

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

Відгук офіційного опонента

докторки технічних наук, професорки Говорущенко Тетяни Олександрівни 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. Актуальність теми дисертаційної роботи

Дослідження ринку нерухомості показують зростаючу потребу у покращенні персоналізації послуг. Завдяки розвитку штучного інтелекту компанії мають можливість створювати детальні профілі користувачів, які дозволяють краще розуміти потреби та уподобання клієнтів та робити більш персоналізованим підхід до клієнта, що є важливим аспектом для забезпечення пропозицій та поліпшення задоволеності клієнтів. Розуміння потреб, вимог та попереднього досвіду клієнтів через аналіз даних є критично важливим для підвищення ефективності комунікації та задоволення клієнтів.

Таким чином, актуальність даної теми полягає у необхідності створення профілів користувачів із допомогою комбінацій методів кластеризації даних для покращення персоналізації послуг, ефективної роботи електронної комерції та вдосконалення обслуговування клієнтів, що відповідає динаміці та вимогам сучасного ринку.

2. Ступінь обґрунтованості наукових положень, висновків і рекомендацій

Обґрунтованість наукових положень та висновків, сформульованих у дисертаційній роботі, є достатньою і базується на детальному аналізі джерел за

даною проблемою, чіткій постановці задач дослідження, використанні сучасних методів дослідження, правильним застосуванням математичного апарату при теоретичному розгляді наукових положень дисертації, а також проявляється у якісному та аргументованому формулюванні висновків.

Наукові тези та висновки, представлені в даній дисертації, є комплексними та витікають із її вмісту, відображаючи результати, отримані дослідником. Наукова достовірність та обґрунтованість результатів обумовлена гармонією між теоретичними висновками та експериментальними даними. Рекомендації щодо застосування результатів дослідження відповідають достатньому рівню обґрунтування, який підкріплений теоретичними та практичними аспектами роботи, проведеної на високому науковому та методологічному рівні, і повною мірою відображають теоретичну та практичну значущість дисертації.

Достовірність та обґрунтованість запропонованих методів і засобів підтверджується результатами експериментальних досліджень та коректним застосуванням методів, які були використані під час виконання роботи.

3. Структура дисертаційної роботи

Дисертаційна робота викладена на 149 сторінках та складається із змісту, вступу, чотирьох основних розділів, списку використаних джерел та додатків.

Дисертація написана логічно, доступно, на високому технічному рівні з використанням сучасної термінології. Матеріали дисертаційної роботи викладено послідовно, доступно для розуміння і сприйняття. Стиль мовлення задовольняє вимоги до текстів науково-технічного змісту. Текст дисертації в достатній мірі проілюстрований таблицями та рисунками, а також фрагментами програмного коду. Здобувач використовує загальноприйнятту термінологію. Дисертаційна робота має логічну структуру.

Дисертаційна робота оформлена відповідно до вимог наказу МОН України від 12 січня 2017 р. № 40 «Про затвердження вимог до оформлення дисертації».

У вступі описано актуальність теми проведеного дослідження, зв'язок роботи з науковими програмами, планами і темами, визначено мету, задачі дисертаційної роботи, описано об'єкт, предмет та методи дослідження, сформовано наукову новизну та практичну цінність роботи.

У першому розділі дисертації зосереджено увагу на аналізі алгоритмів обробки різнотипових даних, таких як K-середніх, DBSCAN, ієрархічна та нечітка кластеризація. В ході аналізу виявлено ряд обмежень цих методів, що включають проблеми масштабування, інтерпретації, універсальності та адаптивності до змінних умов і структур даних. Також зазначається, що не всі сучасні системи ефективно використовують інформацію з напівструктурованих джерел даних. Це відкриває перспективи для подальших досліджень і розробки більш ефективних та гнучких методів кластеризації для різнотипових даних. Такі дослідження можуть покращити здатність систем створювати точні та інформативні профілі користувачів, а також відкрити нові можливості для їх використання у сферах бізнес-аналітики, прогнозування та інших цілях.

