаний із вдосконаленням генеративних моделей, які можуть інтегрувати рекомендації у природну мовну взаємодію, роблячи системи більш інтуїтивними та адаптивними [42, 43].

Системи рекомендацій здатні передбачати оцінки користувачів ще до того, як вони їх поставили, що робить їх потужним інструментом. Основний процес обробки даних у таких системах включає такі основні етапи збору, зберігання, аналізу та фільтрації [44].

Важливу роль у цьому процесі відіграє кероване машинне навчання (supervised learning), яке дозволяє системі навчатися на історичних даних, де кожен приклад містить вхідні ознаки та відповідну цільову змінну (наприклад, оцінку користувача) [45].

Кероване навчання — це підхід у машинному навчанні, при якому алгоритм навчається на розмічених даних, тобто на прикладах, де для кожного вхідного значення вказано правильну відповідь або клас. Назва методу походить від аналогії з навчанням під наглядом вчителя: модель "коригується" на основі відомих правильних результатів. Після навчання на таких даних система здатна аналізувати нові, раніше невідомі приклади та робити точні прогнози або класифікації. Кероване навчання передбачає наявність чітко структурованої навчальної вибірки, де кожен об'єкт має відповідну мітку, що дозволяє алгоритму виявляти закономірності та будувати прогностичні моделі [46].

Наприклад, у рекомендаційних системах це може бути матриця «користувач-елемент» із відомими оцінками, на основі якої модель навчається передбачати невідомі рейтинги. Популярні алгоритми керованого навчання, такі як лінійна регресія або метод опорних векторів (SVM) дозволяють точно моделювати складні залежності між даними, покращуючи якість рекомендацій. Крім того, використання нейронних мереж, особливо глибокого навчання, дає змогу враховувати складні нелінійні взаємозв'язки та приховані закономірності в поведінці користувачів, що значно підвищує ефективність системи [47].

Для більш точного прогнозування у рекомендаційних системах часто застосовують ансамблеві методи, такі як Random Forest, які комбінують переваги різних підходів [48].

Важливим етапом є також оцінка якості моделей за допомогою метрик точності, таких як precision-recall або F1-score для класифікації.

Зокрема, використані в подальшому моделюванні методи класифікації на основі машинного навчання наведено в таблиці нижче.

Таблиця 2.2

## Методи класифікації

Метод	Характеристика
Logistic Regression	Призначена для класифікації бінарних подій (наприклад, так/ні). На відміну від лінійної регресії, вона прогнозує ймовірність належності до певного класу, що робить її ефективною для задач на кшталт виявлення спаму.
K-Nearest Neighbors (KNN)	Класифікує об'єкти на основі їхньої близькості до навчальних прикладів. Незважаючи на простоту реалізації, при роботі з великими даними його ефективність знижується через зростання обчислювальних витрат.
Random Forest	Метод, що поєднує множину некорельованих дерев рішень для підвищення точності прогнозів. Він ефективний як для класифікаційних, так і для регресійних задач, зменшуючи варіативність окремих моделей.
Support Vector Machine (SVM)	Універсальний алгоритм для класифікації та регресії, який знаходить оптимальну гіперплощину для розділення класів. Він максимізує відстань між найближчими точками різних класів, що підвищує точність класифікації.
Naive Bayes	Алгоритм класифікації, що базується на теоремі Байеса та припущенні про незалежність ознак. Це означає, що кожна характеристика впливає на результат незалежно від інших. Розрізняють три основні типи таких класифікаторів:

	мультиноміальний, Бернуллі та Гаусів, які часто застосовують для аналізу текстів, фільтрації спаму та в рекомендаційних системах.
--	---

Джерело: створено автором на основі [40], [50].

### **2.3. Архітектура рекомендаційної системи з використанням методів машинного навчання**

На основі опрацьованих наукових та практичних джерел сформулюємо основні етапи побудови рекомендаційної моделі з використанням методів ШІ. План моделювання полягає в послідовному виконанні наступних пунктів:

- Вибір бази даних;
- Аналіз бази даних, перевірка її на якість;
- Вибір інструмента для моделювання;
- Підготовка до моделювання;
- Навчання моделей та налаштування їх параметрів;
- Підготовка та створення рекомендаційної моделі.

Зокрема, схема нижче відображає загальну структуру роботи рекомендаційної системи, яка поєднує різні техніки машинного навчання для генерації персоналізованих рекомендацій. Загальна структура роботи рекомендаційної системи починається зі збору даних. Наступний етап передбачає застосування різних технік рекомендацій, таких як текстовий аналіз, моделі кластеризації, матрична факторизація та нейронні мережі. Фінальний результат — персоналізований список рекомендованих елементів, який формується на основі аналізу вхідних даних та обраних методів. Схема демонструє логічний потік від збору даних до генерації рекомендацій, підкреслюючи ключові етапи та технології, що використовуються.

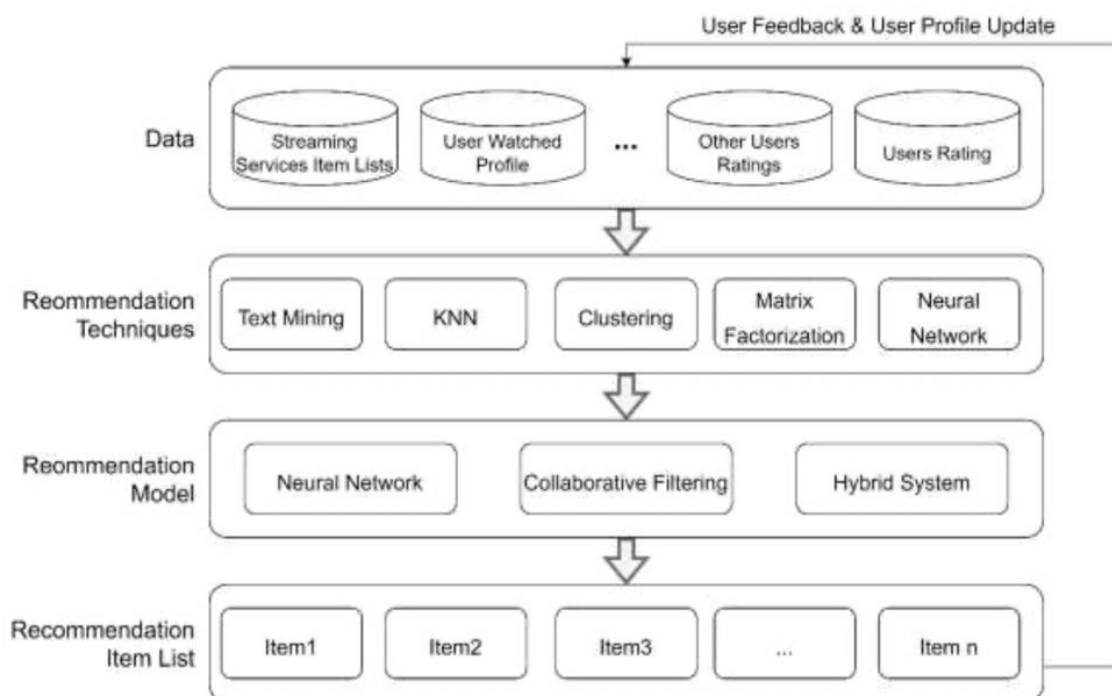


Рис. 2.5. Архітектура гібридної рекомендаційної системи

Джерело: [47]

Рекомендаційна система для підбору вакансій на основі навичок користувача являє собою комплексне рішення, що поєднує методи машинного навчання та обробки природної мови. Система складається з декількох взаємопов'язаних модулів, кожен з яких виконує певну функцію в загальному ланцюжку обробки даних.

1. Вхідний модуль відповідає за отримання та первинну обробку даних від користувача. Користувач вводить перелік своїх професійних навичок у текстовому форматі через консольний інтерфейс. Для підвищення якості подальшої обробки система автоматично виконує приведення тексту до нижнього регістру, видалення зайвих пробілів та базову нормалізацію тексту.

2. Модуль векторного представлення використовує технологію TF-IDF (Term Frequency-Inverse Document Frequency) для перетворення текстової інформації у числові вектори. Цей етап є критично важливим, оскільки кожна навичка користувача отримує вагу, що відображає її важливість, а також створюється компактне числове

представлення текстових даних та забезпечується можливість математичного порівняння різних наборів навичок.

3. Модуль класифікації заснований на алгоритмі класифікаційної моделі, який був обраний після ретельного тестування різних моделей машинного навчання. Особливості роботи цього модуля: модель була попередньо натренована на історичних даних вакансій; для кожного вхідного запиту система визначає ймовірність належності до кожної категорії посад; обирається категорія з найвищою ймовірністю.

Отже, підготовка до навчання моделей, створення алгоритму на написання коду є запорукою успішності процесу навчання моделей. Налаштування усіх параметрів допомагає підвищити якість моделей. Отримана модель може бути використана для побудови рекомендаційної моделі.

Зокрема, варто зазначити, що для довготривалого використання рекомендаційної моделі варто проводити періодичне донавчання для більш точних результатів та підлаштовувати базу даних під потреби конкретного запиту.

## **Висновки до розділу 2**

У даному розділі було досліджено сутність, принципи роботи та основні типи рекомендаційних систем, які є ключовим інструментом у сучасних інформаційних технологіях для персоналізації контенту. Рекомендаційні системи, як спеціалізований підвид машинного навчання, аналізують великі масиви даних, щоб пропонувати користувачам найбільш релевантні варіанти серед численних альтернатив. Було розглянуто три основні підходи: фільтрацію вмісту (Content-Based Filtering), колаборативну фільтрацію (Collaborative Filtering) та гібридні системи, кожен з яких має свої переваги та обмеження.

Також було проаналізовано архітектуру рекомендаційної системи для підбору вакансій, яка включає модулі вхідних даних, векторного представлення (TF-IDF), класифікації (Random Forest) та ранжування (косинусна подібність). Ця архітектура

демонструє, як комбінація методів машинного навчання та NLP дозволяє ефективно обробляти текстові дані і генерувати персоналізовані результати.

Цей розділ став теоретичною основою для практичної реалізації гібридної рекомендаційної системи, яка буде детально розглянута в наступному розділі даної роботи.

## РОЗДІЛ 3. МОДЕЛЮВАННЯ РЕКОМЕНДАЦІЙНОЇ СИСТЕМИ З ВИКОРИСТАННЯМ МЕТОДІВ ШІ

### 3.1. Опис та аналіз бази даних

Для реалізації практичної частини даної роботи та побудови рекомендаційної системи було обрано базу даних «Predicting Job Titles from Resumes», опубліковану у вільному доступі на платформі Kaggle [51]. Використана база даних була створена у 2022 році та містить інформацію про вакансії, включаючи такі ключові параметри, як назва посади, необхідний досвід роботи, ключові навички, категорія посади, функціональна область, галузь та заробітна плата. Вибір саме цього джерела статистичних даних обумовлений його комплексністю та релевантністю для дослідження, оскільки він охоплює широкий спектр характеристик вакансій, що є важливим для подальшого моделювання та побудови точної та ефективної рекомендаційної системи.

