

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ХЕРСОНСЬКИЙ ДЕРЖАВНИЙ УНІВЕРСИТЕТ
Факультет комп'ютерних наук, фізики та математики
Кафедра комп'ютерних наук та програмної інженерії

Розробка освітніх програм з використанням методів
машинного навчання на основі даних ринку праці

Кваліфікаційна робота (проект)
на здобуття ступеня вищої освіти «магістр»

Виконав: студент 2 курсу 261М групи
Спеціальності

126 «Інформаційні системи
та технології»
(шифр, назва)

Освітньо-професійної програми:

«Інформаційні системи та технології»
(назва)

Гулін Олексій Вадимович

Керівник: доктор економічних наук,
професор Кобець В.М.

Рецензент: Савченко С.О.

технічний директор, ТОВ "Клеверстаф"

ЗМІСТ

ВСТУП

РОЗДІЛ 1

ЗАСТОСУВАННЯ ІТ ІНСТРУМЕНТАЦІЮ ДЛЯ УЗГОДЖЕННЯ ВИМОГ РИНКУ ПРАЦІ І ДИЗАЙНУ ОСВІТНІХ ПРОГРАМ

РОЗДІЛ 2

АНАЛІЗ МЕТОДІВ МАШИННОГО НАВЧАННЯ

2.1. Методи машинного навчання з вчителем

2.2. Методи машинного навчання без вчителя

2.3. Оцінка якості методів машинного навчання

РОЗДІЛ 3

КЛАСТЕРНИЙ ПІДХІД ДО ПРОЄКТУВАННЯ ОСВІТНІХ ПРОГРАМ З УРАХУВАННЯМ ВИМОГ РИНКУ ПРАЦІ

3.1. Методологія дослідження

3.2. Дослідницький аналіз даних (EDA)

3.3. Візуалізація методу

3.4. Визначення оптимальної кількості кластерів

3.5. Визначення освітньої програми, що відповідає кластеру позицій фахівця з обробки даних

ВИСНОВКИ

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

ВСТУП

Динамічний характер потреб ринку праці, особливо в інженерних галузях, вимагає постійного розвитку навичок, щоб здобувачі залишалися конкурентоспроможними на ринку праці [1, 2]. Відкриті освітні ресурси (OER) становлять цінне джерело навчальних матеріалів, які можуть використовуватися для персоналізованого навчання, адаптованого до індивідуальних потреб студентів [3].

Техніки штучного інтелекту (ШІ) та аналізу даних дозволяють аналізувати вакансії та визначати вимоги до навичок для конкретних професій [4, 5, 6]. Системи рекомендацій на основі ШІ можуть зіставляти здобувачів з відповідними OER-ресурсами, виходячи з вимог до професійних навичок та індивідуальних навчальних цілей. Ці системи мають такі виклики, як "холодний старт" для нових користувачів або елементів, а також потреба у точному оцінюванні якості контенту. Для цього використовуються методи тематичного моделювання та аналіз метаданих, які сприяють категоризації та оцінці якості освітніх ресурсів [7].

Актуальність дослідження

У сучасному світі інформаційні технології та штучний інтелект дедалі більше впливають на вимоги ринку праці, особливо у сфері науки про дані та інженерії. Це дослідження актуальне через зростання попиту на фахівців, здатних працювати із великими обсягами даних, використовувати методи машинного навчання та володіти актуальними навичками, необхідними для сучасного робочого середовища. У зв'язку зі швидкими технологічними змінами, освітні програми потребують постійного вдосконалення, щоб відповідати сучасним вимогам роботодавців, забезпечуючи випускникам кращі можливості для працевлаштування.

Об'єкт дослідження

Об'єктом дослідження є навички ІТ фахівців з обробки даних узгоджені з вимогами ринку праці.

Предмет дослідження

Предметом дослідження є аналіз потреб ринку праці для підготовки спеціалістів з обробки даних на освітніх програмах з використанням методів машинного навчання.

Мета дослідження

Метою дослідження є розробка підходу до проектування освітніх програм з урахуванням вимог ринку праці для підвищення ефективності підготовки випускників до виконання затребуваних професійних функцій у сфері обробки даних.

Завдання дослідження

1. Провести огляд літератури, що стосується вимог роботодавців у сфері науки про дані та інформаційних технологій.
2. Дослідити методи кластеризації та їх застосування для аналізу вакансій на ринку праці.
3. Розробити методологію кластеризації вакансій за навичками, необхідними для спеціалістів з обробки даних.
4. Визначити, які освітні програми відповідають потребам ринку, використовуючи методи статистичного аналізу та кластеризації.
5. Надати рекомендації щодо оновлення освітніх програм, які б дозволили більш ефективно готувати студентів до вимог ринку праці.

Методи дослідження

У роботі використано методи машинного навчання, зокрема кластеризацію за допомогою алгоритму K-середніх, а також статистичний аналіз для визначення відповідності навчальних програм до вимог ринку праці. Тема моделювання та аналіз метаданих також використовується для оцінки актуальності освітніх ресурсів.

Апробація

Опублікована стаття у фаховому науковому журналі Applied Aspects of Information Technology (категорія Б):

<https://aait.od.ua/index.php/journal/article/view/253/254>.

Структура роботи включає вступ, три розділи, висновки, список використаних джерел. У розділі 1 розглянуте застосування ІТ інструментацію для узгодження вимог ринку праці і дизайну освітніх програм. У розділі 2 проведений аналіз методів машинного навчання. У розділі 3 застосований кластерний підхід до проєктування освітніх програм з урахуванням вимог ринку праці.

З огляду на постійні зміни та впровадження нових технологій, сучасний ринок праці висуває нові вимоги до спеціалістів, які охоплюють як технічні, так і м'які навички. Ринок потребує фахівців, здатних інтегрувати знання у сфері технологій з навичками ефективного спілкування, командної роботи та управління проєктами. Таким чином, роботодавці очікують від кандидатів не лише технічної експертизи, а й уміння адаптуватися до швидкоплинних умов, проявляти креативність та приймати рішення на основі аналізу даних.

Аналіз потреб ринку праці, який базується на штучному інтелекті та кластерному аналізі вакансій, надає можливість зрозуміти, як саме повинні трансформуватися освітні програми. Це включає не лише введення нових дисциплін, але й перегляд існуючих навчальних підходів, щоб надати студентам навички, які відповідатимуть вимогам роботодавців.

ЗАСТОСУВАННЯ ІТ ІНСТРУМЕНТАЦІЮ ДЛЯ УЗГОДЖЕННЯ ВИМОГ РИНКУ ПРАЦІ І ДИЗАЙНУ ОСВІТНІХ ПРОГРАМ

У дослідженні [1] автори розглядають проблему надання персоналізованих освітніх рекомендацій на основі відкритих освітніх ресурсів (OER) і даних ринку праці для учнів, які прагнуть безперервного навчання. Система використовує інформацію про ринок праці для надання рекомендацій, що дозволяє учням здобувати навички, які є найбільш актуальними для їхньої кар'єри. Основна мета полягає в тому, щоб допомогти користувачам орієнтуватися в мінливому середовищі сучасного ринку праці та запропонувати їм навчальні ресурси, які відповідають вимогам роботодавців.

Дослідження підкреслює важливість персоналізації навчального процесу та демонструє, як використання відкритих освітніх ресурсів може сприяти створенню індивідуальних навчальних шляхів для кожного учня. Використання методів машинного навчання дозволило створити систему, що аналізує навички, які користуються найбільшим попитом на ринку праці, та порівнює їх із наявними освітніми матеріалами у вигляді OER. Це дозволяє учням отримати доступ до найбільш релевантних ресурсів, що суттєво підвищує їхні шанси на працевлаштування [1].

У дослідженні [2] автори зосереджуються на процесах передачі емпіричних інженерних знань та вивчають когнітивні фактори, які впливають на цей процес. Використовуючи методи поведінкового аналізу та функціональну ближню інфрачервону спектроскопію (fNIRS), дослідники з'ясували, як нові працівники засвоюють знання в інженерній сфері. Їхнє дослідження показує, що ефективна передача знань залежить не лише від технічних аспектів, а й від когнітивних процесів, таких як увага та пам'ять.

Виявлено, що існує значний когнітивний бар'єр, який заважає швидкому засвоєнню нових знань, що може сповільнювати адаптацію нових співробітників до виробничих умов. Автори також підкреслюють важливість

розробки навчальних програм, які враховують ці когнітивні фактори, та пропонують нові методи для поліпшення передачі знань у виробничих середовищах, що може підвищити продуктивність та ефективність нових кадрів [2].

Tavakoli та його співавтори [3] досліджують роль відкритих освітніх ресурсів у підтримці кар'єрного розвитку працівників. Автори підкреслюють, що сучасний ринок праці вимагає від працівників постійного оновлення знань і навичок, і для цього необхідно використовувати OER як доступний інструмент для самонавчання. Вони розглядають процес надання персоналізованих рекомендацій щодо OER з урахуванням професійних потреб і ринкових вимог.

Система, запропонована авторами, аналізує потреби користувачів і створює індивідуальні навчальні траєкторії на основі доступних відкритих ресурсів. Це дозволяє забезпечити гнучкість і адаптивність у навчальному процесі, що є важливим у контексті швидких змін у технологіях та економіці. Також дослідження демонструє, що використання OER підвищує ефективність самонавчання, оскільки дає змогу користувачам отримувати доступ до актуальних матеріалів у будь-який час і в будь-якому місці [3].

Colombo та його колеги досліджують [4] використання методів машинного навчання для аналізу вакансій на ринку праці, зокрема в Інтернеті. Дослідження [4] зосереджується на аналізі тексту вакансій з метою виявлення ключових навичок, які є найбільш затребуваними у роботодавців. Вони вивчають, як технології штучного інтелекту можуть допомогти визначити ці навички та сприяти кращому розумінню ринкових трендів.

Одним із ключових результатів дослідження стало виявлення значних розбіжностей між навичками, зазначеними у вакансіях, і тими, які надаються навчальними програмами. Це підкреслює необхідність перегляду освітніх стандартів, щоб вони краще відповідали вимогам сучасного ринку праці. Дослідження також підтверджує ефективність використання машинного

навчання для аналізу великих масивів даних, що дозволяє забезпечити точні рекомендації щодо підвищення професійних навичок [4].

У дослідженні [5] автори розглядають проблему передачі емпіричних інженерних знань у контексті швидкої зміни технологій. Вони зосереджуються на тому, як зміна технологічних парадигм впливає на процес концептуального проєктування та як це вимагає нових підходів до передачі знань. Використання експериментальних методів дозволило авторам виявити ключові фактори, що сприяють або заважають передачі знань.

Автори підкреслюють, що ефективна передача знань в умовах змін є критично важливою для забезпечення конкурентоспроможності підприємств. Вони пропонують нові методи для вдосконалення навчальних програм, що враховують технологічні зміни та адаптаційні процеси, які мають бути інтегровані в освітні програми для підготовки майбутніх інженерів [5].

