

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Сумський державний університет
Центр заочної, дистанційної та вечірньої форм навчання
Кафедра комп'ютерних наук

«До захисту допущено»

В.о. завідувача кафедри

Оксана ШОВКОПЛЯС

(підпис)

17 грудня 2024 р.

КВАЛІФІКАЦІЙНА РОБОТА
на здобуття освітнього ступеня магістр

зі спеціальності 122 – Комп'ютерних наук,
освітньо-професійної програми «Інформатика»
на тему: «Інформаційна технологія аналітичних панелей в наукових
дослідженнях»
здобувача групи ІН.мз-31 Барченка Олександра Олександровича

Кваліфікаційна робота містить результати власних досліджень.
Використання ідей, результатів і текстів інших авторів мають посилання
на відповідне джерело.

Олександр БАРЧЕНКО

(підпис)

Керівник, доцент кафедри Наталія БАРЧЕНКО
комп'ютерних наук, к.т.н, доцент

(підпис)

Суми – 2024

Сумський державний університет
Центр заочної, дистанційної та вечірньої форм навчання
Кафедра комп'ютерних наук

«Затверджую»

В.о. завідувача кафедри

Оксана ШОВКОПЛЯС

(підпис)

ІНДИВІДУАЛЬНЕ ЗАВДАННЯ НА КВАЛІФІКАЦІЙНУ РОБОТУ
на здобуття освітнього ступеня магістра

зі спеціальності 122 Комп'ютерні науки, освітньо-професійної програми «Інформатика»
здобувача групи ІН.мз-31 Барченка Олександра Олександровича

1. Тема роботи: «Інформаційна технологія аналітичних панелей в наукових дослідженнях»
затверджую наказом по СумДУ від «05» грудня 2024 р. № 11267-VI
2. Термін здачі здобувачем кваліфікаційної роботи до 10 грудня 2024 року
3. Вхідні дані до кваліфікаційної роботи _____
4. Зміст розрахунково-пояснювальної записки (перелік питань, що їх належить розробити)
1) Аналіз предметної області та постановка завдання дослідження. 2) Огляд методів та технологій створення аналітичних дашбордів для наукових досліджень. 3) Опис інформаційної технології аналітичних панелей в наукових дослідженнях.
5. Перелік графічного матеріалу (з точним зазначенням обов'язкових креслень) _____
6. Консультанти до проекту (роботи), із значенням розділів проекту, що стосується їх _____

Розділ	Консультант	Підпис, дата	
		Завдання видав	Завдання прийняв

7. Дата видачі завдання «18» жовтня 2024 р.

Завдання прийняв до виконання _____

(підпис)

Керівник _____

(підпис)

КАЛЕНДАРНИЙ ПЛАН

№ п/п	Назва етапів кваліфікаційної роботи	Термін виконання	Примітка
1	<i>Аналіз предметної області та постановка завдання дослідження</i>	05.09 – 30.09	Виконано
2	<i>Огляд методів та технологій створення аналітичних дашбордів для наукових досліджень</i>	01.10 – 18.10	Виконано
3	<i>Опис інформаційної технології аналітичних панелей в наукових дослідженнях</i>	19.10 – 11.11	Виконано
4	<i>Аналіз отриманих результатів</i>	12.11 – 14.11	Виконано
5	<i>Оформлення пояснювальної записки до кваліфікаційної роботи</i>	15.11 – 29.11	Виконано

Здобувач вищої освіти _____

(підпис)

Керівник _____

(підпис)

АНОТАЦІЯ

Записка: 44 стор., 10 рис., 1 додаток, 32 використаних джерела.

Обґрунтування актуальності теми роботи – одним з інструментів для ефективного збору, обробки та візуалізації великих обсягів даних є аналітичні дашборди. Дослідження присвячене впровадженню інформаційних технологій для створення аналітичних дашбордів у наукових дослідженнях та є надзвичайно актуальним.

Об’єкт дослідження – процеси обробки та візуалізації даних у наукових дослідженнях, зокрема в розрізі використання аналітичних дашбордів.

Мета роботи – розробити інформаційну технологію створення аналітичних дашбордів, яка забезпечить ефективну візуалізацію даних наукових досліджень та покращить процеси прийняття рішень у наукових дослідженнях.

Методи дослідження – кластерний аналіз, проектування аналітичних дашбордів.

Результати – розроблена інформаційна технологія може бути впроваджена у наукові установи та використовуватись для покращення процесів аналізу та управління даними в різних наукових дисциплінах. Використання аналітичних дашбордів дозволить значно знизити час на обробку даних, підвищить точність та оперативність прийняття рішень.

ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ, НАУКОВІ ДОСЛІДЖЕННЯ,
КЛАСТЕРИЗАЦІЯ, POWER BI, PYTHON

ЗМІСТ

ВСТУП.....	5
1 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ ТА ПОСТАНОВКА ЗАВДАННЯ ДОСЛІДЖЕННЯ.....	7
1.1 Огляд аналітичних панелей.....	7
1.2 Наукові дослідження та доцільність використання дашбордів.....	9
1.3 Розробка прототипу аналітичного дашборду для дослідження процесу цифрової трансформації областей України як приклад використання інформаційної технології.....	12
2 МЕТОДИ ТА ТЕХНОЛОГІЇ СТВОРЕННЯ АНАЛІТИЧНИХ ДАШБОРДІВ ДЛЯ НАУКОВИХ ДОСЛІДЖЕНЬ.....	14
2.1. Аналіз існуючих методів та технологій створення аналітичних дашбордів для наукових досліджень.....	14
2.2 Вимоги до аналітичного дашборду з урахуванням специфіки наукових даних і процесів.....	16
2.3 Середовище розробки Power BI.....	22
2.4 Кластерний аналіз засобами Python.....	23
3 ОПИС ІНФОРМАЦІЙНОЇ ТЕХНОЛОГІЇ АНАЛІТИЧНИХ ПАНЕЛЕЙ В НАУКОВИХ ДОСЛІДЖЕННЯХ.....	26
3.1 Вхідні та вихідні дані.....	26
3.2 Виділення кластерів областей України за схожими показниками цифровізації.....	27
3.3. Прототип аналітичного дашборду для дослідження процесу цифрової трансформації областей України як приклад використання інформаційної технології	32
3.4 Рекомендації щодо впровадження аналітичних дашбордів у наукових установах для покращення моніторингу та управління дослідженнями...	35
ВИСНОВКИ.....	38
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ.....	39
ДОДАТОК.....	43

ВСТУП

Актуальність. Сучасні наукові дослідження все частіше вимагають інструментів, які дозволяють швидко й ефективно обробляти великі обсяги інформації, аналізувати дані та наочно представляти результати. Одним із ключових інструментів для цього є дашборди. Вони забезпечують інтерактивну візуалізацію даних, допомагаючи дослідникам зрозуміти основні показники й тенденції, приймати зважені рішення та покращувати взаємодію в команді.

Дашборди відіграють важливу роль в аналізі великих масивів даних і дослідженні складних систем. Завдяки їм науковці можуть швидко виявляти закономірності, ідентифікувати аномалії та оперативно реагувати на зміни. Це сприяє підвищенню точності досліджень, зменшенню часу, витраченого на рутинні завдання, і поліпшенню якості ухвалення рішень.

Об'єкт дослідження. Процеси обробки та візуалізації даних у наукових дослідженнях, зокрема в розрізі використання аналітичних дашбордів.

Предмет дослідження. Інформаційні технології та методи візуалізації даних, що використовуються для створення аналітичних дашбордів у наукових дослідженнях.

Мета дослідження. Розробити та дослідити інформаційну технологію для створення аналітичних дашбордів, яка забезпечить ефективну візуалізацію даних наукових досліджень, що дозволить автоматизувати процеси збору, обробки та інтерпретації даних, а також покращить прийняття рішень у наукових дослідженнях.

Гіпотеза дослідження. Використання інтегрованих інформаційних технологій аналітичних панелей для обробки та візуалізації наукових даних значно підвищує ефективність процесів аналізу, прийняття рішень і комунікації результатів у наукових дослідженнях, зокрема в умовах великих даних та мультидисциплінарних досліджень.

Наукова новизна. Запропоновано підхід до виділення кластерів схожих регіонів, що дає змогу інтегрувати результати кластеризації в інтерактивні дашборди, дозволяючи науковцям ефективно здійснювати порівняльний аналіз та візуалізацію даних про цифрову трансформацію на рівні окремих регіонів. Така методика забезпечує гнучкість при фільтрації даних для подальших досліджень і прийняття рішень.

Практичне значення отриманих результатів. Створені інструменти для ефективного моніторингу, аналізу та порівняння показників цифрової трансформації регіонів України за допомогою інтеграції методів кластеризації та аналітичних панелей. Розроблені методи кластеризації регіонів за цифровими показниками дозволяють оперативно визначати схожі за рівнем цифровізації регіони. Отримані результати мають значний практичний потенціал для застосування в державному управлінні, наукових дослідженнях та стратегічному плануванні цифрових ініціатив, що дозволить забезпечити більш ефективний розвиток цифрової інфраструктури в Україні та покращити якість наукових і аналітичних процесів у цій сфері.

Структура роботи. Дана робота складається зі вступу, аналітичного огляду, постановки задачі, опису методу вирішення поставленої задачі, опису програмного забезпечення інформаційної системи, висновків, списку використаних джерел та додатку.

1 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ ТА ПОСТАНОВКА ЗАВДАННЯ ДОСЛІДЖЕННЯ

1.1 Огляд аналітичних панелей

Аналітичні панелі або дашборди стали невід’ємною частиною процесу аналізу даних у багатьох сферах, зокрема у наукових дослідженнях. Їх головна мета – надання дослідникам інтуїтивно зрозумілих інструментів для візуалізації, моніторингу та аналізу великих обсягів даних [1, 2].

Сучасні наукові дослідження потребують ефективних інструментів для збору, обробки та візуалізації даних. Дашборди стають важливою складовою аналітики завдяки їх здатності інтерактивно відображати дані, забезпечуючи швидкий доступ до ключової інформації та сприяючи прийняттю обґрунтованих рішень. Дашборд – це інтерактивна панель, яка об’єднує візуальні елементи (графіки, діаграми, таблиці) для відображення даних у зручному форматі. Основною метою є забезпечення користувача швидким доступом до комплексної інформації для моніторингу та аналізу [3].

Розвиток дашбордів розпочався з простих графічних інтерфейсів і еволюціонував у складні програмні рішення, які дозволяють інтегрувати великі масиви даних і застосовувати машинне навчання для аналітики.

Ефективне управління науковими дослідженнями вимагає інтеграції сучасних технологій для аналізу та моніторингу даних. Аналітичні дашборди є потужним інструментом для візуалізації складних даних і прийняття рішень на основі реальних показників.

Аналітичні дашборди в наукових установах можуть вирішувати наступні завдання:

- моніторинг наукової діяльності – відстеження виконання дослідницьких проектів та аналіз ефективності використання фінансових, технічних і людських ресурсів;

- управління даними у вигляді консолідації інформації з різних джерел (таких як лабораторні дані, бібліотеки, системи управління) та контроль за якістю та актуальністю даних;
- прийняття рішень, тобто інтерактивний аналіз показників продуктивності наукових груп та порівняння досягнень у різних напрямках досліджень [4].

Дашборди використовуються для візуалізації результатів досліджень, моніторингу даних у реальному часі, аналізу великих обсягів даних та співпраці між дослідниками.