Розробка програмного коду для моделювання була реалізована за допомогою мови програмування Python у середовищі Google Colab, що дозволило ефективно використати переваги хмарних обчислень для обробки текстових даних та навчання моделей машинного навчання. Використання бібліотек, таких як pandas для маніпуляції даними, scikit-learn для реалізації алгоритмів TF-IDF векторизації, Random Forest та косинусної подібності, а також matplotlib для візуалізації результатів, забезпечило швидку та ефективну розробку системи. Google Colab надав доступ до потужних обчислювальних ресурсів, що було особливо важливо при роботі з великими обсягами текстових даних та навчанні ансамблевих моделей.

Зокрема для досягнення поставлених цілей було використано такі інструменти: pandas для маніпуляції даними, numpy для математичних операцій, seaborn та matplotlib.pyplot для візуалізації, а також бібліотеки sklearn.model\_selection, sklearn.feature\_extraction.text та sklearn.metrics для машинного навчання та аналізу текстових даних.

Набір даних складається з 8 обраних змінних, кожна з яких налічує 27010 унікальних значень, при знаходженні індексів яких маємо наступні значення:

```
Index(['Unnamed: 0', 'Job Salary', 'Job Experience Required', 'Key Skills',
      'Role Category', 'Functional Area', 'Industry', 'Job Title'],
      dtype='object')
```

Рис. 3.1. Змінні бази даних

Джерело: розрахунки автора

Розглянемо детальніше опис змінних у базі даних:

Таблиця 3.1

Опис змінних бази даних

№	Змінна	Опис	Тип
1	Unnamed: 0	ID запису (автогенерується при завантаженні)	Integer
2	Job Salary	Інформація про рівень оплати праці (може містити «Not Disclosed by Recruiter»)	String (Mixed)
3	Job Experience Required	Рівень необхідного стажу роботи	String
4	Key Skills	Перелік основних компетенцій для певної вакансії	String
5	Role Category	Класифікація посади за сферою діяльності	String
6	Functional Area	Спеціалізація посади	String
7	Industry	Сфери економіки, пов'язані з вакансіями	String
8	Job Title	Інформація про професійну посаду	String

Джерело: складено автором на основі [46]

Усі змінні в таблиці, за винятком технічного ідентифікатора (Unnamed: 0), представлені у текстовому форматі (String). Наприклад, «Job Salary» містить не лише конкретні цифрові значення, а й варіанти на кшталт «Not Disclosed by Recruiter», що

унеможливиює автоматичне трактування стовпця як числового. Аналогічно, «Key Skills», «Role Category» або «Industry» складаються з текстових описів, які часто потребують додаткової обробки (наприклад, токенизації або категоризації) для подальшого аналізу. Це характерно для даних, зібраних з рекрутингових платформ, де інформація може містити нестандартизовані формулювання, що ускладнює їх машинну обробку без попереднього очищення та нормалізації.

При перевірці відсутніх значень було виявлено, що всі записи в стовпцях заповнені — жодних пропусків (NaN, NULL або пустих значень) у базі даних не знайдено, що свідчить про повноту та цілісність даних.

	0
<b>Unnamed: 0</b>	0
<b>Job Salary</b>	0
<b>Job Experience Required</b>	0
<b>Key Skills</b>	0
<b>Role Category</b>	0
<b>Functional Area</b>	0
<b>Industry</b>	0
<b>Job Title</b>	0

Рис. 3.2. Перевірка пропущених значень

Джерело: розрахунки автора

Матриця кореляції Спірмена, отримана для аналізу набору даних, демонструє рівень зв'язку між різними категоріальними змінними, перетвореними у числовий формат за допомогою методу Label Encoding. Найвищий рівень кореляції спостерігається між змінними «Role Category» та «Functional Area» (0.46), а також між «Role Category» та «Job Title» (0.67), що свідчить про тісний взаємозв'язок між цими категоріями: посади певної сфери діяльності («Role Category») часто пов'язані з конкретними функціональними областями («Functional Area») та специфічними назвами посад («Job

Title»). Також помітний зв'язок між «Industry» та «Job Title» (0.30) та «Role Category» (0.26), що підтверджує залежність між галуззю економіки, типом посади та її назвою. Змінна «Job Salary» має слабкі кореляції з іншими змінними (найвища — 0.23 з «Industry»), що може бути пов'язано з неоднорідністю даних (наявність значень типу «Not Disclosed by Recruiter»). Кореляції для «Key Skills» та «Job Experience Required» є незначними (близькі до нуля), що вказує на відсутність явної залежності між навичками та досвідом роботи.

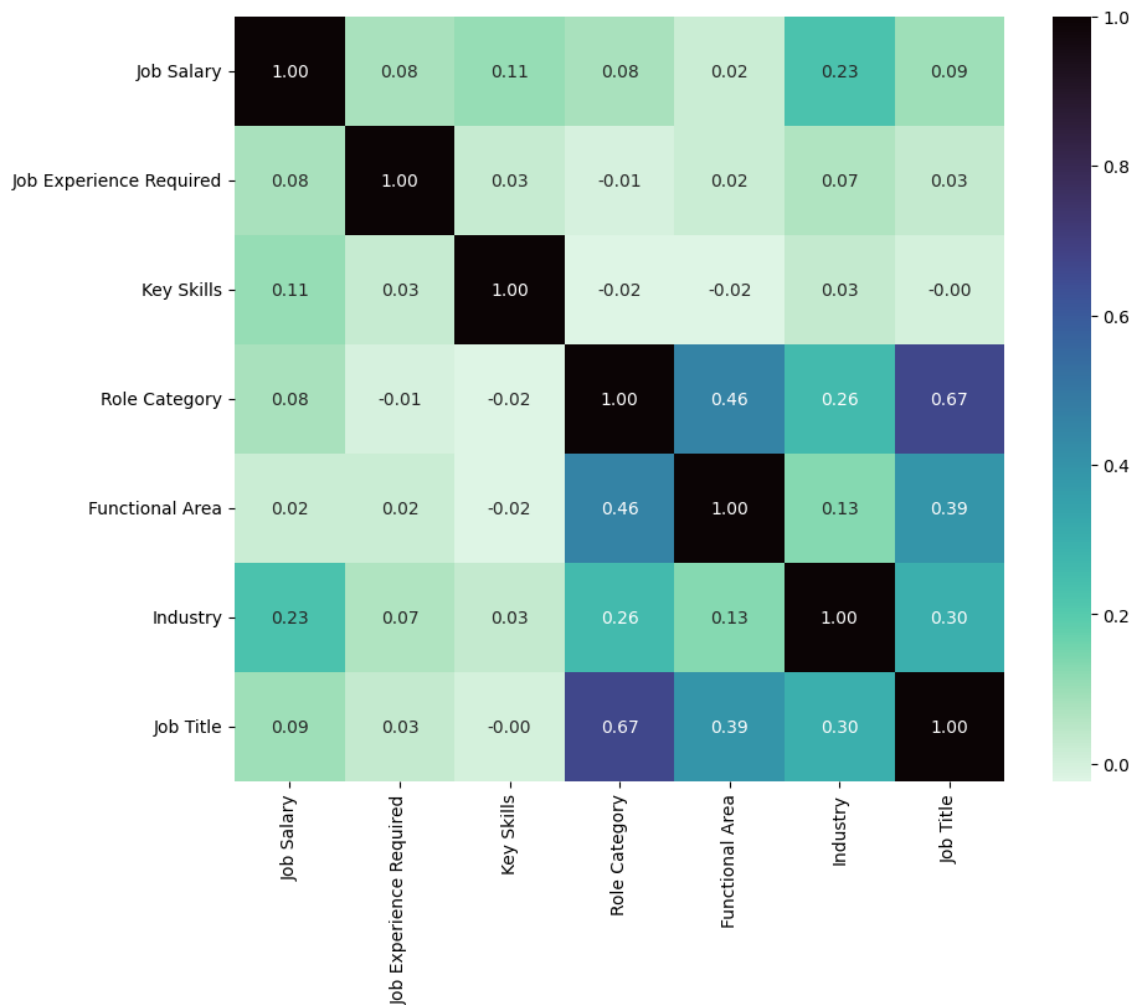


Рис. 3.3. Матриця кореляції Спірмена

Джерело: розрахунки автора

Окрему увагу варто приділити аналізу ключових навичок, оскільки їх частота вказує на попит на ринку праці: стовпчикові діаграми демонструють, що такі навички, як Customer Service, market research, JavaScript та Business Development, зустрічаються

найчастіше. Також візуалізація підкреслює зв'язок між специфічними навичками та професійними ролями (наприклад, навички програмування переважно асоційовані з технічними позиціями, тоді як sales strategy і customer profiling — із посадами в маркетингу).