Робота [6] присвячена розробці концептуальної архітектури освіти в галузі науки про дані. Ermolayev та інші досліджують сучасні вимоги до підготовки фахівців з науки про дані, враховуючи нові виклики, пов'язані з великими даними та аналітикою. У роботі запропоновано модель навчання, що базується на інтеграції теоретичних знань та практичних навичок, необхідних для успішної роботи в сучасних умовах.

Автори підкреслюють важливість адаптації освітніх програм до швидко змінюваних вимог ринку праці, зокрема у сфері науки про дані. Використовуючи запропоновану архітектуру, вони пропонують нові підходи до підготовки фахівців, що дозволяє краще інтегрувати технології в навчальний процес та забезпечити їхню відповідність потребам сучасного ринку [6].

Molavi, Tavakoli і Kismihók досліджують [7] проблематику виділення тем із відкритих освітніх ресурсів (OER) для підтримки індивідуалізованого навчання. Однією з основних проблем, яку вирішує їх дослідження, є надмірність доступних матеріалів, що ускладнює вибір відповідних ресурсів для учнів. Автори використовують методи машинного навчання та текстовий

аналіз для автоматизації процесу виділення релевантних тем із наявних освітніх матеріалів.

За допомогою запропонованих алгоритмів система може виділяти найбільш актуальні теми для кожного користувача залежно від його професійних потреб та ринкових трендів. Це дозволяє користувачам отримати більш точні рекомендації щодо OER, що полегшує процес навчання та сприяє підвищенню кваліфікації в короткі терміни. Дослідження також показує, що автоматизація подібних процесів значно підвищує ефективність системи дистанційного навчання, полегшуючи вибір освітніх матеріалів [7].

У дослідженні [8] Kobets та Osurova зосередилися на ідентифікації ключових факторів, що впливають на якість вищої освіти. Вони проаналізували, як різні елементи освітнього процесу – від викладацьких методик до інфраструктурних ресурсів – впливають на загальну якість освіти для студентів. Автори підкреслюють, що забезпечення високої якості навчання є критично важливим для підготовки конкурентоспроможних фахівців.

Дослідження також підкреслює важливість адаптації освітніх програм до потреб сучасного ринку праці. Автори відзначають, що впровадження новітніх технологій у навчальний процес, зокрема цифрових інструментів та аналітики, може значно підвищити ефективність навчання. Вони також звертають увагу на необхідність забезпечення гнучкості освітніх програм, щоб вони могли швидко адаптуватися до нових викликів та змін на ринку праці [8].

Ketamo та його колеги досліджують [9] застосування штучного інтелекту та аналітики для прогнозування потреб у компетенціях, які будуть затребувані на ринку праці в майбутньому. Основна мета їхнього дослідження полягає у створенні інструментів, що дозволяють передбачити, які знання та навички будуть необхідні для успішного працевлаштування у різних галузях. Вони використовують алгоритми прогнозування, які

аналізують великі обсяги даних з різних джерел, включаючи ринок праці, освітні програми та технологічні тренди.

Дослідження демонструє, що сучасні технології можуть бути використані для адаптації освітніх програм до вимог ринку праці, що дозволяє забезпечити більш точне узгодження між навичками, які отримують студенти, та тими, що затребувані роботодавцями. Крім того, автори наголошують на важливості інтеграції результатів подібних досліджень у національні освітні стратегії, щоб забезпечити довготривалу відповідність навичок робочої сили вимогам ринку [9].

У статті [10] Metrôlho, Ribeiro та Batista досліджують ефективність навчання студентів у рамках реальних проєктів з використанням методології Agile. Їхнє дослідження фокусується на підготовці студентів до роботи в індустрії розробки програмного забезпечення через участь у проєктах типу Full Stack, які імітують роботу в реальних умовах ІТ-галузі. Автори вважають, що навчання через практичні завдання та робота в командах за методологією Agile дозволяє студентам краще підготуватися до реальних викликів, з якими вони зіштовхнуться на ринку праці.

Дослідження показало, що інтеграція подібних практичних проєктів у навчальні програми сприяє підвищенню кваліфікації студентів та їхньої готовності до роботи. Крім того, автори підкреслюють важливість адаптації навчальних програм до вимог сучасного програмного забезпечення та методологій, які використовуються в галузі, зокрема Agile та DevOps, щоб забезпечити більш повну підготовку студентів до професійної діяльності [10].

У дослідженні [11] Kobets, Yatsenko та Buiak аналізують прогалини в підготовці бізнес-аналітиків, зіставляючи потреби ринку праці з освітніми стандартами. Автори звертають увагу на те, що сучасний ринок праці вимагає від бізнес-аналітиків володіння широким спектром навичок, таких як аналітичне мислення, вміння працювати з великими даними, а також глибоке розуміння індустрії. Однак освітні програми часто не відповідають цим

вимогам, що створює розрив між навчанням і практичними потребами роботодавців.

У результаті дослідження автори пропонують ряд рекомендацій для удосконалення освітніх програм з бізнес-аналітики. Зокрема, вони підкреслюють важливість інтеграції сучасних технологій, таких як аналітика даних та штучний інтелект, у навчальний процес. Крім того, вони закликають до більш тісної співпраці між академічними установами та бізнесом, щоб забезпечити більш релевантну підготовку фахівців, які зможуть відповідати актуальним викликам на ринку праці [11].

У роботі [12] Kara та колеги зосереджуються на розробці системи рекомендацій роботи, яка базується на навичках, що витягуються з профілів кандидатів та описів вакансій. Вони використовують методи обробки природної мови (NLP) та машинного навчання для аналізу текстових даних, що дозволяє системі ідентифікувати ключові навички та співвідносити їх із вимогами роботодавців. Такий підхід дозволяє більш точно підібрати вакансії для кандидатів, спрощуючи процес пошуку роботи.

Одним із важливих аспектів цього дослідження є підвищення ефективності ринку праці через використання інноваційних технологій. Автори показують, що автоматизовані системи можуть значно поліпшити процес пошуку роботи та працевлаштування, забезпечуючи більш точне узгодження між кандидатами та роботодавцями. Крім того, дослідження підкреслює важливість безперервного навчання, оскільки технології швидко змінюються, і працівники повинні постійно оновлювати свої знання та навички, щоб залишатися конкурентоспроможними [12].

У дослідженні [13] Ramakrishnan та його колеги зосереджуються на використанні цифрових коммонів для забезпечення відповідності освіти потребам індустрії. Вони стверджують, що цифрові платформи можуть забезпечити відкритий доступ до знань і ресурсів, що сприяє підвищенню рівня підготовки працівників для ринку праці. Автори детально розглядають

кейс використання цифрових коммонів у певній галузі та показують, як це сприяє розвитку професійних навичок, зокрема в технічних спеціальностях.

Одним із ключових аспектів дослідження є вивчення ролі цифрових інструментів у підвищенні ефективності навчання. Автори наголошують на тому, що цифрові платформи дозволяють швидко адаптувати навчальні програми до змін у ринкових умовах та технологіях. Це дослідження показує, що використання подібних інноваційних методів навчання може суттєво покращити якість підготовки фахівців і забезпечити більш точне узгодження між навичками, які отримують учні, та тими, що затребувані на ринку праці [13].

У статті [14] Kobets та його співавтори досліджують розробку системи управління веб-сервісом для пошуку роботи, що базується на компетентнісному підході. Їхня система використовує дані про компетенції кандидатів та вимоги роботодавців для надання персоналізованих рекомендацій щодо вакансій. Автори розглядають, як компетентнісний підхід може сприяти кращому узгодженню між навичками кандидатів та потребами роботодавців, що підвищує шанси на успішне працевлаштування.

Основним результатом дослідження є розробка ефективної моделі веб-сервісу, яка автоматизує процес пошуку роботи та підбору кандидатів. Автори підкреслюють, що компетентнісний підхід дозволяє враховувати не лише формальні кваліфікації, але й м'які навички, що є важливими для успішної кар'єри. Вони також відзначають, що використання сучасних технологій, таких як машинне навчання та аналітика даних, може значно покращити якість рекомендацій та спростити процес підбору персоналу [14].

Wu у роботі [15] аналізує досвід стажувань студентів у галузі науки про дані з метою вдосконалення навчальних програм. Автор досліджує, як стажування впливає на готовність студентів до ринку праці та їхню здатність застосовувати теоретичні знання на практиці. Дослідження показує, що стажування є критично важливим елементом навчання, оскільки воно

дозволяє студентам зануритися у реальні робочі процеси та краще зрозуміти вимоги індустрії.

Одним із ключових висновків дослідження є необхідність інтеграції стажувань у навчальні програми для підготовки фахівців з науки про дані. Автор підкреслює, що співпраця між навчальними закладами та підприємствами може суттєво покращити якість підготовки кадрів, оскільки дозволяє студентам отримати реальний досвід і розвинути навички, які є необхідними для успішної кар'єри у галузі науки про дані [15].

У роботі [16] Kravtsov і Kobets досліджують підходи до вдосконалення навчальних програм у галузі комп'ютерних наук. Вони пропонують еволюційну модель перегляду навчальних планів, яка базується на систематичному оновленні змісту відповідно до вимог ринку праці та новітніх технологічних тенденцій. Ця модель дозволяє гнучко реагувати на швидкі зміни в галузі комп'ютерних наук та підтримувати актуальність навчальних програм.

Автори стверджують, що регулярне оновлення навчальних програм є важливим для підтримки високого рівня підготовки студентів, особливо у сферах, пов'язаних із технологіями, які швидко розвиваються. Вони підкреслюють, що така модель перегляду повинна включати тісну співпрацю з роботодавцями, щоб враховувати їхні потреби у кваліфікованих кадрах. Це дозволяє освітнім закладам випускати фахівців, які володіють сучасними навичками і готові до викликів на ринку праці [16].

У дослідженні [17] Berger та Frey розглядають вплив діджиталізації на ринок праці в Європі та аналізують стратегії подолання прогалів у навичках, що виникають у зв'язку зі швидкими технологічними змінами. Автори звертають увагу на те, що діджиталізація значно змінює структуру зайнятості, збільшуючи попит на працівників із навичками у галузях, пов'язаних із технологіями, водночас скорочуючи кількість робочих місць у традиційних галузях.

Автори пропонують низку рекомендацій для урядів та освітніх інституцій, спрямованих на скорочення прогалів у навичках. Серед таких рекомендацій – розвиток програм професійного навчання, спрямованих на перепідготовку працівників, інвестиції в цифрову освіту та стимулювання співпраці між бізнесом і академічними установами. Вони підкреслюють важливість створення навчальних програм, які дозволяють швидко адаптуватися до нових викликів, що виникають у зв'язку з розвитком автоматизації та штучного інтелекту [17].