Застосування дашбордів у науковій діяльності має низку суттєвих переваг, що сприяють покращенню процесів аналізу та презентації результатів. Однією з основних переваг є інтерактивність та гнучкість. Дашборди дозволяють користувачам змінювати параметри, фільтрувати дані та динамічно взаємодіяти з інформацією, що дає змогу дослідникам адаптувати аналіз під конкретні вимоги та умови. Це особливо корисно при роботі з великими масивами даних, де швидка зміна критеріїв є необхідною для гнучкої і точної обробки інформації [5].

Іншою важливою перевагою є простота аналізу. Візуалізація складних даних у вигляді графіків, діаграм або карток значно спрощує їхнє сприйняття та аналіз, надаючи дослідникам можливість швидше виявляти закономірності, тренди та аномалії. Це також дозволяє забезпечити доступність результатів для ширшого кола осіб, які можуть не мати глибоких знань у специфічній області досліджень, але потребують зрозумілої презентації даних [6].

Оперативність є ще однією важливою характеристикою дашбордів. Ці інструменти дозволяють отримувати результати в режимі реального часу, що є важливим для коригування напрямків досліджень та прийняття рішень на основі поточних даних. У наукових проектах, де швидка реакція на нову інформацію може суттєво вплинути на подальший хід експериментів, така оперативність є надзвичайно корисною [7].

Дашборди можуть бути застосовані в різних наукових напрямках. У біологічних дослідженнях вони дозволяють відстежувати динаміку експериментальних даних, наприклад, зміни в стані організмів або ефективність різних лікувальних методів. У соціології дашборди використовуються для аналізу результатів опитувань, соціальних тенденцій і поведінкових патернів, що дозволяє здійснювати більш точне прогнозування соціальних процесів. В фізиці та математиці ці інструменти можуть бути використані для обробки результатів чисельних моделювань, експериментальних вимірювань та інших складних даних [8].

Для створення дашбордів використовуються різноманітні програмні засоби. Tableau є потужним інструментом для створення інтерактивних дашбордів, який надає широкий набір можливостей для аналізу і візуалізації даних. Power BI від Microsoft є популярною платформою для бізнес-аналітики, але також широко використовується для наукових досліджень завдяки своїй гнучкості та інтеграційним можливостям. Також широко використовуються Python та R, що надають безліч бібліотек, таких як Plotly та Dash, для створення кастомізованих дашбордів, що дозволяє науковцям розробляти інструменти, що точно відповідають специфікаціям їхніх досліджень [9].

Таким чином, дашборди є потужним інструментом для візуалізації та аналізу даних у різних наукових галузях, сприяючи підвищенню ефективності та оперативності досліджень.

1.2 Наукові дослідження та доцільність використання дашбордів

Основні типи даних у наукових дослідженнях та способи їх інтеграції для побудови дашбордів. Сучасні наукові дослідження потребують аналізу значних обсягів даних, які можуть бути різного типу та походження. Для ефективного використання цих даних у аналітичних дашбордах необхідно розуміти їх природу, структуру, а також методи

інтеграції. Для побудови аналітичних дашбордів у наукових дослідженнях необхідно враховувати різноманітність типів даних та особливості їх інтеграції. Вибір методів і технологій інтеграції залежить від джерела, формату даних і вимог до системи. Поєднання структурованих і неструктурованих даних, підтримка потокової обробки та використання API сприяє створенню дашбордів, які є ефективним інструментом для аналізу складних наукових даних [10].

Основними типами даних у наукових дослідженнях є:

- структуровані дані, які мають чітку структуру та організовані у вигляді таблиць, з полями, що відповідають конкретним атрибутам. В якості прикладів можна навести результати опитувань, лабораторні вимірювання, фінансові показники у форматах CSV, Excel, SQL-таблиць;
- неструктуровані дані без визначеної структури, які важко організувати в традиційні таблиці. Приклади таких даних – текстові документи, зображення, відео, аудіофайли відповідно у таких форматах, як TXT, PDF, JPG, MP4;
- напівструктуровані дані, тобто дані з частковою структурою, які не вписуються повністю у традиційну реляційну модель. Такими даними можуть бути, наприклад, логи систем, JSON- та XML-файли;
- дані у режимі реального часу, які генеруються та оновлюються безперервно, часто у реальному часі, наприклад, дані сенсорів IoT, інформація стосовно фінансових ринків, метеорологічні показники. Щодо форматів – це можуть бути потокові дані з API, MQTT-протоколи;
- геопросторові дані, які містять інформацію про географічне розташування. Прикладами можуть виступати дані GPS, карти землекористування, супутникові зображення у таких форматах, як GeoJSON, KML, Shapefile;

- часові ряди – дані, впорядковані за часовими інтервалами, що відображають зміни певних показників у часі. Це можуть бути фінансові показники, кліматичні зміни, показники виробництва у форматах таблиць з мітками часу, SQL-таблиць, спеціалізованих бібліотек (HDF5) [11].

Способи інтеграції даних для побудови дашбордів. Підключення до баз даних відбувається шляхом інтеграції з реляційними (MySQL, PostgreSQL) та нереляційними (MongoDB, Cassandra) базами даних з використанням таких інструментів, як Tableau, Power BI, SQLAlchemy. Перевагами цього способу є простота роботи з великими масивами даних та підтримка складних запитів.

Імпорт файлів відбувається шляхом завантаження даних з файлів, наприклад, CSV, Excel, JSON, а в якості інструментів виступають Python (Pandas, JSON), Excel, OpenRefine. Головною перевагою є імпорту є легкість обробки статичних даних.

Використання API для інтеграції з веб-сервісами з метою отримання даних у реальному часі, наприклад, використання API метеорологічних служб або фінансових платформ за допомогою таких інструментів, як REST API, GraphQL. Перевага – доступ до актуальних даних без необхідності їх зберігання.

Інтеграція потокових даних реалізується шляхом використання платформ для збору та обробки даних у реальному часі за допомогою такої інструментарію, як Apache Kafka та Spark Streaming. Перевагою є можливість роботи з великим обсягом швидкоплинних даних.

Обробка нестандартних форматів з використанням спеціалізованих бібліотек для обробки текстових, геопросторових даних або зображень з використанням NLTK (текст), GDAL (геодані), OpenCV (зображення). Перевагою такої інтеграції є гнучкість у роботі з різноманітними даними [12, 13].

Дашборди стають невід'ємним інструментом у наукових дослідженнях, допомагаючи вирішувати низку важливих проблем.

Сучасні наукові дослідження часто оперують великими обсягами даних, що потребують систематизації та швидкого аналізу. Дашборди забезпечують можливість візуалізації та миттєвого доступу до ключових показників, що полегшує роботу з великими наборами даних.

Завдяки інтерактивним елементам дашбордів, дослідники можуть оперативіно реагувати на зміни у даних та швидко адаптувати свої стратегії, зменшуючи затримки у прийнятті рішень [14].

Дашборди допомагають стандартизувати представлення даних та забезпечують одне джерело правди (single source of truth). Це дозволяє усунути суперечливі інтерпретації даних і підвищує узгодженість висновків.

Дослідницькі групи часто складаються з учасників із різними компетенціями. Дашборди, які чітко візуалізують дані та ключові метрики, сприяють кращому розумінню міждисциплінарних команд і полегшують обмін результатами. Використання дашбордів дозволяє науковцям зосередитися на аналізі та інтерпретації даних замість трудомісткої підготовки, тим самим прискорюючи цикл досліджень і підвищуючи їхню ефективність [16].

Всі ці переваги обумовлюють доцільність використання інформаційної технології аналітичних панелей в наукових дослідженнях.

1.3 Розробка прототипу аналітичного дашборду для дослідження процесу цифрової трансформації областей України як приклад використання інформаційної технології

Цифрова трансформація є одним із ключових напрямів розвитку суспільства та економіки України. Для оцінки ефективності процесів цифровізації на регіональному рівні необхідний інструмент, який дозволяє зручно аналізувати динаміку змін за основними показниками щодо цифрової

трансформації областей України. У процесі цифрової трансформації регіонів важливо враховувати не лише індивідуальні показники областей, але й їхні подібності, які можуть свідчити про спільні тренди, виклики та перспективи розвитку. Кластерний аналіз дозволяє групувати області за схожими показниками цифровізації, сприяючи розробці регіонально орієнтованих стратегій та ефективнішому управлінню процесами цифрової трансформації.

Мета дослідження – розробити та дослідити інформаційну технологію для створення аналітичних дашбордів, яка забезпечить ефективну візуалізацію даних наукових досліджень, що дозволить автоматизувати процеси збору, обробки та інтерпретації даних, а також покращить прийняття рішень у наукових дослідженнях щодо цифрової трансформації областей України.

Для досягнення поставленої мети необхідно виконати наступні завдання:

1. Провести аналіз існуючих методів та технологій створення аналітичних дашбордів для наукових досліджень.
2. Описати Вимоги до аналітичного дашборду з урахуванням специфіки наукових даних і процесів.
3. Виділити кластери областей України за схожими показниками цифровізації та дослідити їх особливості.
4. Розробити прототипу аналітичного дашборду для дослідження процесу цифрової трансформації областей України як приклад використання інформаційної технології.
5. Підготувати рекомендації щодо впровадження аналітичних дашбордів у наукових установах для покращення моніторингу та управління дослідженнями.

2 МЕТОДИ ТА ТЕХНОЛОГІЙ СТВОРЕННЯ АНАЛІТИЧНИХ ДАШБОРДІВ ДЛЯ НАУКОВИХ ДОСЛІДЖЕНЬ

2.1. Аналіз існуючих методів та технологій створення аналітичних дашбордів для наукових досліджень

Вибір методів та технологій для створення дашбордів визначається специфікою дослідження, вимогами до інтерфейсу користувача, обсягом даних, що обробляються, та необхідною гнучкістю системи.

Методи створення аналітичних дашбордів. Метод шаблонного підходу, який базується на використанні готових шаблонів для створення дашбордів. Такі інструменти, як Tableau, Power BI або QlikView, пропонують широкий вибір шаблонів для швидкої інтеграції даних. Основними перевагами цього методу є швидкість створення та мінімальні вимоги до програмування, але він може бути обмеженим у налаштуванні специфічних функцій або інтеграції нестандартних даних [18, 19].

Кастомізований підхід із використанням мов програмування, коли для створення більш гнучких дашбордів використовуються мови програмування, зокрема Python (з бібліотеками, як-от Plotly, Dash, Matplotlib), R (Shiny), або JavaScript (D3.js). Такий підхід дозволяє розробляти рішення з високим ступенем адаптації під специфічні потреби, але вимагає значних знань у програмуванні [20, 21].

Метод інтеграції з базами даних. Дашборди, створені на основі інтеграції з базами даних (наприклад, MySQL, PostgreSQL, MongoDB), дозволяють в реальному часі обробляти великі обсяги даних. Це особливо важливо для наукових досліджень із великими наборами експериментальних даних [22].

Методи машинного навчання та штучного інтелекту. У сучасних аналітичних дашбордах дедалі частіше використовуються технології штучного інтелекту (ШІ) для прогнозування та аналізу даних. Такі

інструменти, як Azure Machine Learning, Google AI Platform, дозволяють інтегрувати моделі машинного навчання у дашборди, роблячи їх ще потужнішими для аналізу даних [23, 24].

Технології створення аналітичних дашбордів. Хмарні технології, такі як Google Cloud, Amazon Web Services (AWS), або Microsoft Azure, забезпечують масштабованість і доступність дашбордів, тобто дозволяють дослідникам отримувати доступ до дашбордів з будь-якого пристрою та оперативно оновлювати дані.

