

Голові разової спеціалізованої вченої ради
Національного університету
«Львівська політехніка»
д.т.н., професору Камінському Роману Миколайовичу

ВІДГУК РЕЦЕНЗЕНТА

Доктора технічних наук, професора Жежничка Павла Івановича
на дисертаційну роботу **Ткачика Олександра Андрійовича**
«Методи та засоби кластеризації різнотипових даних»,
подану до захисту на здобуття наукового ступеня **доктора філософії**
з галузі знань 12 «Інформаційні технології»
та спеціальності 122 «Комп'ютерні науки»

1. Загальна характеристика роботи

Дисертація доктора філософії фокусується на розробці методів штучного інтелекту для створення користувацьких профілів на онлайн-платформі нерухомості, метою якої є оптимізація взаємодії між менеджерами системи та користувачами та підвищення задоволеності клієнтів. Робота включає аналіз та вдосконалення алгоритмів кластеризації даних, розробку алгоритму підготовки даних, включаючи обробку пропущених значень та викидів, а також розробку інноваційної архітектури інформаційної системи для автоматизованого профілювання користувачів, з використанням безсерверного підходу для зниження вартості розгортання.

Загалом, ця дисертація представляє собою важливе дослідження в області застосування штучного інтелекту в онлайн-бізнесі, зокрема у сегменті нерухомості, зосереджуючись на підвищенні ефективності і задоволеності користувачів через інноваційні технологічні рішення та є завершеним науковим дослідженням.

2. Оцінка актуальності теми дисертації

Актуальність теми дисертації полягає у відповідності сучасним тенденціям ринку нерухомості, де необхідні інноваційні підходи до персоналізації обслуговування клієнтів та електронної комерції. З урахуванням швидкого розвитку технологій та доступності великих обсягів даних, існує унікальна можливість використання методів кластеризації даних для створення детальних профілів користувачів, зацікавлених у нерухомості.

Персоналізація послуг є ключовим завданням для компаній у сфері нерухомості, оскільки клієнти сьогодні очікують індивідуалізованого підходу та послуг, які відповідають їхнім потребам і уподобанням. Використання методів

кластеризації даних для створення профілів користувачів дозволяє компаніям надавати персоналізовані пропозиції, що підвищує задоволеність клієнтів.

Розвиток електронної комерції в галузі нерухомості, що характеризується швидким зростанням та популярністю, вимагає адаптації до великої кількості інформації та конкуренції. Створення індивідуалізованих профілів користувачів і їх класифікація за допомогою кластеризації даних дозволяє надавати більш точні рекомендації та пропозиції, відповідні унікальним потребам кожного клієнта.

Поліпшення обслуговування клієнтів залишається основною метою для бізнесу в сфері нерухомості, де важливим є розуміння потреб, вимог та досвіду клієнтів. Використання профілів користувачів та методу кластеризації даних сприяє кращій систематизації інформації про клієнтів, що покращує комунікацію, індивідуальний підхід та задоволеність клієнтів.

Враховуючи вищезазначене, можна зробити висновок, що підвищення рівня персоналізації послуг, розвиток електронної комерції у нерухомості та покращення якості обслуговування клієнтів є актуальними та важливими аспектами в сучасному ринку нерухомості, що підтверджує актуальність теми дослідження.

Робота виконувалася в рамках держбюджетної теми кафедри систем штучного інтелекту «Методи та засоби інформаційної безпеки та гігієни на основі інтерпретованого штучного інтелекту» (№ державного реєстру 0123U101687).

3. Наукова цінність дисертаційного дослідження

Наукова цінність дисертаційного дослідження, зосередженого на розвитку методів кластеризації та профілювання клієнтів у сфері онлайн-нерухомості та включає декілька аспектів:

- Автор проводить детальний порівняльний аналіз сучасних алгоритмів кластеризації, що дозволяє оцінити їх ефективність та придатність для різнотипових даних. Це порівняння є важливим для визначення оптимальних методів, які можуть бути використані в подальшому дослідженні.
- На основі проведеного аналізу дослідження розширює існуючі методи кластеризації, впроваджуючи підхід поділу на пакети та зважування характеристик під час препроцесингу даних.
- Розробка нової моделі профілю клієнта, яка включає поведінкові та психографічні характеристики, є важливим кроком у покращенні індивідуального підходу до клієнтів.

