

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

Відгук офіційного опонента
доктора технічних наук, професора Анатолія САЧЕНКА 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

Вибраний напрямок дисертаційного дослідження обумовлений розвитком штучного інтелекту в контексті ринку нерухомості. З огляду на стрімкий прогрес технологій та збільшення обсягів доступних даних, все більше актуальним стає розроблення нових методів персоналізації послуг у сфері нерухомості. При цьому застосування методів кластеризації даних для створення деталізованих профілів користувачів, які зацікавлені у покупці або оренді нерухомості, відкриває перспективи для підвищення задоволеності клієнтів та ефективності бізнес-процесів. Тому тема дисертаційного дослідження, присвячена розробленню методів та алгоритмів кластеризації різнотипових даних, є, безумовно, актуальною.

2. Ступінь обґрунтованості наукових положень, висновків і рекомендацій

Наукові положення, висновки та рекомендації, представлені в дисертації, мають теоретичне підґрунтя та апробовані на міжнародних і національних конференціях. Результати роботи, отримані автором, мають достатнє обґрунтування, а їх достовірність підтверджена експериментально.

3. Структура дисертаційної роботи

Структура дисертаційної роботи Ткачика О. А. містить чотири основних розділи.

У першому розділі проведено огляд алгоритмів обробки різнотипових даних, включаючи K-середні, DBSCAN, ієрархічну кластеризацію та нечітку

кластеризацію. Крім того, розглянуті питання масштабування, інтерпретації, універсальності та адаптивності до змінюваних умов і структур даних.

У другому розділі проведено огляд методів кластеризації даних. Запропоновано формальне визначення профілів користувачів, включаючи демографічні, поведінкові, психографічні, мотиваційні та знаннєво-інформаційні критерії. Крім того, описано процес класифікації профілів на основі атрибутів користувачів та рівень задоволеності клієнтів за допомогою різних метрик.

У третьому розділі розроблено алгоритм підготовки даних. Розглянуто методи обробки пропущених значень, виявлення та видалення дублікатів і виконано аналіз методів зменшення розмірності даних та виділення нових ознак. Запропоновано метод кластеризації для роботи з потоковими даними на основі поділу на пакети.

У четвертому розділі представлено розробку архітектури інформаційної системи для автоматизованого профілювання користувачів на базі різнотипових даних клієнтів. Вказано, що застосування без серверної концепції на основі Google Cloud Functions дозволяє суттєво знизити вартість розгортання системи порівняно із стандартною архітектурою.

4. Наукова новизна результатів досліджень

До основних наукових результатів дисертаційної роботи слід віднести:

1. На основі удосконаленого методу кластеризації різнотипових даних розширено сферу застосування методів кластеризації даних, що дозволяє працювати ефективно зі структурованими та напівструктурованими даними.
2. Розроблена концепція профілювання клієнтів включає в себе поведінкові та психографічні характеристики, що дозволяє точніше визначати потреби та уподобання клієнтів, а також оцінювати їх рівень задоволеності.
3. Запропонований класифікатор профілів клієнтів, який відрізняється від існуючих зважуванням відгуків на першому етапі та кластеризацією на другому етапі, що дозволяє пришвидшити обслуговування клієнтів, і, в результаті, підвищити ефективність бізнес-процесів.

5. Значення результатів дослідження для практики.

Розробка нової архітектури інформаційної системи для автоматизованого профілювання користувачів на основі різнотипових даних клієнтів є важливим

кроком у покращенні процесів обробки даних, так як призводить до суттєвого зниження вартості розгортання системи порівняно із традиційними методами. Розроблений метод кластеризації, що використовує пакети даних та перцентильні аналізи для підвищення ефективності обробки, має практичне значення для широкого спектру застосувань у сучасному динамічному діловому середовищі. Дисертаційне дослідження проводились в рамках кафедральної держбюджетної теми «Методи та засоби інформаційної безпеки та гігієни на основі інтерпретованого штучного інтелекту».

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

Наукові результати та положення повністю відображені у семи опублікованих працях: одна стаття у міжнародному періодичному виданні квартилю Q2, що входить до наукометричних баз Web of Science та Scopus; три статті у фахових наукових виданнях України; три тези доповідей у збірниках міжнародних конференцій. Основні положення та результати дисертаційного дослідження апробовані на міжнародних і національних наукових конференціях.