Інструменти візуалізації даних, з яких найбільш поширеними є Tableau (забезпечує інтерактивну візуалізацію великих обсягів даних), Power BI (підтримує інтеграцію з багатьма джерелами даних), D3.js (пропонує можливості кастомізації для розробників).

Такі інструменти, як платформи з відкритим кодом, а саме Dash, Shiny, Metabase, є популярними серед дослідників завдяки низьким витратам і можливості кастомізації.

Використання таких фреймворків, як Apache Hadoop або Apache Spark, дозволяє створювати дашборди для роботи з великими обсягами даних.

Сучасні методи та технології створення аналітичних дашбордів пропонують широкий спектр можливостей для дослідників. Вибір оптимального підходу залежить від потреб проєкту, складності даних та технічних можливостей команди, яка над ним працює. Тенденція до використання інтеграції з ІІІ та хмарними платформами вказує на подальший розвиток і вдосконалення аналітичних інструментів, що сприяє ефективності наукових досліджень [25 – 27].

2.2 Вимоги до аналітичного дашборду з урахуванням специфіки наукових даних і процесів

Оскільки наукові дані зазвичай мають особливості, які відрізняються від бізнес-даних, важливо розробити дашборд, який може ефективно обробляти й візуалізувати саме такі дані.

Ключовими аспектами для розробки вимог є:

- 1) аналіз специфіки наукових даних;
- 2) моделювання взаємодії з даними;
- 3) типи візуалізацій для наукових даних;
- 4) інтерфейс і зручність користування;
- 5) безпека та доступ до даних.

Аналіз специфіки наукових даних. Наукові дані становлять основу досліджень у багатьох галузях і класифікуються за кількома типами залежно від їхньої природи, структури та способу використання. Основними типами наукових даних є кількісні, якісні, тимчасові та просторові. Кожен із цих типів має свої особливості, методи збору, аналізу та представлення.

Кількісні дані виражаються числовими значеннями й застосовуються для проведення математичних і статистичних обчислень. Вони є результатом фізичних вимірювань, експериментів або моделювання. Прикладами кількісних даних є числові показники температури, тиску, швидкості, а також результати досліджень у вигляді графіків або таблиць. Такі дані широко використовуються для аналізу закономірностей, перевірки гіпотез та прогнозування [22].

Якісні дані, на відміну від кількісних, мають описовий характер і використовуються для класифікації або аналізу властивостей об'єктів. Вони можуть включати результати інтерв'ю, характеристики явищ, категоризацію за певними параметрами або наявність тегів. Прикладами якісних даних є відповіді респондентів у соціологічних опитуваннях або позначення об'єктів

у наукових дослідженнях за допомогою категорій, таких як «висока», «середня» або «низька активність».

Тимчасові дані представляють інформацію, що змінюється з часом. Їхнє вивчення дозволяє аналізувати динаміку явищ, зміни у процесах або тенденції. Наприклад, дослідження зміни клімату базуються на тривалих серіях спостережень температури та атмосферних явищ. Експериментальні дані, які збираються у різні часові періоди, також є прикладом тимчасових даних.

Просторові дані характеризують інформацію, пов'язану з географічним розташуванням. Вони включають координати, картографічні дані, тривимірні моделі або інші форми просторового представлення. Такі дані широко застосовуються у географії, екології, містобудуванні та суміжних дисциплінах. Наприклад, географічні інформаційні системи (ГІС) використовують просторові дані для створення карт, моделювання міграції тварин або прогнозування землетрусів.

Наукові дані можуть зберігатися у різних форматах, які забезпечують їх доступність для аналізу та використання у дослідженнях. Серед найпоширеніших форматів – CSV і Excel, які підходять для зберігання таблиць; JSON та XML, що використовуються для структурованих даних у веб-додатках; HDF5, оптимізований для великих обсягів даних; а також бази даних SQL та NoSQL для складніших структур. Крім того, для отримання та обробки наукових даних використовуються API – інтерфейси, які дозволяють автоматизовано взаємодіяти з базами даних і платформами.

Таким чином, специфіка наукових даних визначає методи їхнього збирання, обробки та аналізу. Різноманітність типів і форматів даних вимагає відповідного підходу до їхнього опрацювання, забезпечуючи ефективність досліджень та інновацій [20].

Моделювання взаємодії з даними. У сучасних наукових дослідженнях дашборди відіграють ключову роль у зборі, інтеграції, обробці та візуалізації даних. Вони слугують інтерактивними платформами для відображення

складної інформації у зрозумілому вигляді, забезпечуючи дослідникам можливість швидкого аналізу даних та прийняття рішень на основі виявлених закономірностей [16, 17].

Одним із головних аспектів створення дашбордів є забезпечення підтримки інтеграції з різними джерелами даних. Це включає можливість підключення до баз даних наукових інститутів, публікацій, а також результатів експериментів. Така інтеграція дозволяє об'єднувати розрізнені дані в єдине інформаційне середовище, спрощуючи їхній аналіз та інтерпретацію.

Ще однією важливою функцією дашбордів є автоматизація збору даних. У наукових дослідженнях критично важливо, щоб дані регулярно оновлювалися з актуальних джерел. Для цього необхідно розробити механізми, які забезпечують автоматичне підключення до баз даних, отримання нових результатів експериментів чи завантаження інформації з публічних API. Такий підхід дозволяє мінімізувати людський фактор у процесі збору даних і гарантує їхню актуальність у реальному часі [15].

Наукові дослідження часто працюють із великими обсягами даних, які потребують спеціалізованих підходів до їхньої обробки, зберігання та візуалізації. Так звані «великі дані» (Big Data) вимагають використання потужних алгоритмів і технологій для обробки, таких як розподілені обчислення, хмарні платформи або системи зберігання даних із високою пропускну здатністю. Дашборди повинні бути здатні ефективно працювати з такими обсягами інформації, надаючи зручні інструменти для її сегментації, фільтрації та деталізації. Крім того, візуалізація даних повинна бути адаптована до великих обсягів, забезпечуючи зрозуміле подання складних структур і взаємозв'язків [28].

Таким чином, сучасні дашборди для наукових досліджень повинні поєднувати функції інтеграції даних, їхньої автоматизованої актуалізації та обробки великих обсягів інформації. Такий підхід забезпечує дослідникам

необхідні інструменти для роботи з комплексними даними, сприяючи покращенню якості аналізу та прискоренню процесу прийняття рішень.

Типи візуалізацій для наукових даних. У наукових дослідженнях візуалізація даних є ключовим елементом для представлення результатів, аналізу складних взаємозв'язків і виявлення трендів.

Графіки та діаграми відіграють основну роль у цьому процесі. Серед них статистичні графіки, гістограми, кругові діаграми, лінійні графіки та діаграми розсіяння є найбільш популярними завдяки своїй здатності чітко відображати числові показники та зміни в часі. Вони дозволяють дослідникам спостерігати закономірності, порівнювати різні набори даних і виявляти тренди. Наприклад, лінійні графіки часто застосовуються для демонстрації динаміки явищ, тоді як діаграми розсіяння корисні для оцінки взаємозв'язків між змінними [15].

Окремої уваги заслуговують мапи та геопросторові візуалізації, які є незамінними в географічних дослідженнях і суміжних галузях. Вони дозволяють відображати просторові дані у вигляді карт або графіків розподілу, що дозволяє аналізувати географічні закономірності, наприклад, поширення явищ у регіонах чи глобальні тренди. Візуалізація просторових даних за допомогою карт значно підвищує доступність і наочність інформації, дозволяючи швидко ідентифікувати регіональні особливості.

Інтерактивні елементи візуалізації надають дослідникам додаткові можливості для роботи з даними. Завдяки фільтрам і налаштуванням користувачі можуть обирати певні критерії для аналізу, налаштовувати параметри відображення або виділяти конкретні ділянки даних для деталізації. Наприклад, у дослідженнях економічних показників це може бути фільтрування за роками, регіонами чи галузями. Інтерактивність також дозволяє порівнювати результати різних сценаріїв, надаючи дослідникам інструменти для багатовимірного аналізу [17, 25].

Таблиці з розширеною аналітикою є важливим компонентом у роботі з великими масивами даних. Вони надають можливість сортувати, фільтрувати

та агрегувати інформацію, дозволяючи отримувати узагальнену картину або занурюватися в деталі окремих аспектів. Інтерактивні таблиці є зручними для порівняння даних, відстеження змін і виділення ключових показників. Завдяки цим функціям таблиці стають потужним інструментом для глибокого аналізу.

Таким чином, графіки, діаграми, геопросторові мапи, інтерактивні елементи та таблиці є невід'ємною частиною сучасної наукової аналітики. Вони забезпечують наочне подання складних даних, спрощують процес аналізу та сприяють отриманню нових інсайтів у дослідницьких проєктах [11 – 13].

Інтерфейс і зручність користування. Сучасні дашборди, які використовуються у наукових дослідженнях, повинні відповідати вимогам зручності, адаптивності та доступності, щоб забезпечити ефективну взаємодію з користувачами різного рівня підготовки.

Одним із найважливіших аспектів є простота навігації, адже інтерфейс дашборду має бути інтуїтивно зрозумілим для науковців, аналітиків, студентів чи інших користувачів. Важливо, щоб навіть ті, хто не має досвіду роботи з подібними інструментами, могли швидко зрозуміти, як знайти потрібну інформацію, використовувати доступні функції та працювати з даними. Чітка структура, логічна організація елементів і зрозумілі піктограми допомагають зменшити криву навчання та підвищити ефективність використання.

Адаптивний дизайн є ще однією обов'язковою характеристикою дашборду. В умовах зростання використання мобільних пристроїв дашборд повинен коректно відображатися на різних екранах, включаючи персональні комп'ютери, планшети та смартфони. Це забезпечує дослідникам доступ до даних у будь-який момент і з будь-якого місця, що особливо важливо для колаборативних проєктів або роботи в польових умовах. Адаптивність дозволяє зберігати функціональність та зручність інтерфейсу незалежно від розміру пристрою [10].

Для наукових досліджень, які проводяться на міжнародному рівні, мультимовність стає критично важливою. Можливість зміни мови інтерфейсу дозволяє зробити дашборд доступним для користувачів із різних країн. Це сприяє більш широкому залученню фахівців до спільних дослідницьких проєктів, підвищує зручність використання та покращує сприйняття результатів незалежно від мовного бар'єра.

Наявність покрокової допомоги у дашборді є важливим аспектом для нових користувачів або тих, хто працює з інструментом епізодично. Інструкції, інтерактивні підказки чи навчальні матеріали допомагають зрозуміти, як ефективно використовувати функції дашборду, аналізувати дані або налаштовувати візуалізації під свої потреби. Це зменшує час на освоєння системи та допомагає уникати помилок у роботі [8].

Таким чином, сучасні дашборди повинні поєднувати простоту навігації, адаптивний дизайн, мультимовність та підтримку у вигляді покрокової допомоги. Ці характеристики роблять їх універсальними інструментами для роботи з науковими даними, забезпечуючи зручність і ефективність у використанні незалежно від досвіду чи технічних можливостей користувачів.

Підтримка колаборації. Забезпечення спільного доступу до результатів є важливим аспектом наукової співпраці, оскільки дозволяє дослідникам спільно працювати з даними та здійснювати їх аналіз у реальному часі. Для цього необхідно створити умови для надання доступу до дашбордів, що використовуються для візуалізації результатів досліджень, усім членам дослідницької групи. Такий підхід сприяє ефективному обміну інформацією та підтримує інтерактивність процесу аналізу даних [5].