- Введення нового класифікатора, який використовує зважування відгуків на першому етапі та кластеризацію на другому, є інноваційним рішенням, яке сприяє підвищенню швидкості та ефективності обслуговування клієнтів.

4. Оцінка практичного значення результатів роботи

Практичне значення роботи, що фокусується на розробці інформаційних технологій для профілювання користувачів у сфері онлайн-нерухомості, виявляється у вирішенні ключових науково-прикладних завдань у галузі комп'ютерних наук. Розроблені методи та інструменти спрямовані на покращення персоналізації послуг та підвищення задоволеності клієнтів, що є важливими факторами в сучасній сфері нерухомості.

Крім того, результати дослідження сприяють проектуванню та розгортанню сучасних інформаційних систем для автоматизації процесу профілювання користувачів, що значно зменшує витрати та підвищує ефективність роботи компаній у сфері нерухомості. Таким чином, ця робота вносить вагомий вклад у розвиток інформаційних технологій у галузі нерухомості.

Запропоновані методи і моделі впроваджені у навчальний процес Національного університету «Львівська політехніка» при викладанні дисципліни «Системний аналіз». Розроблені методи та моделі впроваджено у ТЗОВ Reintech.

Результати впровадження підтверджені відповідними актами.

5. Оцінка достовірності та обґрунтованості основних положень і висновків дисертації

Наукові положення, висновки та пропозиції у дисертації базуються на теоретичній основі та підкріплені експериментальними дослідженнями, що забезпечує їхню цілісність та надійність.

Теоретична частина дослідження ґрунтується на фундаментальних концепціях прикладної математики та комп'ютерних наук, використовуючи загальні та спеціалізовані методи математичної статистики, що є важливим для аналізу великих даних та кластеризації. Впровадження сучасних методів кластеризації, таких як потокова кластеризація та інші спеціалізовані техніки, демонструє глибоке розуміння потреб і викликів, пов'язаних із обробкою різнотипових даних.

Експериментальні дослідження, проведені автором, включають ретельний аналіз і порівняння різних методів кластеризації, що дозволяє оцінити їх ефективність у реальних сценаріях. Розроблені методи для покращення

швидкодії та якості кластеризації підтверджуються за допомогою математичної статистики та логічних аргументів.

Усі теоретичні припущення та аналітичні залежності в дисертації є коректними і не містять внутрішніх протиріч, що забезпечує обґрунтованість висновків. Додатково, апробація результатів на міжнародних конференціях та наукових семінарах підтверджує їх практичну цінність та актуальність. Отже, можна визнати, що дисертаційна робота є надійною та вагомою в контексті онлайн-нерухомості.

Нові наукові результати та положення повністю відображені у 7 роботах: 1 стаття у науковому виданні квартилю Q1, що входить до наукометричної бази Scopus; 3 статті у фахових наукових виданнях України; 3 тези доповідей у збірниках міжнародних конференцій. Кількість публікацій, їх повнота та обсяг у достатній мірі відображають особистий внесок автора і відповідають вимогам, що висуваються до дисертації на здобуття наукового ступеня доктора філософії.

6. Оцінка змісту й оформлення дисертації

Дисертація оформлена коректно, згідно чинних вимог.

У вступі обґрунтовано актуальність задачі дослідження, виділено мету, задачі та структуру дисертаційної роботи.

Перший розділ присвячений аналізу алгоритмів обробки даних, таких як K-середніх, DBSCAN, ієрархічна та нечітка кластеризація. Виявлено обмеження цих методів, що включають масштабування, інтерпретацію, універсальність та адаптивність до змінних умов. Розкрито потенціал подальшого розвитку кластеризаційних методів для різнотипових даних.

