

## **ВІДГУК РЕЦЕНЗЕНТА**

доктора технічних наук, професора кафедри систем штучного інтелекту  
Національного університету «Львівська політехніка»

**Мельникової Наталії Іванівни**

на дисертаційну роботу ТКАЧИКА ОЛЕКСАНДРА АНДРІЙОВИЧА  
на тему «МЕТОДИ ТА ЗАСОБИ КЛАСТЕРИЗАЦІЇ РІЗНОТИПОВИХ  
ДАНИХ»

подану на здобуття наукового ступеня доктора філософії  
за спеціальність спеціальністю 122 – Комп'ютерні науки

### **1. Актуальність теми дисертаційної роботи та зв'язок з науковими планами і програмами**

Покращення персоналізації є одним з найважливіших завдань для компаній у сфері нерухомості. Сьогодні клієнти очікують індивідуального підходу та послуг, які відповідають їхнім потребам та уподобанням. За допомогою створення профілів користувачів та використання методів кластеризації даних, можна отримати унікальну можливість розуміти більш глибоко потреби та уподобання клієнтів, що дозволить компаніям забезпечити персоналізовані пропозиції та покращити задоволеність клієнтів.

Електронна комерція нерухомості швидко розвивається та стає все більш популярною. Онлайн-платформи надають споживачам зручність та доступність при пошуку, виборі та придбанні нерухомості. Проте, великий обсяг інформації та зростаюча конкуренція вимагають впровадження персоналізованих підходів. Шляхом створення профілів користувачів та їх класифікації за допомогою методу кластеризації даних, компанії можуть пропонувати індивідуальні рекомендації та пропозиції, що найкраще відповідають унікальним потребам та побажанням кожного клієнта.

Покращення обслуговування клієнтів завжди було і залишається головним завданням у будь-якому бізнесі. У сфері нерухомості, де взаємодія з клієнтами має велике значення, розуміння їхніх потреб, вимог та

попереднього досвіду стає критичним. Створення профілів користувачів та використання методу кластеризації даних допомагає уточнити та систематизувати інформацію про клієнтів, що в свою чергу сприяє покращенню комунікації, індивідуального підходу та задоволеності клієнтів.

Тема дисертації Ткачика Олександра Андрійовича відповідає сучасним перспективним напрямкам розвитку науки та техніки України, а дисертаційна робота виконана в межах пріоритетного напрямку розвитку науки і техніки.

## **2. Аналіз змісту дисертації. Ступінь обґрунтованості наукових положень, висновків і рекомендацій, сформульованих у дисертації**

Результати дисертаційного дослідження є достовірними та завершеними. Достовірність отриманих результатів підтверджується: коректністю основних припущень і положень, на яких ґрунтуються методи машинного навчання, математичного та об'єктно-орієнтованого моделювання програмного забезпечення; відповідністю до існуючих результатів в прикладах використання запропонованих методів; представленням результатів апробації запропонованих методів інформаційної технології та програмно-апаратного комплексу.

Усі етапи проведеного дисертаційного дослідження мають теоретичне та практичне обґрунтування, що свідчать про достовірність тверджень автора.

Дисертаційна робота складається зі вступу, чотирьох розділів, висновків, списку використаних джерел із 103 найменувань. Загальний обсяг роботи складає 124 сторінок друкарського тексту, з них 135 сторінка основного тексту.

У **вступі** розкрита тема дисертаційного дослідження та її актуальність, а також сформовано наукову новизну та практичне значення одержаних результатів, мету та завдання дослідження. Крім того наведено свідчення проведеної апробації та перелік опублікованих праць за темою дисертації із зазначенням особистого внеску автора у роботах виконаних у співавторстві.

У **першому** розділі було проведено аналіз методів обробки різноманітних видів даних, таких як K-середніх, DBSCAN, ієрархічна кластеризація та нечітка кластеризація. У цьому аналізі були виявлені різні виклики та обмеження, включаючи питання щодо масштабування, інтерпретації, універсальності та адаптивності до змінних умов і структур даних. Також було відзначено, що не всі існуючі системи, які присутні на ринку, використовують інформацію з соціальних мереж або інших неструктурованих або напівструктурованих джерел даних для покращення своїх послуг. Це відкриває можливості для подальшого розширення та вдосконалення таких систем. На основі цих спостережень можна стверджувати, що необхідно проводити додаткові дослідження з метою розробки більш ефективних і гнучких методів кластеризації для обробки різноманітних видів даних. Це не тільки сприятиме підвищенню точності та інформативності користувацьких профілів, але також відкриє нові можливості для використання цих даних у сферах бізнес-аналітики, прогнозування та інших цілях.