Одним з важливих елементів у цьому контексті є інтеграція з іншими інструментами, які активно використовуються в наукових колабораціях. Наприклад, платформи GitHub та GitLab надають можливості для спільного доступу до коду та даних, що дозволяє дослідникам працювати над проєктами у розподіленому середовищі. Інтеграція дашбордів з такими платформами дозволяє забезпечити безперервний доступ до всіх необхідних

ресурсів, спрощуючи процес спільного аналізу та публікації результатів досліджень. Це, в свою чергу, сприяє підвищенню прозорості, повторюваності та ефективності наукової роботи.

Враховання специфіки наукової діяльності та потреб дослідників допоможе створити дашборд, який не тільки задовольняє вимоги до візуалізації даних, але й забезпечить високу ефективність використання, безпеку, доступність та інтеграцію з іншими системами.

2.3 Середовище розробки Power BI

Power BI – це потужний інструмент бізнес-аналітики, який стає дедалі популярнішим у наукових дослідженнях завдяки своїм можливостям інтеграції, аналізу та візуалізації даних. Розроблений корпорацією Microsoft, Power BI дозволяє дослідникам працювати з великими масивами даних різного типу, що є критично важливим у сучасних міждисциплінарних дослідженнях [17, 25]. Його гнучкий інтерфейс та інтерактивний характер роботи дозволяють ефективно адаптуватися до різних задач, зокрема аналізу числових, текстових, просторових або часових даних.

Однією з головних переваг Power BI є його здатність інтегрувати дані з численних джерел. Дослідники можуть працювати з локальними файлами у форматах CSV, Excel, JSON, а також отримувати інформацію з баз даних SQL чи хмарних сервісів, таких як Azure, Google BigQuery чи Amazon S3. Така інтеграція полегшує роботу з різними типами даних, що часто зустрічаються у наукових проєктах. Наприклад, у біомедичних дослідженнях це можуть бути клінічні записи, отримані через API медичних платформ, а у кліматичних дослідженнях – великі набори спостережень за погодою.

Зручність Power BI полягає в його здатності автоматизувати процеси обробки даних та їхньої візуалізації. Мова DAX (Data Analysis Expressions), яка використовується для створення складних формул і обчислень, дозволяє виконувати глибокий аналіз, створювати нові метрики та виявляти приховані

закономірності. Ці функції особливо корисні у ситуаціях, коли потрібно проаналізувати взаємозв'язки між численними змінними або простежити динаміку показників у часі.

Інтерактивні панелі Power BI надають можливість у реальному часі візуалізувати результати у вигляді графіків, карт або діаграм, що значно полегшує інтерпретацію складних даних.

Важливою перевагою Power BI для наукових досліджень є можливість працювати з геопросторовими даними. Це особливо актуально для дослідників у галузі географії, екології чи урбаністики. Інструмент дозволяє будувати інтерактивні карти, візуалізувати географічні координати та аналізувати просторові закономірності. У поєднанні з часовими рядами, Power BI дозволяє вивчати динаміку змін просторових параметрів, таких як міграція населення, розподіл температур або поширення захворювань.

У науковій спільноті Power BI також цінується за його можливості сприяти колективній роботі. Завдяки хмарній платформі Power BI Service дослідники можуть публікувати результати своїх аналізів, обмінюватися звітами та співпрацювати з колегами з різних установ. Такий підхід підвищує ефективність роботи у великих дослідницьких проєктах і полегшує поширення знань.

Таким чином, Power BI є надзвичайно корисним інструментом для наукових досліджень, який поєднує інтеграцію даних, автоматизацію аналізу та зручну візуалізацію. Його універсальність і адаптивність роблять його особливо ефективним для роботи з міждисциплінарними даними, забезпечуючи дослідників цінною інформацією для формулювання гіпотез, прийняття рішень та оприлюднення результатів.

2.4 Кластерний аналіз засобами Python

Кластеризація є одним із ключових методів аналізу даних у наукових дослідженнях, що використовується для виявлення структур та

закономірностей у великих наборах інформації. Цей метод належить до некерованого машинного навчання і спрямований на групування об'єктів у класи або кластери на основі їхньої схожості чи близькості. Кластеризація застосовується у різних галузях, зокрема в біоінформатиці, соціології, маркетингу, екології та фізиці, дозволяючи дослідникам групувати дані за ознаками, які раніше могли бути невідомими. У науці це часто використовується для виявлення нових підтипів у наборі даних, наприклад, класифікації видів у біології чи аналізу кліматичних зон.

Одним із найпотужніших інструментів для кластеризації є Python, який пропонує широкий набір бібліотек і функцій для аналізу даних. Python є популярним завдяки своїй гнучкості, простоті та багатству інструментів для роботи з кластеризацією, що дозволяє виконувати як базові, так і складні задачі.

Для реалізації кластеризації дослідники можуть використовувати бібліотеки, такі як scikit-learn, SciPy, NumPy, pandas та matplotlib, які забезпечують повний цикл роботи з даними: від їхньої підготовки до візуалізації результатів.

Scikit-learn є однією з найбільш популярних бібліотек для кластеризації в Python, що надає реалізації різних алгоритмів, таких як K-means, DBSCAN, агломеративна кластеризація та спектральна кластеризація. Наприклад, метод K-means дозволяє розділити дані на заздалегідь визначену кількість кластерів, визначаючи центри цих кластерів шляхом мінімізації відстаней між об'єктами та їхніми центроїдами. DBSCAN, навпаки, не вимагає задавання кількості кластерів і ефективно працює з даними складної форми, дозволяючи виділяти окремі структури навіть у наявності шуму [29].

Бібліотека SciPy також широко застосовується для виконання завдань кластеризації, особливо в агломеративних підходах. Вона забезпечує інструменти для побудови дендрограм, які дозволяють візуально оцінювати ієрархію групування об'єктів.

NumPy і pandas забезпечують роботу з багатовимірними масивами та таблицями даних, що є критично важливим для попередньої обробки даних перед кластеризацією. Підготовка даних включає нормалізацію, видалення пропущених значень, кодування категоріальних змінних, оскільки якість кластеризації залежить від правильності та консистентності вихідних даних.

Результати кластеризації в Python часто візуалізують за допомогою matplotlib або seaborn. Це дозволяє дослідникам зрозуміти структуру даних, побачити взаємодію між кластерами та підтвердити достовірність отриманих результатів. Наприклад, у дво- або тривимірному просторі можна графічно відобразити розташування кластерів і їхній взаємозв'язок, що полегшує інтерпретацію результатів [30, 31].

Застосування кластеризації в наукових дослідженнях на базі Python дозволяє не лише обробляти великі масиви даних, але й оптимізувати їх структуру, що підвищує ефективність подальшого аналізу. Python забезпечує дослідників гнучкими інструментами для розробки індивідуальних алгоритмів, що відповідають специфічним задачам дослідження, роблячи цей метод незамінним у багатьох дисциплінах.

Для створення програми кластеризації даних з використанням алгоритму K-means та графічного інтерфейсу на Python, можна використати кілька потужних бібліотек, таких як sklearn (для виконання кластеризації за допомогою алгоритму K-means), tkinter (для створення графічного інтерфейсу), matplotlib (для візуалізації результатів кластеризації) та numpy (для обробки числових даних) [32].

3 ОПИС ІНФОРМАЦІЙНОЇ ТЕХНОЛОГІЇ АНАЛІТИЧНИХ ПАНЕЛЕЙ В НАУКОВИХ ДОСЛІДЖЕННЯХ

3. 1. Вхідні та вихідні дані

При зборі даних була використана інформація з відкритих джерел, включаючи Державну службу статистики та Міністерство цифрової трансформації.

Для формалізації і опису системи побудована діаграма в нотації IDEF0. Головний процес, який містить підсистеми, які взаємодіють із зовнішніми факторами – «Розробка та використання інформаційної технології аналітичних панелей в наукових дослідженнях» (рис. 3.1). На входах – статистичні дані (звіти Мінцифри та інших установ) та запити користувача. Механізми – програмне забезпечення для аналітичних панелей (Power BI, Python) та користувачі системи (науковці, аналітики, розробники). Керування процесом – алгоритми обробки даних (методи статистичного аналізу, машинне навчання) та користувачі системи (науковці, аналітики, розробники).



Рисунок 3.1 – Контексна діаграма процесу «Розробка та використання інформаційної технології аналітичних панелей в наукових дослідженнях» в нотації IDEF0

3.2 Виділення кластерів областей України за схожими показниками цифровізації

В якості алгоритму кластеризації обрано K-means (як популярний метод для виділення кластерів у даних). Даними для дослідження (табл. 3.1) є підсумкові показники цифрової трансформації регіонів України.

Таблиця 3.1 – Вхідні дані

Область	Інституційна спроможність	Розвиток інтернету	Розвиток ЦНАП	Режим «без паперів»	Цифрова освіта	Візитівка області	Проникнення базових електронних послуг	Галузева цифрова трансформація
Вінницька	16	0,743	0,852	0,894	0,778	0,4	0,696	0,721
Волинська	0,6	0,733	0,911	0,928	0,256	0	0,772	0,716
Дніпропетровська	1	0,993	0,972	0,952	0,694	1	0,754	0,924
Донецька	0,325	0,252	0,631	0,527	0,396	0,68	0,596	0,345
Житомирська	0,38	0,909	0,895	0,859	0,349	0,4	0,646	0,403
Закарпатська	0,571	0,868	0,841	0,85	0,372	1	0,766	0,504
Запорізька	0,331	0,405	0,385	0,063	0,155	0,4	0,595	0,261
Івано-Франківська	0,81	0,896	0,853	0,162	0,124	0,8	0,78	0,338
Київська	0,71	0,377	0,824	0,711	0,258	1	0,61	0,21
Кіровоградська	0,213	0,388	0,664	0,531	0,163	0	0,614	0,154
Луганська	0,093	0,1	0,743	0,459	0,262	0,4	0,623	0,224
Львівська	0,89	0,833	0,822	0,924	0,562	0,8	0,68	0,82
Миколаївська	0,11	0,53	0,467	0,497	0,509	0,6	0,427	0,3
Одеська	0,9	0,841	0,705	0,958	0,702	1	0,877	0,808
Полтавська	0,8	0,894	0,905	0,967	0,88	0,6	0,737	0,56
Рівненська	1	0,907	0,875	0,891	0,195	0,8	0,705	0,573
Сумська	0,44	0,471	0,763	0,569	0,605	0	0,622	0,321
Тернопільська	1	0,992	0,83	0,933	0,718	1	0,834	0,999
Харківська	0,794	0,615	0,703	0,553	0,152	0,48	0,544	0,325
Херсонська	0,441	0,428	0,787	0,693	0,589	0,4	0,438	0,066
Хмельницька	0,861	0,609	0,719	0,653	0,165	0,4	0,596	0,458
Черкаська	0,744	0,771	0,855	0,803	0,312	0,6	0,759	0,387
Чернівецька	0,37	0,633	0,852	0,39	0,13	0,4	0,596	0,211
Чернігівська	0,493	0,612	0,629	0,596	0,267	0,4	0,4	0,507

Виконано серію експериментів, спрямованих на визначення оптимальної кількості кластерів. Інструментами для аналізу слугували мова програмування Python та бібліотеки Scikit-learn, Matplotlib. Програмний код наведений в додатку.

Опис програмного коду

Використання K-means. У функції `perform_kmeans(data, n_clusters)` створюється екземпляр класу `KMeans` з бібліотеки `sklearn.cluster`, де задається кількість кластерів та фіксується випадковий стан генератора для відтворюваності результатів. Далі виконується метод `fit`, який підлаштовує модель до введених даних.