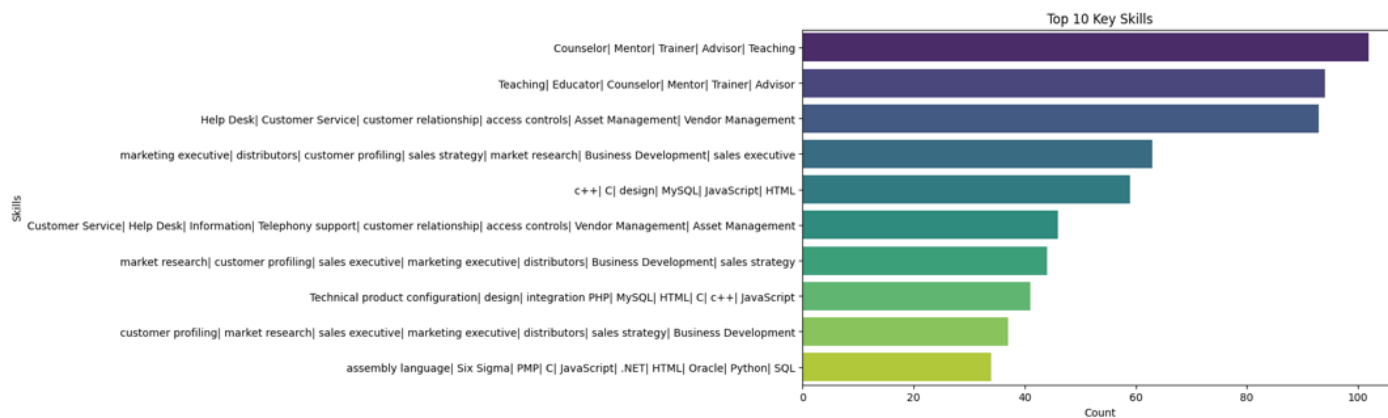


Рис. 3.4. Розподіл ключових навичок

Джерело: розрахунки автора

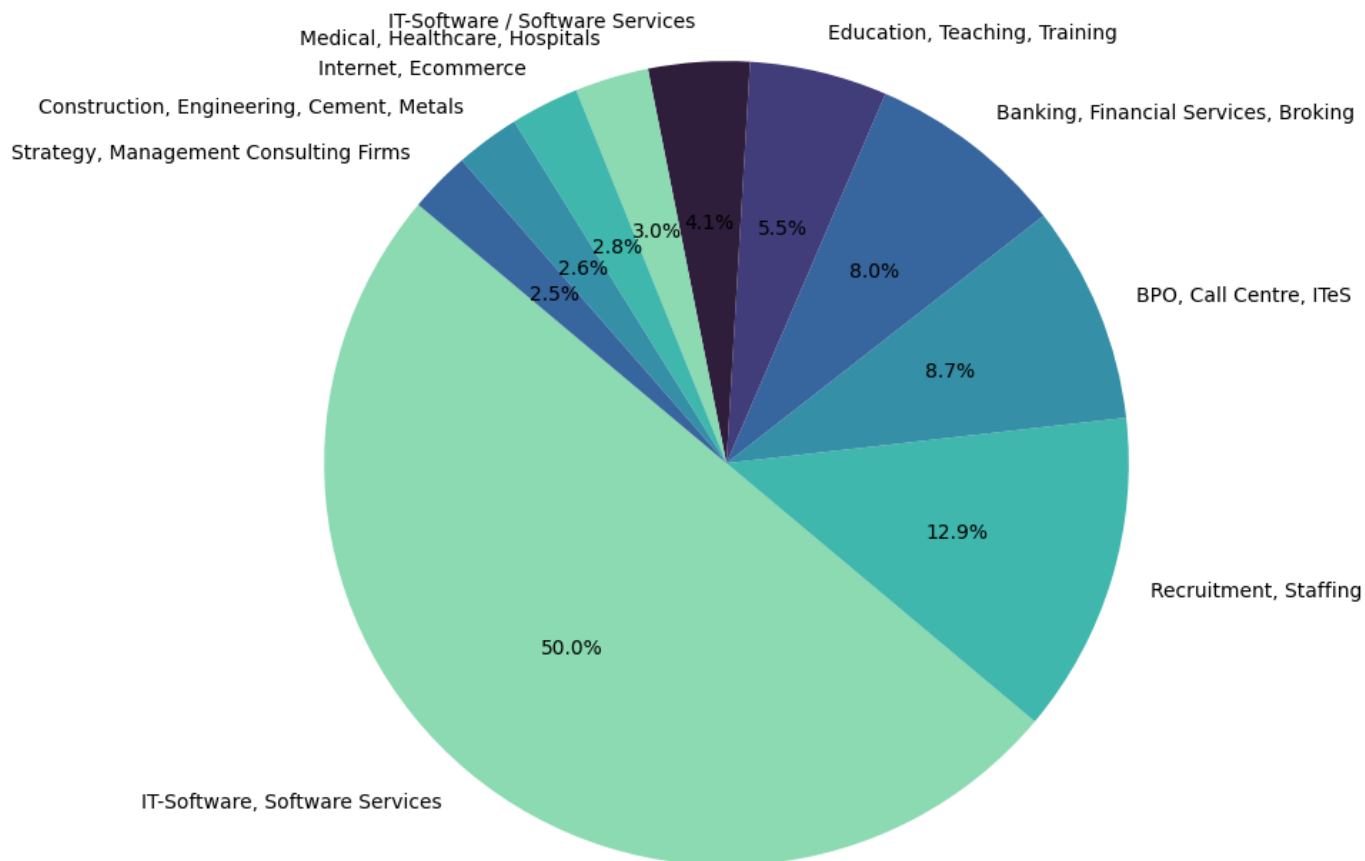


Рис. 3.5. Розподіл кількості вакансії за типом галузі

Джерело: розрахунки автора

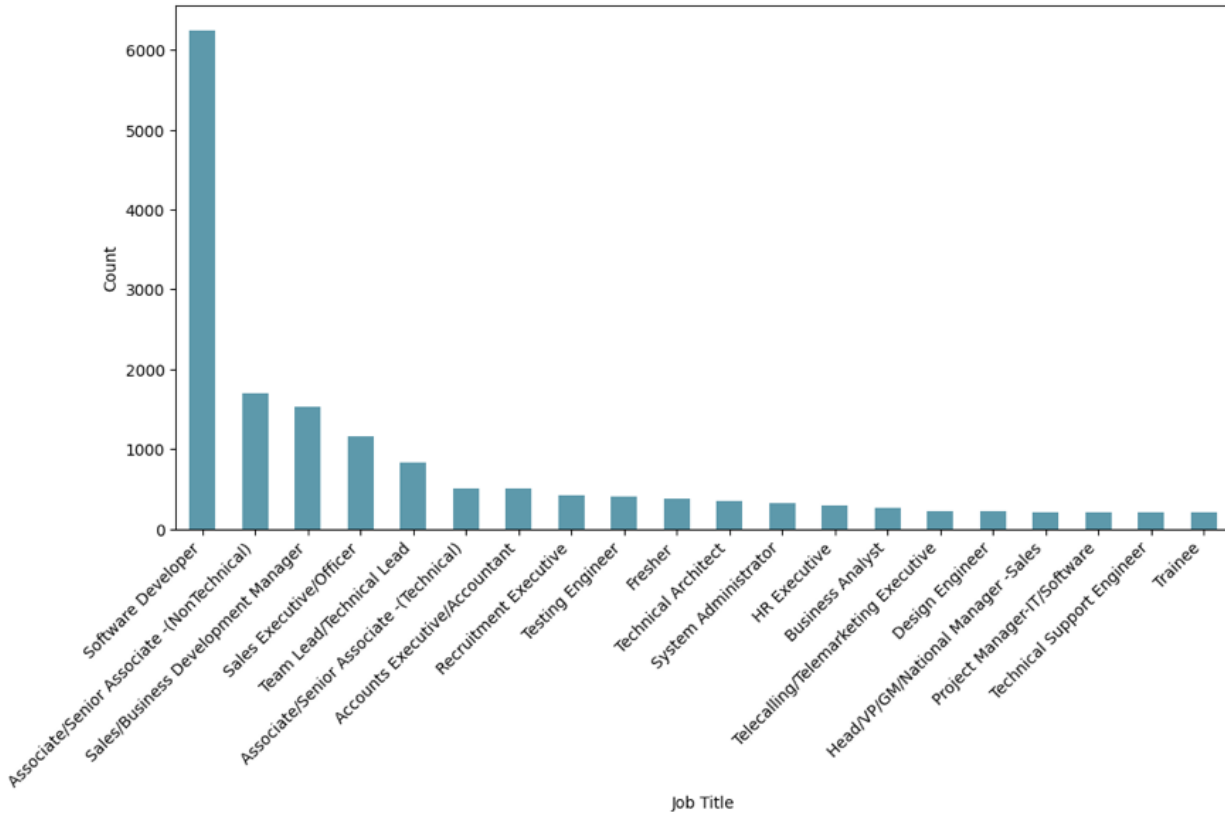


Рис. 3.6. Розподіл посад за частотою

Джерело: розрахунки автора

Найпопулярнішими навичками є JavaScript, HTML, SQL та Python, що підкреслює значення технічних компетенцій у сучасних вакансіях. Окрім того, часто зустрічаються навички, пов'язані з управлінням (Management, Monitoring) та продажами (Sales, Business Executive), що узгоджується з високим попитом на менеджерські та комерційні позиції.

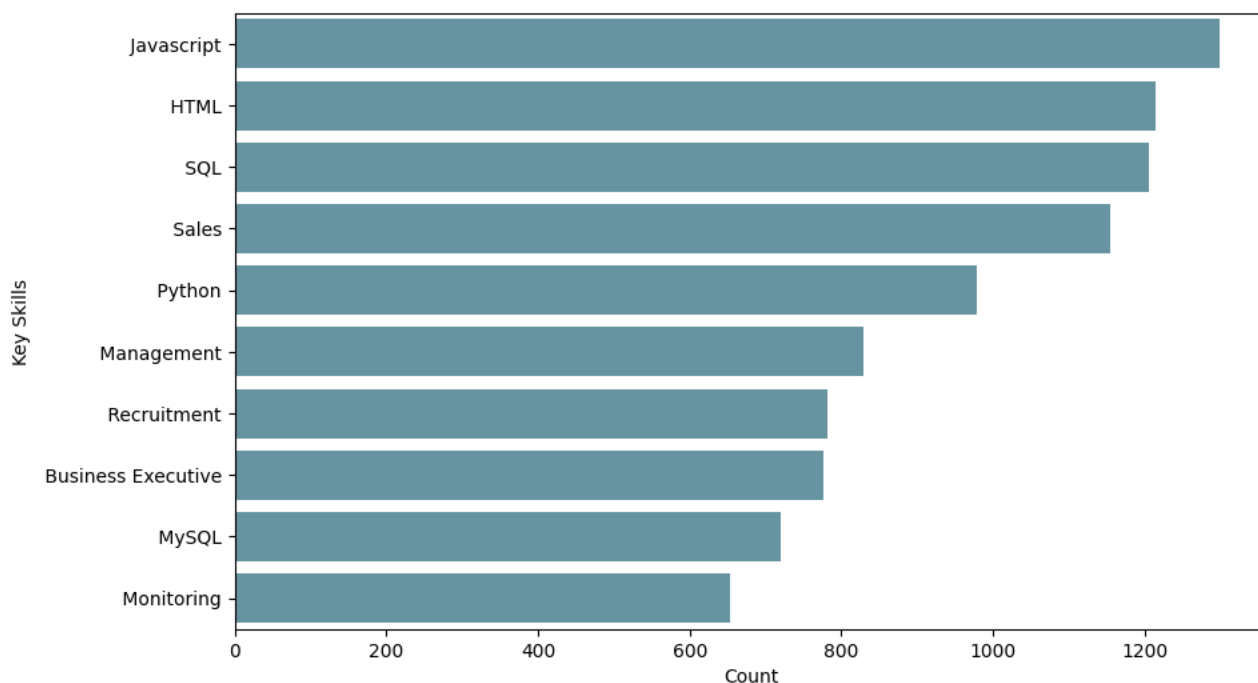


Рис. 3.7. Розподіл кількості ключових навичок у оголошеннях

Джерело: розрахунки автора

Отже, проведений аналіз даних дозволив виявити ключові тенденції сучасного ринку праці на основі досліджуваного набору даних, що містить приклади посад в ІТ-сфері. Найбільш затребуваними галузями виявилися ІТ-сектор, фінансові послуги та продажі, що підтверджується розподілом вакансій та популярністю технічних навичок (JavaScript, SQL, Python) та управлінських компетенцій. Частотний аналіз посад показав значний попит як на технічних фахівців (Software Developer, Technical Support), так і на менеджерські ролі (Sales Executive, HR Executive).

### 3.2. Розробка моделей машинного навчання та аналіз результатів моделювання

Для подальшої побудови рекомендаційної системи одним з кроків є вибір оптимальної моделі класифікації. Застосування моделей класифікації необхідне для того, щоб автоматично визначати найбільш релевантні варіанти на основі наявних даних в межах окремих кластерів, що дозволяє рекомендаційній системі приймати обґрунтовані рішення та забезпечувати персоналізовані рекомендації користувачам.

