ВІДГУК

офіційного опонента доктора технічних наук, професора
Цмоця Івана Григоровича
на дисертаційну роботу **Польгуль Тетяни Дмитрівни**
«Інформаційна технологія виявлення шахрайства при інсталюванні мобільних додатків з використанням інтелектуального аналізу даних»,
поданої на здобуття наукового ступеня кандидата технічних наук за спеціальністю 05.13.06 – інформаційні технології.

Актуальність теми дисертаційної роботи. Дисертаційна робота Польгуль Тетяни Дмитрівни, присвячена одній з важливих і актуальних задач – виявлення шахрайства при інсталюванні мобільних додатків. Існуючі системи-аналоги виконують рейтингування користувачів на основі не всіх, а вибіркових вхідних даних, а також, використовують існуючі бази з шахрайськими даними, тому відбувається упущення шахраїв, які мають інші властивості, шаблони, поведінку.

Тому розробка інформаційної технології виявлення шахрайства при інсталюванні мобільних додатків з використанням інтелектуального аналізу даних, яка б відстежувала та визначала шаблони шахраїв, які непомітні людині, є без сумніву актуальною та важливою у сучасному світі.

Обрана тема та мета дисертації, що спрямовані на підвищення точності та швидкодії процесів виявлення шахрайства при інсталюванні мобільних додатків, виявляють свою актуальність та перспективність.

У дисертації наведено теоретичні узагальнення і нові вирішення наукової задачі розробки інформаційної технології виявлення шахрайства при інсталюванні мобільних додатків, а також відповідних програмних засобів її реалізації.

Таким чином, тематика дисертаційної роботи є актуальною та важливою для сучасного рівня науки й техніки.

Зв'язок роботи з науковими програмами, темами, планами. Актуальність теми підтверджується тим, що дисертаційне дослідження проводилось згідно з планами науково-дослідних робіт кафедри комп'ютерних наук Вінницького національного технічного університету, в тому числі в межах: наукового проекту «Методологія побудови високопродуктивних інтелектуалізованих паралельно-ієрархічних систем на основі сучасних мережевих обчислювальних комплексів з гетерогенною архітектурою» (2015 р.), при виконанні якого автор брала участь як виконавець окремих підрозділів; кафедральної теми 47 К2 «Моделі, методи, технології та пристрої інтелектуальних інформаційних систем управління, економіки, навчання та комунікацій» (2016-2018 р.), при виконанні якої автор брала участь як відповідальний виконавець; кафедральної теми 22 К1 «Розробка спеціалізованих засобів штучного інтелекту на основі інтелектуального аналізу даних та

машинного навчання» (2019 р.), при виконанні якої автор брала участь як виконавець окремих підрозділів.

Ступінь обґрунтованості наукових положень, висновків і рекомендацій, сформульованих у дисертаційній роботі. Наукові положення, результати і висновки дисертації, отримані автором, в цілому є достатньо доказовими та обґрунтованими.

В роботі використані сучасні методи шкалювання, теорія множин, а також, методи класифікації, статистичні методи, методи машинного навчання, інтелектуальний аналіз даних, методи кластеризації, нейромережеві методи для розробки інформаційної технології виявлення шахрайства при інсталюванні мобільних додатків, що також обґрунтовано досвідом їх застосування для створення інформаційної технології виявлення шахрайства при інсталюванні мобільних додатків. Основні наукові положення дисертаційної роботи представлено моделями, методами та алгоритмами. Наведені в роботі теоретичні положення та твердження викладено у логічній послідовності та в достатній мірі аргументовано. Адекватність запропонованих методів, моделей, алгоритмів підтверджується результатами експериментальних досліджень.

Достовірність отриманих результатів забезпечено коректністю постановки завдання та формалізації процесу виявлення шахрайства як аномалії в даних (розділ 2), здійсненому аналізу та класифікації різнорідних даних при інсталюванні мобільних додатків (розділ 2) та аналізу та підготовки вхідних даних для проведення експериментальних досліджень з використанням інформаційної технології виявлення шахрайства (розділ 4), розробці методики проведення експериментальних досліджень шахрайства при інсталюванні мобільних додатків (розділ 4), розробці відповідних методів, моделей та алгоритмів (розділ 2, 3, 4), що стали основою розроблної інформаційної технології виявлення шахрайства при інсталюванні мобільних додатків з використанням інтелектуального аналізу даних; а також, здійсненого дослідження адекватності моделей, точності та швидкодії методу виявлення шахрайства та аналізу результатів тестування (розділ 4).