Функція для візуалізації. У функції `plot_clusters(data, kmeans)` використовується бібліотека `matplotlib` для створення графіка розподілу точок за кольорами кластерів, а також червоний хрестик для відображення центрів кластерів.

Графічний інтерфейс. Основне вікно створено за допомогою `tkinter`. Користувач вводить кількість кластерів і самі дані (кожен рядок – це набір ознак для одного об'єкта, розділений комами). При натисканні кнопки «Run K-means» запускається кластеризація, результат виводиться у вигляді повідомлення, а також відображається графік.

Обробка введених даних. Дані вводяться як текст, де кожен рядок містить значення ознак об'єкта, розділені комами. Програма перевіряє, чи введено дані, і чи мають вони відповідну структуру для кластеризації (щонайменше два стовпці).

Результати кластеризації. Обраний підхід характеризується необхідністю попереднього завдання кількості кластерів N . У рамках експерименту було виконано кластеризацію для $N = 3, 4, 5$, а результати наведено в таблиці 3.2.

Таблиця 3.2 – Експериментальні виділення кластерів

N=3		N=4		N=5	
Область	Кластер	Область	Кластер	Область	Кластер
Донецька	1	Івано-Франківська	1	Івано-Франківська	1
Запорізька	1	Київська	1	Київська	1
Івано-Франківська	1	Харківська	1	Харківська	1
Київська	1	Хмельницька	1	Хмельницька	1
Кіровоградська	1	Черкаська	1	Черкаська	1
Луганська	1	Вінницька	2	Вінницька	2
Миколаївська	1	Волинська	2	Волинська	2
Харківська	1	Житомирська	2	Житомирська	2
Херсонська	1	Сумська	2	Дніпропетровська	3
Чернівецька	1	Дніпропетровська	3	Закарпатська	3
Чернігівська	1	Закарпатська	3	Львівська	3
Вінницька	2	Львівська	3	Одеська	3
Волинська	2	Одеська	3	Полтавська	3
Житомирська	2	Полтавська	3	Рівненська	3
Сумська	2	Рівненська	3	Тернопільська	3
Хмельницька	2	Тернопільська	3	Запорізька	4
Черкаська	2	Донецька	4	Кіровоградська	4
Дніпропетровська	3	Запорізька	4	Луганська	4
Закарпатська	3	Кіровоградська	4	Чернівецька	4
Львівська	3	Луганська	4	Донецька	5
Одеська	3	Миколаївська	4	Миколаївська	5
Полтавська	3	Херсонська	4	Сумська	5
Рівненська	3	Чернівецька	4	Херсонська	5
Тернопільська	3	Чернігівська	4	Чернігівська	5

Для наочного аналізу даних кластерів були розроблені діаграми, які наведені на рис. 3.2 – 3.4.

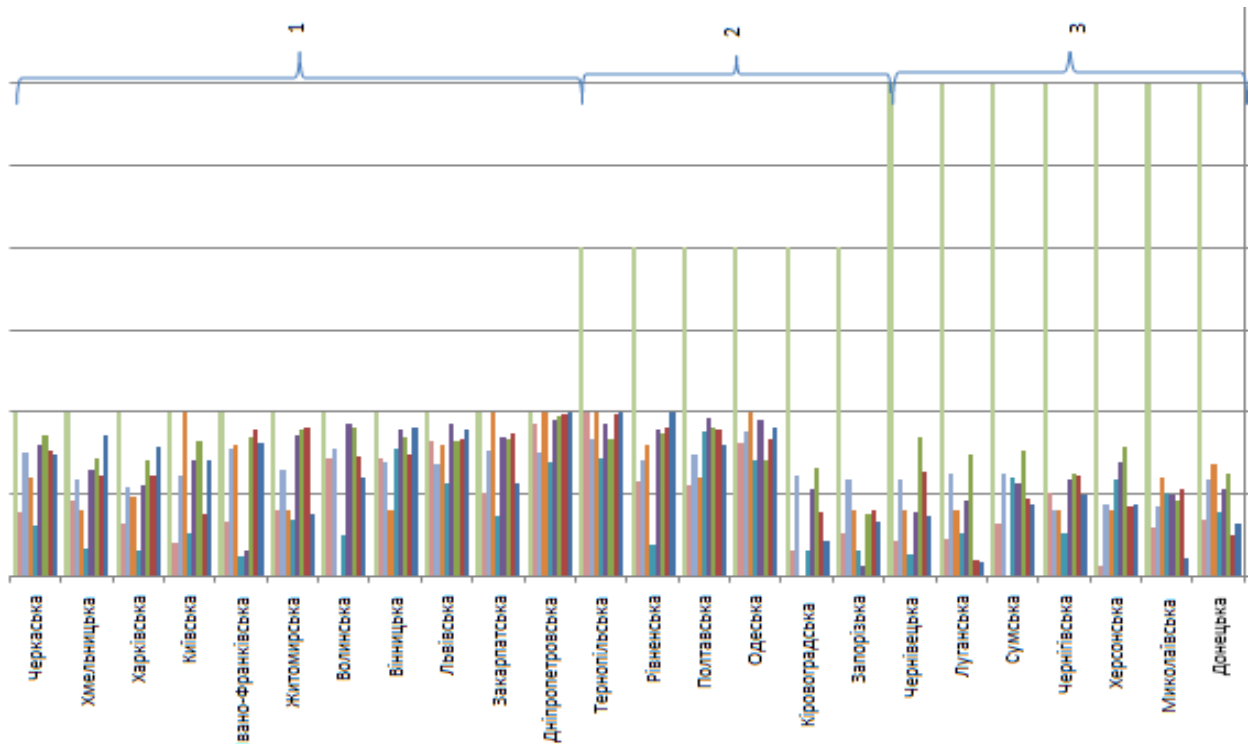


Рисунок 3.2 – Розбиття на кластери, N=3

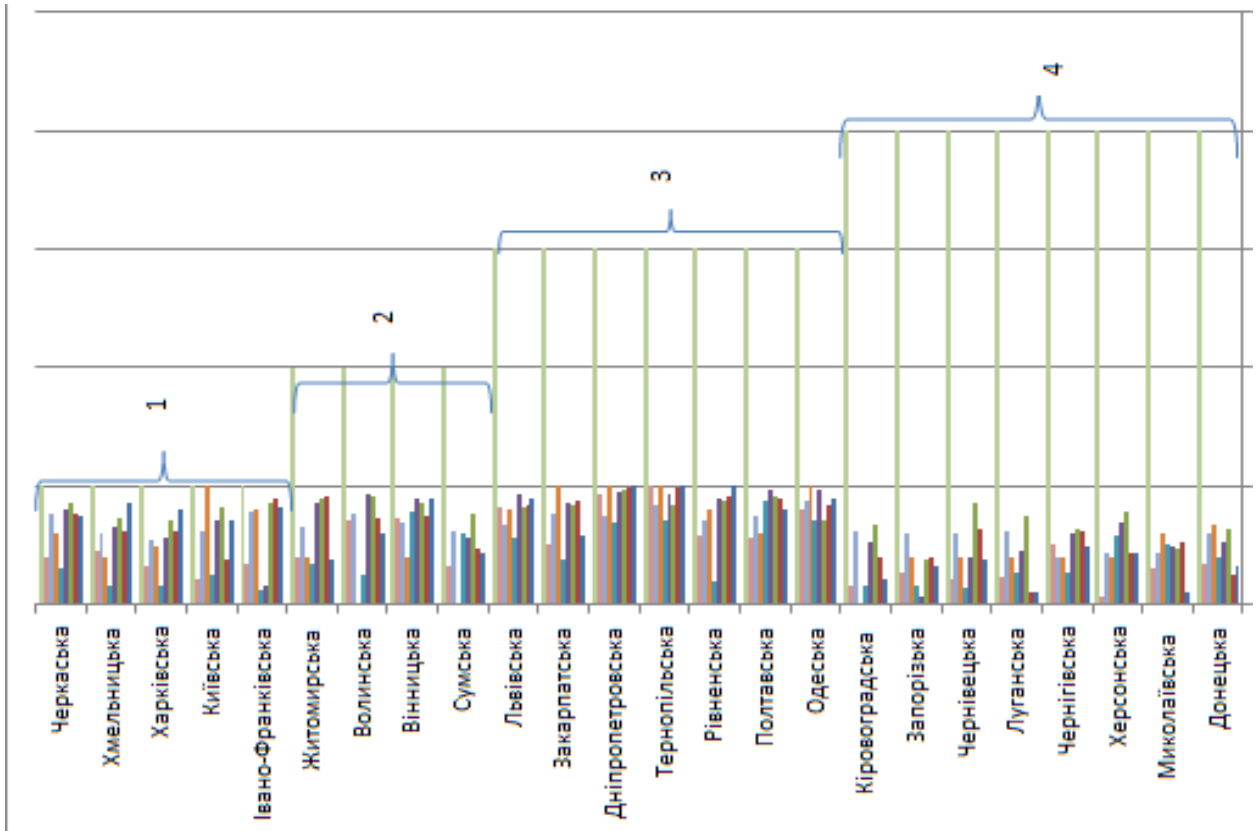


Рисунок 3.3 – Розбиття на кластери, N=4

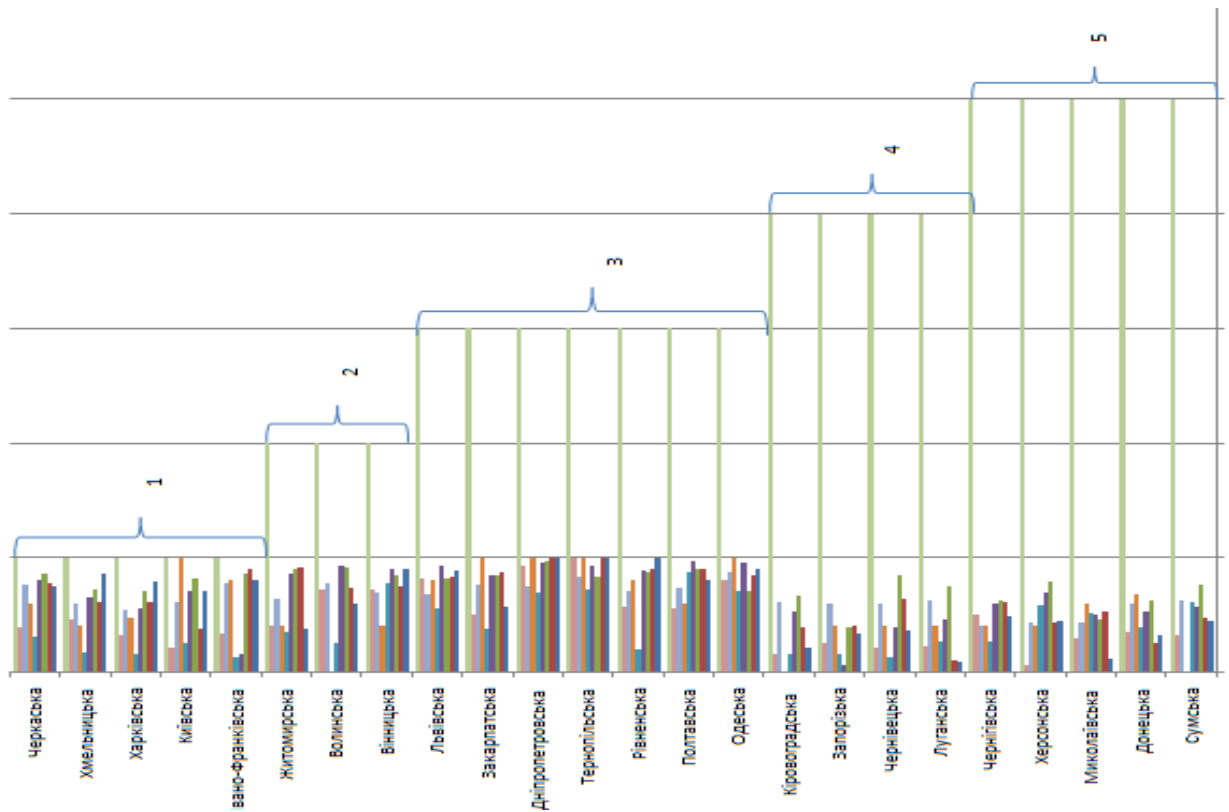


Рисунок 3.4 – Розбиття на кластери, N=5

У другому кластері має місце доволі неоднорідний розподіл областей за показниками, що відображено на рис. 3.2. Незначно покращує ситуацію збільшення кількості кластерів, незважаючи на неоднорідність кластерів 3 і 4 (рис. 3.3). Найкраща картина щодо однорідності спостерігається при розподілі на п'ять кластерів, що візуально дозволяє виділити об'єднання областей за рівнями: високий, вище середнього, середній, нижче середнього, низький (рис. 3.4).