У другому розділі зосереджено увагу на методах кластеризації даних та формуванні профілів користувачів. Розглянуто різні критерії, включаючи демографічні, поведінкові, психографічні, мотиваційні та інформаційні. Описано процес класифікації та визначення рівня задоволеності клієнтів за допомогою метрик CSS, NPS, CES, CSI.

Третій розділ присвячений розробці алгоритму підготовки даних, включаючи обробку пропущених значень, виявлення та усунення дублікатів і викидів, а також методів зменшення розмірності даних. Розроблено методіку для кластеризації різнотипових даних, яка ефективно працює з потоковими даними.

У четвертому розділі представлено архітектуру інформаційної системи для автоматизації профілювання користувачів. Запропоновано використання безсерверного підходу на основі Google Cloud Functions, що знижує вартість розгортання системи у чотири рази порівняно з традиційними методами.

7. Зауваження до дисертаційної роботи

Попри позитивне враження, в роботі помічено певні недопрацювання, зокрема:

1. Тема роботи сформульована невдало, оскільки основний акцент досліджень направлено на створення профілів користувачів та управління ними за допомогою кластеризації різнотипових даних.
2. У роботі некоректно використовується поняття «різнотипові дані», визначення якого у п.1.2 описується значно ширший усталений термін «heterogeneous data», тобто «неоднорідні дані».
3. У практичному значенні одержаних результатів зустрічаються такі оцінки як «стабільно швидкі результати» (п.2) та «збереження якості кластеризації» (п.3), які не мають явних підтверджень в основному тексті роботи.
4. У п.2.4 некоректно вживається поняття «критерій профілю користувача». В контексті викладеного матеріалу доцільно було б вживати термін «критерій відповідності профілю користувача» або «критерій задоволеності клієнта». Також з п.2.4.2 не зрозуміло як «визначення сили зв'язку між об'єктами» впливає на «профіль користувача».
5. У роботі зустрічаються невідповідності між «методами» та «алгоритмами», наприклад: п.2.1 та має заголовок «Аналіз методів кластеризації даних», а розділ 3 має заголовок «Розроблення алгоритмів опрацювання різнотипових даних», хоча у вмісті мова йде як про «алгоритми», так і про «методи».
6. Для наведених схем інформаційної системи доцільно було вибрати один варіант нотації, наприклад UML. Також у роботі зустрічаються орфографічні помилки, зокрема у формулах п.3.2.2.

Наведені зауваження не знижують загального враження та позитивної оцінки представленого дисертаційного дослідження.

8. Загальна оцінка дисертаційної роботи

Дисертаційна робота Ткачика Олександра Андрійовича «Методи та засоби кластеризації різнотипових даних» є завершеною науковою працею і має важливе значення онлайн-платформ ринків нерухомості і для розвитку комп'ютерних наук загалом. Вирішено актуальну задачу, яка полягає у пришвидшенні обслуговування клієнтів у системі ринку нерухомості.

Одержані нові розв'язки науково-прикладної проблеми. Актуальність, практичне значення, новизна та завершеність досліджень, обґрунтування висновків заслуговують позитивної оцінки.


Зміст дисертаційної роботи, отримані основні наукові положення та висновки відповідають паспорту спеціальності 122 «Комп'ютерні науки».

Вказані у п. 7 цього відгуку зауваження щодо представленого дослідження не знижують вагомість отриманих у роботі наукових та практичних результатів і не змінюють її позитивної оцінки.

Робота відповідає вимогам вимогам «Порядку присудження ступеня доктора філософії», затвердженого постановою Кабінету Міністрів України від 12 січня 2022 р. № 44, а її автор **Ткачик Олександр Андрійович** заслуговує на присудження наукового ступеня доктора філософії за спеціальністю 122 «Комп'ютерні науки».

Рецензент – Проректор з науково-педагогічної роботи та інформатизації

доктор технічних наук, професор,

Національний Університет «Львівська Політехніка»  Павло ЖЕЖНИЧ

«Підпис Жежнич П.І. засвідчую»:

Вчений секретар

Національного університету «Львівська політехніка»,

к.т.н., доц.



Роман БРИЛИНСЬКИЙ