Зокрема, було побудовано та проведено порівняльний аналіз п'яти різних моделей керованого машинного навчання: логістичну регресію, метод k-найближчих сусідів (KNN), Random Forest, метод опорних векторів (SVM) та Naive Bayes. Кожна з цих моделей навчена на підготовлених даних, а їх ефективність оцінюється за такими метриками якості: accuracy, recall, precision та F1-середнє.

В результаті моделювання було отримано наступні оцінки якості класифікаційних моделей машинного навчання для обробки природної мови:

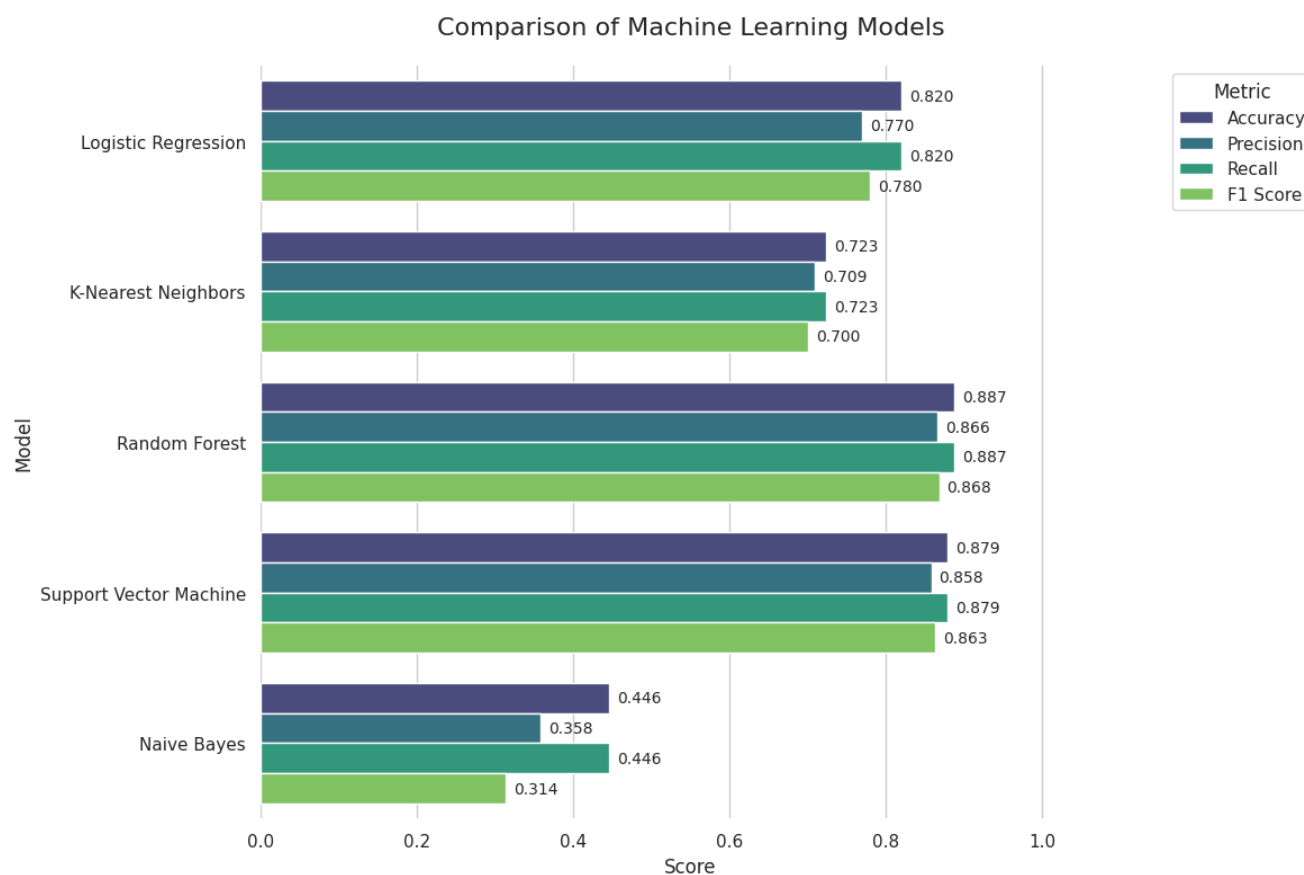


Рис. 3.8. Метрики оцінки якості моделей

Джерело: розрахунки автора

Наведений графік порівняння демонструє ефективність п'яти різних моделей машинного навчання за чотирма оцінковими метриками: Accuracy, Precision, Recall та F1 Score. Ці показники дозволяють комплексно оцінити якість прогнозування професійних позицій на основі комбінації таких ознак, як посада та ключові навички:

- Random Forest - продемонстрував найкращі результати серед усіх моделей за всіма метриками. Досяг найвищої точності (88,7%), а також високих значень Precision, Recall та F1-міри, що робить його найефективнішим алгоритмом для побудови рекомендаційної системи. Його здатність ефективно обробляти нелінійні залежності та уникати перенавчання забезпечує стабільність прогнозів.

- Модель Support Vector Machine (SVM) показала досить високі результати з точністю 88,3% та збалансовані результати за всіма метриками, що свідчить про його гарну придатність для класифікації даних із високою розмірністю. Ефективність методу обумовлена здатністю знаходити оптимальні межі між класами.

- Логістична регресія - продемонструвала високу ефективність з точністю 82,5%. Значення Precision та Recall дещо нижчі порівняно з Random Forest та SVM, проте модель залишається прийнятною для використання. Перевагою моделі є простота інтерпретації результатів, що може бути корисним для аналізу важливих ознак.

- K-Nearest Neighbors (KNN) – метод показав нижчу точність (72,4%) у порівнянні з іншими моделями. Precision та F1-міра також виявилися дещо слабшими, що може бути пов'язано з високим рівнем шуму. Чутливість до масштабування даних та обчислювальна складність роблять його менш привабливим для даної задачі.

- Naive Bayes – метод продемонстрував найнижчі результати серед усіх алгоритмів. Точність склала лише 44,1%, а F1-міра опустилася до 30,7%, що свідчить про низьку якість класифікації. Це пояснюється припущенням про незалежність ознак, яке не виконується в даному наборі даних, де назви посад та навички тісно пов'язані.

Отже, Random Forest виявився найефективнішим алгоритмом для побудови системи рекомендацій завдяки високій точності та стійкості до перенавчання. SVM також показав гарні результати, що робить його додатковим варіантом для подальших досліджень. Naive Bayes продемонстрував незадовільну якість класифікації, що підкреслює важливість вибору моделі, яка відповідає специфіці даних.

Отримані результати моделювання свідчать про те, що для побудови системи рекомендації вакансій на основі текстових ознак найкраще підходять ансамблеві методи

(Random Forest) або SVM, тоді як простіші алгоритми (KNN, Naive Bayes) вимагають додаткового налаштування або вибору інших підходів.

### 3.3. Розробка та імплементація рекомендаційної системи

Рекомендаційна система для підбору вакансій на основі навичок користувача реалізована за допомогою машинного навчання та методів обробки природної мови.

Інтерфейс користувача для взаємодії з рекомендаційною системою був реалізований у хмарному середовищі Google Colaboratory за допомогою інтерактивних елементів IPython. Це включало створення текстових полів для введення запитів, запуску аналізу та візуалізації результатів у вигляді відсортованого списку рекомендацій. Такий підхід дозволив уникнути необхідності розробки окремого веб-інтерфейсу, зберігаючи при цьому всі переваги інтерактивної роботи з системою. Використання вбудованих можливостей Google Colab для відображення результатів (таких як matplotlib для графіків та pandas DataFrame для структурованого виводу) забезпечило зручний спосіб представлення рекомендацій кінцевим користувачам системи.

Основний процес роботи рекомендаційної системи можна описати наведеним у таблиці нижче алгоритмом:

Таблиця 3.2

Побудова рекомендаційної системи

Етап	Опис
Ініціалізація та завантаження даних	<ul style="list-style-type: none"> <li>- Завантажується попередньо підготовлений датасет вакансій;</li> <li>- Завантажуються натреновані моделі (Random Forest, TF-IDF vectorizer);</li> <li>- Ініціалізуються допоміжні структури даних.</li> </ul>
Обробка вхідних даних користувача	<ul style="list-style-type: none"> <li>- Очікування вводу від користувача;</li> <li>- Трансформація введених даних;</li> <li>- Створення вектора ознак для подальшої обробки.</li> </ul>

Прогнозування категорії посади	- Безпосередня класифікація за допомогою моделі Random Forest; - Перетворення числового коду категорії у текстову назву; - Підготовку до подальшого кроку.
Пошук та ранжування вакансій	- Фільтрація вакансій за обраною категорією; - Обчислення міри схожості для кожної вакансії за допомогою коефіцієнта Cosine similarity; - Створення рейтингу вакансій за ступенем відповідності.
Підготовка та візуалізація результатів	- Результати сортуються за спаданням релевантності; - Відбираються лише найбільш відповідні вакансії; - Готуються дані для виводу користувачеві.

Джерело: побудовано автором

Система починає роботу з отримання вхідних даних, які включають назви вакансій (Job Title) та ключові навички (Key Skills), що зберігаються у структурованому вигляді у формі таблиці (DataFrame). На етапі попередньої обробки текстових даних відбувається об'єднання цих двох полів у єдине текстове поле «combined\_features» шляхом конкатенації, що дозволяє враховувати всі текстові характеристики вакансій у подальшому аналізі. Для уникнення проблем з відсутніми даними використовується метод fillna(""), який замінює відсутні значення порожніми рядками.

Далі перейдемо до етапу векторизації тексту за допомогою методу TF-IDF, який перетворює текстові дані у числовий формат, зручний для машинного навчання. Векторизатор налаштований на обробку максимум 7000 найважливіших слів (max\_features=7000) та автоматично видаляє стоп-слова англійської мови (stop\_words='english'), такі як артикли та сполучники, що не несуть значної семантичної навантаження. Результатом цього етапу є розріджена матриця X, де рядки відповідають окремим вакансіям, а стовпці - словам з їхніми ваговими коефіцієнтами TF-IDF, що відображають важливість кожного слова у конкретній вакансії відносно всього корпусу документів.

```

TF-IDF Vector (first and last 5 rows and columns):
      00 000 00pm 01 02249254263  zs zsm zss zte  eädaily
0      0.0 0.0 0.0 0.0          0.0 0.0 0.0 0.0 0.0      0.0
1      0.0 0.0 0.0 0.0          0.0 0.0 0.0 0.0 0.0      0.0
2      0.0 0.0 0.0 0.0          0.0 0.0 0.0 0.0 0.0      0.0
3      0.0 0.0 0.0 0.0          0.0 0.0 0.0 0.0 0.0      0.0
4      0.0 0.0 0.0 0.0          0.0 0.0 0.0 0.0 0.0      0.0
27005 0.0 0.0 0.0 0.0          0.0 0.0 0.0 0.0 0.0      0.0
27006 0.0 0.0 0.0 0.0          0.0 0.0 0.0 0.0 0.0      0.0
27007 0.0 0.0 0.0 0.0          0.0 0.0 0.0 0.0 0.0      0.0
27008 0.0 0.0 0.0 0.0          0.0 0.0 0.0 0.0 0.0      0.0
27009 0.0 0.0 0.0 0.0          0.0 0.0 0.0 0.0 0.0      0.0

```

```
Vector shape: (27010, 7000)
```

Рис. 3.9. Векторизація тексту за допомогою методу TF-IDF

Джерело: розрахунки автора

Паралельно здійснюється підготовка цільової змінної для навчання моделі класифікації. Категорії вакансій, такі як "Data Scientist" чи "Web Developer", перетворюються у числові мітки за допомогою категоріального кодування, що дозволяє алгоритмам машинного навчання ефективно працювати з цими даними. Для безпосередньої класифікації вакансій за категоріями використовується ансамблевий метод Random Forest, який будує множину дерев рішень на основі випадкових підмножин даних та ознак. Кожне дерево в ансамблі навчається незалежно, а підсумковий прогноз формується шляхом голосування всіх дерев, що забезпечує високу точність та стійкість моделі до перенавчання.

