

## МЕНЕДЖМЕНТ

УДК 005.95.96

Б. М. Бойко<sup>1</sup>, І. С. Процик<sup>2</sup>

ORCID ID: <sup>1</sup> 0009-0002-9850-2764, <sup>2</sup>0000-0002-6370-1344

Національний університет “Львівська політехніка”

Кафедра менеджменту та міжнародного підприємництва

### ЗАСТОСУВАННЯ СИСТЕМИ ТЕСТОВОГО ОПИТУВАННЯ НА ОСНОВІ КЛАСТЕРНОГО АНАЛІЗУ ТА МАШИННОГО НАВЧАННЯ В ЗАДАЧАХ ПРОФВІДБОРУ ФАХІВЦІВ

<https://doi.org/10.23939/smeu2024.02.001>

© Бойко Б.М., Процик І.С., 2024

Дослідження спрямоване на розробку тестового опитування для ефективного відбору фахівців в ІТ-галузі, що базується на використанні сучасних методів машинного навчання, зокрема кластерного аналізу з застосуванням методу k-середніх. Зважаючи на обмежений доступ до існуючих тестувальних платформ, які зазвичай доступні лише для великих компаній на платній основі, було прийнято рішення створити альтернативний веб-додаток. Цей додаток стане доступним інструментом для більш широкого кола користувачів та дозволить автоматизувати процес оцінки навичок кандидатів. Ключовою особливістю дослідження є застосування кластерного аналізу для групування користувачів за їхніми професійними навичками, когнітивними здібностями та психологічними характеристиками. Це забезпечує можливість точнішої оцінки відповідності кандидатів вимогам роботодавців та сприяє кращій організації даних для подальшого аналізу. Дослідження також підкреслює важливість кластерного аналізу у випадках, коли попередня гіпотеза про структуру даних відсутня, що робить цей підхід універсальним інструментом для класифікації даних. Окрім технічних аспектів, дослідження охоплює перспективи застосування адаптивних тестів, які можуть змінювати рівень складності в реальному часі залежно від відповідей користувача. Це дозволяє підвищити точність оцінювання, а також зменшити вплив суб'єктивних чинників під час відбору. Додатково розглянуто можливість аналізу поведінкових та емоційних характеристик кандидатів, таких як стійкість до стресу та комунікативні здібності, що є важливими для успішної роботи в команді.

**Ключові слова:** інформаційна система, тестування, профвідбір, кластеризація.

#### Постановка проблеми

Завданням дослідження є розробка тестового опитування для відбору фахівців у сфері ІТ. Існуючі тести цього типу зазвичай доступні лише для компаній і вимагають платної підписки, тому було вирішено створити альтернативний веб-додаток. Для аналізу результатів тестування користувачів

буде використовуватися машинне навчання, зокрема кластерний аналіз за допомогою центроїдної моделі (метод k-середніх) [1, 4].

Багато досліджень спрямовані на те, щоб організувати отримані дані у зручні для огляду структури. Кластерний аналіз є потужним інструментом, оскільки включає різні алгоритми класифікації. Цей метод широко застосовується в різних галузях, коли потрібно розділити великі обсяги інформації на групи для подальшого аналізу. Особливо цікавим є його використання в психологічних дослідженнях групових процесів і явищ. Суттєва особливість кластерного аналізу полягає в тому, що він не є традиційним статистичним методом, оскільки більшість процесів перевірки статистичної значущості до нього непридатні. Проте кластерний аналіз забезпечує найбільш значиме рішення, тому його часто застосовують, коли існують дані, але немає попередньої гіпотези про класи цих даних.

### **Актуальність дослідження**

Актуальність даного дослідження полягає у зростаючій необхідності вдосконалення процесів відбору фахівців в ІТ-галузі. Сучасний ринок праці характеризується високою конкуренцією серед кандидатів, і компанії все більше покладаються на тести для оцінки не лише технічних знань, але й когнітивних та особистісних характеристик. Більшість існуючих платформ для тестування доступні лише великим компаніям та на платній основі, що обмежує можливості для малого та середнього бізнесу. Тому створення доступної альтернативи у вигляді веб-додатку з використанням машинного навчання та кластерного аналізу є важливим завданням.

Застосування кластерного аналізу, зокрема методу **k-середніх**, дозволяє ефективно аналізувати великі обсяги даних кандидатів та групувати їх за рівнем професійних навичок, інтелектуальних здібностей і навіть психологічних характеристик. Це забезпечує точнішу оцінку компетенцій кандидатів, що допоможе підвищити якість процесу відбору та сприятиме кращому відповідності кандидатів потребам компаній.

Крім того, зростає попит на автоматизовані системи оцінювання, які можуть адаптуватися до швидко мінливих вимог ринку ІТ-послуг. Це робить дослідження особливо актуальним в умовах цифрової трансформації, коли технологічні рішення відіграють вирішальну роль у підвищенні ефективності бізнес-процесів.

### **Формулювання мети та завдань статті**

Метою даного дослідження є розробка тестового опитування для відбору фахівців в ІТ-сфері. На сьогоднішній день існують подібні тести, але вони зазвичай доступні лише для компаній за передплатою. У зв'язку з цим було прийнято рішення створити альтернативний сервіс — веб-додаток. Для аналізу результатів тестування користувачів використовується машинне навчання, зокрема метод k-середніх у кластерному аналізі.

### **Аналіз останніх досліджень і публікацій**

Аналіз літератури по даному дослідженню включає огляд джерел, що стосуються кластерного аналізу, машинного навчання та сучасних методів тестування навичок для відбору фахівців в ІТ-галузі.

Кластерний аналіз широко застосовується для класифікації даних у різних галузях. Метод k-середніх є одним із найпоширеніших алгоритмів кластеризації, який ефективно поділяє дані на групи, мінімізуючи внутрішньогрупові відмінності. Роботи таких авторів, як М. Мак-Квін (MacQueen, 1967) та Дж. Ллойд (Lloyd, 1982), заклали фундамент для розвитку цього методу, описавши математичну основу та алгоритмічну реалізацію [2].

Машинне навчання відіграє важливу роль у сучасних рішеннях для класифікації та аналізу даних. У літературі значну увагу приділяють адаптивним моделям, які здатні навчатися на основі великих наборів даних, таких як глибокі нейронні мережі та методи класифікації, які описані в роботах таких авторів, як Ян Лекун (Yann LeCun, 2015) та Дж. Хінтон (Geoffrey Hinton, 2006) [5].

Тести для відбору кадрів на ринку праці є важливим інструментом для оцінки професійних компетенцій. У дослідженнях показано, що такі тести дозволяють роботодавцям ефективніше оцінювати навички кандидатів. Роботи А. Шермана та Д. Мелло (Sherman & Mello, 2020) підкреслюють важливість використання комбінованих тестів, які оцінюють як технічні, так і "м'які" навички, такі як вміння вирішувати проблеми та здатність до навчання [6].