У **другому** розділі виконано аналіз методів кластеризації даних. Проведено формальне визначення профілів користувачів, яке охоплює різні аспекти, включаючи демографічні характеристики, психографічні особливості, мотиваційні фактори, інформаційні вподобання та інші показники. Кожен з цих критеріїв дозволяє відібрати підмножину даних, зібраних за конкретними характеристиками. Отримані підмножини даних включають в себе різноманітні типи інформації, такі як числові дані, текстові дані, оцінки в порядковій шкалі, категоріальні дані та інше. Детально описано процес класифікації профілів користувачів, що базується на атрибутах користувачів. Також виконано оцінку рівня задоволеності користувачів за допомогою різних метрик, таких як CSS, NPS, CES та CSI. Кожна з цих метрик допомагає оцінити ступінь задоволеності клієнтів різними аспектами, такими як якість послуг, спілкування з брокерами та агентами нерухомості, процес транзакції, рівень професіоналізму, ціноутворення, умови контракту і

загальний досвід співпраці з компанією або агентством нерухомості. Ці метрики дозволяють отримати більш глибоке розуміння потреб користувачів.

У **третьому** розділі розглянуто алгоритм підготовки даних. Виконано аналіз і порівняння різних методів обробки відсутніх даних. Також проведено аналіз методів виявлення і видалення дублікатів, розглянуто та зіставлено різні підходи до виявлення та вилучення викидів. Досліджено роботу методів зменшення розмірності даних і виділення нових ознак. Статистичний метод перцентилів був використаний для обчислення початкових центроїдів. Також розроблено метод кластеризації різноманітних типів даних, що дозволяє працювати з потоковими даними, розділяючи їх на пакети.

У **четвертому** розділі дисертаційного дослідження була розроблена архітектура інформаційної системи для автоматизованого профілювання користувачів на основі різних типів даних клієнтів.

У **висновках** висвітлена тема та завдання дисертаційного дослідження у повній мірі.

### **3. Наукова новизна отриманих результатів**

**Мета дисертаційної роботи** є розроблення методів та моделей аналізу різнотипових даних, зокрема, потокових даних та напівструктурованих даних для підвищення рівня задоволеності клієнта.

У роботі досягнуто таких науковиз результатів:

Вперше розроблено метод кластеризації різнотипових даних, який дозволяє працювати з потоковими даними на основі поділу на пакети та врахування зважування характеристик під час препроцесингу даних.

Вперше формалізовано профіль клієнта, яка відрізняється від існуючих тим, що включає поведінкові та психографічні характеристики, що дає можливість визначити рівень задоволеності клієнта.

Вперше розроблено ієрархічний класифікатор профілів клієнтів, який, на відміну від існуючих рішень, застосовує кластеризацію на першому рівні та

зважування відгуків на другому, що дає можливість пришвидшити обслуговування клієнтів.

#### **4. Оформлення дисертації, дотримання вимог академічної доброчесності та повнота викладу наукових результатів в опублікованих працях**

Дисертаційну роботу написано українською мовою, на хорошому стилістичному рівні. Застосована в роботі наукова термінологія є загальноновизнаною, стиль викладення результатів теоретичних і практичних досліджень, нових наукових положень, висновків і рекомендацій забезпечує доступність їх сприйняття та використання. Стиль викладу матеріалів досліджень і наукових положень забезпечує їх належне сприйняття. Оформлення дисертації відповідає усім необхідним атестаційним вимогам. Проведена перевірка дисертації на наявність академічного плагіату, отримані результати свідчать про високу індивідуальність роботи. По всьому тексту дисертації простежується авторський стиль. У дисертації не виявлено використання результатів досліджень інших науковців без посилань на відповідні джерела.

Основні результати дисертаційного дослідження опубліковано 7 наукових праць, зокрема 3 статті у фахових наукових виданнях, 1 стаття з Q1, 3 публікацій у матеріалах конференцій. Обсяг друкованих робіт, їх зміст і кількість відповідають вимогам щодо публікацій основного змісту дисертації на здобуття наукового ступеня доктора філософії.