Розподіл даних показав, що деякі області, такі як Київська, мали вищі показники, тоді як менш урбанізовані регіони демонстрували нижчий рівень цифровізації.

Встановлена наступна відповідність між кластером та рівнем цифровізації: кластер 1 – середній рівень цифровізації, кластер 2 – вище середнього, кластер 3 – високий рівень, кластер 4 – найнижчий рівень, кластер 5 – нижче середнього.

Ці регіони вирізняються високою концентрацією ІТ-компаній, активним використанням електронних послуг та високим рівнем цифрової освіти населення.

Інтерпретація кластерів. Кластери 2, 3 – ці області можуть бути використані як приклад для впровадження цифрових ініціатив у інших регіонах. Основний виклик – підтримка темпів зростання та розширення цифрових послуг. Кластер 1 – перехідні регіони, в яких необхідно сфокусуватися на усуненні таких бар'єрів для розвитку цифровізації, як інфраструктурні обмеження. Кластери 4, 5 – це регіони, для яких важливе залучення державних і приватних інвестицій для покращення доступу до цифрових технологій та підвищення цифрової грамотності населення.

Можна зробити висновок, що кластеризація областей України за показниками цифровізації дозволяє краще зрозуміти їхній поточний стан, визначити регіони, які потребують додаткової уваги, та спланувати регіонально-орієнтовані стратегії цифрової трансформації. Виявлені кластери можуть стати основою для подальшого дослідження факторів, які впливають

на рівень цифровізації, та розробки практичних рекомендацій для прискорення цього процесу.

3.3. Прототип аналітичного дашборду для дослідження процесу цифрової трансформації областей України як приклад використання інформаційної технології

Розроблена структура та макети аналітичних панелей наведена на рис. 3.5.

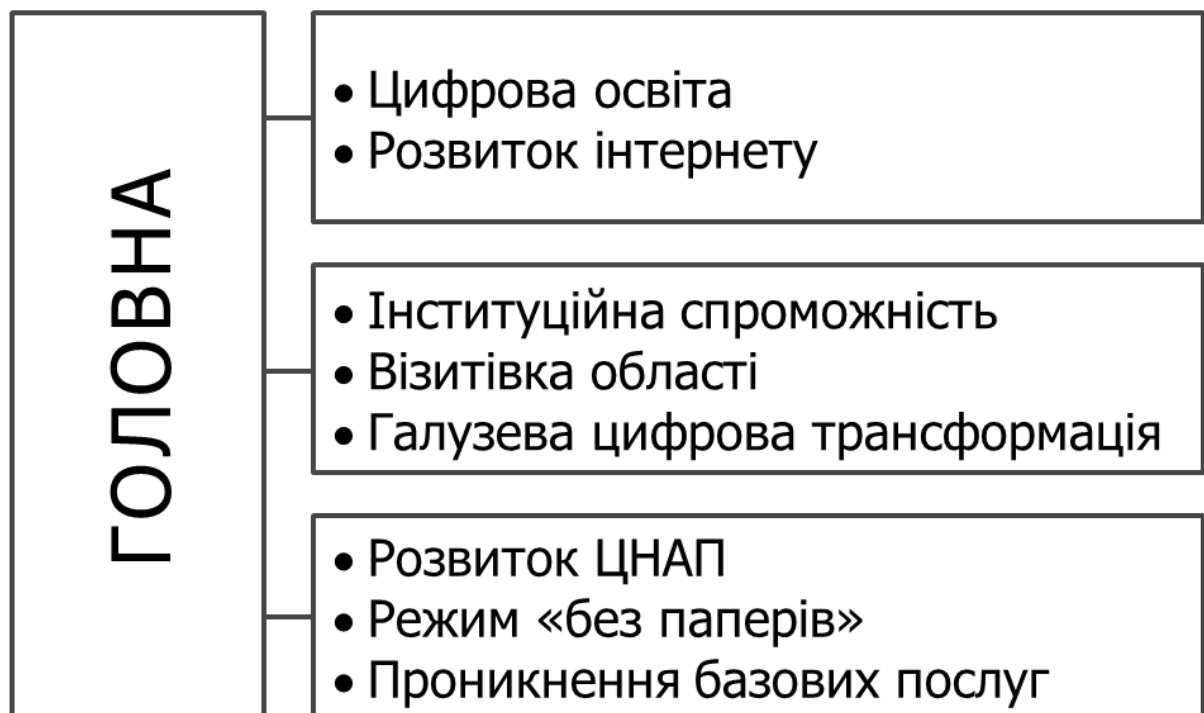


Рисунок 3.5 – Структура та макети аналітичних панелей

Аналітичний дашборд, розроблений за наведеною вище схемою, представлено на рис. 3.6 – 3.9. Головна сторінка містить елементи керування для фільтрації даних, картки основних показників та гістограму індексу (рис. 3.6).

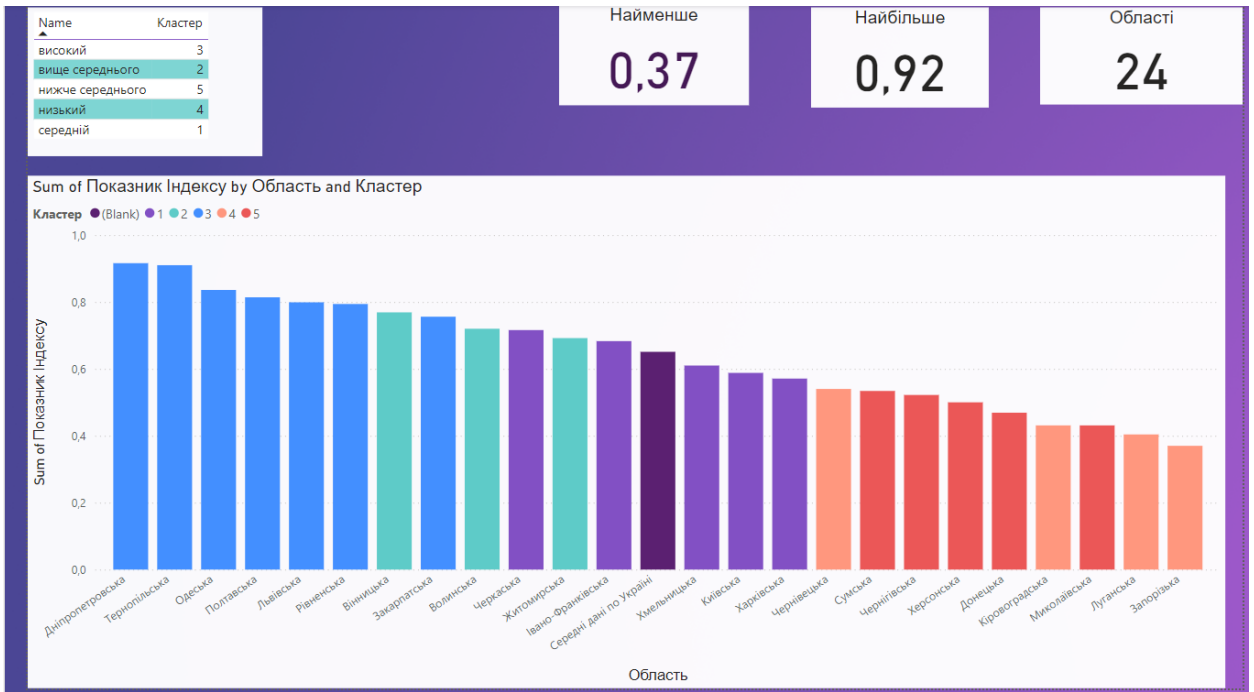


Рисунок 3.6 – Головна сторінка

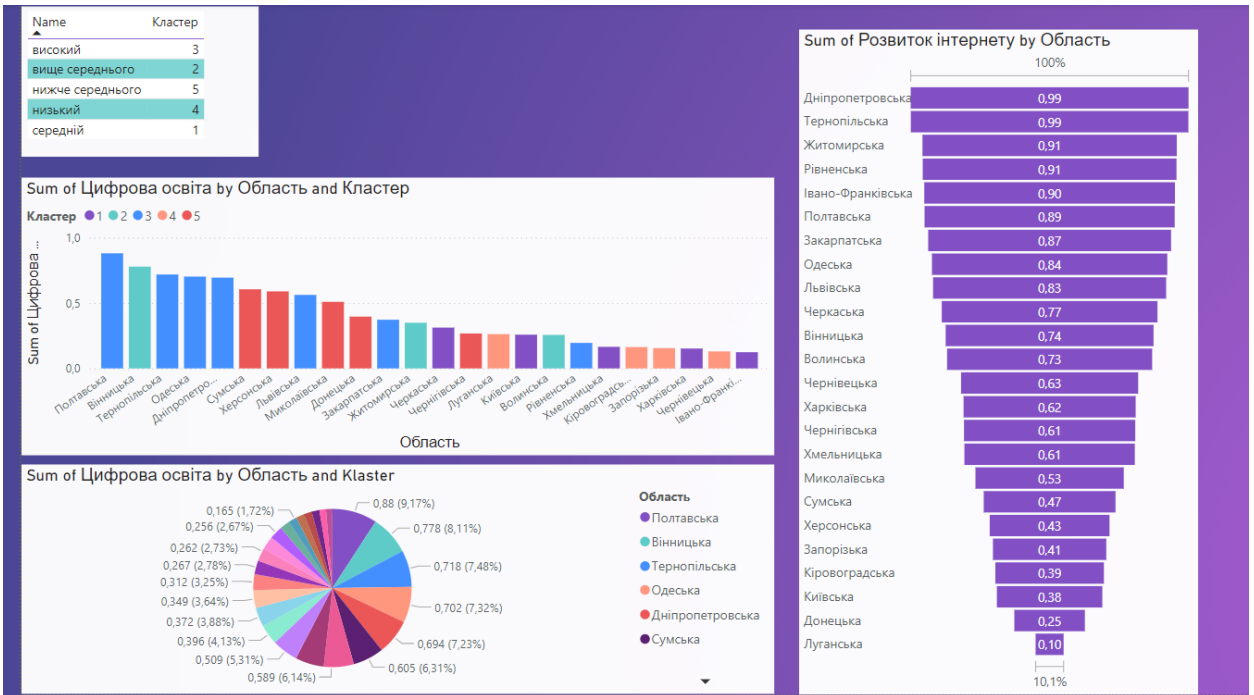


Рисунок 3.7 – Сторінка 1 (показники Цифрова освіта та Розвиток Інтернету)