Згідно з дослідженнями таких платформ, як LinkedIn та компаній, що займаються HR-аналітикою, навички вирішення проблем та здатність до навчання визначаються як найважливіші у сучасній економіці знань. Наприклад, в аналітичних звітах LinkedIn (2022) робиться висновок, що адаптивність до нових знань і критичне мислення є ключовими для фахівців у технічних галузях, таких як IT [7].

Загальний аналіз літератури показує, що кластерний аналіз є ефективним інструментом для групування великих обсягів даних і може бути успішно застосований у створенні тестів для відбору фахівців в IT. Метод k-середніх є одним з найбільш популярних підходів у цьому контексті завдяки своїй простоті та ефективності. Крім того, існує значний інтерес до психологічного тестування кандидатів, що робить кластеризацію важливим інструментом для оцінки особистісних характеристик.

### **Виклад основного матеріалу**

Однією з ключових проблем рекрутингу є обрання з великої кількості кандидатів людини, яка найбільше підходить на певну посаду. Під час процесу відбору застосовують різні види тестування: інтелектуальні, кваліфікаційні, особистісні та інші. Завдання найму ще складніше, якщо мова йде про професію галузі IT, де кандидати повинні мати технічні та когнітивні навички, які є необхідні для роботи. HR-менеджери можуть використовувати тести для визначення потенціалу кандидата, а саме його вміння навчатися та використовувати знання в специфіці посади компанії, на яку вони подаються [2].

Як правило, виділяють чотири складові потенціалу:

- Інтелектуальні можливості

Це вміння обробляти інформацію різного виду, аналізувати та синтезувати дані, підтримувати взаємозв'язки, приходити до логічного правильного висновку. Для оцінки інтелектуальних можливостей розробляються спеціальні тести. Сьогодні найкращою практикою вважається використання окремих тестів для оцінювання кожної здатності, оскільки інтегральний «інтелектуальний тест» (наприклад, тест IQ), недостатньо інформативний. Тести на IQ дуже добре оцінюють певні розумові здібності, включаючи логіку, абстрактні міркування, здатність до навчання та можливості робочої пам'яті - скільки інформації ви можете пам'ятати [1, 3].

Але тести не працюють, коли йдеться про вимірювання тих здібностей, які мають вирішальне значення для прийняття правильних рішень у реальних ситуаціях. Це тому, що вони не в змозі оцінити такі речі, як здатність людини критично зважувати інформацію, чи може людина відкидати інтуїтивні когнітивні упередження.

- Особливості особистості

Це риси характеру, схильності, переваги та установки. Ці фактори дозволяють зрозуміти, які компетенції є внутрішньо притаманні людині, яка поведінка в реальних робочих умовах буде для неї природною і характерною. Зрозуміло, люди здатні вести себе по-різному, але в напружених умовах — наприклад, на новій посаді — співробітник в першу чергу буде проявляти ту поведінку, яка йому властива від природи.

- Навички, досвід і знання

Іншими словами, набуті вміння та засвоєна інформація. Цю складову потенціалу оцінити найлегше. Варто врахувати те, що знання і навички — найбільш мінлива складова потенціалу: їх легко отримати і часто легко втратити, якщо не використовувати, тому документальні свідчення і стаж в галузі навряд чи можуть безпосередньо прогнозувати майбутню успішність і ефективність співробітника.

- Мотивація

Предмет, який викликає постійні суперечки серед психологів і фахівців з управління персоналом. Для цілей бізнесу найбільш важливим елементом оцінки мотивації є виявлення значимих для кандидата факторів мотивації під час співбесіди і зіставлення цих даних з реальністю роботи в організації і на даній посаді: чим більше збігів, тим більше ефективним буде кандидат. При цьому набір характерних для людини факторів мотивації може іноді змінюватися.

Відносна вага складових потенціалу може змінюватися від однієї компетенції до іншої: наприклад, для аналітичних компетенцій найважливіше інтелектуальні здібності, а для комунікативних (здатність вступати в письмовій та усній формах у спілкуванні та підтримувати його не нижче тієї межі, поза якої констатується ізоляція індивіда від соціуму) на перший план виходять особистісні особливості. Наявність потенціалу не гарантує успішного і впевненого прояви тієї чи іншої компетенції. Однак низький рівень потенціалу на ділі майже завжди означає, що кандидат буде зазнавати труднощів у цій галузі — іншими словами, ледве зможе проявляти необхідну компетенцію на потрібному рівні і з потрібною ефективністю [8].

Оцінка перерахованих чотирьох компонентів потенціалу є загальноприйнятим підходом, кращою практикою в оцінці зовнішніх і внутрішніх кандидатів, яка дозволяє забезпечити наявність «правильних людей на правильних місцях» в організації.

Існує декілька основних груп тестів, які застосовують для підбору майбутніх працівників:

1. Тести на розумові здібності, які поділяються на:
  - Тести на розумовий розвиток(словниковий запас, пам'ять, швидкість мовлення)
  - Тести здібностей (вміння працювати з текстом, цифрами, мислення)
2. Оцінка особистості та інтересів (емоційна гнучкість, свідомість, екстравертність чи інтровертність, здатність до навчання)
3. Тести на рухові та фізичні можливості
4. Тести на досягнення (знання в конкретній області)
5. Робочі завдання для відбору персоналу

Тести на профпридатність оцінюють критичне мислення, здатність вирішувати проблеми, вчитися та застосовувати нову інформацію. Згідно з дослідженням проведеним LinkedIn двома найважливішими якостями, які прагнуть роботодавці є: навички вирішення проблем (65%) та можливість освоювати нові знання (64%) [1]. Ці здібності складно оцінити, виходячи лише з резюме та співбесіди, і саме тести на профпридатність можуть допомогти з вирішенням даної проблеми. Дані тестування можна використовувати майже в будь-якому професійному контексті, але вони особливо корисні для професій середнього та вищого рівня, де вони дають найточніший прогнози ефективності роботи кандидата.

При оцінці ефективності роботи персоналу враховуються:

- обсяг виконаної роботи;
- складність поставлених завдань;
- особливості покладених функціональних обов'язків;
- результати праці.