Moldovan у дослідженні [18] аналізує сучасні вимоги до знань і навичок у контексті Індустрії 4.0, акцентуючи увагу на впровадженні новітніх технологій, таких як автоматизація, штучний інтелект і інтернет речей (IoT). Вона досліджує, як ці технології змінюють вимоги до робочої сили і які нові компетенції стають необхідними для успішної кар'єри в умовах четвертої промислової революції.

Одним із головних результатів цього дослідження є визначення ключових навичок, необхідних для роботи в умовах Індустрії 4.0, зокрема навичок програмування, аналізу даних, управління автоматизованими системами та адаптивності до технологічних змін. Авторка також розглядає сучасні методи навчання, що включають практичне навчання на робочих місцях, і підкреслює необхідність інтеграції цих методів у традиційні навчальні програми для забезпечення актуальності підготовки кадрів [18].

У роботі [19] Short і Keller-Bell досліджують ключові навички, які повинні мати працівники у 21 столітті для того, щоб залишатися конкурентоспроможними на ринку праці. Автори підкреслюють, що технологічні зміни та глобалізація значно змінюють вимоги до робочої сили, і тепер важливими є не лише технічні знання, але й м'які навички, такі як комунікація, критичне мислення, лідерство та здатність до навчання протягом усього життя.

Вони також акцентують увагу на тому, що сучасний ринок праці вимагає від працівників високого рівня гнучкості та адаптивності, оскільки

технології швидко змінюються. Автори наводять приклади освітніх програм, які допомагають розвивати ці навички, і пропонують впровадження інтегрованих підходів до навчання, які поєднують технічні знання з розвитком м'яких навичок, щоб забезпечити повну підготовку до викликів сучасної робочої сили [19].

Kobets, Yatsenko і Buiak зосередили свою увагу на аналізі прогалин у підготовці бізнес-аналітиків в Україні, порівнюючи вимоги ринку праці з існуючими освітніми стандартами.[20] Їхнє дослідження виявило, що, попри зростаючий попит на кваліфікованих бізнес-аналітиків, існує значний розрив між тим, що вимагають роботодавці, і тим, що пропонують освітні програми. Автори підкреслюють важливість модернізації навчальних програм для підготовки бізнес-аналітиків, включаючи більше фокусу на практичні навички, такі як робота з великими даними та аналітичні інструменти.

Вони також пропонують впровадження більш тісної співпраці між університетами та компаніями, щоб краще узгоджувати освітні стандарти з ринковими потребами. Зокрема, автори вважають за необхідне збільшити кількість курсів з використанням реальних кейсів і проєктів, що дозволить студентам отримати більш практичний досвід і підготуватися до викликів реального ринку [20].

У дослідженні [21] Manuika та інші аналізують вплив автоматизації на ринок праці та трансформацію робочих місць. Автори підкреслюють, що швидкий розвиток технологій, таких як штучний інтелект і роботизація, суттєво змінює структуру зайнятості. Хоча автоматизація сприяє зростанню продуктивності та створенню нових робочих місць у певних галузях, вона також призводить до втрати багатьох традиційних робочих місць, особливо в тих секторах, де людська праця може бути легко замінена машинами.

Автори прогнозують, що в найближчі десятиліття автоматизація створить потребу в перепідготовці мільйонів працівників, які повинні будуть розвивати нові навички, такі як робота з технологіями, критичне мислення та креативність. Вони також акцентують увагу на необхідності розробки нових

освітніх стратегій і програм, які будуть відповідати вимогам швидко змінюваного ринку праці, та на важливості підтримки безперервного навчання для підтримання конкурентоспроможності працівників у майбутньому [21].

У дослідженні [22] Agarwal і Ojha досліджують вплив технологій Індустрії 4.0 на досягнення цілей сталого розвитку у виробничих операціях. Автори використовують аналітичний ієрархічний процес (АНР) для оцінки пріоритетності різних технологій Індустрії 4.0, таких як штучний інтелект, Інтернет речей (ІоТ) та автоматизація, з точки зору їх впливу на сталий розвиток.

Дослідження показує, що Індустрія 4.0 має великий потенціал для сприяння досягненню екологічних, економічних та соціальних цілей сталого розвитку через підвищення ефективності виробничих процесів, зменшення впливу на довкілля та створення нових можливостей для працевлаштування. Водночас автори підкреслюють, що для повного використання цього потенціалу необхідно впроваджувати інновації у сфері освіти та підготовки кадрів, щоб забезпечити наявність достатньо кваліфікованих фахівців, здатних працювати з новими технологіями [22].

У роботі [23] Li та інші зосереджуються на аналізі прогалин у знаннях і навичках, необхідних для роботи в галузі науки про дані у виробничій промисловості. Автори підкреслюють, що з розвитком Індустрії 4.0 та автоматизації у виробництві виникає зростаюча потреба у фахівцях, які можуть ефективно працювати з великими даними та аналітичними інструментами. Вони аналізують, як сучасні освітні програми відповідають вимогам виробничої галузі, і виявляють значні прогалини в підготовці фахівців.

Автори пропонують переглянути підходи до навчання у сфері науки про дані, зокрема впровадження практичних кейсів та проєктів, які дозволять студентам розвивати прикладні навички, необхідні для роботи в умовах реального виробництва. Крім того, вони звертають увагу на необхідність

інтеграції новітніх технологій у навчальні програми для забезпечення актуальності знань та компетенцій випускників, щоб відповідати вимогам сучасного ринку праці [23].

Novak і Kobets досліджують вплив штучного інтелекту (ШІ) на забезпечення продовольчої безпеки держав у світі. У роботі [24] вони аналізують, як технології ШІ можуть бути використані для оптимізації агропромислових процесів, зокрема у сфері моніторингу врожаїв, прогнозування погодних умов, управління ресурсами та боротьби з хворобами рослин. Автори підкреслюють, що впровадження ШІ у сільське господарство може суттєво підвищити ефективність виробництва та допомогти країнам справлятися з викликами, пов'язаними з продовольчою безпекою.

Дослідження також звертає увагу на можливі ризики, пов'язані з використанням ШІ, зокрема залежність від технологій та потенційні загрози кібербезпеці. Автори підкреслюють важливість розробки регуляторних механізмів для мінімізації цих ризиків та забезпечення відповідального використання ШІ в агропромисловій сфері. Крім того, вони акцентують увагу на необхідності навчання фермерів та інших працівників сільськогосподарського сектору для успішної інтеграції технологій ШІ у їхню повсякденну роботу [24].

У дослідженні [25] Vista аналізує навички, які будуть найбільш затребуваними у 21-му столітті, використовуючи підходи на основі даних. Автор підкреслює, що сучасний ринок праці вимагає від працівників не лише технічних навичок, але й вміння адаптуватися до нових умов, критично мислити та вирішувати складні проблеми. Враховуючи швидкі зміни у технологічній сфері, важливим є розвиток навичок, які дозволяють працівникам ефективно використовувати інноваційні технології та знаходити творчі рішення.

Дослідження акцентує увагу на важливості безперервного навчання, оскільки багато професій зазнають значних змін під впливом автоматизації та

цифрових технологій. Автор пропонує розробити освітні програми, які допоможуть розвивати м'які навички поряд із технічними компетенціями, щоб підготувати працівників до викликів майбутнього ринку праці. Vista також наводить приклади успішної адаптації освітніх стратегій у різних країнах для розвитку таких навичок [25].

У статті [26] Kobets et al. представлено систему управління веб-сервісом, що сприяє пошуку роботи для виробничих підприємств, використовуючи компетентнісний підхід. Автори наголошують на важливості врахування специфічних компетенцій кандидатів у процесі підбору персоналу, що часто не враховується традиційними методами. Запропонований сервіс дозволяє створювати профілі, що детально відображають професійні навички, що підвищує точність відповідності між вакансіями та кандидатами.

Розроблена система демонструє високу ефективність у тестуванні, що підтверджує її здатність покращити процес підбору кадрів. Завдяки автоматизації та систематизації даних, компанії отримують змогу швидше знаходити підходящих кандидатів, що відповідають специфічним вимогам виробництв. Цей підхід може суттєво зменшити час, витрачений на відбір резюме, та підвищити якість підбору кадрів.

Стаття також підкреслює можливість впровадження даної системи в різних галузях, що робить її універсальною. У подальших дослідженнях автори рекомендують розглянути функціональність системи для оцінки компетенцій через онлайн-тести, а також провести порівняльний аналіз з іншими HR-рішеннями, що існують на ринку. Це дозволить ще більше оптимізувати процес підбору кадрів та підвищити задоволеність як роботодавців, так і працівників [26].

РОЗДІЛ 2

АНАЛІЗ МЕТОДІВ МАШИННОГО НАВЧАННЯ

2.1 Методи машинного навчання з вчителем

Методи машинного навчання з учителем називають так тому, що вони працюють з **навчальним набором даних**, де для кожного прикладу відомі як **вхідні значення (ознаки)**, так і **правильний результат або мітка**. Ця мітка діє як своєрідний «вчитель», який направляє модель на шляху до правильного прогнозу або класифікації. У такому середовищі модель працює над визначенням залежності між вхідними даними та їх правильним результатом, навчаючись розпізнавати закономірності та зв'язки в цих даних.

У процесі навчання модель аналізує приклади з набору даних, намагаючись зрозуміти, як різні значення ознак (або ж різні комбінації цих значень) пов'язані з певними результатами. Для кожного випадку алгоритм передбачає результат, порівнює його з правильним значенням і, якщо є розбіжності, коригує свої параметри, щоб максимально зменшити цю помилку в майбутніх передбаченнях. Таким чином, модель вчиться на прикладах, отримуючи зворотний зв'язок від «вчителя» (правильної мітки), щоб зрозуміти, які особливості даних ведуть до конкретного результату.

Такий підхід спрямований на те, щоб модель, навчившись на позначених даних, змогла надалі передбачати результати для нових, невідомих випадків, спираючись на отримані знання. Машинне навчання з учителем є особливо корисним для задач, де необхідно чітко визначити закономірності у великих наборах даних, забезпечуючи стабільні й точні результати, коли дані мають достатню кількість ознак та прикладів для ефективного навчання.

Методи машинного навчання з учителем націлені на те, щоб модель могла не лише «запам'ятати» дані, але й навчитися **узагальнювати закономірності**, що містяться в них, щоб робити точні передбачення на нових даних.

Метод Регресії

Лінійна регресія є базовим методом для прогнозування числових значень і аналізу залежностей між змінними. В основі цього методу лежить побудова лінії, яка найкраще відображає залежність між незалежною змінною XXX і залежною змінною YYY . Формула для лінійної регресії виглядає як: $y = mx + by = mx + by = mx + b$ де mmm — це кут нахилу (градієнт), що вказує, як зміна XXX впливає на YYY , а bbb — це перетин з віссю uuu . Для знаходження параметрів mmm і bbb найчастіше використовується метод найменших квадратів, що мінімізує суму квадратів відстаней між точками даних і побудованою лінією.