7. Мова та стиль дисертаційної роботи

Текст дисертаційної роботи викладено з використанням сучасної професійної термінології. Матеріал дисертації проілюстрований рисунками, графіками й таблицями. Мова і стиль викладення та оформлення дисертації, в основному, відповідають вимогам, які ставляться до наукових праць.

8. Дискусійні положення та зауваження до дисертаційної роботи

У дисертаційній роботі виявлено цілий ряд недоліків та зауважень, основними з яких є наступні.

1. Зміст сформованих задач дисертаційної роботи не цілком відповідає змісту отриманих висновків.

2. Структура дисертаційної роботи потребує суттєвого доопрацювання, зокрема:

- Оглядовий Розділ 1 займає більше 35% загального обсягу рукопису і міг би бути скороченим
- Розділ 2, який би мав описувати теоретичний вклад автора, містить, здебільшого, оглядові матеріали. У назві розділу фігурують «Алгоритми опрацювання різнотипових даних», розробці яких присвячений Розділ 3

- Виходячи з тематики дисертації та логіки викладу матеріалу треба було б послідовно розглянути методи та засоби кластеризації різнотипових даних
- Назва Розділу 4 «Розроблення архітектури...» носить незакінчений характер

3. Два положення наукової новизни «Вперше побудовано модель профілю клієнта...» та «Вперше розроблено класифікатор профілів клієнтів...», сформульовані некоректно. По-перше вони в явному виді навіть не згадані у Змісті дисертації, по-друге- вони безпосередньо не витікають із тексту дисертації.

4. У дисертаційній роботі відсутні схеми алгоритмів, що дещо утруднює її сприйняття.

5. Розроблена архітектура інформаційної системи апробована у рамках специфічної задачі створення профілів користувачів. Водночас, у дисертації було б корисно розширити сферу застосування інформаційної системи з врахуванням різних сценаріїв використання.

6. Добре було показати, як результати проведеного порівняльного аналізу вартості між запропонованою безсерверною архітектурою інформаційної системи та її стандартною серверною архітектурою можуть бути використані для інших різновидностей інформаційних систем.

7. Не зовсім обґрунтованим є використання метрик оцінки рівня задоволеності користувача, оскільки вони значно звужують сферу застосування профілів користувачів.

8. Практичне значення дисертаційної роботи не підтверджене актами впровадження.

9. Висновки потребують конкретизації та уточнення. У загальних висновках говориться про зниження обчислювальних витрат в 4 рази, а у Висновках до розділу 4 – про зменшення кінцевої вартості системи.

10. У дисертації зустрічаються невдалі вирази («Стандартний метод k-means вважає...», «Модель намагається ідентифікувати кластери», тощо), граматичні та синтаксичні помилки по тексту.

Зазначені зауваження та недоліки впливають, в помітній мірі, на якість отриманих автором результатів. Однак, беручи до уваги актуальність і завершеність дисертаційної роботи, повноту, та апробацію опублікованих результатів, дисертацію, в цілому, можна оцінити позитивно.

9. Загальний висновок.

Дисертаційна робота Олександра Ткачика на тему «Методи та засоби кластеризації різнотипових даних» є завершеним науковим дослідженням, що

стосується вирішення важливого наукового завдання - розроблення інформаційної системи профілювання користувачів для пришвидшення обслуговування клієнтів у системі ринку нерухомості. Основні положення дисертації характеризуються актуальністю, науковою новизною та практичною цінністю, а сформульовані висновки є достатньо обґрунтованими. Тема та зміст роботи відповідають паспорту спеціальності 122 «Комп'ютерні науки» та нормам академічної доброчесності

Тому вважаю, що дана дисертаційна робота відповідає 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 йому наукового ступеня доктора філософії за спеціальністю 122 «Комп'ютерні науки».

Офіційний опонент – Заслужений винахідник України, доктор технічних наук, професор, професор кафедри інформаційно-обчислювальних систем і управління, директор НДІ інтелектуальних комп'ютерних систем, Західноукраїнський національний університет

Анатолій САЧЕНКО



Підпис *Анатолія САЧЕНКА*
Підпис *Анни СЕЦЕНКО*