Людина може справлятися зі своїми обов'язками, однак ніколи не вкладається в терміни, постійно відволікати колег для допомоги з роботою, в той час як ефективність складається з двох показників:

- Час, витрачений на досягнення результату
- Витрачені ресурси

## Застосування системи тестового опитування на основі кластерного аналізу ....

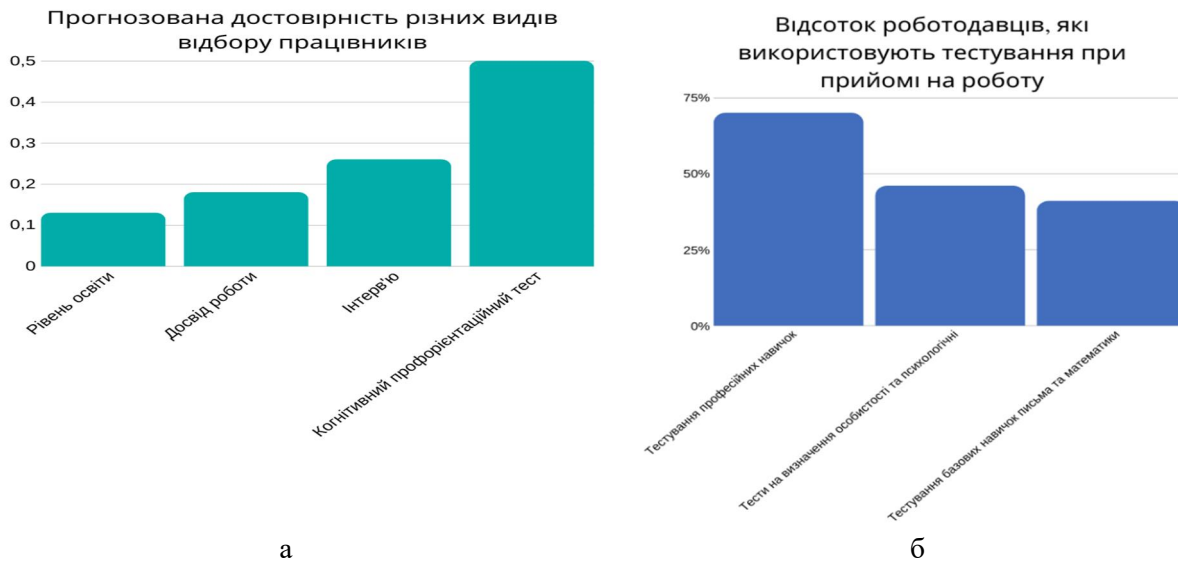


Рис.1. а) Діаграма прогнозованої достовірності різних видів відбору працівників згідно дослідження LinkedIn; б) Діаграма відсотку роботодавців, які використовують тестування при прийомі на роботу

Частота використання тестування перед працевлаштуванням зростає з кожним роком. Завдяки збільшенню кількості кандидатів на позицію через простоту подавання резюме онлайн, HR-менеджери та рекрутери починають більше покладатися на data-driven методи для процесу найму. Згідно опитування проведеного American Management Association (AMA) використання тестування для профвідбору зростає протягом останніх 15 років[2]. Дані AMA показали, що (Рис. 1. б)):

- 70 % роботодавців застосували певний вид тестування навичок роботи
- 46 % роботодавців використовують психологічні тести та / або тести для визначення особистості кандидатів чи теперішніх працівників
- 41 % роботодавців перевіряють базову грамотність та математичні навички кандидатів

Яких результатів слід очікувати від використання тестів до працевлаштування? Важливо мати реалістичні цілі та очікування щодо того, чого може досягати програма тестування перед працевлаштуванням. Використовуючи тестування, роботодавці отримують об'єктивні дані щодо процесу найму. Використання тестів повинно сприяти покращенню результатів найму та мінімізувати ризик негативного досвіду та зайвих витрат.

Однак не менш важливо бути реалістом і розуміти, чого не чекати від тестування при прийому на роботу. Включення тестів у процес найму не означає, що роботодавці після цього ніколи не зроблять помилки а, лише те, що їх буде менше. Жоден тест не є ідеальним. Незважаючи на те, що дослідження показують, що тести є значно більш точними та надійними в якості прогнозів, ніж резюме чи співбесіда, роботодавці повинні усвідомлювати, що не існує єдиної методології відбору, яка б була на 100% точною для прогнозування ефективності.

Як результат, тести до працевлаштування повинні бути лише одним елементом у всебічному наборі критеріїв, що використовуються для оцінювання претендентів, включаючи резюме, співбесіди, досвід роботи, освіту та все інше, що має значення для посади. Тести до працевлаштування надають найбільшу цінність, коли застосовуються у верхній частині процесу найму, щоб відстежувати кандидатів, які не підходять. Однак у кінцевому рахунку організації, які використовують тестування, приймають свої остаточні рішення на основі багатьох факторів, серед яких тестування має бути важливою складовою. Компанії повинні розуміти, що тести впорядковують та вдосконалюють процес найму, а не замінюють [3, 5].

Для опису структури організації даних побудовано дві UML-діаграми, а саме використання та діяльності (Рис. 2-3).

Згідно з першою діаграмою (Рис. 2) в системі є користувач, який має змогу зареєструватися та авторизуватися. Користувач може вибрати тестування і розпочати його. Після проходження тестування користувач відправляє відповіді на тесту. Після аналізу відповідей система тестування відправляє результат користувачу. На основі цього результату користувач може зробити висновки в якому напрямку навчання чи працевлаштування йому варто орієнтуватись (Рис. 3).

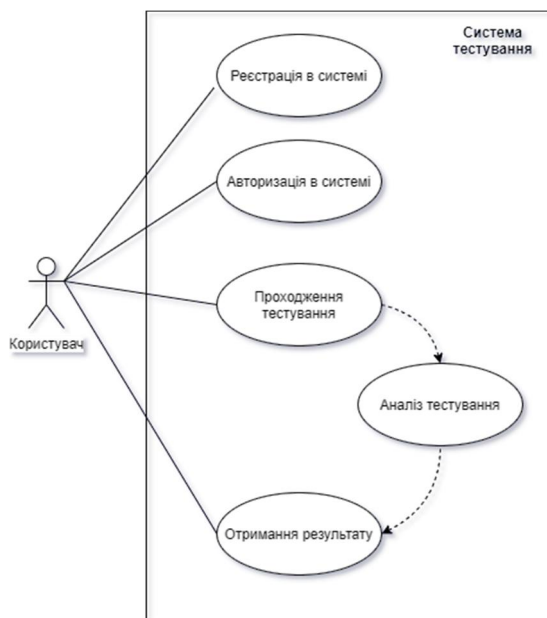


Рис. 2. UML-діаграма використання

Тест особистості — це оцінка, яка використовується роботодавцями, щоб допомогти знайти кандидата, чії риси характеру найкраще відповідають конкретній посаді. Тестування до працевлаштування покликане розкрити конкретні аспекти особистості кандидата та оцінити ймовірність того, що він чи вона підходить цій посаді.

У сучасній культурі праці на основі метрики роботодавці шукають інструмент підбору персоналу, який дає їм кількісно оцінювані результати, на яких можна базувати рішення. Тести особистості на роботі до працевлаштування проходять онлайн, де вони обробляються миттєво. Потім результати перевіряються та нормуються стосовно тисяч інших кандидатів, прискорюючи процес найму та гарантуючи, що кандидати, які рухаються вперед, сумісні з компанією.