Приклад: уявімо, що ми хочемо спрогнозувати ціну будинку на основі площі його житлової площі. Використовуючи лінійну регресію, ми визначаємо, як зміна площі (незалежна змінна) впливає на ціну (залежна змінна), що дозволяє передбачити ціну для будинків з невідомою площею на основі трендів.

Метод опорних векторів (Support Vector Machine, SVM)

Метод опорних векторів (SVM) орієнтований на класифікацію, хоча його також можна застосовувати для задач регресії. Його основна ідея полягає в пошуку гіперплощини, яка розділяє дані так, щоб відстань між найближчими точками обох класів (опорними векторами) і гіперплощиною була максимальною. Це допомагає забезпечити найкращу можливу роздільну здатність між класами.

У випадках, коли дані неможливо розділити лінійно, SVM використовує ядрові методи, які дозволяють переносити дані в простір більш високої розмірності, де їх можна поділити лінійно. Наприклад, ядро Радіальної базисної функції (RBF) часто застосовується для даних, які розподілені у складних, нелінійних структурах.

Приклад: уявіть, що ви маєте дані про електронну пошту, де кожне повідомлення позначено як "спам" або "не спам". Метод SVM знайде

оптимальну лінію або поверхню, що розділяє ці два класи, таким чином, що максимізує відстань між гіперплощиною і найближчими точками обох класів.

Метод kkk-найближчих сусідів (k-Nearest Neighbors, k-NN)

Метод **kkk-найближчих сусідів (k-NN)** є простим та інтуїтивним алгоритмом для задач класифікації та регресії. Він передбачає, що для нового зразка потрібно знайти kkk найближчих сусідів з тренувальних даних і, залежно від класів сусідів, призначити клас або значення.

Для цього методу важливим є вибір кількості сусідів kkk, а також спосіб обчислення відстані (евклідова або манхеттенська). Якщо kkk занадто мале, модель може підлаштуватися під шум (перенавчання); якщо велике, модель може не реагувати на дрібні зміни в даних.

Приклад: уявімо, що ми хочемо визначити клас клієнта банку (наприклад, потенційний позичальник або вкладник) на основі доходу та віку. Метод kkk-NN шукає найближчі точки до нового клієнта і визначає клас залежно від класів цих точок.

Дерева рішень

Дерева рішень представляють ієрархічну модель, яка будується за принципом "якщо-то" правил. Модель дерева рішень розділяє дані на основі певних критеріїв, що дозволяє дійти до кінцевого рішення, яке відповідає певному класу або значенню. Кожен вузол дерева представляє певну ознаку, а його гілки відображають можливі значення цієї ознаки.

Дерева рішень корисні, коли потрібно зрозуміти логіку процесу ухвалення рішень, оскільки результат легко пояснити, пройшовши по дереву від кореня до листа.

Приклад: розглянемо завдання класифікації клієнтів інтернет-магазину за можливістю повторної покупки. Дерево рішень може використовувати такі ознаки, як "кількість покупок у минулому місяці" і "кількість відвідувань сайту", щоб розподілити клієнтів на "постійних" та "одноразових".

Випадковий ліс (Random Forest)

Випадковий ліс — це ансамблевий метод, що базується на сукупності дерев рішень. Ідея полягає в тому, щоб створити декілька дерев рішень на різних підмножинах даних і ознак, а потім об'єднати їх результати для отримання остаточного прогнозу або класифікації. Для класифікації кожне дерево “голосує” за клас, а для регресії обчислюється середнє значення передбачень усіх дерев.

Перевагою випадкового лісу є зменшення ризику перенавчання і підвищення стабільності та точності моделі порівняно з окремими деревами рішень.

Приклад: уявімо, що ми прогнозуємо кредитоспроможність клієнтів банку. Випадковий ліс будує декілька дерев рішень на різних комбінаціях ознак, таких як дохід, вік, наявність позик, і робить прогноз на основі середнього результату або "голосування" всіх дерев. Це допомагає зменшити вплив випадкових змін і отримати більш надійний результат.

Ці методи є основою для різних задач, де потрібно передбачити певний результат на основі навчальних даних. Кожен з них має свої сильні сторони та сфери застосування, що дозволяє підібрати оптимальний підхід до розв'язання конкретної задачі.

Зазначені методи є основою для вирішення широкого спектра задач, де потрібно передбачити певний результат на основі навчальних даних. Їх вибір залежить від кількох важливих чинників, зокрема типу задачі (класифікація, регресія), розміру і характеристик даних, а також обчислювальних ресурсів. Кожен з методів має свої сильні сторони, недоліки та сфери застосування, що дозволяє підібрати оптимальний підхід для конкретного завдання.

2.2 Методи машинного навчання без вчителя

Методи машинного навчання без учителя використовуються для аналізу даних, коли немає попередньо відомих міток, які могли б бути орієнтирами для навчання. Їх головна мета — знаходження **прихованих**

шаблонів і залежностей в даних, що дозволяє системі самостійно структурувати дані, групувати об'єкти або виділяти найважливіші характеристики без людського втручання. Це особливо корисно в умовах, коли даних багато, але їхня структура або характеристики не до кінця зрозумілі. Такий підхід дозволяє виявляти нові знання або закономірності, які можуть бути приховані при звичайному аналізі.

Методи машинного навчання без учителя намагаються створити внутрішнє представлення даних, орієнтуючись на подібності і відмінності між прикладами, що дозволяє знаходити групи, сегменти або інші структури, які повторюються в даних. Це працює тому, що реальні дані часто мають деякі природні закономірності — наприклад, певні ознаки можуть бути статистично схильними до з'являтися разом або, навпаки, рідко трапляються в комбінації. В таких випадках методи без учителя можуть «виявити» ці закономірності, не маючи попередньої інформації про те, що вони мають знайти.

Метод К-середніх

Алгоритм К-середніх (K-Means) є одним із найпоширеніших методів кластеризації. Цей метод часто використовують для розділення даних на певну кількість кластерів, задану заздалегідь. Його основний принцип — знаходження центрів (середніх значень) для кожного кластера, куди призначаються точки, найближчі до цього центру. Ідея алгоритму полягає в тому, щоб мінімізувати суму квадратів відстаней від кожної точки до центру кластера, до якого вона належить. Процес повторюється до тих пір, поки точки не перестануть переходити між кластерами, а центри не стабілізуються.

Ключова особливість К-середніх — він працює особливо ефективно на даних, де є приблизно однакова кількість точок для кожного кластера, і кластерні структури добре розділені. Проте він менш ефективний для складних або нерівномірних даних, коли кластери мають різні розміри або форми.

Ієрархічна кластеризація

Ієрархічна кластеризація відрізняється тим, що будує дерево або дендрограму, представляючи багаторівневу структуру кластерів. На початку кожен об'єкт вважається окремим кластером, а потім вони об'єднуються у більші групи на основі їх подібності. Цей метод не потребує визначення кількості кластерів заздалегідь, що робить його більш гнучким. Зрештою, користувач може вибрати рівень поділу (глибину дерева), щоб виділити певну кількість кластерів.

Наприклад, дендрограма може показати, як групи клієнтів розділені на різні рівні в залежності від їхніх звичок покупок. Вибравши рівень дерева, компанія може, наприклад, зупинитися на чотирьох основних типах клієнтів або більш детально розділити їх на підгрупи.

Аналіз головних компонент (PCA)

Аналіз головних компонент (PCA) — це метод зниження розмірності даних, який дозволяє зберегти максимум інформації, зменшивши при цьому кількість ознак. PCA знаходить новий набір координат (головні компоненти), який орієнтований на основі варіацій у даних. Перші кілька головних компонент зберігають найбільше варіативності даних, що дозволяє відобразити складні структури у більш простому вигляді.

Цей метод широко застосовується в задачах, де дані мають велику кількість характеристик, таких як аналіз зображень, обробка звуку та тексту. Наприклад, для розпізнавання облич PCA може зменшити кількість пікселів, які обробляються, залишаючи лише найважливіші, що робить обробку швидшою та ефективнішою.

Метод гаусових змішаних моделей (GMM)

Метод гаусових змішаних моделей (GMM) є розширенням кластеризації, який дозволяє працювати з даними, що мають складніші структури. GMM припускає, що дані можуть бути розподілені за декількома нормальними розподілами (тобто, кожен кластер має нормальний розподіл). Використовуючи ймовірності належності кожної точки до кожного з

кластерів, модель обчислює, наскільки ймовірно, що конкретна точка належить до певного кластера, враховуючи середнє і дисперсію для кожного розподілу. GMM дозволяє класифікувати дані з м'якими межами між кластерами.

Цей метод підходить для ситуацій, коли дані можуть перетинатися між кластерами або мають складну форму. Наприклад, у задачах сегментації зображень він дозволяє визначати об'єкти з нечіткими межами.

Метод ап'іорі (Apriori)

Apriori є методом навчання без учителя, що використовується для пошуку асоціацій і правил в даних. Він особливо корисний у ритейлі для виявлення моделей покупки. Наприклад, вивчаючи історію покупок клієнтів, метод Apriori може виявити, що «якщо покупець купує хліб, то, ймовірно, він також купить молоко». Алгоритм працює, знаходячи часті набори об'єктів (комбінації продуктів, що часто купуються разом), і генерує правила, які мають певну підтримку та впевненість.

Цей метод є основою для багатьох систем рекомендацій і дозволяє визначати зв'язки між елементами, які на перший погляд можуть здаватися неочевидними.

Переваги методів навчання без учителя

Методи машинного навчання без учителя дозволяють знаходити закономірності в необроблених даних, де немає наперед відомих результатів або відповіді. Вони особливо корисні для аналізу великих обсягів даних, де може бути складно зрозуміти структуру або тенденції без автоматизованого підходу. Вони дозволяють виявляти групи схожих об'єктів, знайомитися з різними сегментами, які можуть бути важливими для бізнесу, аналізувати складні структури даних і зменшувати обсяг інформації до найбільш значущих характеристик.

Цей тип навчання є основою для багатьох інноваційних застосувань, таких як системи рекомендацій, аналіз зображень, обробка природної мови, управління ризиками, виявлення шахрайства, та персоналізований маркетинг.

Без необхідності в ручній розмітці даних методи навчання без учителя стають надзвичайно важливими у випадках, коли швидкість, обсяг даних і складність є вирішальними.

2.3 Оцінка якості методів машинного навчання

Оцінка якості методів машинного навчання — це один з найважливіших етапів, оскільки він дозволяє зрозуміти, наскільки добре модель здатна передбачати або класифікувати нові дані. Для оцінки якості моделей існує багато метрик, кожна з яких має своє призначення і підходить для певних типів задач. Вибір відповідних метрик залежить від того, чи займається модель класифікацією або регресією.