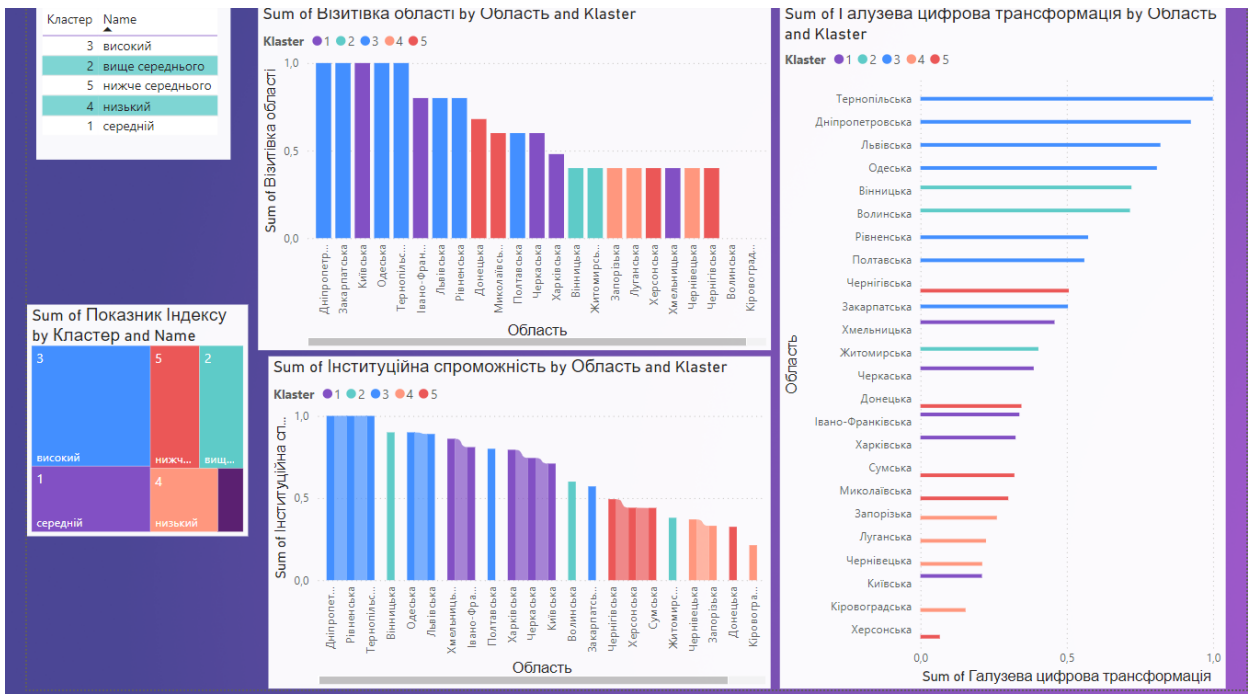


Рисунок 3.8 – Сторінка 2 (Інституційна спроможність, Візитівка області та Галузева цифрова трансформація)

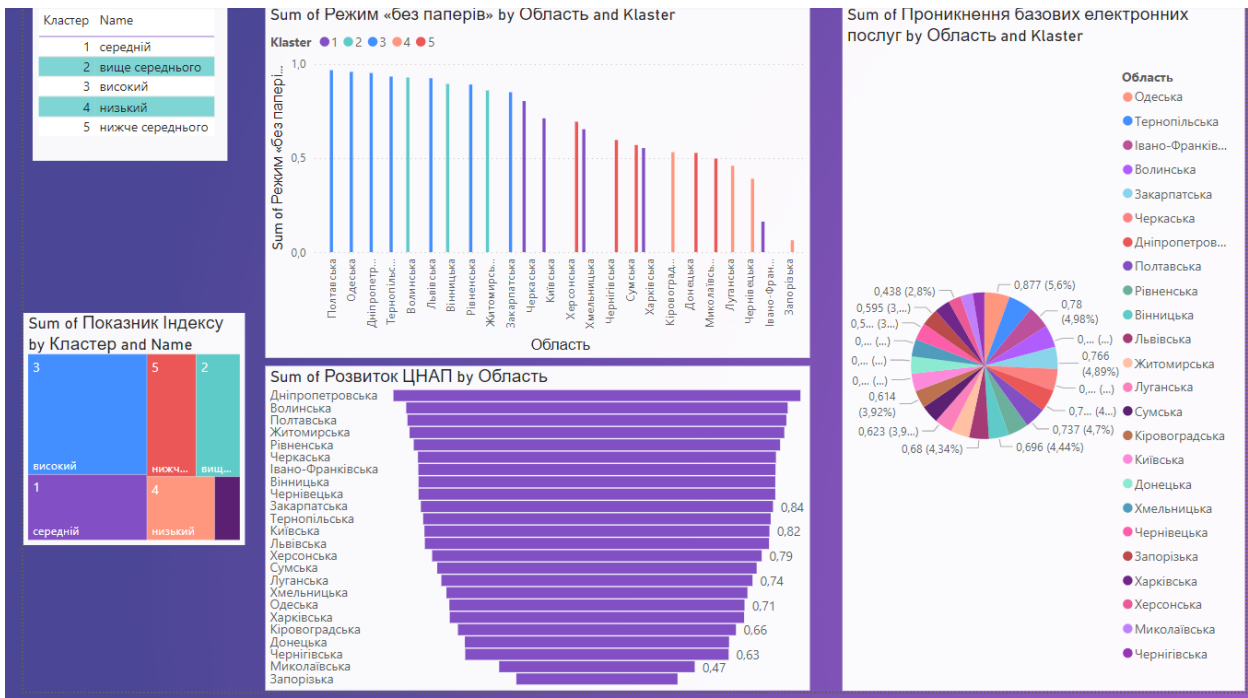


Рисунок 3.9 – Сторінка 2 (Розвиток ЦНАП, Режим «без паперів», Проникнення базових послуг)

Розроблені панелі дають можливість проаналізувати наведені показники. Наприклад, на першій сторінці при виділенні кластерів можна помітити певну залежність між розвитком Інтернету і цифрової освіти (рис. 3.10).

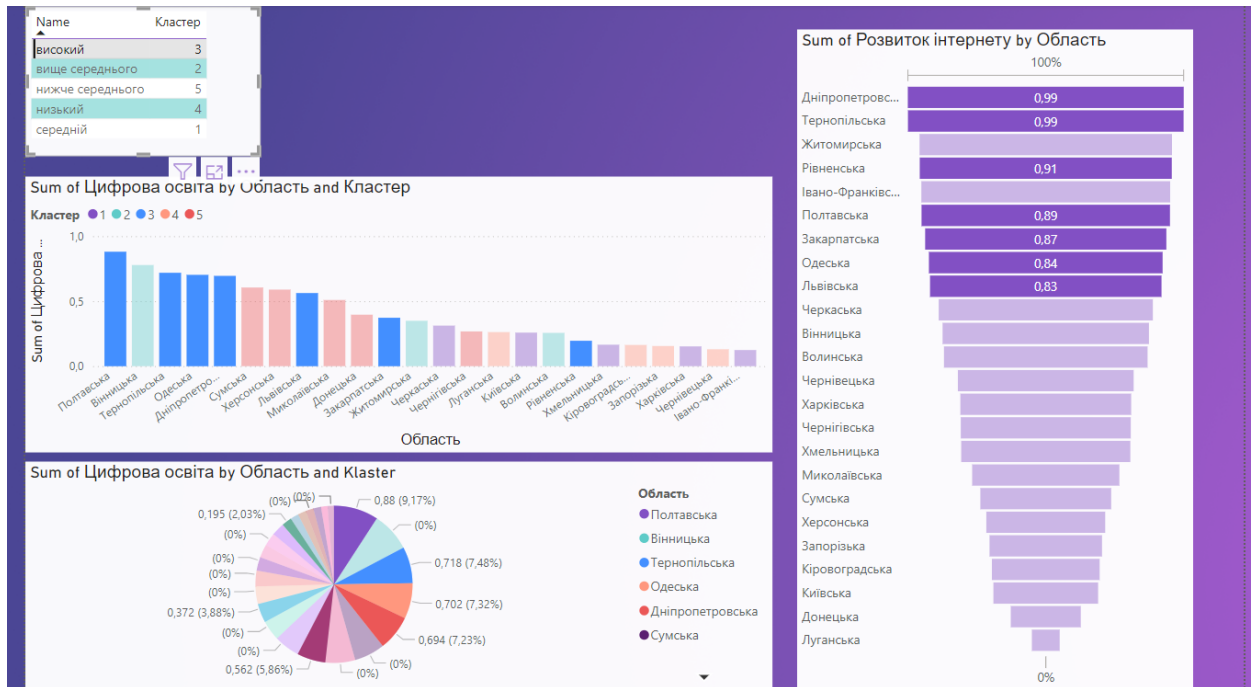


Рисунок 3.10 – Високі показники Розвитку Інтернету та високі показники Цифрової освіти

3.4 Рекомендації щодо впровадження аналітичних дашбордів у наукових установах для покращення моніторингу та управління дослідженнями

Основні етапи впровадження аналітичних дашбордів. Аналіз потреб установи, який включає визначення цілей (наприклад, моніторинг проєктів, аналіз грантової активності чи оцінка публікаційної продуктивності) та вибір ключових показників (наприклад, кількість публікацій, кількість патентів, успішність грантів, темпи виконання проєктів).

Збір і підготовка даних передбачає створення єдиного сховища, яке інтегрує внутрішні джерела, такі як результати досліджень і бюджети, та

зовнішні, включно з бібліометричними даними й рейтингами, а також очищення і нормалізацію цих даних за допомогою інструментів ETL (Extract, Transform, Load) для автоматизації процесів.

Вибір платформи для дашборду залежить від технічних можливостей установи та передбачає аналіз популярних інструментів, зокрема Power BI, Tableau, Python (Dash) і Google Data Studio.

Розробка прототипу включає створення базової версії дашборду з основними функціями та залучення кінцевих користувачів для його тестування.

Реалізація та масштабування передбачає впровадження готового рішення на рівні підрозділів або всієї установи з одночасним забезпеченням інтеграції з існуючими системами управління.

Навчання персоналу включає проведення тренінгів для дослідників і адміністративного персоналу з користування дашбордом, а також розробку інструкцій та супровідної документації.

Практичні рекомендації. Орієнтація на користувача передбачає розробку інтуїтивного дизайну інтерфейсу, який відповідає потребам цільової аудиторії, зокрема науковців, керівників та адміністраторів, а також забезпечення можливості адаптації дашборду під конкретні завдання.

Забезпечення актуальності даних включає налагодження автоматичного оновлення через API або регулярні імпорти, а також використання механізмів перевірки їхньої коректності.

Інтеграція зі сторонніми системами передбачає підключення до баз даних публікацій, таких як Scopus і Web of Science, а також інтеграцію з фінансовими та проєктними системами управління установи.

Забезпечення безпеки включає впровадження механізмів контролю доступу для захисту конфіденційної інформації та використання резервного копіювання для запобігання втратам даних.

Моніторинг і вдосконалення передбачають регулярний збір зворотного зв'язку від користувачів та постійне оновлення дашборду відповідно до нових вимог і технологій.

Очікувані результати впровадження аналітичних дашбордів у наукових установах включають підвищення прозорості та ефективності управління науковими дослідженнями, скорочення часу на пошук і обробку даних, можливість приймати обґрунтовані рішення на основі реальних показників та оптимізацію використання ресурсів установи.