Існує кілька поширених версій цих тестів:

#### 1. The Caliper Profile

The Caliper Profile вимірює, як риси особистості індивіда співвідносяться з його роботою. Тест складається з декількох типів запитань. Найпоширеніший тип представляє ряд тверджень, і завдання полягає у визначенні, який вислів найкраще узгоджується з вашою точкою зору.

І навпаки, є також питання, які вимагають визначити твердження, які найменше відображають вашу точку зору. Можна зіткнутися з істинними / неправдивими запитаннями та запитаннями з численним вибором, щоб відповісти за шкалою "ступеня згоди", починаючи від "сильно погоджуючись" до "сильно не згоден". The Caliper Profile відрізняється від інших тестів особистості тим, що він вивчає як позитивні, так і негативні якості, забезпечуючи тим самим повну картину кандидата.

#### 2. The Myers-Briggs Type Indicator

Одним з найвідоміших інструментів для відображення особистості працівників є індикатор типу Маєрс-Бріггс (МВТІ). За версією СРІ, видавця тесту, 89 компаній із Fortune 100 використовують МВТІ, перш ніж наймати нового працівника.

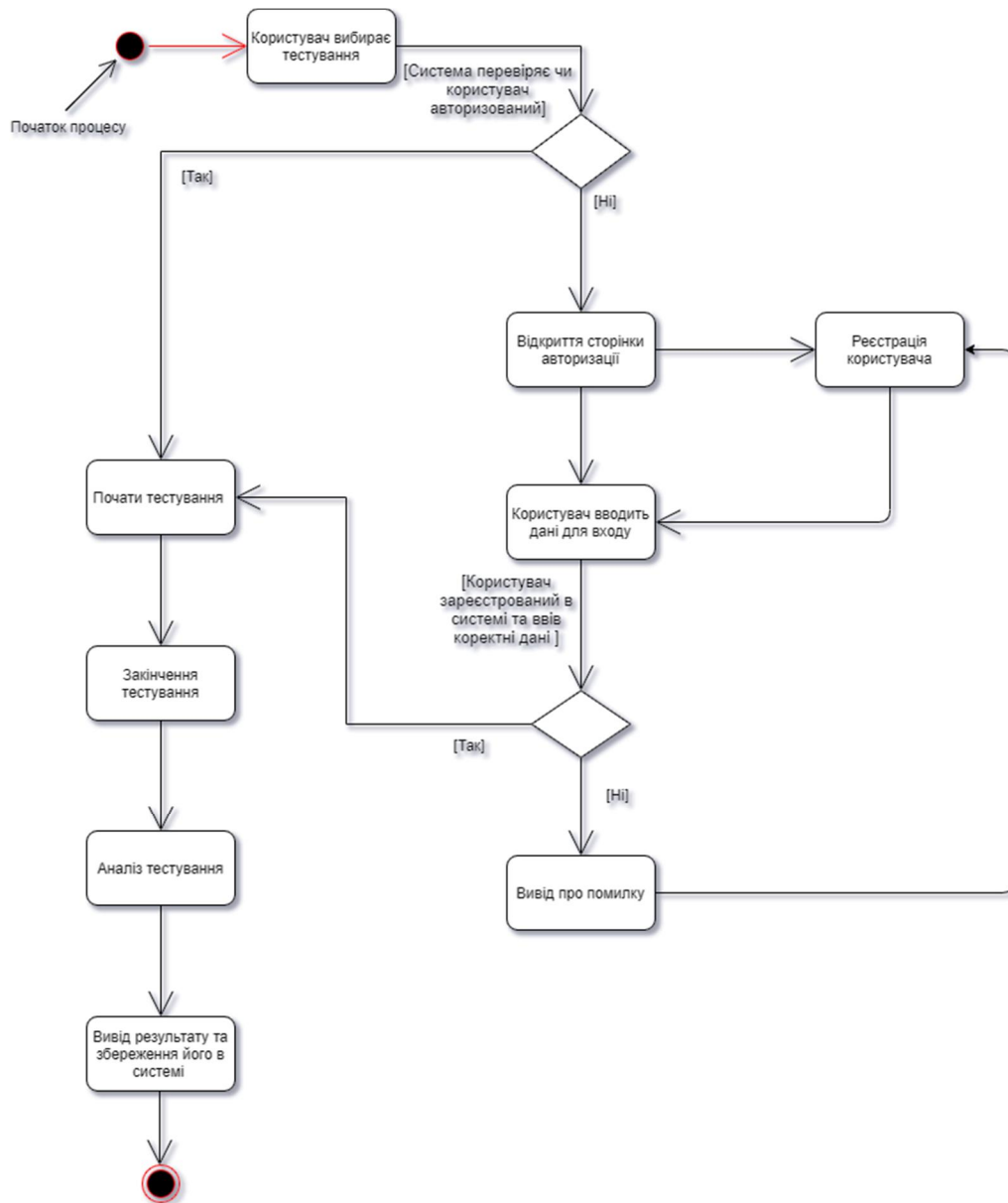


Рис. 3. UML-діаграма діяльності

МВТІ визначає, чи особистість працівника схиляється до однієї з двох тенденцій у наступних групах: «Екстраверсія проти інтроверсії», «Інтуїція проти чутливості», «Мислення проти відчуття» та «Судження проти сприйняття». В результаті індивід може потрапити до одного з 16 типів особистості. Індикатор типу Майерса-Брігга часто використовується роботодавцями для вирішення питання про те чи зможе кандидат згодом перейти на роботу з командою. МВТІ складається з 93 питань. Відповідаючи на кожне запитання, ви отримуєте два варіанти тверджень : або А, або В , які визначають, до яких тенденцій ви схиляєтесь.

### 3. The SHL Occupational Personality Questionnaire

SHL Occupational Personality Questionnaire або OPQ32 створена для того, щоб представити компаніям уявлення про те, як певна поведінка впливає на роботу кандидата. Тест складається з 104 питань, які вимірюють 32 різні характеристики. Кандидати оцінюються за трьома основними сферами: «Відносини з людьми», «Мислення та почуття» та «Емоції». Під час тесту кандидатам пропонують чотири твердження, і вони повинні вибрати, яке твердження найкраще їх описує, а яке твердження

найменше описує. OPQ32 був спеціально розроблений для того, щоб гарантувати, що його масштаби є відповідними та придатними для робочого місця.

#### 4. The Hogan Personality Inventory (HPI)

The Hogan Personality Inventory (HPI) був розроблений у 1980-х роках і спочатку використовувався в соціально-аналітичному контексті, але зараз він використовується для прогнозування ефективності роботи. Він був застосований на більш ніж 500 000 кандидатів по всьому світі та затверджений на більш ніж 200 професіях, що підтверджує, що це послідовний та надійний інструмент, який використовується для оцінки темпераменту людини та того, як він відповідає вимогам певної позиції. HPI заснований на п'яти факторній моделі. Він складається з 206 правдивих / неправдивих питань, які необхідно виконати приблизно за 15 до 20 хвилин. HPI оцінює сім первинних шкал та шість професійних шкал ("Орієнтація на обслуговування", "Стрес толерантність", "Надійність", "Кліричний потенціал", "Потенціал збуту" та "Управлінський потенціал"), крім того, що має 42 підсклади. Він проводиться в режимі онлайн за допомогою миттєвого звіту.