У другому розділі дисертації розглядаються методи кластеризації даних і здійснюється формальне визначення профілів користувачів. Аналіз охоплює різні критерії, які включають демографічні, поведінкові, психографічні, мотиваційні, та знаннево-інформаційні аспекти користувачів. Ці критерії допомагають у формуванні підмножин даних з наявного датасету, які містять різноманітні типи даних, включно із числовими, текстовими, ординальними та категоріальними. Розділ також включає опис процесу класифікації профілів користувачів, заснований на їх атрибутах. Окрема увага приділяється визначенню рівня задоволеності користувача з використанням таких метрик як CSS (Customer Satisfaction Score), NPS (Net Promoter Score), CES (Customer Effort Score) та CSI (Customer Satisfaction Index).

У третьому розділі дисертації розроблено новий метод кластеризації різнотипових даних, який адаптований для роботи з потоковими даними. Цей метод базується на поділі даних на пакети, що дозволяє ефективно обробляти

великі обсяги інформації у реальному часі. Автор також використовує статистичний метод перцентилів для розрахунку початкових центроїдів, що є основою для кластеризації даних. Окрему увагу приділено розробці алгоритму підготовки даних, методам зменшення розмірності даних і виділення нових ознак, які є ключовими для ефективного обробки великих наборів даних та виявлення значущих патернів.

У четвертому розділі дисертації описується розробка архітектури інформаційної системи, призначеної для автоматизованого профілювання користувачів на основі різнотипових даних клієнтів. Основна мета цієї архітектури полягає у створенні ефективною та економічно вигідною системи, яка здатна ефективно обробляти великі обсяги даних для формування профілів користувачів. Завдяки впровадженню цієї архітектури, автор відзначає значне зниження кінцевої вартості системи – в чотири рази, порівняно з вартістю стандартної архітектури, яка базується на використанні дроплетів.

4. Наукова новизна результатів досліджень

Аналіз змісту дисертаційної роботи і відповідних наукових публікацій дозволив визначити ключові наукові тези, висновки, а також зроблені здобувачем пропозиції та рекомендації. Наукові аспекти, висновки та рекомендації, представлені в дисертації, є добре обґрунтованими. Ця новизна підкріплюється точним формулюванням наукових завдань, адекватністю обраних методів для їх вирішення, використанням як загальних, так і спеціалізованих наукових методик дослідження, а також залученням досягнень вітчизняної та міжнародної літератури у галузі комп'ютерних наук і використанням великого обсягу даних.

Новизна найбільш значущих наукових результатів цього дослідження, що відображають персональний внесок автора дисертації, включає наступне:

- автор розвинув метод кластеризації різнотипових даних, який дозволяє працювати зі структурованими та напівструктурованими даними на основі поділу на пакети та врахування зважування характеристик під час препроцесингу даних;

- автор вперше побудував модель профілю клієнта, яка відрізняється від існуючих тим, що включає поведінкові та психографічні характеристики, що дає можливість визначити рівень задоволеності клієнта;
- автор вперше розробив класифікатор профілів клієнтів, який, на відміну від існуючих рішень, застосовує зважування відгуків на першому етапі та кластеризацію на другому, що дає можливість пришвидшити обслуговування клієнтів.

5. Значення результатів дослідження для науки і практики.

Дослідження робить важливий внесок у галузь комп'ютерних наук, зокрема у розвиток методів обробки різнотипових даних та кластеризації. Впровадження розробленої архітектури інформаційної системи, заснованої на безсерверному підході, демонструє значне зниження витрат на ІТ-інфраструктуру, що має важливе економічне значення для компаній, особливо тих, які мають обмежені ресурси. Актуальність дисертаційної роботи також обґрунтована тим, що окремі її розділи виконувались в рамках науково-дослідної теми «Методи та засоби інформаційної безпеки та гігієни на основі інтерпретованого штучного інтелекту» (№ державного реєстру 0123U101687).