Аналітичні дашборди є ключовим інструментом для моніторингу та управління науковими дослідженнями. Їх впровадження вимагає комплексного підходу, включаючи визначення потреб, підготовку даних, вибір відповідної платформи та навчання персоналу. Дотримання наведених рекомендацій забезпечить науковим установам можливість ефективно використовувати дані для досягнення стратегічних цілей та підвищення конкурентоспроможності у сфері науки.

ВИСНОВКИ

Розроблена та досліджена інформаційна технологія для створення аналітичних дашбордів, яка забезпечує ефективну візуалізацію даних наукових досліджень та дозволяє автоматизувати процеси збору, обробки та інтерпретації даних, а також здатна покращити прийняття рішень у наукових дослідженнях щодо цифрової трансформації областей України.

Для досягнення поставленої мети виконані наступні завдання:

1. Проведено аналіз існуючих методів та технологій створення аналітичних дашбордів для наукових досліджень.
2. Описані вимоги до аналітичного дашборду з урахуванням специфіки наукових даних і процесів.
3. Виділені кластери областей України за схожими показниками цифровізації та досліджені їх особливості.
4. Розроблений прототип аналітичного дашборду для дослідження процесу цифрової трансформації областей України як приклад використання інформаційної технології.
5. Підготовлені рекомендації щодо впровадження аналітичних дашбордів у наукових установах для покращення моніторингу та управління дослідженнями.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Susnjak, T., Ramaswami, G. S., & Mathrani, A. (2022). Learning analytics dashboard: a tool for providing actionable insights to learners. *International Journal of Educational Technology in Higher Education*, 19(1), 12.
2. Valle, N., Antonenko, P., Dawson, K., & Huggins-Manley, A. C. (2021). Staying on target: A systematic literature review on learner-facing learning analytics dashboards. *British Journal of Educational Technology*, 52(4), 1724-1748.
3. Paramayoga, A. L. P., & Kristianto, B. (2024, August). Development of an Analytical Dashboard for Enhancing Digital Data Performance Analysis of PT. Bintang Toedjoe. In *2024 3rd International Conference on Creative Communication and Innovative Technology (IC3INA)* (pp. 1-7). IEEE.
4. Gutiérrez-Braojos, C., Rodríguez-Domínguez, C., Daniela, L., & Carranza-García, F. (2023). An analytical dashboard of collaborative activities for the knowledge building. *Technology, Knowledge and Learning*, 1-27.
5. Sedrakyan, G., Malmberg, J., Verbert, K., Järvelä, S., & Kirschner, P. A. (2020). Linking learning behavior analytics and learning science concepts: Designing a learning analytics dashboard for feedback to support learning regulation. *Computers in Human Behavior*, 107, 105512.
6. Jayashanka, R., Hettiarachchi, E., & Hewagamage, K. P. (2022). Technology Enhanced Learning Analytics Dashboard in Higher Education. *Electronic Journal of e-Learning*, 20(2), 151-170.
7. Alam, M. I., Malone, L., Nadolny, L., Brown, M., & Cervato, C. (2023). Investigating the impact of a gamified learning analytics dashboard: Student experiences and academic achievement. *Journal of Computer Assisted Learning*, 39(5), 1436-1449.
8. Rahman, M. Z. N., Prakasa, E., & Bijaksana, M. A. (2024, October). Developing a Visual Dashboard to Improve Research Activity Assessment. In *2024 International Conference on Computer, Control, Informatics and its Applications (IC3INA)* (pp. 366-371). IEEE.
9. Paramesha, M., Rane, N. L., & Rane, J. (2024). Big data analytics, artificial intelligence, machine learning, internet of things, and blockchain for

enhanced business intelligence. *Partners Universal Multidisciplinary Research Journal*, 1(2), 110-133.

10. Tory, M., Bartram, L., Fiore-Gartland, B., & Crisan, A. (2021). Finding their data voice: Practices and challenges of dashboard users. *IEEE Computer Graphics and Applications*, 43(1), 22-36.

11. Al-Sulaiti, A., Mansour, M., Al-Yafei, H., Aseel, S., Kucukvar, M., & Onat, N. C. (2021, April). Using data analytics and visualization dashboard for engineering, procurement, and construction project's performance assessment. In *2021 IEEE 8th International Conference on Industrial Engineering and Applications (ICIEA)* (pp. 207-211). IEEE.

12. Zare, S. (2023). *Application of data science and visualization for making interactive dashboards in market research* (Doctoral dissertation, Politecnico di Torino).

13. Arputharaj, J. V., Yakub, M. E., Haruna, A. A., & Senthil Kumar, A. (2023, June). Review and Design of Integrated Dashboard Model for Performance Measurements. In *International Conference on Soft Computing and Signal Processing* (pp. 1-10). Singapore: Springer Nature Singapore.

14. Han, J., Kim, K. H., Rhee, W., & Cho, Y. H. (2021). Learning analytics dashboards for adaptive support in face-to-face collaborative argumentation. *Computers & Education*, 163, 104041.

15. Verbert, K., Ochoa, X., De Croon, R., Dourado, R. A., & De Laet, T. (2020, March). Learning analytics dashboards: The past, the present and the future. In *Proceedings of the tenth international conference on learning analytics & knowledge* (pp. 35-40).

16. Sahin, M., & Ifenthaler, D. (2021). Visualizations and dashboards for learning analytics: A systematic literature review. *Visualizations and dashboards for learning analytics*, 3-22.

17. Барченко Н. Л., Любчак В. О., Карінцева О. І., Ковальов Б. Л., Пономаренко І. О. Моделі опису індикаторів прогресу цифрової трансформації економіки // *Вісник СумДУ. Серія Економіка*. 3, 2022. С 42-50.

18. Batt, S., Grealis, T., Harmon, O., & Tomolonis, P. (2020). Learning Tableau: A data visualization tool. *The Journal of Economic Education*, 51(3-4), 317-328.

19. Mittal, M., & Raheja, N. G. (2024). *Data Visualization and Storytelling with Tableau*. CRC Press.
20. Dabbas, E. (2021). *Interactive Dashboards and Data Apps with Plotly and Dash: Harness the power of a fully fledged frontend web framework in Python—no JavaScript required*. Packt Publishing Ltd.
21. Lavanya, A., Gaurav, L., Sindhuja, S., Seam, H., Joydeep, M., Uppalapati, V., ... & SD, V. S. (2023). Assessing the performance of python data visualization libraries: a review. *Int J Comput Eng Res Trends*, 10(1), 29-39.
22. Costas-Jauregui, V., Oyelere, S. S., Caussin-Torrez, B., Barros-Gavilanes, G., Agbo, F. J., Toivonen, T., ... & Tenesaca, J. B. (2021, October). Descriptive analytics dashboard for an inclusive learning environment. In *2021 IEEE Frontiers in Education Conference (FIE)* (pp. 1-9). IEEE.
23. Salas-Pilco, S. Z., Xiao, K., & Hu, X. (2022). Artificial intelligence and learning analytics in teacher education: A systematic review. *Education Sciences*, 12(8), 569.
24. Kerssens, N., & Van Dijck, J. (2022). Governed by edtech? Valuing pedagogical autonomy in a platform society. *Harvard Educational Review*, 92(2), 284-303.
25. Gonçalves, C. T., Gonçalves, M. J. A., & Campante, M. I. (2023). Developing Integrated Performance Dashboards Visualisations Using Power BI as a Platform. *Information*, 14(11), 614.
26. Singh, R. K. (2024). Developing a big data analytics platform using Apache Hadoop Ecosystem for delivering big data services in libraries. *Digital Library Perspectives*, 40(2), 160-186.
27. Jovanoska, D., Nechkoska, R. P., & Mancheski, G. Metabase cockpits as a base for BI in Strategic management. In *International Scientific Conference Strategic Management and Decision Support Systems in Strategic Management, University of Novi Sad*, 103-110..
28. Alam, A., & Mohanty, A. (2023, January). From Bricks to Clicks: The Potential of Big Data Analytics for Revolutionizing the Information Landscape in Higher Education Sector. In *International Conference on Data Management, Analytics & Innovation* (pp. 721-732). Singapore: Springer Nature Singapore.

29. Raschka, S., Liu, Y. H., & Mirjalili, V. (2022). *Machine Learning with PyTorch and Scikit-Learn: Develop machine learning and deep learning models with Python*. Packt Publishing Ltd.

30. Fuhrer, C., Solem, J. E., & Verdier, O. (2021). *Scientific Computing with Python: High-performance scientific computing with NumPy, SciPy, and pandas*. Packt Publishing Ltd..

31. Sial, A. H., Rashdi, S. Y. S., & Khan, A. H. (2021). Comparative analysis of data visualization libraries Matplotlib and Seaborn in Python. *International Journal*, 10(1), 277-281.

32. Géron, A. (2022). *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow*. " O'Reilly Media, Inc."

ДОДАТОК

```

import tkinter as tk
from tkinter import messagebox
from sklearn.cluster import KMeans
import numpy as np
import matplotlib.pyplot as plt
from matplotlib.backends.backend_tkagg import FigureCanvasTkAgg

# Функція для виконання кластеризації K-means
def perform_kmeans(data, n_clusters):
    kmeans = KMeans(n_clusters=n_clusters, random_state=42)
    kmeans.fit(data)
    return kmeans

# Функція для відображення результатів кластеризації на графіку
def plot_clusters(data, kmeans):
    plt.figure(figsize=(6, 6))
    plt.scatter(data[:, 0], data[:, 1], c=kmeans.labels_, cmap='viridis',
marker='o')
    plt.scatter(kmeans.cluster_centers_[:, 0], kmeans.cluster_centers_[:, 1],
c='red', marker='x', s=200)
    plt.title("K-means Clustering")
    plt.xlabel("Feature 1")
    plt.ylabel("Feature 2")
    plt.colorbar(label='Cluster')
    plt.show()

# Функція для обробки введених даних і виконання кластеризації
def on_run_button_click():
    try:
        n_clusters = int(cluster_count_entry.get())
        data_str = data_entry.get().strip()

        # Перевірка на порожній ввід
        if not data_str:
            raise ValueError("Data is empty.")

        # Конвертація введених даних в масив чисел
        data = np.array([list(map(float, row.split(','))) for row in
data_str.split('\n')])

```

```

# Перевірка на відповідність кількості ознак
if len(data.shape) != 2 or data.shape[1] < 2:
    raise ValueError("Data must have at least 2 features (columns).")

# Виконання кластеризації
kmeans = perform_kmeans(data, n_clusters)

# Виведення результатів
messagebox.showinfo("Clustering Done", f"K-means completed!
{n_clusters} clusters formed.")

# Відображення результатів на графіку
plot_clusters(data, kmeans)
except Exception as e:
    messagebox.showerror("Error", f"An error occurred: {e}")

# Створення графічного інтерфейсу за допомогою tkinter
root = tk.Tk()
root.title("K-means Clustering")

# Введення кількості кластерів
tk.Label(root, text="Enter number of clusters (K):").pack(pady=10)
cluster_count_entry = tk.Entry(root)
cluster_count_entry.pack(pady=10)
cluster_count_entry.insert(0, "3") # Значення за замовчуванням

# Введення даних
tk.Label(root, text="Enter data (comma-separated for each feature, rows
separated by newlines):").pack(pady=10)
data_entry = tk.Text(root, height=10, width=40)
data_entry.pack(pady=10)

# Кнопка для запуску кластеризації
run_button = tk.Button(root, text="Run K-means", command=on_run_button_click)
run_button.pack(pady=20)

# Запуск GUI
root.mainloop()

```