#### 5. The DiSC Behavior Inventory

The DiSC Behavior Inventory (DiSC) вимірює основні риси кандидата на основі чотирьох типів особистості. Ця модель поведінки у чотирьох стилях — це найдавніший стиль перевірки особистості; це було ще з часів Гіппократа, близько 400 до н.е. Профіль особистості DiSC представлений у багатьох версіях, кожен з яких включає в себе варіацію чотирьох основних факторів DiSC: "Домінант (D)", "Вплив (I)", "Сталий (S)" і "Сумісний (C)". Компанії використовують DiSC як інструмент, щоб допомогти зрозуміти стиль професійної поведінки працівника та його здатність працювати як частина команди. DiSC є надзвичайно зручним для користувача тестом, і він значно коротший, ніж інші тести, коливаючись від 12 до 30 питань. Кандидати отримують прикметники чи фрази, і їх просять вибрати, які, на їх думку, стосуються їх найбільше і найменше.

Керуючи підприємствами, прийнято використовувати традиційні форми концептуального моделювання. В них враховуються інформаційні потоки, які проходять між основними ланками, і лише певні потоки здатні взаємодіяти між собою, при цьому утворюючи основну інформаційну систему (IC) для підтримки і прийняття рішень. Нові завдання застосування IC дають змогу масштабувати інформаційні процеси в бік обробки транзакцій і зв'язків між ними. Для ефективного інформаційного управління варто застосовувати процес моделювання, як спосіб покращення інформативності процесів керування, а також як засіб відокремлення інформаційних рівнів організаційної структури підприємства.

Моделювання даних - це спосіб документування системи програмного забезпечення за допомогою діаграм взаємозв'язків між сутностями (ER Diagram), що є представленням структури даних у таблиці для бази даних компанії. Моделі даних застосовуються для досягнення багатьох цілей, від концептуальних моделей високого рівня, логічних до фізичних моделей даних і зазвичай представлених діаграмою взаємозв'язків сутності. Модель даних є інструкцією, яка використовується аналітиками баз даних та розробниками програмного забезпечення при розробці та впровадженні системи з вибраною базою даних.

Модель «сутність-зв'язок» (ERD) - це модель даних, яка зображує інформацію концептуальних схем, використовуючи конструкції блоків. Таке «зображення» служить для двох цілей. Вона дозволяє розробникам бази даних коротко, але точно описати загальний дизайн. Діаграма ER може бути легко перетворена в реляційну схему. Всього існує три компоненти в ERD: сутності, атрибути та відносини.

Кількість таблиць, необхідних для бази даних — Entities(сутності) — це основні об'єкти ERD. Конкретний приклад сутності називається екземпляром. Кожен екземпляр стає записом або рядком у таблиці. У наведені нижче ERD-діаграмі є 3 сутності: User, Result, Test (Рис. 4).



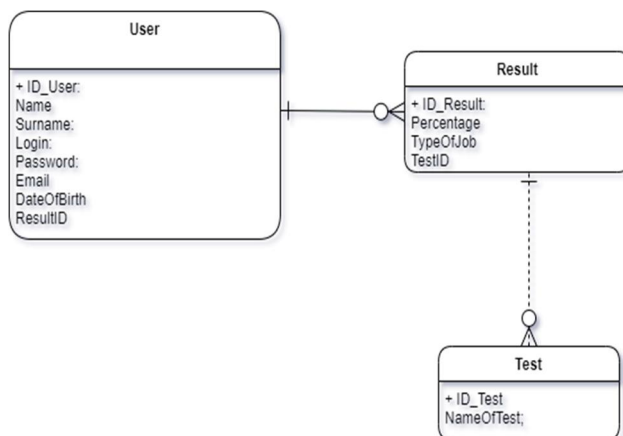


Рис. 4. Модель «сутність-зв'язок» (ERD)

Суть трьох моделей даних полягає в тому, що бізнес-аналітик використовує концептуально-логічну модель для моделювання, існуючих в системі бізнес-об'єктів. Тоді як розробник баз даних або інженер бази даних розробляє концептуальну та логічну модель ER для створення фізичної моделі, яка представляє фізичну структуру бази даних готову для реалізації. У Таблиці 1 нижче показана різниця між трьома моделями даних. Модель ER зазвичай складається на трьох рівнях абстракції:

- Концептуальна модель ERD / концептуальні дані
- Логічна модель ERD / Логічні дані
- Фізична ERD / модель фізичних даних

Хоча всі три рівні моделі ER містять сутності з атрибутами та взаємозв'язками, вони відрізняються між собою цілями, для яких вони створені, та аудиторіями, на які вони призначені.

Таблиця 1

**Порівняння процесів трьох моделей даних**

ERD характеристика	Концептуальна	Логічна	Фізична
Організація(назва)	Так	Так	Так
Відношення	Так	Так	Так
Стовпець		Так	Так
Тип стовпця		Вибірково	Так
Первинний ключ			Так
Зовнішній ключ			Так

Логічна ERD - це детальна версія концептуальної ERD. Логічна ER модель розробляється для деталізації концептуальної моделі шляхом чіткого визначення стовпців у кожному об'єкті та введення операційних та транзакційних сутностей [5].

Фізична ERD являє собою фактичний проект реляційної бази даних. Фізична модель даних розробляється на основі логічної модель даних шляхом присвоєння кожному стовпцю типу, довжини, значенню NULL і т.д. Нижче зображено фізичну ERмодель даних, в якій було добавлено ще одну таблицю“Category”.В згаданій таблиці містяться назва категорії та ідентифікатор [3].

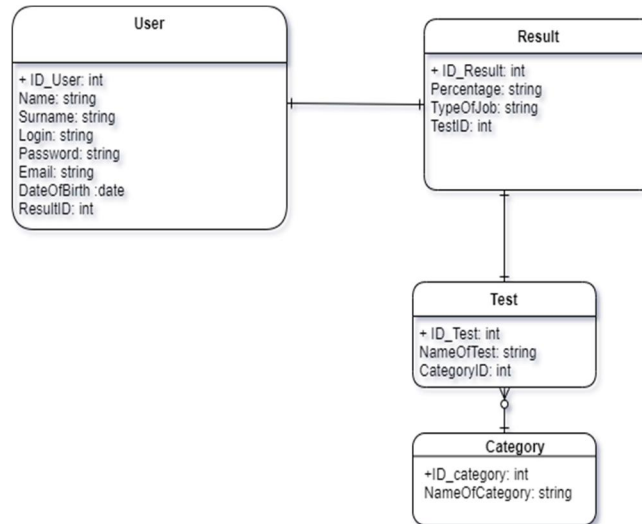


Рис. 5. Фізична ERD

Кластерний аналіз (КА) будує систему класифікації досліджуваних об'єктів і змінних у вигляді дерева (дендрограми) або здійснює розбиття об'єктів на задане число віддалених один від одного класів.