Після навчання моделі система здатна класифікувати нові текстові запити, такі як "Python Developer with Django". Запит спочатку перетворюється у TF-IDF вектор за допомогою того ж векторизатора, що використовувався при навчанні, а потім передається на вхід моделі Random Forest для визначення відповідної категорії. Цей етап важливий для подальшого пошуку рекомендацій, оскільки дозволяє значно звужувати простір пошуку шляхом фільтрації вакансій за прогнозованою категорією. На завершальному етапі система використовує метрику косинусної подібності для знаходження найбільш схожих вакансій до вхідного запиту серед відфільтрованого

набору. Косинусна подібність обчислюється між вектором запиту та векторами вакансій цільової категорії, враховуючи кут між векторами у багатовимірному просторі ознак, що дозволяє кількісно оцінити семантичну близькість текстових описів.

Кінцевим результатом роботи системи є ранжований список рекомендацій, які не лише належать до тієї ж категорії, що й вхідний запит, але й мають найвищу міру подібності за текстовими характеристиками. Така архітектура ефективно поєднує методи обробки природної мови та машинного навчання, забезпечуючи точність і релевантність рекомендацій при оптимальній швидкості роботи. Використання Random Forest для попередньої класифікації дозволяє значно зменшити обсяг даних, що підлягають подальшому аналізу подібності, що є особливо важливим при роботі з великими наборами даних. В той же час, застосування косинусної подібності на останньому етапі гарантує, що рекомендовані вакансії будуть не лише формально належати до потрібної категорії, але й максимально відповідатимуть конкретному текстовому запиту за своїм змістом і ключовими характеристиками.

Отже, в результаті моделювання гібридної рекомендаційної системи з використанням методу машинного навчання та обробки природної мови було отримано ефективний інструмент для автоматизованого підбору вакансій, який демонструє високу точність класифікації (88,7% для моделі Random Forest) та здатність генерувати персоналізовані рекомендації на основі аналізу ключових навичок користувача. Система успішно інтегрує TF-IDF для векторизації текстових даних, алгоритм Random Forest для категоризації посад та косинусну подібність для ранжування результатів, що дозволяє не лише точно визначати відповідні категорії вакансій, але й оцінювати ступінь їх відповідності введеним навичкам.

Enter your skills:

Рис. 3.10. Приклад введення даних користувачем

Джерело: розрахунки автора

Enter your skills: sql

You may look into DBA jobs

Here are the most relevant jobs in this category:

Found 105 matching jobs. Top 5 most relevant:

Job Title	Key Skills	Job Salary	Job Experience Required	Role Category	Functional Area	Industry	cosine_similarity	
2311	DBA	SQL Server Administration SQL Data...	2,00,000 - 3,00,000 PA.	3 - 4 yrs	Admin/Maintenance/Security/Datawareh...	IT Software - DBA, Datawarehousing	IT-Software, Software Services	0.660598
12412	DBA	SQL Server ms sql server sql serve...	4,00,000 - 7,00,000 PA.	3 - 8 yrs	Admin/Maintenance/Security/Datawareh...	IT Software - Application Programmin...	IT-Software, Software Services	0.542007
13364	DBA	MS SQLI Db2 L2I SQL DBA	Not Disclosed by Recruiter	4 - 8 yrs	Admin/Maintenance/Security/Datawareh...	IT Software - DBA, Datawarehousing	IT-Software, Software Services	0.477529
11843	DBA	database design t sql cursors l tu...	Not Disclosed by Recruiter	2 - 5 yrs	Admin/Maintenance/Security/Datawareh...	IT Software - Application Programmin...	IT-Software, Software Services	0.465112
8618	DBA	DBAI SQL Server dbal SQL	Not Disclosed by Recruiter	2 - 5 yrs	Admin/Maintenance/Security/Datawareh...	IT Software - DBA, Datawarehousing	IT-Software, Software Services	0.464855

Рис. 3.11. Результати роботи рекомендаційної системи

Джерело: розрахунки автора

```
Salary Statistics for these positions:
count                105
unique                14
top      Not Disclosed by Recruiter
freq                  83
Name: Job Salary, dtype: object
```

```
Experience Requirements:
Job Experience Required
4 - 9 yrs                9
5 - 8 yrs                8
2 - 5 yrs                7
3 - 6 yrs                6
5 - 10 yrs              6
2 - 4 yrs                5
3 - 4 yrs                4
4 - 8 yrs                4
2 - 6 yrs                4
3 - 5 yrs                3
3 - 8 yrs                3
6 - 10 yrs              3
10 - 20 yrs             3
```

Рис. 3.12. Вивід додаткової інформації щодо рівня заробітної плати та досвіду

Джерело: розрахунки автора.

Реалізована система демонструє ефективність у підборі вакансій на основі навичок користувача. Використання Random Forest та TF-IDF забезпечує достатню точність, а сортування за косинусною подібністю дозволяє персоналізувати результати. Подальша оптимізація може покращити якість рекомендацій за рахунок сучасних NLP-підходів.

### **Висновки до розділу 3**

В даному розділі кваліфікаційної роботи було проаналізовано обрану базу даних та побудовано такі моделі класифікації, як: Logistic Regression, KNN, Random Forest, SVM, Naive Bayes. Для оцінки якості моделей використано метрики accuracy, precision, recall, F1-score. Порівняльний аналіз дозволив визначити найефективніший алгоритм для побудови рекомендаційної системи, а також з'ясувати, які підходи найкраще підходять для обробки даних із заданими характеристиками (наприклад, нерівномірний розподіл класів або наявність шуму).

Головним результатом моделювання стало створення функціональної гібридної рекомендаційної системи, яка здатна ефективно аналізувати вхідні навички користувача та надавати персоналізовані рекомендації щодо найбільш релевантних вакансій.

Зокрема, побудована рекомендаційна модель має наступні характеристики:

- Висока точність класифікації – модель Random Forest продемонструвала точність 88,7%, що робить її оптимальним вибором для даної задачі.

- Гнучкість обробки даних – використання TF-IDF для векторизації текстових навичок дозволило ефективно порівнювати різні набори даних.

- Релевантність рекомендацій – застосування косинусної подібності забезпечило точне ранжування вакансій за ступенем відповідності навичкам користувача.

- Додаткові аналітичні можливості – система надає корисну інформацію про середні зарплати, вимоги до досвіду та інші параметри, що допомагає користувачам приймати обґрунтовані рішення.

## ВИСНОВОК

Сучасні технології ШІ та машинного навчання відіграють ключову роль у розвитку інтелектуальних рекомендаційних систем, які трансформують спосіб взаємодії користувачів з інформацією, товарами та послугами. Такі системи широко застосовуються у сфері електронної комерції, соціальних мереж, а також у HR-технологіях для автоматичного підбору вакансій. Вони аналізують великі обсяги даних, виявляють закономірності в поведінці користувачів і на основі цих даних генерують персоналізовані рекомендації.

У даній дипломній роботі було розроблено гібридну рекомендаційну систему для підбору вакансій на основі аналізу професійних навичок користувача. Проведене дослідження охопило всі етапи створення системи - від аналізу предметної області та збору даних до реалізації алгоритмів машинного навчання та оцінки ефективності роботи системи.

У ході дослідження було виконано всі поставлені завдання, що дозволило розробити ефективну рекомендаційну систему для рекрутингу на основі методів ШІ та машинного навчання:

1. Аналіз сучасних підходів до побудови рекомендаційних систем у рекрутингу. Досліджено основні методики, що використовуються в сучасних рекомендаційних системах для підбору персоналу, зокрема колаборативну фільтрацію, контентно-орієнтовані підходи та гібридні моделі. Встановлено, що найбільш перспективним є гібридний підхід, який поєднує переваги різних методів для підвищення точності та релевантності рекомендацій.

2. Дослідження ефективності алгоритмів машинного навчання для класифікації. Проведено порівняльний аналіз різних алгоритмів машинного навчання (Logistic Regression, KNN, Random Forest, SVM, Naive Bayes) для класифікації вакансій на основі навичок користувачів. Найкращі результати продемонструвала модель Random Forest з точністю 88,7%, що підтверджує її ефективність у задачах обробки текстових даних.

3. Розробка архітектури гібридної рекомендаційної системи. Запропоновано архітектуру системи, яка інтегрує:

- метод TF-IDF для векторизації текстових даних (ключових навичок),
- модель Random Forest для класифікації вакансій,
- косинусну подібність для ранжування результатів.

Така комбінація дозволила забезпечити точний підбор вакансій з урахуванням індивідуальних характеристик користувача.

4. Реалізація та оцінка точності рекомендаційної системи. Розроблено систему, що демонструє високу якість персоналізованих рекомендацій. Оцінка ефективності показала, що система досить точно класифікує вакансії та надає релевантні результати, що підтверджує її практичну придатність для використання в HR-процесах.

Отримані результати свідчать про перспективність подальших досліджень у напрямку вдосконалення рекомендаційних систем з використанням сучасних методів ШІ та обробки природної мови. Робота вносить внесок у розвиток прикладних рішень у галузі HR-технологій та автоматизації підбору персоналу.

## СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Montesa M. AI Recruiting in 2025: The Definitive Guide. Phenom. 2025. URL: <https://www.phenom.com/blog/recruiting-ai-guide>
2. Artificial Intelligence in HR Global Market Report 2025. The Business Research Company. 2025. URL: <https://www.thebusinessresearchcompany.com/report/artificial-intelligence-in-hr-global-market-report>
3. McGill J. AI in Recruitment — Statistics & Trends. ContentDetector.AI. 2024. URL: <https://contentdetector.ai/articles/ai-in-recruitment-statistics-trends/>
4. Techlabs M. Types of Recommendation Systems & Their Use Cases. Medium. 2021. URL: <https://marutitech.medium.com/what-are-the-types-of-recommendation-systems-3487cbafa7c9>
5. Chang H., Shi T. Research and Design of a Job Search Service Platform Based on Recommendation Algorithm. The Frontiers of Society, Science and Technology. 2023. Vol. 5. No 14. PP. 78-83. URL: <https://francispress.com/uploads/papers/EUiK5Mm7YgiD8KLW7CMsMNPZ8rEHc41T9MqDcnTy.pdf>
6. Саджид Х. Системи рекомендацій, що використовують LLM та векторні бази даних. Unite.AI. 2024. URL: <https://www.unite.ai/uk/%D1%81%D0%B8%D1%81%D1%82%D0%B5%D0%BC%D0%B8-%D1%80%D0%B5%D0%BA%D0%BE%D0%BC%D0%B5%D0%BD%D0%B4%D0%B0%D1%86%D1%96%D0%B9%2C-%D1%89%D0%BE-%D0%B2%D0%B8%D0%BA%D0%BE%D1%80%D0%B8%D1%81%D1%82%D0%BE%D0%B2%D1%83%D1%8E%D1%82%D1%8C-llms-%D1%82%D0%B0-%D0%B2%D0%B5%D0%BA%D1%82%D0%BE%D1%80%D0%BD%D1%96-%D0%B1%D0%B0%D0%B7%D0%B8-%D0%B4%D0%B0%D0%BD%D0%B8%D1%85/>
7. Akash Takyar. AI in talent acquisition: Use cases, development and benefits. 2025. URL: <https://www.leewayhertz.com/ai-in-talent-acquisition/>

8. AI For Recruiting - A Complete Guide. 2025. URL: <https://cvviz.com/ai-recruiting/>
9. Machine Learning In Recruitment & How To Do It Right. 2024. URL: <https://harver.com/blog/machine-learning-in-recruitment/>
10. How to Use AI to Recruit Faster in 2024? 2024. URL: <https://www.ditrc.com/how-to-use-ai-to-recruit-faster-in-2024/>
11. Bilin Sh., Xiaojun L., Genqing B. A survey of research hotspots and frontier trends of recommendation systems from the perspective of knowledge graph. *Expert Systems with Applications*. Vol. 165. 2021. URL: <https://doi.org/10.1016/j.eswa.2020.113764>
12. Drishti De. Personalized Job Recommendation System using Collaborative Filtering. *Medium*. 2021. URL: <https://de-drishti.medium.com/personalized-job-recommendation-system-using-collaborative-filtering-6d663046692a>
13. Скібська К., Панасюк О. (2023). Рекрутинг як складова цифрової стратегії менеджменту персоналу. *Herald of Khmelnytskyi National University. Economic Sciences*. 2023. № 316(2). <https://doi.org/10.31891/2307-5740-2023-316-2-1>
14. Євдокимов В.В., Морозов А. В., Вакалюк Т. А. формалізовані методи та алгоритми для створення рекомендаційних систем. актуальні проблеми автоматизації та інформаційних технологій. 2024. С. 120-133. URL: <http://dx.doi.org/10.15421/432411>
15. Чередніченко О.Ю., Янголенко О.В, Іващенко О.В, Матвеев О.М. Моделі формування рекомендацій у інтелектуальних системах електронної комерції. *Системи обробки інформації*. 2020. № 1 (160). С. 32-39. doi: 10.30748/soi.2020.160.04.
16. Widodo R. I. H., Herdiyanto R. F., Thoriq M., Rahman I. S. Job Recommendation System Combining Collaborative Filtering and Content Based Filtering. 2024. URL: <https://doi.org/10.20944/preprints202407.1700.v1>
17. Charan S. N., Suhas G. K., Yathisha L., Devananda S. N. Job Recommendation System: Content-Based and Collaborative Filtering for Predictive Job Recommendation Systems. *Journal of Information Systems Engineering and Management*. 2025. URL: <https://doi.org/10.52783/jisem.v10i2.2302>

18. Лобур М. В., Шварц М. Є., Стех Ю. В. Моделі і методи прогнозування рекомендацій для колаборативних рекомендаційних систем. 2018. URL: <https://science.lpnu.ua/sites/default/files/journal-paper/2019/feb/15581/181912maket-68-75.pdf>
19. Mulay A., Sutar S., Patel J., Chhabria A., Mumbaikar S. Job Recommendation System Using Hybrid Filtering. ITM Web of Conferences. 2022. P. 44. URL: [https://www.researchgate.net/publication/360418721\\_Job\\_Recommendation\\_System\\_Using\\_Hybrid\\_Filtering](https://www.researchgate.net/publication/360418721_Job_Recommendation_System_Using_Hybrid_Filtering)
20. Dascălu M-I., Hang A., Puskás I-F., Bodea C-N. CareProfSys: a job recommender system based on machine learning and ontology to support learners' employability at regional level. Issues in Information Systems. 2023. Vol. 24. PP. 71-82. DOI: [https://doi.org/10.48009/3\\_iis\\_2023\\_107](https://doi.org/10.48009/3_iis_2023_107)
21. Hu Y., Koren Y., Volinsky C. Collaborative Filtering for Implicit Feedback Datasets. AT&T Labs – Research. 2025. URL: <http://yifanhu.net/PUB/cf.pdf>
22. Білорус Т.В. Управління персоналом: 800+ запитань та відповідей. Навч. посіб. Київ: Видавництво Ліра-К. 2021. 320 с. (18,6 д.а) (навчальний посібник).
23. Білорус Т.В. Методичний інструментарій організації пошуку і відбору персоналу. Вісник Київського національного університету імені Тараса Шевченка. Серія “Економіка”. 2015. № 7(172). С. 20-29.
24. Chornous G., Lem T. Developing hybrid recommendation systems: Ukrainian dimension. Access to science, business, innovation in digital economy, ACCESS Press. 2022. Vol. 3(2). P. 89-106.
25. Chornous G., Nikolskyi I., Wyszynski M., Kharlamova G., Stolarczyk P. A hybrid user-item-based collaborative filtering model for e-commerce recommendations. Journal of International Studies. 2021. Vol. 14(4). P. 157-173. DOI:10.14254/2071-8330.2021/14-4/11
26. Chornous G., Sviatiuk Yu., Usenko D. Hybridization of Recommender Systems in E-Commerce Based on Demographic Filtering. Contemporary technologies and society: innovations, artificial intelligence, and challenges. Scientific edition: V. 2023. P. 351-357.

27. Recommendation System. NVIDIA. 2025. URL: <https://www.nvidia.com/en-us/glossary/recommendation-system/>
28. Pham K. What are Recommendation Systems? 2022. URL: <https://medium.com/@khang.pham.exxact/what-are-recommendation-systems-6bb5036042db>
29. What are Recommender Systems? 2024. URL: <https://www.geeksforgeeks.org/what-are-recommender-systems/>
30. Caballar R. D. What is a recommendation engine? 2024. URL: <https://www.ibm.com/think/topics/recommendation-engine>
31. Dee C. What role does AI play in recommendation systems and engines? 2024. URL: <https://www.algolia.com/blog/ai/what-role-does-ai-play-in-recommendation-systems-and-engines>
32. Jonna V. Types of Recommendation Systems. 2024. URL: <https://ellow.io/types-of-recommendation-systems/>
33. Murel J. What is collaborative filtering? 2024. URL: <https://www.ibm.com/think/topics/collaborative-filtering>
34. Deutschman Z. Recommender Systems: Machine Learning Metrics and Business Metrics. 2023. URL: <https://neptune.ai/blog/recommender-systems-metrics>
35. Mgarbi H., Chkouri M.Y., Tahiri A. Building a recommendation system based on the job offers extracted from the web and the skills of job seekers. International Journal of Electrical and Computer Engineering (IJECE). 2023. URL: <https://www.semanticscholar.org/paper/Building-a-recommendation-system-based-on-the-job-Mgarbi-Chkouri/a433d1730756c173418264c0186af20ca9f385c#extracted>
36. What is AI (artificial intelligence)? 2025. URL: <https://www.mckinsey.com/featured-insights/mckinsey-explainers/what-is-ai>
37. Supervised Learning: What It Is and How It Works. Grammarly. 2024. URL: <https://www.grammarly.com/blog/ai/what-is-supervised-learning/>

38. Rupali S., Rupam M. Enhancing Job Recommendation Systems through Machine Learning: A Comprehensive Analysis of Skill Sync Job Recommendation System. IJSRET 741 International Journal of Scientific Research & Engineering Trends. 2024. Vol. 10. URL: [https://ijsret.com/wp-content/uploads/2024/05/IJSRET\\_V10\\_issue3\\_219.pdf](https://ijsret.com/wp-content/uploads/2024/05/IJSRET_V10_issue3_219.pdf)
39. What Is AI/ML? Are You Using It Correctly? Boomi. 2025. URL: <https://boomi.com/blog/what-is-ai-ml/>
40. Talebi M. The difference between Artificial Intelligence (AI), Machine Learning (ML), Deep Learning (DL). 2023. URL: [https://www.linkedin.com/posts/majid-talebi-7b007571\\_the-difference-between-artificial-intelligence-activity-7017970572668063744-ka\\_I?trk=public\\_profile\\_like\\_view](https://www.linkedin.com/posts/majid-talebi-7b007571_the-difference-between-artificial-intelligence-activity-7017970572668063744-ka_I?trk=public_profile_like_view)
41. Kim, T., Lim, J. Developing an Intelligent Recommendation System for Non-Information and Communications Technology Major University Students. Appl. Sci. 2023. URL: <https://doi.org/10.3390/app132312774>
42. 9 Machine Learning Algorithms For Recommendation Engines. 2024. URL: <https://insights.daffodilsw.com/blog/machine-learning-algorithms-for-recommendation-engines>
43. McDonald A. Will AI make our information intelligent? 2019. URL: <https://intelligent-information.blog/en/will-ai-make-our-information-intelligent/>
44. 9 Real-World Problems that can be Solved by Machine Learning. Maruti Techlabs. 2024. URL: [https://marutitech.com/problems-solved-machine-learning/?utm\\_source=medium&utm\\_medium=content\\_promotion&utm\\_campaign=Types\\_Of\\_Recommendation\\_Systems](https://marutitech.com/problems-solved-machine-learning/?utm_source=medium&utm_medium=content_promotion&utm_campaign=Types_Of_Recommendation_Systems)
45. Jang D.H. SVM for Recommendation Systems: Unveiling the Power of Support Vector Machines in Personalized Recommendations. Medium. 2024. URL: <https://medium.com/@jangdaehan1/svm-for-recommendation-systems-unveiling-the-power-of-support-vector-machines-in-personalized-569e8a75dcec>
46. Supervised and Unsupervised learning. GeeksforGeeks. 2025. URL: <https://www.geeksforgeeks.org/supervised-unsupervised-learning/>