6. Повнота відображення наукових положень, висновків і рекомендацій в опублікованих автором дисертаційних працях

Результати дослідження опубліковані достатньо повно. Основні положення дисертації опубліковані автором у 7 наукових працях, із них 1 стаття – у виданні, що індексуються в наукометричній базі даних Scopus (журнал квартилю Q2), 3 статті – у фахових виданнях України, 3 публікації – у збірниках наукових праць конференцій. Основні положення та результати дослідження, представлені у дисертації, оприлюднені та обговорені на міжнародних науково-практичних конференціях, а також на науковому семінарі кафедри Системи штучного інтелекту Національного університету “Львівська політехніка”.

Таким чином, наукові результати, описані в дисертаційній роботі, повністю висвітлені у наукових публікаціях здобувача.

7. Мова та стиль дисертаційної роботи

Дисертація написана згідно із вимогами до науково-технічної термінології. Виклад матеріалу відзначається лаконічністю і відсутністю зайвих деталей. Тема, зміст та отримані наукові результати роботи відповідають спеціальності 122 «Комп'ютерні науки», галузі знань 12 «Інформаційні технології».

8. Дискусійні положення та зауваження до дисертаційної роботи

1. На мою думку, здобувач некоректно формулює мету дослідження (розроблення методів і засобів як самоціль), адже метою даного дослідження насправді є підвищення швидкодії кластеризації даних, пришвидшення обслуговування клієнтів та збільшення персоналізування даних клієнта шляхом розроблення методів та засобів кластеризації різнотипових даних;

2. При формулюванні практичного значення одержаних результатів здобувач констатує факт стабільно швидких результатів потокової кластеризації та втрати якості кластеризації через розбиття даних із застосуванням запропонованого ним підходу, проте незрозумілі кількісні виміри ані збільшення швидкодії кластеризації, ані втрати якості кластеризації, внаслідок чого незрозуміло, чи втрати якості є припустимими за наявного збільшення швидкодії;

3. У першому розділі, на мою думку, варто було б поглибити аналіз досліджень, присвячених кластеризації у сфері нерухомості;

4. У другому розділі описано лише один підхід для визначення зв'язків між критеріями профілю користувача, проте незрозуміло, чому саме цей підхід було використано здобувачем;

5. Класифікація профілів користувачів, на мою думку, описана досить поверхнево;

6. У підрозділі 3.1.4. недостатньо обґрунтованим є вибір методу виявлення та видалення викидів;

7. У дисертаційній роботі не наведено статистику та дані про зміну/оновлення профілів користувачів під час апробації результатів.

Вважаю, що висловлені зауваження не є визначальними і не применшують загальну наукову новизну та практичну значущість результатів та не впливають на позитивну оцінку дисертаційної роботи.

9. Загальний висновок.

Вважаю, що дисертаційна робота здобувача ступеня доктора філософії Ткачика Олександра Андрійовича на тему « Методи та засоби кластеризації різнотипових даних » виконана на високому науковому рівні, не порушує принципів академічної доброчесності та є закінченим науковим дослідженням, сукупність теоретичних та практичних результатів якого розв'язує наукове завдання, що має істотне значення для галузі інформаційних технологій. Дисертаційна робота за актуальністю, практичною цінністю та науковою новизною повністю відповідає вимогам чинного законодавства України, що передбачені в п. 6 – 9 «Порядку присудження ступеня доктора філософії та скасування рішення разової спеціалізованої вченої ради закладу вищої освіти, наукової установи про присудження ступеня доктора філософії», затвердженого Постановою Кабінету Міністрів України від 12 січня 2022 р. № 44.

Здобувач Ткачик Олександр Андрійович заслуговує на присудження ступеня доктора філософії в галузі знань Інформаційні технології за спеціальністю 122 Комп'ютерні науки.

Офіційний опонент
докторка технічних наук, професорка,
завідувачка кафедри
комп'ютерної інженерії
та інформаційних систем
Хмельницького національного університету

Тетяна ГОВОРУЩЕНКО