Необхідність у застосуванні методів кластерного аналізу з'являється в тому випадку, коли задано безліч характеристик, за якими тестується багато випробовуваних; суть завдання полягає у виділенні класів (груп) випробовуваних, які є близькими по безлічі характеристик (профілю). На першому етапі матриця змішування (оцінки людей по різних характеристиках) перетворюється в матрицю відстаней. Для підрахунку матриці відстані здійснюється підбір метрики, або методу обчислення відстані між об'єктами в багатовимірному просторі. Якщо об'єкт описується  $k$  ознаками, то він може бути представлений як точка в  $k$ -вимірному просторі. Можливість вимірювання відстаней між об'єктами в  $k$ -вимірному просторі вводиться через поняття метрики.

Нехай об'єкти  $i$  і  $j$  належать множині  $M$  і кожен об'єкт описується  $k$  ознаками, тоді будемо говорити, що на множині  $M$  задана метрика, якщо для будь-якої пари об'єктів, що належать множині  $M$ , визначено невід'ємне число  $d_{ij}$ , яке задовольняє наступним умовам (аксіомам метрики) [1]:

1. Аксіома тотожності:  $d_{ij} = 0 \Leftrightarrow i \equiv j$ .
2. Аксіома симетричності:  $d_{ij} = d_{ji} \forall i, j$ .
3. Нерівність трикутника:  $\forall i, j, z \in M$ , виконується нерівність  $d_{iz} \leq d_{ij} + d_{zi}$ .

Відстані, обчислювані на основі коефіцієнта кореляції, відображають узгодженість коливань оцінок, на відміну від метрики Евкліда, яка визначає схожість в середньому. Вибір метрики визначається завданням дослідження і типом даних. Крім наведених вище методів, розроблені метрики для рангових і дихотомічних змінних і т.д. (У всіх вище наведених формулах  $i, j$  - номери стовпців;  $k$  - номер рядка;  $d_{ij}$  - елемент матриці відстаней;  $x_{ik}, x_{jk}$  - елементи вихідної матриці;  $n$  - кількість об'єктів).

Крім агломеративних ієрархічних методів існує також велика кількість ітеративних методів кластерного аналізу. Основна відмінність полягає в тому, що процес класифікації починається з завдання початкових умов: це може бути число класів, критерії завершення класифікації і т.д. До таких методів відносяться, методи  $k$ -means та інші, які вимагають від дослідників інтуїції та творчого підходу. Ще до проведення класифікації необхідно уявляти, скільки класів повинно бути утворено, коли закінчити процес класифікації і т.д. Від правильно вибраних початкових умов буде залежати результат класифікації, оскільки некоректно обрані умови можуть привести до «розмитості» класів. Таким

чином, ці методи використовуються, якщо є теоретичне обґрунтування, наприклад, кількості очікуваних класів, а також після проведення ієрархічних методів класифікації, які дозволяють виробити найбільш оптимальну стратегію дослідження.

Метод *k*-means можна віднести до ітеративним методам еталонного типу. Назва йому було дано Дж. Мак-Куїном. В даному випадку для аналізу набору даних було використано метод *k*-means. Метод *k*-середніх - це метод кластерного аналізу, метою якого є поділ *m* спостережень (з простору) на *k* кластерів, при цьому кожне спостереження відноситься до того кластеру, до центру (центроїду) якого воно найближче.

В якості міри близькості використовується Евклідова відстань, яку було вище згадано.

Метод *k*-means розділяє *m* спостережень на *k* груп (або кластерів) ( $k \leq m$ ),  $S = \{S_1, S_2, \dots, S_k\}$ , щоб мінімізувати сумарне квадратичне відхилення точок кластерів від центроїдів цих кластерів [1]:

$$\min \left[ \sum_{i=1}^k \sum_{x \in S_i} \|x^{(j)} - \mu_i\|^2 \right], \text{ де } x^{(j)} \in R^n, \mu_i \in R^n$$

$\mu_i$  – центроїд для кластеру  $S_i$

Дані спостереження були поділені на 4 кластери, які відповідають за non-IT(людина, яка не підходить), SE(SoftwareEngineer), QA(Quality Assurance engineer), PM(Project Manager).

Для реалізації було використано бібліотеки pandas, numpy, matplotlib.pyplot. Також був застосовано модуль metrics, для Евклідової відстані і готову функцію sklearn.cluster.KMeans. Нижче зображено результат роботи алгоритму на основі датасету і частинку самого датасету. В датасеті містяться номери варіантів відповідей на 62 питання (Рис. 6).

NON-IT	1	1	1	3	1	3	3	2	1	0	2	1	0	0	0	4
NON-IT	1	0	0	4	2	2	2	3	2	0	0	0	2	0	1	4
NON-IT	0	2	1	4	0	4	3	4	0	2	0	0	1	2	2	3
NON-IT	1	2	0	3	1	2	3	4	2	2	1	2	2	1	0	3
NON-IT	4	2	2	2	2	2	2	2	2	0	2	1	2	3	2	2
NON-IT	0	3	2	3	0	3	2	3	2	1	0	0	2	0	2	4
PM	3	4	3	2	3	1	1	2	3	4	3	4	3	4	4	0
PM	4	3	3	1	4	0	2	1	4	4	4	3	4	2	3	1
PM	4	4	4	0	3	0	0	0	3	3	3	4	3	4	4	0
PM	4	3	4	1	3	1	1	1	4	4	2	4	4	4	4	0
PM	3	4	4	0	4	1	0	0	4	4	3	3	4	4	4	0
PM	4	3	4	1	2	0	1	1	4	4	4	4	4	4	3	1
PM	3	4	3	0	3	0	3	1	4	4	4	3	4	4	4	0

Рис. 6. Частина датасету



Рис. 7. Результат *k*-means кластеризації

Для кластеризації даних було використано датасет з 300 рядками, в яких містяться відповіді кандидатів на 62 питання тестування. Внаслідок чого дані було поділено на 4 кластери: поп-ІТ, SE, QA, РМ. Було використано мову програмування Python та її бібліотеки з функціями для аналізу даних.

Отже, розглянуто математичну складову проекту, де результатом є поділ даних відповідей тестування на кластери відповідно до напрямку, який варто обрати кандидату.

Архітектура MVC дозволяє розробити роботу з візуальним представленням та бізнес-логікою додатків. При роботі з Django використовується термін MVT - Model-View-Template або модель представлення-шаблон. Компоненти MVT можна використовувати незалежно один від іншого (Рис. 8).