47. Portugal I., Alencar P., Cowan D. The Use of Machine Learning Algorithms in Recommender Systems: A Systematic Review. Expert Systems with Applications. 2015. URL: [10.1016/j.eswa.2017.12.020](https://doi.org/10.1016/j.eswa.2017.12.020)
48. Moez A. Supervised Machine Learning. DataCamp. 2022. <https://www.datacamp.com/blog/supervised-machine-learning>
49. Belcic I., Stryker C. What is supervised learning? IBM. 2024. URL: <https://www.ibm.com/think/topics/supervised-learning>
50. Supervised Machine Learning. GeeksforGeeks. 2025. URL: <https://www.geeksforgeeks.org/supervised-machine-learning/>
51. Predicting Job Titles from Resumes. Kaggle. 2022. URL: <https://www.kaggle.com/datasets/thedevastator/predicting-job-titles-from-resumes/data>

## ДОДАТКИ

### Додаток А

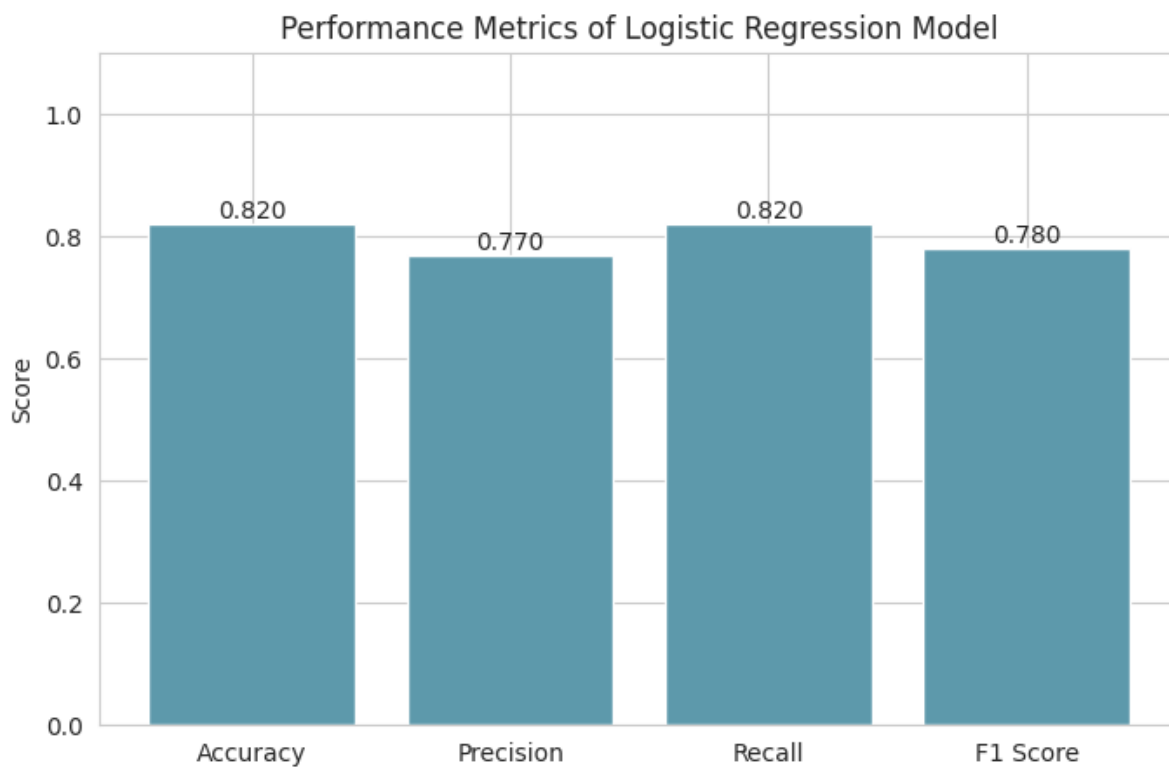


Рис. 3.8. Метрики оцінки якості моделі логістичної регресії

Джерело: розрахунки автора

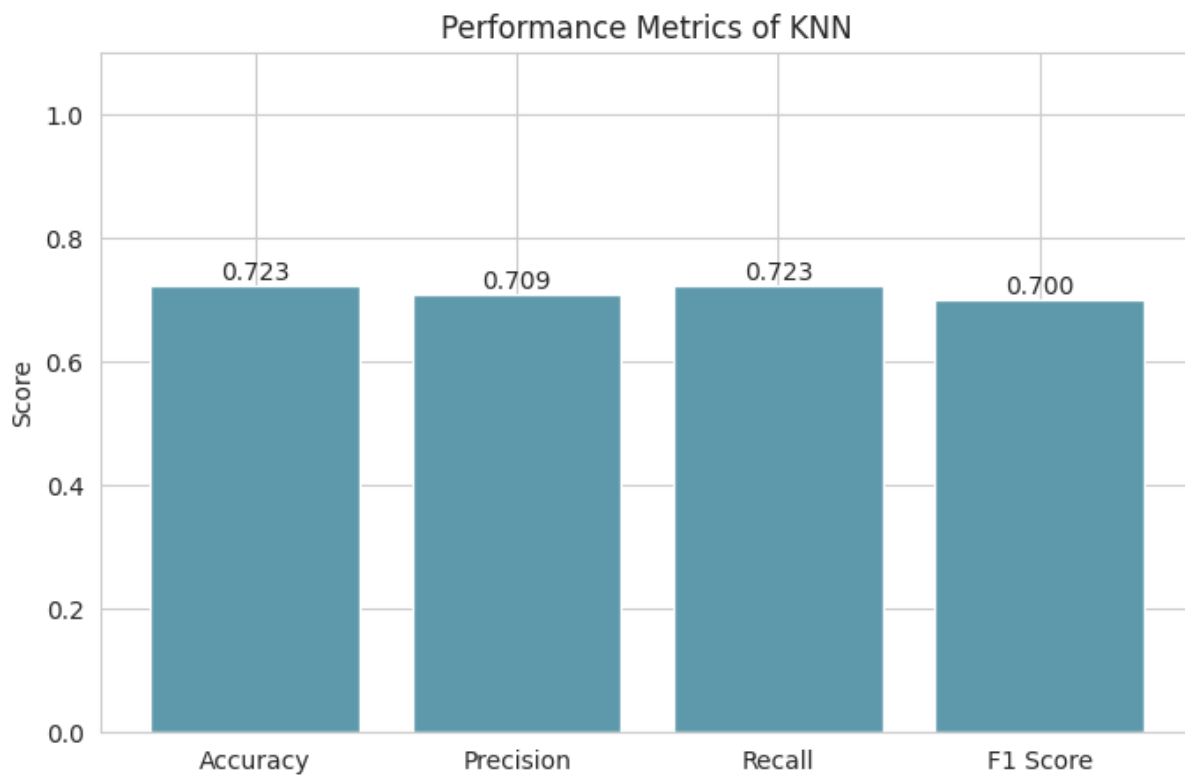


Рис. 3.9. Метрики оцінки якості моделі KNN

Джерело: розрахунки автора

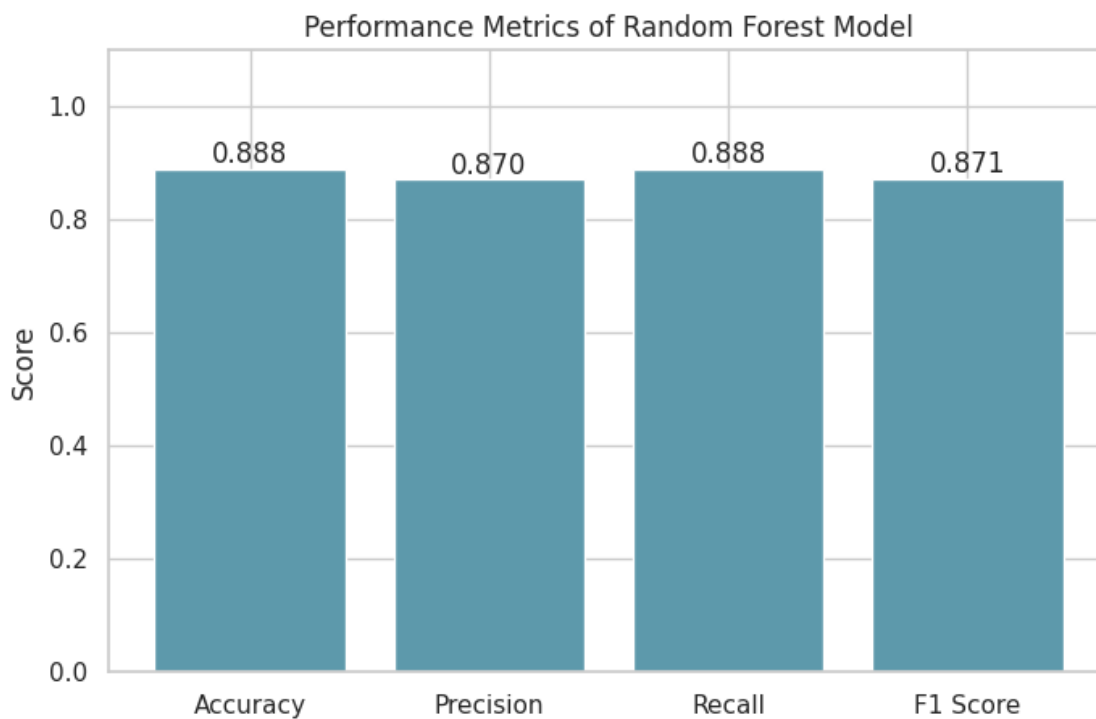


Рис. 3.9. Метрики оцінки якості моделі Random Forest

Джерело: розрахунки автора

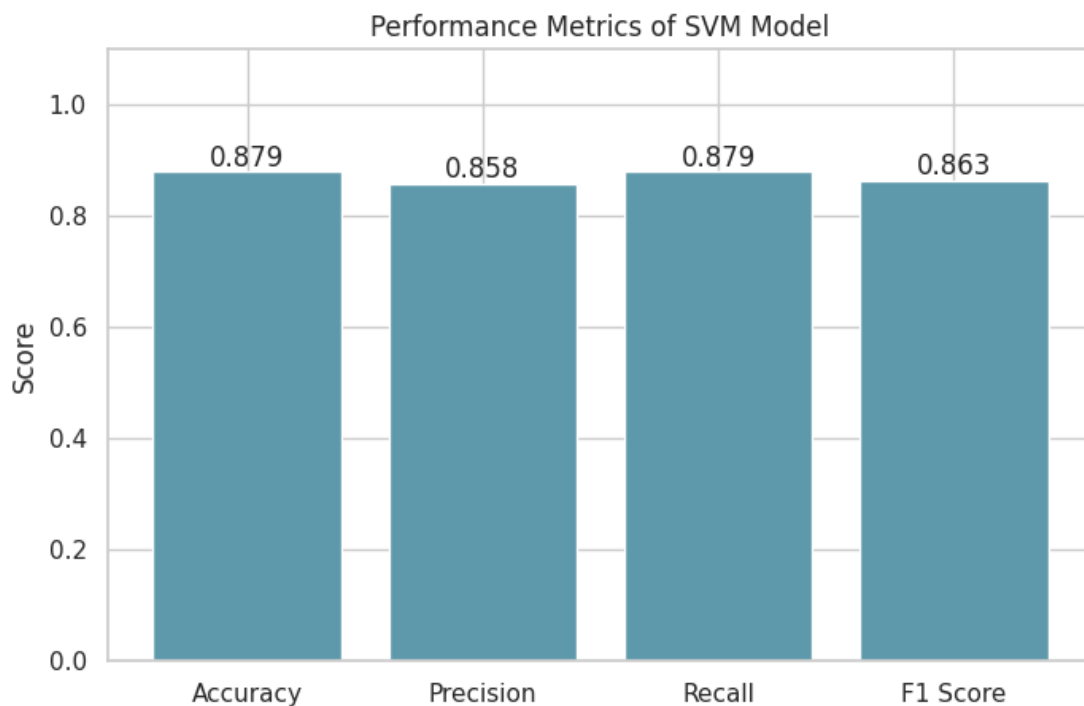


Рис. 3.9. Метрики оцінки якості моделі SVM

Джерело: розрахунки автора

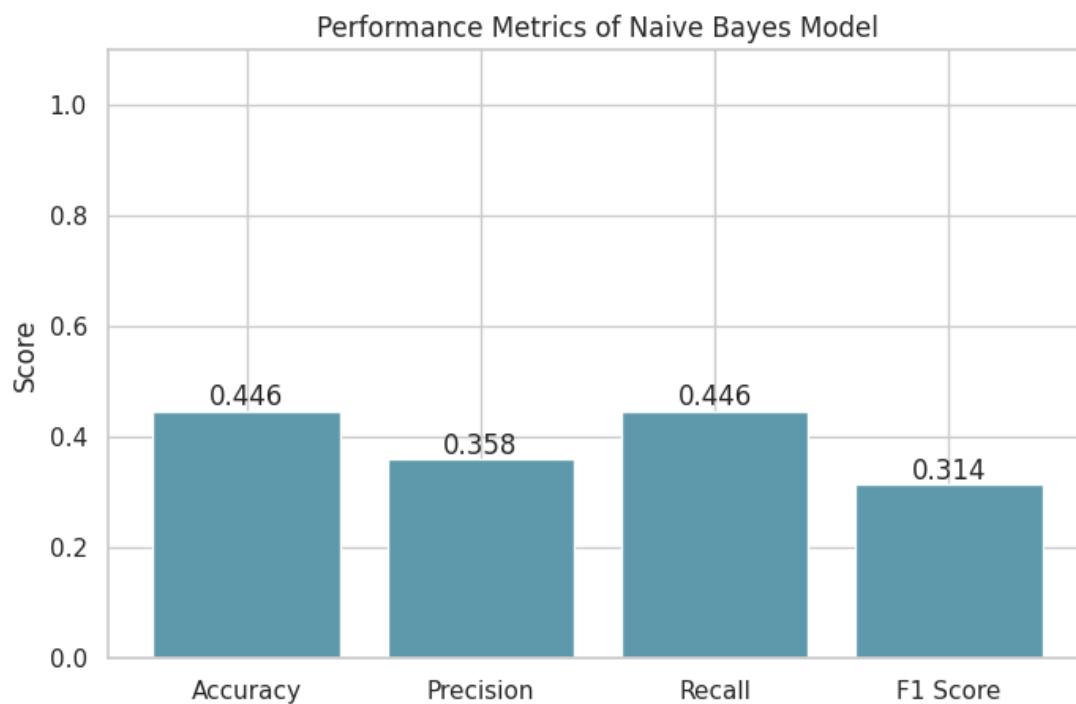


Рис. 3.9. Метрики оцінки якості моделі Naive Bayes

Джерело: розрахунки автора

## Додаток Б

## Програмний код системи

```
# Import necessary libraries
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.metrics import accuracy_score, classification_report
import matplotlib.pyplot as plt
import seaborn as sns

# Load the dataset
jobs_df = pd.read_csv('jobs.csv')

# Приклад перетворення категоріальних даних (Label Encoding)
jobs_df_encoded = jobs_df.drop(columns=['Unnamed: 0']).apply(lambda x: x.factorize()[0] if
x.dtype == 'object' else x)

corr_matrix = jobs_df_encoded.corr(method='spearman') # метод Спірмена для нелінійних
зв'язків

plt.figure(figsize=(10, 8))
sns.heatmap(corr_matrix, annot=True, cmap='mako_r', fmt=".2f")
plt.title("")
plt.show()

# Combine 'Job Title' and 'Key Skills' for use in the model
jobs_df['combined_features'] = jobs_df['Job Title'] + " " + jobs_df['Key Skills']
jobs_df['combined_features'].fillna("", inplace=True)
```

```
# Initialize the TF-IDF Vectorizer
tfidf = TfidfVectorizer()
tfidf

# Fit and transform the combined features into vectors
X = tfidf.fit_transform(jobs_df['combined_features'])

jobs_df['Job Title'] = jobs_df['Job Title'].astype('category')
y = jobs_df['Job Title'].cat.codes # Convert job titles to numeric codes

# Split the data into training and test sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import accuracy_score, confusion_matrix, precision_score, recall_score,
f1_score, classification_report

# Initialize the Random Forest model
rf = RandomForestClassifier()

# Train the model
rf.fit(X_train, y_train)

# Predict on the test set
y_pred_rf = rf.predict(X_test)

# Evaluate the model
accuracy_rf = accuracy_score(y_test, y_pred_rf)
```

```
conf_matrix_rf = confusion_matrix(y_test, y_pred_rf)
precision_rf = precision_score(y_test, y_pred_rf, average='weighted') # 'weighted' for
multiclass
recall_rf = recall_score(y_test, y_pred_rf, average='weighted')
f1_rf = f1_score(y_test, y_pred_rf, average='weighted')

userInput = input("Enter your skills: ")
pred = tfidf.transform([userInput.lower()])

# Predict using Random Forest and get the job title
output_code = rf.predict(pred)[0]
job_title = jobs_df['Job Title'].cat.categories[output_code]

print(f"\nYou may look into {job_title} jobs")
print(f"Here are the most relevant jobs in this category:\n")

cos = []
# Create a copy of the filtered data to avoid SettingWithCopyWarning
labelData = jobs_df[jobs_df['Job Title'] == job_title].copy()

for index, row in labelData.iterrows():
    skills = [row['Key Skills']]

    skills = [" if pd.isna(skill) else skill for skill in skills]

    skillVec = tfidf.transform(skills)
    cos_lib = cosine_similarity(skillVec, pred)
    cos.append(cos_lib[0][0])
```

```
labelData.loc[:, 'cosine_similarity'] = cos
```

```
all_possible_columns = [
```

```
    'Job Title',
```

```
    'Key Skills',