Метрики для задач регресії

У задачах регресії модель передбачає числові значення, і завдання полягає в тому, щоб оцінити точність цих прогнозів. Однією з найпоширеніших метрик є **корінь середньоквадратичної помилки (Root Mean Squared Error, RMSE)**. RMSE вимірює середню відстань між передбаченим значенням і фактичним результатом, і чим менше це значення, тим точніша модель. RMSE обчислюється як корінь квадратного середнього відхилення передбачених значень від фактичних:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

де y_i — фактичне значення, \hat{y}_i — передбачене значення, а n — загальна кількість точок. RMSE чутлива до великих відхилень, тому висока RMSE свідчить про те, що модель має значні помилки для деяких прикладів.

Іншою важливою метрикою в регресії є **коефіцієнт детермінації (R-Squared або R^2)**. Ця метрика показує, наскільки добре модель пояснює варіативність даних. Вона має значення від 0 до 1, де 1 означає ідеальну відповідність даних, а 0 вказує, що модель не пояснює жодної варіативності. R^2 обчислюється за формулою:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

де \bar{y} — середнє значення фактичних значень. Високє значення R^2 вказує на те, що модель добре описує дані, хоча це не завжди означає точні передбачення для кожного окремого значення.

Метрики для задач класифікації

У задачах класифікації модель передбачає клас, до якого належить кожен приклад. Тут корисною є **матриця неточностей (Confusion Matrix)**, яка дозволяє оцінити, наскільки добре модель відокремлює класи. Матриця неточностей містить чотири основні показники:

- **True Positives (TP)** — кількість випадків, коли модель правильно передбачила позитивний клас.
- **True Negatives (TN)** — кількість випадків, коли модель правильно передбачила негативний клас.
- **False Positives (FP)** — кількість випадків, коли модель передбачила позитивний клас помилково.
- **False Negatives (FN)** — кількість випадків, коли модель передбачила негативний клас помилково.

На основі матриці неточностей обчислюються метрики, такі як **точність (Accuracy)**, **повнота (Recall)**, **точність (Precision)** і **F1-міра**. Точність показує частку правильно передбачених результатів серед усіх випадків:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Однак точність може бути недостатньо інформативною у випадках, коли класи сильно незбалансовані. Наприклад, якщо клас «0» значно переважає над класом «1», то навіть модель, яка завжди передбачає «0», матиме високу точність, хоча вона насправді не здатна розрізняти класи. У таких випадках корисними є інші метрики.

Повнота (Recall) показує частку правильно передбачених позитивних випадків серед усіх фактичних позитивних випадків:

$$Recall = \frac{TP}{TP + FN}$$

Повнота важлива, коли критично виявити всі позитивні випадки, наприклад, у виявленні захворювань, де пропущений випадок може мати серйозні наслідки.

Точність (Precision), навпаки, показує, яка частка позитивних передбачень є правильними:

$$Precision = \frac{TP}{TP + FP}$$

Ця метрика особливо корисна, коли важливо мінімізувати помилкові спрацьовування, наприклад, у системах виявлення шахрайства.

Для балансування між точністю і повнотою використовується **F1-міра**, яка є середнім гармонійним між точністю і повнотою:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

F1-міра забезпечує збалансовану оцінку як для задач, де важливі і повнота, і точність.

Підбір метрик залежно від задачі

Метрика оцінки якості моделі повинна відповідати цільовій задачі. Наприклад, у задачах регресії RMSE та R^2 забезпечують загальне уявлення про точність передбачень, де RMSE допомагає зрозуміти розмір помилки для кожного передбачення, а R^2 вказує, наскільки добре модель описує загальну тенденцію даних. У задачах класифікації точність може бути корисною для збалансованих класів, тоді як для незбалансованих краще використовувати точність, повноту або F1-міру. Усі ці метрики дозволяють отримати комплексне уявлення про ефективність моделі, що є ключем до її подальшого вдосконалення та налаштування.

Оцінка якості моделі — це не лише процес перевірки точності передбачень, а й важливий етап у процесі розробки, що допомагає визначити, наскільки модель здатна **узагальнювати знання і адаптуватися до нових даних**. Оцінка якості стає основою для вибору відповідних моделей та їх параметрів, а також підвищення надійності у реальних умовах використання.

Для **регресійних задач** RMSE і R^2 часто використовуються разом. Наприклад, навіть при високому значенні R^2 низька RMSE може свідчити про наявність кількох точок з великими відхиленнями, які потребують додаткового аналізу. Низьке значення RMSE свідчить про те, що модель надійно прогнозує кожне значення, а високий R^2 показує загальну відповідність тенденцій. Але важливо пам'ятати, що деякі регресійні моделі, зокрема ті, що включають нелінійні зв'язки або роботу з великим обсягом ознак, можуть потребувати додаткових метрик, таких як **MAE (Mean Absolute Error)**, який вимірює середню абсолютну різницю між передбаченим та реальним значенням, незалежно від напрямку відхилення. MAE підходить для задач, де важливий реальний розмір помилки, наприклад, для оцінки вартості нерухомості або прогнозування витрат.

Для **класифікаційних задач** також важливо враховувати метрики, які забезпечують точнішу картину щодо відмінностей між класами. Наприклад, **ROC-крива (Receiver Operating Characteristic)** і **AUC (Area Under Curve)** є потужними інструментами для візуалізації та оцінки якості класифікатора на всіх порогах чутливості. ROC-крива відображає співвідношення між **чутливістю (Sensitivity)** і **специфічністю (Specificity)**, тобто між можливістю правильно визначати позитивні випадки і здатністю уникати помилкових позитивних результатів. AUC, площа під цією кривою, показує загальну якість моделі незалежно від порогу; значення ближче до 1 свідчить про високу точність моделі.

Наприклад, у задачах медичної діагностики модель повинна мати високу чутливість, щоб знизити ризик пропуску важливих випадків, навіть якщо це трохи знизить специфічність. Водночас у завданнях, де важливіше

уникнути помилкових позитивних випадків (наприклад, у визначенні шахрайства у фінансових транзакціях), специфічність може мати більший пріоритет. Це дозволяє налаштувати модель під конкретні потреби, регулюючи поріг чутливості.

Також варто враховувати **перехресну валідацію (Cross-Validation)** як метод для більш точної оцінки моделі. Це техніка, яка розбиває дані на кілька підвбірок (фолдів), де кожна підвбірка використовується для тестування моделі, навченої на решті вибірок. Середнє значення показників якості (наприклад, RMSE або точності) на всіх фолдах дозволяє зменшити залежність результатів від випадкових особливостей одного набору даних і отримати більш об'єктивну оцінку моделі.

Ефективне оцінювання якості також залежить від **розуміння контексту застосування моделі**. Наприклад, у задачах прогнозування часу доставки товарів незначна похибка може бути не такою важливою, але в задачах, пов'язаних із медичними діагнозами або фінансовими ризиками, навіть невеликі помилки можуть мати серйозні наслідки. Вибір метрик повинен враховувати ці ризики та вимоги, що дозволяє вибрати модель, яка найкраще відповідає не лише точності, але й надійності в конкретній прикладній сфері.

Таким чином, оцінка якості моделей машинного навчання включає різноманітні метрики і підходи, кожен з яких розкриває різні аспекти моделі — від точності прогнозів і здатності до узагальнення до стабільності та чутливості до особливостей конкретних задач. Глибоке розуміння цих методів і їх комбінування дозволяє більш ефективно керувати моделями на кожному етапі життєвого циклу — від навчання до впровадження та підтримки.

РОЗДІЛ 3

КЛАСТЕРНИЙ ПІДХІД ДО ПРОЄКТУВАННЯ ОСВІТНІХ ПРОГРАМ З УРАХУВАННЯМ ВИМОГ РИНКУ ПРАЦІ

3.1 Методологія дослідження

Кластеризація методом К-середніх — це алгоритм неконтрольованого навчання, розроблений для кластеризації даних на основі їхньої схожості. Неконтрольоване навчання означає, що немає конкретного результату (кількості кластерів), який потрібно передбачити, і алгоритм намагається знайти шаблони в даних. У методі К-середніх потрібно вказати кількість кластерів, на які, на нашу думку, доцільно поділити дані. Алгоритм випадково призначає кожне спостереження певному кластеру і знаходить центр кожного з них. Далі алгоритм ітерується через два кроки:

- Перепризначає точки даних до кластеру, центр якого найближчий.
- Обчислює новий центроїд для кожного кластера.

Ці два кроки повторюються доти, доки варіація всередині кластеру не може бути зменшена понад певний поріг. Внутрішньокластерна варіація обчислюється як сума евклідових відстаней між точками даних і відповідними центроїдами кластерів.

Ми будемо використовувати відкритий набір даних про професії та пов'язані навички для фахівців з обробки даних, що складається з 3744 ІТ-професій («інженер з даних», «аналітик даних», «вчений з даних», «програміст»). Для кожної з цих професій вимоги сформульовані у вигляді навичок (SQL, Python, Scala/Spark, Data Engineering, AWS, Azure, ETL, Data Analysis, Snowflake, Kafka, Java, Data Modeling, Airflow, Data Warehousing, Machine Learning, DevOps, Kubernetes, Docker, Databricks, Git). Профіль вимог для кожної професії визначено у вигляді бінарних значень (1 - якщо

навичка включена у вимоги до цієї професії; 0 - в іншому випадку). Перші значення цієї відповідності між професіями та навичками продемонстровані у Табл. 3.1 за допомогою функції Python `head(clustered_job_postings)`.

Таблиця 3.1.