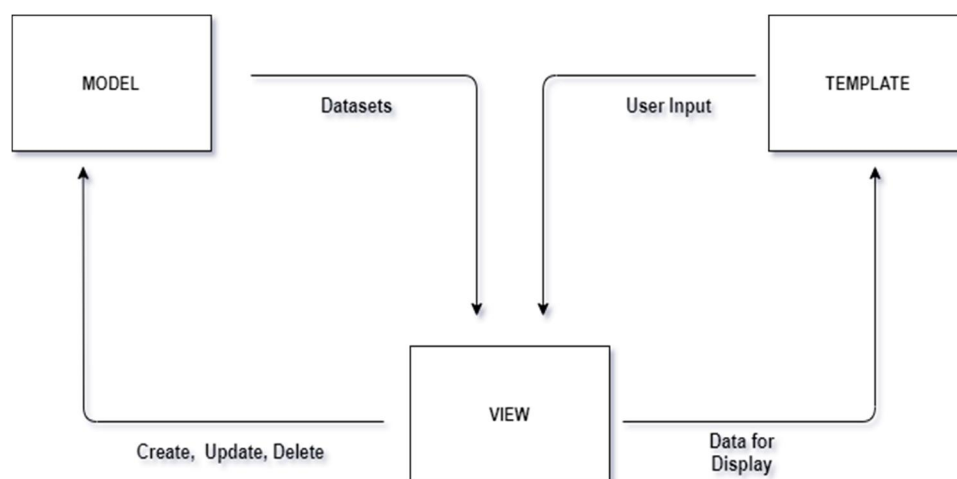


Рис. 8. Схема архітектури MVT в Django

Моделі містять інформацію про дані. Ці дані представлені атрибутами або полями. Модель відповідає за бізнес-логіку, методи, властивості та інші елементи, пов'язані з маніпуляцією даними. Такі моделі дозволяють розробникам створені, читати, оновлювати та знімати об'єкти в базах даних.

Метою роботи є розробка і аналіз тестового опитування, доведення доцільності використання вищезгаданої інформаційної системи при прийомі на роботу, при вступі в навчальний заклад або для профорієнтації учня старшої школи.

Оцінювання проводять за двома етапами:

1. Попереднє оцінювання НТР
2. Оцінювання готовності НТР до комерціалізації.

Для виконання першого етапу - попереднього оцінювання необхідно дати відповіді на такі питання.

1. Назва НТР:

Розробка системи тестового опитування в задачах профвідбору фахівців ІТ.

2. Галузь промисловості:

Сфера послуг (ІТ-компанії).

3. Загальне призначення НТР (зокрема потреби, яку вона повинна задовольнити).

Визначення потенціалу учасника тестування.

4. Чи існують аналоги (прототипи) Вашої НТР (якщо так, то їх зазначити):

WikiJob, JobTestPrep, The balance careers, Mettl, CareerExplorer.

5. Опишіть конкурентні переваги Вашої НТР, завдяки яким вона може бути привабливою для комерціалізації.

Спеціалізація на сфері ІТ, розподіл по професіях на основі тестування.

6. Вкажіть можливі сфери використання Вашої НТР (вид економічної діяльності, наприклад, охорона здоров'я, кондитерська промисловість, будівництво, оборонна промисловість, туризм тощо):

Надання інформаційних послуг

7. Вкажіть прогноз масштабів можливої реалізації результатів вашої НТР (можливі кілька варіантів відповідей):
- для галузі на рівні країни;
  - для конкретного підприємства;
8. Зазначте тип/типи ринків, на якому можливе використання результатів Вашої НТР (можливі кілька варіантів відповідей):
- ринок кінцевих споживачів;
  - ринок підприємств та організацій;
  - ринок посередників.
9. Вкажіть рівень новизни ідеї, яка покладена в основу Вашої НТР:
- нова для розробника (організації);
10. Вкажіть рівень глибини дослідження предмету Вашої НТР:
- ідея модифікується на підставі попередніх досліджень.

Отож, було описано створення веб-сайту для тестування. Результатом цієї моделі є веб-додаток, який містить тестування, їхні категорії, можливість авторизації. Після проходження тестування користувачу виводиться результат у відсотках. Всі особисті дані користувача є зашифровані.

Отже, було створено веб-сайт для проходження тестування і отримання результату. В майбутньому можливий розвиток його з розширенням категорій та варіантів тестувань для інших професій та напрямів (не тільки ІТ).

### **Висновки**

В сучасному світі всі процеси підприємств людство намагається пришвидшити та автоматизувати. І в сфері рекрутингу існує проблема відбору кандидатів не тільки по технічних навичках, а і когнітивних. Все частіше можна почути таке словосполучення, як емоційний інтелект, який відіграє не меншу роль, ніж IQ та досвід людини. При відборі кандидатів у сфері ІТ також перевіряють емоційний інтелект, проводячи співбесіду. Даний процес можна автоматизувати і скороти витрати часу та ресурсів при рекрутингу.

У дослідженні було проаналізовано актуальність розробки тестування кандидатів для відбору в ІТ-галузь. На основі даних статистики опитування компаній можна зрозуміти, що даний проект є затребуваний. І вже існує широкий спектр підприємств, які застосовують схожі види тестування.

В роботі було проведено аналіз готових рішень тестувань кандидатів і проведено порівняння з аналогами. В описі готових рішень було згадано, що існують платні веб-сервіси, як орієнтуються на великі компанії та річну підписку. Тому було прийнято рішення про створення безкоштовного веб-сервісу для тестування кандидатів в ІТ.

При вирішенні завдання в роботі показано розробку веб-додатку для тестування, який створений на основі мови програмування Python та фреймворку Django. Дана мова програмування і фреймворк були обрані через їхню простоту та ефективність. Крім цього був приведений аналіз даних, а саме відповідей на тестування для кластеризації учасників тестування по напрямках. Кластеризація даних була здійснена, використовуючи метод k-means.

Таким чином, завдання вирішені в повному обсязі, поставлена вище мета досягнута. Було розроблено і проведено аналіз тестового опитування, доведено доцільність використання вищезгаданої інформаційної системи при прийомі на роботу, при вступі в навчальний заклад або для профорієнтації учня старшої школи. В ході проведення дослідження доведено гіпотезу, яка полягає в тому, що проблему тестування можна вирішити за допомогою створення веб-додатку, який аналізує відповіді

та відправляє користувачу результат. На основі цього результату учасник тестування може оцінити свій рівень профпридатності в сфері ІТ. В майбутньому можна додати нові категорії та тестування так, як розроблена система це дозволяє.