#### **5. Теоретичне та практичне значення одержаних результатів**

Теоретичне значення одержаних результатів полягає у розробленні методу кластеризації для обробки різнотипових даних, який дозволяє працювати з потоковими даними, розділяючи їх на пакети та враховуючи вагомість характеристик під час попередньої обробки даних, формалізованого профілю клієнта, який відрізняється від існуючих підходів, оскільки він

включає в себе як поведінкові, так і психографічні аспекти, що дозволяє оцінити рівень задоволеності клієнта та ієрархічного класифікатора профілів клієнтів, який, відмінно від існуючих рішень, використовує класифікацію на першому рівні та зважування відгуків на другому, що призводить до прискорення обслуговування клієнтів

Практичне значення одержаних результатів представлено у вигляді інформаційної системи, яка призначена для створення та управління профілями користувачів, що цікавляться інвестиціями, продажем або орендою нерухомості. Обрано поширений тип інформаційної системи – веб-застосунок, який забезпечує взаємодію з користувачами через API-сервіси [88] та зберігання необхідних даних у базі даних. Ця система має потенціал підтримувати різноманітні клієнтські потреби, аналізуючи їхню поведінку, інтереси та вподобання. Використання веб-застосунку дозволяє забезпечити широкий охоплюючий доступ до системи для користувачів з будь-якого місця, де є доступ до Інтернету. Веб-інтерфейс дозволяє зручно взаємодіяти з системою та забезпечує зручність і легкість використання для користувачів різного рівня технічної грамотності..

Впровадження запропонованої архітектури призвело до зниження кінцевої вартості системи в 4 рази порівняно зі стандартною архітектурою, що використовує дроплети..

## **6. Зауваження та дискусійні положення щодо змісту дисертації**

Загалом результати досліджень представлені в дисертації подані на достатньому науковому рівні. Проте, деякі його положення є дискусійними, що пов'язано з складністю досліджуваної предметної області, варто зробити певні уточнення, зауваження та рекомендації з метою подальшого удосконалення досліджуваної проблематики.

1. Необхідно доопрацювати визначення об'єкту, предмету та методів дослідження.

2. У роботі доцільно було визначити основні характеристики різнотипності даних, що впливають на процес їхньої кластеризації.
3. Доцільно щоб визначені задачі дисертаційної дослідження корелювалися із задачами, що описані в 4 п. розділу 1.
4. Недостатньо проведено порівняльний аналіз існуючих методів кластеризації для роботи з визначеним типом даних, не наведені метрики оцінки роботи методів.
5. У розділі 3 було б доцільно структурувати виклад пропонованого методу ієрархічного класифікатора профілів клієнтів.
6. У роботі зустрічається багато описок та пунктуаційних неточностей.

## **7. Загальний висновок**

Дисертація Ткачика Олександра Андрійовича на тему «Методи та засоби кластеризації різнотипових даних» є завершеною науковою працею, у якій розв'язано конкретне науково-прикладне завдання створення математичного забезпечення та програмно-технічного комплексу із власною архітектурою та удосконаленою інформаційною технологією, яка б забезпечила надійну роботу усіх складових системи, як єдиного пристрою для вирішення задачі класифікації профілів користувача та покращення рівня задоволеності клієнта у сфері надання послуг.

Дисертація виконана на належному науково-методичному рівні, написана науковим стилем, матеріал вкладений у логічній послідовності, висновки науково обґрунтовані та підтверджуються результатами досліджень. Опубліковані здобувачем роботи, їх апробація повною мірою відображають основні положення дисертації. Анотація розкриває основні її положення та висновки, є ідентичною дисертації за структурою та змістом, а також не містить інформації, відсутньої у дисертації. Висловлені зауваження та дискусійні положення не впливають на загальну позитивну оцінку роботи.

Зміст дисертаційної роботи відповідає задекларованій темі, забезпечує досягнення поставленої мети і вирішення завдань дослідження та відповідає вимогам, встановленим Порядком проведення експерименту з присудження ступеня доктора філософії, затвердженого постановою Кабінету міністрів України від 06.03.2019 р. № 167, а її автор Ткачик Олександр Андрійович заслуговує на присудження наукового ступеня доктора філософії за спеціальністю 122 – Комп'ютерні науки.

Рецензент



Наталія МЕЛЬНИКОВА

професор кафедри систем штучного інтелекту

Національного університету «Львівська політехніка»

доктор технічних наук

Вчений секретар

Національного університету «Львівська політехніка»,

к.т.н., доц.



Роман БРИЛИНСЬКИЙ