Відповідність між професіями та навичками

job_title	SQL	Python	Scala/Spark	Data Engineering	AWS	Azure	ETL	Data Analysis	Snowflake
Data Engineer 2	1	0	0	0	0	1	0	0	0
Staff Data Engineer	1	1	0	0	0	0	1	0	1
Senior Data Engineer, Public Company	1	1	0	0	0	0	1	0	1
Senior Data Engineer, Public Company	1	1	0	1	0	0	1	0	1
Senior Staff AI Data Engineer	1	1	0	1	1	1	1	0	1

3.2 Дослідницький аналіз даних (EDA)

Дані складаються з оголошень про вакансії для фахівців з обробки даних з різними вимогами до навичок. Щоб визначити, які навички випускників закладів вищої освіти краще сприятимуть працевлаштуванню на конкретні посади, ми ідентифікуємо основні типи посад для фахівців з обробки даних, кластеризуючи ці посади. Для кластеризації методом К-середніх ми використовуватимемо стовпці з навичками для формування кластерів за допомогою Python. Приклад коду на Python для кластеризації методом К-середніх та знаходження центрів кластерів для позицій фахівців з обробки даних наведено нижче (Лістинг 3.1) [28].

```
import matplotlib.pyplot as plt
import numpy as np
from sklearn.cluster import KMeans

# Load the data
file_path = '/mnt/data/clustered_job_postings (1).csv'
data = pd.read_csv(file_path)
```



```

# Select feature columns (all columns except 'job_title' and 'cluster')
features = data.columns[1:-1]
X = data[features].values

# Perform K-means clustering
kmeans = KMeans(n_clusters=3, random_state=0) # Adjust n_clusters as needed
kmeans.fit(X)
y_kmeans = kmeans.predict(X)

# Plot the clustered data
plt.scatter(X[:, 0], X[:, 1], c=y_kmeans, cmap='viridis', edgecolor='k')
centers = kmeans.cluster_centers_
plt.scatter(centers[:, 0], centers[:, 1], c='red', s=200, alpha=0.5, label='Centers')

plt.xlabel('First Feature')
plt.ylabel('Second Feature')
plt.title('K-means Clustering of Job Postings')
plt.legend()
plt.show()

```

Лістинг 3.1. Розподіл вимог для позицій фахівців з обробки даних

Після кластеризації ми візуалізуємо результати кластеризації методом К-середніх на Рис. 3.1, використовуючи метод t-SNE. Точки даних пофарбовані відповідно до призначеного кластеру, а червоні точки представляють центри кластерів. Таким чином, у наборі даних професій фахівців з обробки даних було виявлено чотири кластери (Рис. 3).



Рис. 3.1. Кластеризація позицій фахівців з обробки даних

3.3 Візуалізація методу

Було візуалізовано кластеризацію даних за навичками, використовуючи метод t-SNE (Лістинг 3.2) [28]. Результати візуалізації показані на Рис. 3.2, де метод стохастичного вкладення сусідів (t-SNE) застосовується як нелінійний метод зниження розмірності, який часто використовується для візуалізації багатовимірних даних у двох або трьох вимірах. Цей метод особливо корисний для ідентифікації кластерів у даних. Таким чином, результати кластеризації дозволили отримати розподіл кластерів.

```
import pandas as pd
from sklearn.cluster import KMeans
from sklearn.manifold import TSNE
import matplotlib.pyplot as plt
import seaborn as sns

# Завантаження даних з CSV файлу
df = pd.read_csv('/content/clustered_job_postings (1).csv')
# Видалення стовпця 'job_title' для кластеризації
X = df.drop('job_title', axis=1)
# Виконання кластеризації методом k-середніх
```

```

kmeans = KMeans(n_clusters=4, random_state=20) # Ви можете змінити кількість кластерів
kmeans.fit(X)
# Додавання результатів кластеризації до оригінального DataFrame
df['cluster'] = kmeans.labels_
# Використання t-SNE для зниження розмірності даних
tsne = TSNE(n_components=2, perplexity=30, random_state=20)
X_tsne = tsne.fit_transform(X)
# Додавання t-SNE компонентів до DataFrame
df['tsne-2d-one'] = X_tsne[:, 0]
df['tsne-2d-two'] = X_tsne[:, 1]
# Візуалізація результатів
plt.figure(figsize=(16, 10))
sns.scatterplot(
x='tsne-2d-one', y='tsne-2d-two',
hue='cluster',
palette=sns.color_palette('hsv', len(df['cluster'].unique())),
data=df,
legend="full",
alpha=0.8
)

plt.title('Кластеризація Data Engineers за навичками з використанням t-SNE')
plt.show()

# Збереження результатів у новий CSV файл
df.to_csv('clustered_data.csv', index=False)

```

Лістинг 3.2. Візуалізація кластерів позицій фахівців з обробки даних



Рис. 3.2. Візуалізація кластерів позицій фахівців з обробки даних за допомогою методу t-SNE

Розподіл кластерів:

- Жовтий кластер (кластер 0) розташований з лівого боку графіка. Він містить кілька щільних груп точок.
- Зелений кластер (кластер 1) знаходиться з правого боку графіка. Він містить найбільше точок і є більш розсіяним, ніж інші кластери.
- Синій кластер (кластер 2) зосереджений у верхній частині графіка. Точки в ньому досить компактні.
- Фіолетовий кластер (кластер 3) розташований у нижній частині графіка. Точки утворюють щільну групу.

Щільність точок:

- Деякі кластери мають області з високою щільністю точок, що вказує на більшу подібність навичок між нашими даними в цих областях.
- Інші кластери, такі як зелений кластер, мають більш широко розподілені точки, що може вказувати на більшу різноманітність навичок у цьому кластері.

Особливості кластерів:

- Кластери можуть відповідати різним групам навичок або спеціалізаціям у наших даних.
- Жовтий і синій кластери можуть представляти групи з дуже специфічними навичками, враховуючи їхню щільність і компактність.
- Фіолетовий кластер також може представляти високоспеціалізовану групу, але дещо більшу за розміром.
- Зелений кластер може охоплювати ширший набір навичок або мати більш загальну спеціалізацію.

Визначення оптимальної кількості кластерів

Для визначення оптимальної кількості кластерів ми застосовуємо метод «лікоть». Графік методу «лікоть» для визначення оптимальної кількості кластерів показує, що інерція (різниця між елементами кластерів) різко

знижується від 2 до 4 кластерів, а потім знижується менш значно (Рис. 3.3). Це вказує на те, що оптимальна кількість кластерів знаходиться в діапазоні від 3 до 4.

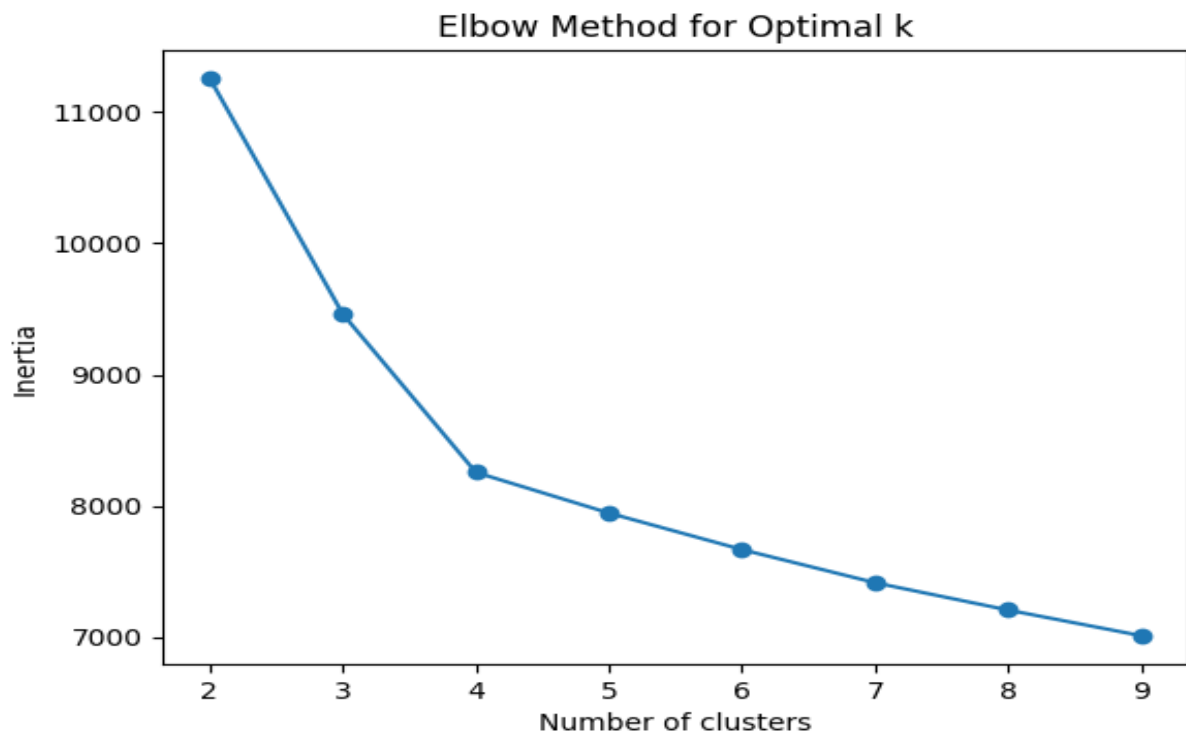


Рис. 3.3. Визначення оптимальної кількості кластерів за допомогою методу «лікоть»

3.4 Визначення оптимальної кількості кластерів

Оптимальна кількість кластерів за допомогою методу «лікоть» зазвичай обирається там, де на графіку спостерігається «лікоть», тобто точка, після якої зниження інерції стає менш значним. На Рис. 4 видно, що такою точкою є чотири кластери.

Три кластери: Виділяють окремі групи для інженерів з даних та інших позицій, включаючи аналітиків даних і фахівців з науки про дані.

Чотири кластери: Дають ще більш детальний поділ між професіями, зокрема, розділяючи фахівців з науки про дані на різні кластери.

Рекомендації

- Ми можемо використовувати 4 кластери для оптимальної кластеризації. Це надасть більш детальний розподіл професій і спеціалізацій.

- Якщо потрібна ще більш детальна сегментація, можна розглянути 5 кластерів, але слід враховувати, що значущість зниження інерції зменшується після 4 кластерів.

Відповідно до методу «лікоть» для визначення оптимальної кількості кластерів найбільш значне зниження коефіцієнта інерції, яке показує різницю між кластерами, відбулося при кількості кластерів 4 (Рис. 3.3). Подальше збільшення кількості кластерів призводить до незначного покращення якості кластеризації, що стало основою для вибору 4 кластерів. На основі результатів розподілу позицій між кластерами ми отримуємо наступну таблицю 3.2.

Таблиця 3.2

Розподіл позицій (професій) між кластерами

Clusters	data engineer	data analyst	data scientist	software engineer
0	763	42	25	46
1	527	503	22	34
2	894	350	115	56
3	361	0	1	5

Ця таблиця показує, що назви посад не завжди однозначно визначають необхідні навички та компетенції кандидатів на вакансії. І навпаки, посади з різними назвами можуть бути схожими за набором необхідних навичок. Це означає, що кандидату варто в першу чергу звертати увагу на набір навичок, які він або вона повинні мати, щоб відповідати вимогам бажаної посади.

Розподіл навичок фахівців з обробки даних у кожному кластері показано на Рис. 3.4.