### Перспективи подальших досліджень

Перспективи подальших досліджень у сфері створення тестового опитування для відбору фахівців в ІТ-галузі з використанням машинного навчання та кластерного аналізу є досить широкими. Основні напрями, які можуть бути розвинуті в майбутніх дослідженнях, включають:

#### 1. Покращення алгоритмів кластеризації

Майбутні дослідження можуть зосередитися на вдосконаленні методів кластеризації, зокрема оптимізації алгоритму k-середніх. Можливо також дослідити інші підходи, такі як ієрархічна кластеризація, метод DBSCAN або гібридні методи, які можуть підвищити точність класифікації кандидатів, особливо при обробці великих і неоднорідних даних.

#### 2. Інтеграція з адаптивними системами тестування

Існує перспектива розробки адаптивних тестів, які можуть змінювати рівень складності в реальному часі залежно від відповідей користувача. Така система могла б збирати додаткові дані для більш точного аналізу навичок кандидатів, що підвищить ефективність відбору.

#### 3. Аналіз поведінкових та емоційних даних

Окрім технічних навичок, майбутні дослідження можуть включати використання кластерного аналізу для оцінки поведінкових та емоційних характеристик кандидатів, таких як стійкість до стресу, комунікативні здібності та емоційний інтелект. Це дозволить більш цілісно оцінювати фахівців.

#### 4. Розробка рекомендаційних систем

На основі результатів кластерного аналізу можна створювати системи, які не тільки групують кандидатів за певними характеристиками, але й автоматично рекомендують їх для певних вакансій або ролей, відповідно до їхнього профілю. Це сприятиме скороченню часу на пошук та відбір кадрів.

#### 5. Вивчення психологічних аспектів тестування

Психологічні тести стають важливим елементом у відборі персоналу. Майбутні дослідження можуть розширити використання кластеризації для аналізу особистісних характеристик кандидатів та їх впливу на успішність роботи в ІТ-середовищі. Це допоможе створити більш комплексні моделі відбору.

#### 6. Адаптація до специфічних галузей ІТ

Дослідження можуть бути зосереджені на розробці спеціалізованих тестів і кластеризаційних моделей для різних напрямів ІТ: розробка програмного забезпечення, кібербезпека, аналіз даних, DevOps тощо. Кожна з цих галузей потребує специфічних навичок, і оптимізація тестування під конкретні завдання є перспективним напрямом.

#### 7. Міжнародні стандарти та інтеграція з ринком праці

Дослідження можуть вивчати можливість інтеграції створених тестів і результатів кластерного аналізу з глобальними платформами для пошуку роботи, такими як LinkedIn або спеціалізованими ІТ-ресурсами. Це сприятиме стандартизації оцінки фахівців та підвищенню мобільності кадрів на міжнародному ринку.

Ці напрями дозволять значно розширити можливості застосування кластерного аналізу та машинного навчання у тестуванні кандидатів і створять більш ефективні інструменти для відбору кваліфікованих фахівців в ІТ-галузі.

### Список літератури

1. Aldenderfer, M. S., & Blashfield, R. K. (1984). *Cluster analysis*. Sage Publications.
2. Chan, D., & Schmitt, N. (2004). An agenda for future research on personnel selection. *Journal of Applied Psychology*, 89(4), 627–643. <https://doi.org/10.1037/0021-9010.89.4.627>

3. Everitt, B. S., Landau, S., Leese, M., & Stahl, D. (2011). *Cluster analysis* (5th ed.). Wiley.
4. Hogan, J., & Holland, B. (2003). Using theory to evaluate personality and job-performance relations: A socio-analytic perspective. *Journal of Applied Psychology*, 88(1), 100–112. <https://doi.org/10.1037/0021-9010.88.1.100>
5. Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: A review. *ACM Computing Surveys*, 31(3), 264–323. <https://doi.org/10.1145/331499.331504>
6. MacQueen, J. B. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* (Vol. 1, pp. 281–297). University of California Press.
7. Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, 124(2), 262–274. <https://doi.org/10.1037/0033-2909.124.2.262>
8. Thibodeaux, T., & Moore, S. (2015). The rise of skill testing in the hiring process. *Human Resource Management International Digest*, 23(4), 22–25. <https://doi.org/10.1108/HRMID-04-2015-0064>

#### References

1. Aldenderfer, M. S., & Blashfield, R. K. (1984). *Cluster analysis*. Sage Publications.
2. Chan, D., & Schmitt, N. (2004). An agenda for future research on personnel selection. *Journal of Applied Psychology*, 89(4), 627–643. <https://doi.org/10.1037/0021-9010.89.4.627>
3. Everitt, B. S., Landau, S., Leese, M., & Stahl, D. (2011). *Cluster analysis* (5th ed.). Wiley.
4. Hogan, J., & Holland, B. (2003). Using theory to evaluate personality and job-performance relations: A socio-analytic perspective. *Journal of Applied Psychology*, 88(1), 100–112. <https://doi.org/10.1037/0021-9010.88.1.100>
5. Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: A review. *ACM Computing Surveys*, 31(3), 264–323. <https://doi.org/10.1145/331499.331504>
6. MacQueen, J. B. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* (Vol. 1, pp. 281–297). University of California Press.
7. Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, 124(2), 262–274. <https://doi.org/10.1037/0033-2909.124.2.262>
8. Thibodeaux, T., & Moore, S. (2015). The rise of skill testing in the hiring process. *Human Resource Management International Digest*, 23(4), 22–25. <https://doi.org/10.1108/HRMID-04-2015-0064>

**B. Boiko, I. Protsyk**

Lviv Polytechnic National University

Department of Management and International Entrepreneurship

#### APPLICATION OF A TEST SURVEY SYSTEM BASED ON CLUSTER ANALYSIS AND MACHINE LEARNING IN THE TASKS OF PROFESSIONAL SELECTION OF SPECIALISTS

© Boiko B., Protsyk I., 2024

The research is aimed at developing a test survey for the effective selection of specialists in the IT field, based on the use of modern machine learning methods, particularly cluster analysis using the k-means method. Given the limited access to existing testing platforms, which are typically available only to large companies on a paid basis, the decision was made to create an alternative web application. This application will become an accessible tool for a wider range of users and will allow automating the process of evaluating candidates' skills. A key feature of the research is the application of cluster analysis for grouping users based on their professional skills, cognitive abilities, and psychological characteristics. This enables a more precise assessment of candidates' suitability for employers' requirements and facilitates better data organization for further analysis. The research also highlights the importance of cluster analysis in cases where there is no prior hypothesis about the data structure, making this approach a universal tool for data classification. In addition to technical aspects, the study covers the prospects of using adaptive tests that can adjust the level of difficulty in real time depending on the user's responses.

*Бойко Б. М., Процик І. С.*

**This improves the accuracy of the evaluation and reduces the influence of subjective factors during selection. Additionally, the possibilities of analyzing candidates' behavioral and emotional characteristics, such as stress resilience and communication skills, are considered, as they are important for successful teamwork.**

**Keywords: information system, testing, professional selection, clustering.**