```

```
    'Job Salary',
```

```
    'Job Experience Required',
```

```
    'Role Category',
```

```
    'Functional Area',
```

```
    'Industry',
```

```
    'cosine_similarity'
```

```
]
```

```
available_columns = [col for col in all_possible_columns if col in labelData.columns]
```

```
essential_columns = ['Job Title', 'Key Skills', 'cosine_similarity']
```

```
for col in essential_columns:
```

```
    if col not in available_columns:
```

```
        available_columns.insert(0, col)
```

```
top_5 = labelData.sort_values("cosine_similarity", ascending=False)[available_columns]
```

```
pd.set_option('display.max_columns', None)
```

```
pd.set_option('display.width', 1000)
```

```
pd.set_option('display.max_colwidth', 40)
```

```
print(f"Found {len(top_5)} matching jobs. Top 5 most relevant:")
```

```
display(top_5.head())
```

```
# Additional summary statistics
```

```
if 'Job Salary' in top_5.columns:
```

```
    print("\nSalary Statistics for these positions:")
```

```
    print(top_5['Job Salary'].describe())
```

```
if 'Job Experience Required' in top_5.columns:
```

```
    print("\nExperience Requirements:")
```

```
    print(top_5['Job Experience Required'].value_counts())
```

**Київський національний університет імені Тараса Шевченка**

**Економічний факультет**  
**Кафедра економічної кібернетики**

**ЗАВДАННЯ**

**на кваліфікаційну роботу магістра**

студентки 2 курсу магістратури спеціальності 051 «Економіка», ОНП

«Економічна кібернетика»

Кожановської Ольги Сергіївни

1. Тема роботи: «Розробка системи рекомендацій у рекрутингу на основі методів штучного інтелекту».
2. Термін завершення роботи: 16 травня 2025 року.
3. Об'єкт дослідження: рекомендаційна система в сфері рекрутингу.
4. Предмет дослідження: методи та алгоритми машинного навчання, що використовуються для побудови ефективної рекомендаційної системи.
5. Мета дослідження: моделювання та розробка ефективної рекомендаційної системи для підбору вакансій на основі аналізу навичок користувача з використанням методів машинного навчання та обробки природної мови.
6. Завдання дослідження:
  - 6.1. дослідження впливу цифровізації та впровадження технологій ШІ на ринку праці та в рекрутингу;
  - 6.2. аналіз сучасних підходів до побудови рекомендаційних систем у процесах рекрутингу;
  - 6.3. аналіз наукових і практичних підходів до створення рекомендаційних систем з впровадженням машинного навчання;

- 6.4. дослідження теоретико-методологічних аспектів моделювання рекомендаційних систем;
- 6.5. дослідження ефективності різних алгоритмів машинного навчання для класифікації;
- 6.6. розробка архітектури гібридної системи, що поєднує різні методи та інструменти машинного навчання;
- 6.7. реалізація рекомендаційної системи, оцінка точності та релевантності рекомендацій.

Науковий керівник: д.е.н., проф. Черноус Галина Олександрівна

Студент: Кожановська Ольга Сергіївна \_\_\_\_\_

(підпис)

Затверджено на засіданні кафедри економічної кібернетики  
протокол № \_\_ від \_\_ 2025 р.

### Календарний план виконання кваліфікаційної роботи магістра

№	Етапи роботи	Терміни виконання	Відмітка керівника про виконання
1	Вибір теми кваліфікаційної роботи магістра.	01.10.2024- 15.11.2024	
2	Розробка та затвердження завдання кваліфікаційної роботи магістра.	16.11.2024- 25.12.2024	
3	Ознайомлення з науково-інформаційними джерелами за обраним напрямом дослідження, виявлення наукової проблематики.	26.12.2024- 15.01.2025	
4	Підготовка вступу та I розділу кваліфікаційної роботи.	16.01.2025- 16.02.2025	
5	Підготовка II розділу кваліфікаційної роботи.	17.02.2025- 10.03.2025	
6	Підготовка III розділу кваліфікаційної роботи.	11.03.2025- 15.04.2025	
7	Оформлення магістерської роботи згідно з вимогами.	16.04.2025- 09.05.2025	
8	Доопрацювання та виправлення зауважень та рекомендацій наукового керівника.	10.05.2025- 16.05.2025	
9	Підготовка та отримання документів (відгук керівника, довідка про унікальність).	16.05.2025- 19.05.2025	
10	Подання роботи в електронному вигляді.	19.05.2025	
11	Підготовка презентації та захист кваліфікаційної роботи.	26.05.2025	

**Науковий керівник:** д.е.н., проф. Черноус Галина Олександрівна

**Студент:** Кожановська Ольга Сергіївна