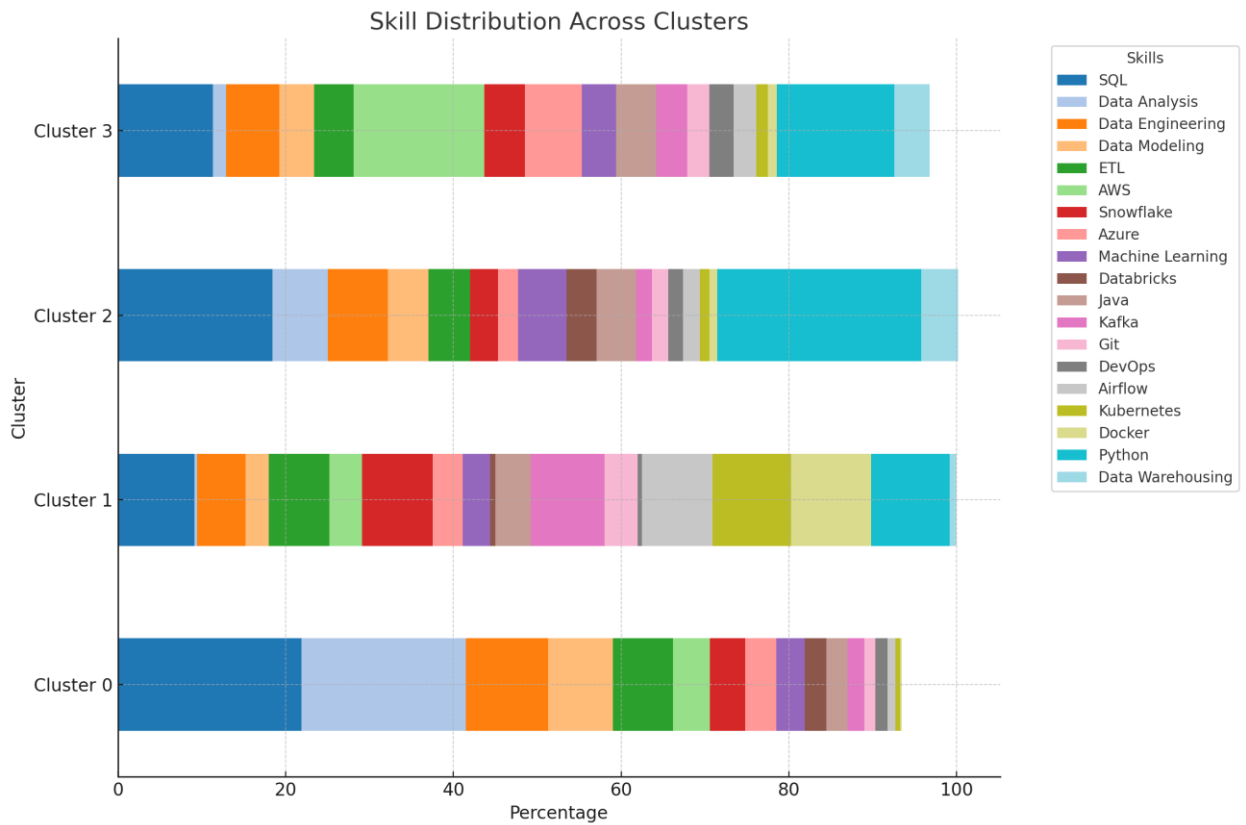


Рис. 3.4. Розподіл навичок фахівців з обробки даних у професійних кластерах

На основі аналізу розподілу навичок було визначено наступні назви для відповідних кластерів (Таблиця 3).

Таблиця 3.3

Опис кластерів

Назва кластеру	Опис кластеру
Data Analyst & Engineer	Навички SQL (21,9%) та Аналізу Даних (19,6%) домінують, підкреслюючи акцент на аналітиці даних та роботі з базами даних. Інженерія Даних (9,8%) та Моделювання Даних (7,7%) вказують на важливість інженерних завдань, пов'язаних з управлінням даними. ETL (7,2%) демонструє роботу з процесами вилучення, трансформації та завантаження даних, що є важливою частиною обов'язків аналітика та інженера даних. Назва відображає інтеграцію аналітичних та інженерних функцій у цьому кластері.
Data Platform Engineer	Важливість навичок контейнеризації та оркестрації: Docker (9,5%) і Kubernetes (9,4%). Програмування та бази даних: Python (9,4%) і SQL (9,1%) є ключовими навичками, які підкреслюють важливість програмування та роботи з базами даних. Наявність Kafka (8,8%),

	Snowflake (8,4%), Airflow (8,4%) та ETL (7,2%) фокусується на обробці та переміщенні даних.
Data Science & Engineering Specialist	Навички Python (24,3%) та SQL (18,4%) домінують, що підкреслює ключову роль програмування та обробки даних. Навички Інженерії Даних (7,2%), ETL (5%) та Моделювання Даних (4,8%) пов'язані з інженерією та управлінням даними. Машинне Навчання (5,8%) фокусується на аналізі даних та застосуванні моделей машинного навчання. Аналіз Даних (6,6%) та Зберігання Даних (4,4%) визначають аналітичні навички та роботу зі сховищами даних, які доповнюють технічний аспект.
Cloud Data Engineer	AWS (15,6%) та Azure (6,8%) разом складають понад 22% вимог, що свідчить про сильний фокус на хмарних платформах. Програмування та управління даними: Python (14,0%) та SQL (11,3%) є другою та третьою найважливішими навичками, підкреслюючи важливість навичок програмування та роботи з базами даних. Різні технології для роботи з даними: Snowflake (4,8%), Kafka (3,7%), Airflow (2,7%) вказують на потребу працювати з сучасними інструментами обробки та аналізу даних у хмарному середовищі. Назва відображає комбінацію хмарних технологій та навичок інженерії даних.

На основі отриманих кластерів професій для фахівців з обробки даних потенційному кандидату на відповідну вакансію необхідно визначити відповідну освітню програму, яка дозволить йому/їй отримати бажану посаду, враховуючи навички, що користуються попитом на ринку праці в кластері відповідних позицій.

3.5 Визначення освітньої програми, що відповідає кластеру позицій фахівця з обробки даних

Підхід до вибору посади на основі набору навичок дозволяє побудувати індивідуальний навчальний шлях, який здобувач роботи може пройти у вищих навчальних закладах або самостійно як неформальну освіту на освітніх платформах (Coursera, Udemy), сертифікати яких визнаються роботодавцями, або як частину стажування в компанії.

Основна ідея вибору освітньої програми, яка дозволила б отримати бажану посаду на ринку праці, полягає в порівнянні частки навичок, представлених як вимоги до жорстких навичок у кластері вакансій, із часткою професійних компетенцій, які формуються в рамках конкретної освітньої програми. Далі необхідно підтвердити надійність відповідності жорстких навичок кластеру вакансій професійним компетенціям освітньої програми, використовуючи статистичний критерій χ^2 .

Розглянемо освітню програму "Інформаційні системи та технології" магістерського рівня Херсонського державного університету (Україна) (<https://ksu24.kspu.edu/s/m69no>) у Таблиці 3.4.

Таблиця 3.4

Освітні компоненти магістерської програми з інформаційних систем і технологій (ОК – обов'язковий компонент/курс, ВК – вибіркового компонент/курс)

№	Дисципліни освітньої програми	ECTS
ОК 3	Формальні методи інженерії програмного забезпечення	3
ОК 4	Управління технологіями розробки фінансових інструментів	3
ОК 5	Цифрові валюти та блокчейн-технології	3
ОК 6	Моделі прогнозування часових рядів для бізнес-аналітики	6,5
ОК 7	Моделювання та проектування інформаційних систем	3
ВК	ЕС 7, ЕС 8, ЕС 9, ЕС 10	3*4

Порівняємо частку навичок кластера 2 "Спеціаліст з науки та інженерії даних" із професійними компетенціями, що відповідають цій освітній програмі. Для визначення частки професійних компетенцій ми використаємо кількість кредитів за освітніми компонентами. Результати порівняння наведено в Таблиці 3.5.

Порівняємо частку навичок кластера 2 "Спеціаліст з науки та інженерії даних" із професійними компетенціями, що відповідають цій освітній програмі. Для визначення частки професійних компетенцій ми використаємо кількість кредитів за освітніми компонентами. Результати порівняння наведено у Таблиці 3.5.

Для перевірки гіпотези про те, що фактичний і очікуваний розподіли не є випадково подібними, застосуємо критерій χ^2 . Кроки для обчислення критерію χ^2 :

1. Обчислення фактичного χ^2 для кожної категорії навичок за формулою:

$\chi^2 = \frac{(SS - Av)^2}{Av} = 22.1$ де SS — частка навичок, CS — частка компетенцій для конкретних предметів, $Av = \frac{SS + CS}{2}$ — середнє (очікуване) значення між часткою навичок і компетенцій для кожної категорії у таблиці 4.

2. Порівняємо фактичне значення з критичним значенням:

$$\chi_{cr}^2 = CHISQ.INV.RT(5\%; 18 - 1) = 27.587$$

Оскільки $\chi^2 = 22.1 < 27.587 = \chi_{cr}^2$, ми не можемо відхилити нульову гіпотезу. Це означає, що різниця між фактичним і очікуваним розподілом часток навичок не є значущою, і відстань між частками можна пояснити випадковими коливаннями, тобто ці частки однакові.

Таблиця 3.5

Розрахунок часток навичок і компетентностей

№	Cluster 2	Number	Skill share, SS	ECTS	Competences share, CS	Average, Av	χ_{fact}^2	Courses
1	Python	1409	24.3%	5	13.5%	18.9	1.54	OC 3, EC 7.1, EC 8.4
2	SQL	1067	18.4%	4	10.8%	14.6	0.99	OC 4, EC 8.3
3	Data Engineering	416	7.2%	6	16.2%	11.7	1.74	OC 3, EC 7.1, EC 8.1, EC 9.2
4	Data Analysis	384	6.6%	5.5	14.9%	10.7	1.59	OC 6, EC 7.4, EC 9.1.
5	Machine Learning	335	5.8%	5.5	14.9%	10.3	1.99	OC 6, EC 7.4, EC 10.1.
6	ETL	290	5.0%	4.5	12.2%	8.6	1.49	OC 6, EC 9.1
7	Data Modelling	278	4.8%	4.5	12.2%	8.5	1.60	OC 7, OC 5
8	Java	272	4.7%	0	0.0%	2.4	2.35	
9	Data Warehousing	254	4.4%	1	2.7%	3.6	0.20	EC 8.1
10	Databricks	208	3.6%	0	0.0%	1.8	1.80	
11	Snowflake	193	3.3%	0	0.0%	1.7	1.65	
12	Azure	137	2.4%	0	0.0%	1.2	1.20	

13	Airflow	115	2.0%	0	0.0%	1.0	1.00	
14	Git	110	1.9%	1	2.7%	2.3	0.07	OC 9
15	Kafka	112	1.9%	0	0.0%	1.0	0.95	
16	DevOps	104	1.8%	0	0.0%	0.9	0.90	
17	Kubernetes	61	1.1%	0	0.0%	0.6	0.55	
18	Docker	57	1.0%	0	0.0%	0.5	0.50	

Це означає, що розподіл специфічних предметних компетенцій освітньої програми з інформаційних систем і технологій відповідає вимогам кластера "Спеціаліст з науки та інженерії даних", і цю освітню програму можна обрати для розвитку відповідних навичок з метою отримання бажаної посади на ринку праці.

ВИСНОВКИ

Оптимальна кількість кластерів була визначена на основі відкритих даних про позиції фахівців з обробки даних, методу кластеризації та методу «лікоть»: Аналітик і Інженер з Даних, Інженер Платформи Даних, Спеціаліст з Науки та Інженерії Даних, Хмарний Інженер з Даних. Кожен із цих кластерів має унікальне поєднання від 17 до 20 навичок для фахівців з обробки даних. У цілому кластеризація дозволяє визначити найбільш затребувані групи фахівців з обробки даних з найбільш актуальними навичками. На основі порівняння часток навичок у профілі фахівців кожного кластеру та часток професійних компетенцій в освітніх програмах запропоновано підхід для визначення відповідності між вимогами ринку праці та освітньою програмою для підготовки фахівців з обробки даних. Представлено приклад застосування цього підходу до кластера "Спеціаліст з Науки та Інженерії Даних" і магістерської програми "Інформаційні системи та технології" Херсонського державного університету. Використовуючи статистичний критерій, підтверджено відповідність між структурою кластера

за навичками та структурою освітньої програми за професійними компетенціями.

У майбутньому планується дослідження формування індивідуальної освітньої траєкторії для кандидатів на основі навичок, які їм потрібно розвинути для працевлаштування або підвищення кваліфікації на ринку праці, з використанням системи штучного інтелекту.

Зважаючи на сучасні виклики та вимоги ринку праці, використання кластерного підходу для оцінки професійних компетенцій і складання навчальних програм дає змогу гнучко реагувати на зміни в галузі технологій і надає студентам можливість розвивати найбільш актуальні навички. Такий підхід допомагає університетам і навчальним закладам тісніше співпрацювати з індустрією, розробляючи програми, які відповідають не лише поточним, але й майбутнім потребам ринку праці. Застосування штучного інтелекту у формуванні індивідуальних навчальних траєкторій додатково дозволить автоматизувати процес адаптації освітніх програм, забезпечуючи індивідуальний підхід до кожного студента, орієнтуючись на його професійні цілі.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Tavakoli, M., Kismihok, G., Mol, S.T. «Система персоналізованих рекомендацій на основі інформації ринку праці для безперервних учнів з використанням відкритих освітніх ресурсів (OER)». Праці Міжнародної конференції з підтримки освіти за допомогою комп'ютерів (CSEDU). SciTePress, 2020. DOI: 10.48550/arXiv.2005.07465.
2. Wang, F., Jiang, Z., Li, X., Li, G. «Когнітивні фактори передачі емпіричних інженерних знань: поведінкове та fNIRS-дослідження». *Advanced Engineering Informatics* 2021; 47: 101207. DOI: 10.1016/j.aei.2020.101207.
3. Tavakoli, M., Faraji, A., Mol, S. T., Kismihók, G. «Рекомендації OER для підтримки кар'єрного розвитку». Праці Конференції IEEE Frontiers in Education (FIE): 21-24 жовтня 2020 року, Уппсала, Швеція, 2020. том 50. IEEE. DOI: 10.1109/FIE44824.2020.9274175.
4. Colombo, E., Mercorio, F., Mezzanzanica, M. «Застосування інструментів машинного навчання для аналізу вакансій в Інтернеті щодо ринку праці та навичок. В «Термінатор чи Джетсони?». Економічні та політичні наслідки штучного інтелекту 2018.
5. Li, X., Jiang, Z., Guan, Y., Li, G., Wang, F. «Сприяння передачі емпіричних інженерних знань в умовах зміни технологічної парадигми: експериментальне дослідження на етапі концептуального проектування». *Advanced Engineering Informatics* 2019; 41: 100927. DOI: 10.1016/j.aei.2019.100927.
6. Ermolayev, V., Suarez-Figueroa, M. C., Molchanovskyi, O. «Архітектура освіти з науки про дані». Праці 14-ї Міжнародної конференції з інформаційних та комунікаційних технологій в освіті, науці та промисловості. ICTERI 2018. CEUR-WS, 2018. том 2104, сс. 734-746, Університет Ахена. Доступно за адресою: https://ceur-ws.org/Vol-2104/paper_266.pdf.

7. Molavi, M., Tavakoli, M., Kismihók, G. «Виділення тем з відкритих освітніх ресурсів». Праці Європейської конференції з технологій, що підсилюють освіту. Springer, 2020, сс. 455-460. DOI: 10.1007/978-3-030-57717-9_44.
8. Kobets, V., Osypova, N. V. «Ідентифікація факторів забезпечення якості вищої освіти для студентів». International Journal for Quality Research 2023; 17(1): 195–208. DOI: 10.24874/ijqr17.01-12.
9. Ketamo, H., Moisiu, A., Passi-Rauste, A., Alamäki, A. «Картування майбутніх навчальних програм: Використання штучного інтелекту та аналітики для прогнозування потреб у компетенціях». Праці 10-ї Європейської конференції з нематеріальних активів та інтелектуального капіталу 2019, сс. 144-153.
10. Metrólho, J.C., Ribeiro, F.R., Batista, R. «Підготовка студентів до індустрії програмного забезпечення: кейс на прикладі проекту Full Stack з використанням Agile». ICSEA 2022: Праці Сімнадцятої міжнародної конференції з питань прогресу в розробці програмного забезпечення 2022, сс. 75-80.
11. Kobets, V., Yatsenko, V., Buiak, L. «Подолання прогалин у компетенціях бізнес-аналітиків: потреби ринку праці проти освітніх стандартів». Праці конференції ICTERI 2020. CCIS, 2021. том 1308, сс. 22-45. Springer, Cham, DOI: 10.1007/978-3-030-77592-6_2.
12. Kara, A., Daniş, F.S., Orman, G.K., Turhan, S.N., Özlü, A. «Рекомендації роботи на основі витягнутих навичок». Lecture Notes in Networks and Systems 2022. сс. 497-507. DOI: 10.1007/978-3-031-16075-2_35.
13. Ramakrishnan, M., Gregor, S., Shrestha, A., Soar, J. «Досягнення освіти, що відповідає вимогам індустрії, через цифрові коммони: кейс». Journal of Computer Information Systems 2022; 63(4): 950-964. DOI: 10.1080/08874417.2022.2115955.
14. Kobets, V.; Tsiuriuta, N.; Lytvynenko, V.; Mykhaylova, V. «Система управління веб-сервісом для пошуку роботи з використанням компетентнісного підходу». Праці 16-ї Міжнародної конференції з

- інформаційних та комунікаційних технологій в освіті, науці та промисловості. ICTERI 2020. CEUR-WS, 2020. том 2732, сс. 290-302. Університет Ахена. Доступно за адресою: <https://ceur-ws.org/Vol-2732/20200290.pdf>.
15. Wu, W. «Дослідження досвіду стажувань студентів з науки про дані для вдосконалення навчальних програм». Праці 27-ї конференції ACM з інновацій та технологій у комп'ютерній науці 2022. том 1, сс. 505-511. DOI: 10.1145/3502718.3524741.
16. Kravtsov, H., Kobets, V. «Еволюційна модель перегляду для покращення навчальної програми з комп'ютерних наук». Праці конференції ICTERI 2018. CCIS, 2019. том 1007, сс. 127-147. Springer, Cham. DOI: 10.1007/978-3-030-13929-2_7.
17. Berger, T., Frey, S.V. «Діджиталізація, робочі місця та конвергенція в Європі: стратегії закриття прогалів у навичках». Oxford Martin School (2016). Доступно за адресою: https://oms-www.files.svdcdn.com/production/downloads/reports/SCALE_Digitalisation_Final.pdf. Дата доступу: 14.09.2024.
18. Moldovan, L. «Аналіз сучасного стану знань і навичок на тему Індустрії 4.0 та вимог до навчання на основі роботи». Procedia Manufacturing 2019; 32: 294-301. DOI: 10.1016/j.promfg.2019.02.217.
19. Short, M.N., Keller-Bell, Y. «Основні навички для робочої сили 21 століття». В IGI Global eBooks, сс. 97-110 (2021). Доступно за адресою: <https://doi.org/10.4018/978-1-7998-3022-1.ch006>.
20. Kobets, V.; Yatsenko, V.; Vuiak, L. «Визначення прогалів у підготовці бізнес-аналітика між вимогами ринку праці та стандартами навчальних програм: приклад України». Праці 16-ї Міжнародної конференції з інформаційних та комунікаційних технологій в освіті, науці та промисловості. ICTERI 2020, CEUR-WS, 2020. том 2732, сс. 499-514, Університет Ахена. Доступно за адресою: <https://ceur-ws.org/Vol-2732/20200499.pdf>.

21. Manyika, J., Lund, S., Chui, M., Bughin, J., Woetzel, J., Batra, P., Sanghvi, S. «Втрачені робочі місця, нові робочі місця: трансформації робочої сили в епоху автоматизації». McKinsey Global Institute (2017). Доступно за адресою:
<https://www.mckinsey.com/~media/BAB489A30B724BECB5DEDC41E9BB9FAC.ashx>, дата останнього доступу 08/11/2024.
22. Agarwal, A., Ojha, R. «Пріоритетизація впливу Індустрії 4.0 на цілі сталого розвитку: перспектива аналітичного ієрархічного процесу в виробничих операціях». *Journal of Cleaner Production* 2024; 444: 141189. DOI: 10.1016/j.jclepro.2024.141189.
23. Li, G., Yuan, C., Kamarthi, S., Moghaddam, M., Jin, X. «Навички науки про дані та вимоги до знань у виробничій промисловості: аналіз прогалин». *J. of Manufact. Systems* 2021; 60: 692-706. DOI: 10.1016/j.jmsy.2021.07.007.
24. Novak, O., Kobets, V. «Вплив штучного інтелекту на продовольчу безпеку держав у світі». Праці Міжнародної конференції з інформаційних та комунікаційних технологій в освіті, науці та промисловості. *ICTERI 2023. CCIS, 2023. том 1980, сс. 240-251. Springer, Cham. DOI: 10.1007/978-3-031-48325-7_18.*
25. Vista, A. «Ідентифікація навичок майбутнього на основі даних: навички 21-го століття для робочої сили 21-го століття». *SAGE Open* 2020; 10(2): 215824402091590. DOI: 10.1177/2158244020915904.
26. Kobets, V., Tsiuriuta, N., Lytvynenko, V., Novikov, M., Chizhik, S. «Система управління веб-сервісом для пошуку роботи з використанням компетентнісного підходу для виробничих підприємств». Праці *Advances in Design, Simulation and Manufacturing II. DSMIE 2019. Lecture Notes in Mechanical Engineering, 2020, сс. 138–148. Springer, Cham. DOI: 10.1007/978-3-030-22365-6_14.*