Достовірність наукових положень, висновків і рекомендацій, що сформульовані в дисертаційній роботі, підтверджується апробацією та впровадженням результатів досліджень на реальних об’єктах професійної сфери діяльності, про що свідчать відповідні документи.

Наукова новизна отриманих результатів. До основних наукових результатів, одержаних здобувачем особисто, належить:

1. Вперше розроблений метод усунення різнорідності вхідних даних, що являє собою сукупність процедур вибору ознак, зниження розмірності та нормалізації даних, відмінність якої полягає у новій моделі усунення різнорідності даних шляхом шкалювання за інформативністю, що дозволяє всю множину різнорідних даних про користувачів звести до вектору уніфікованих ознак без зменшення діагностичної цінності інформації.
2. Вперше розроблений узагальнений метод виявлення шахрайства при інсталюванні мобільних додатків, відмінність якого полягає у використанні

запропонованої моделі класифікації користувачів та методу усунення різнорідності початкових даних, що дозволяє визначити профілі шахраїв та підвищити достовірність виявлення шахрайства під час інсталюванні мобільних додатків.

3. Удосконалена модель класифікації користувачів на основі глибинних нейронних мереж у частині зниження розмірності та нормалізації даних згідно запропонованого методу усунення різнорідності даних, яка може бути основою для портретування шахраїв з метою спрощення процесів їх виявлення.

Наукові положення, сформульовані в дисертації, досить повно обґрунтовані. Кожен пункт наукової новизни достатньою мірою підтверджений теоретичними, а також експериментальними дослідженнями.

Практичні результати роботи, їх рівень та ступінь впровадження. Практичне значення отриманих результатів роботи полягає у наступному: здійснено класифікацію різнорідних даних, що дозволило спростити процес аналізу різнорідних за метриками, розмірностями і шаблонами даних та автоматизувати його; розроблено алгоритм пошуку аномалій в даних, алгоритми процесу подолання різнорідності вхідних даних, алгоритм виявлення шахрая при інсталюванні мобільних додатків, алгоритм створення узагальненого портрету шахрая та алгоритм мінімізації часу виявлення шахраїв на основі розпаралелення обчислювальних процесів, які покладені в основу інформаційної технології, що підвищило точність та швидкодію виявлення шахраїв; запропоновано інформаційну технологію виявлення шахрайства при інсталюванні мобільних додатків, яка використовує запропоновані метод виявлення шахрайства при інсталюванні мобільних додатків, метод подолання різнорідності вхідних даних, модель класифікації даних, і, на відміну від існуючих систем Antifraud, дозволила підвищити точність класифікації користувачів до 99,14 %, зокрема точність класифікації шахраїв – до 82,76 %; розроблено програмне забезпечення “Mobile App Install Fraud Detection System” для виявлення шахрайства при інсталюванні мобільних додатків.

Результати дисертаційної роботи впроваджені на іноземному (Garuda AI B. V.) та українських (ТОВ «ВІН ІНТЕРАКТИВ», ТОВ «4ХайТек», ПП «Літсофт») підприємствах, а також навчальних закладах: у навчальний процес кафедри комп’ютерних наук Вінницького національного технічного університету (ВНТУ) та у навчальний процес кафедри інформатики, програмної інженерії та економічної кібернетики Херсонського державного університету.

Повнота викладення результатів досліджень в опублікованих працях. Основні положення дисертації опубліковані у 20 працях, серед яких 5 статей надруковано у наукових виданнях, які входять до переліку фахових видань з технічних наук, затверджених МОН України (одна з яких також входить до наукометричної бази даних Scopus). Крім того, 6 статей опубліковано в міжнародних наукових виданнях, п’ять з яких входять до міжнародної наукометричної бази Scopus (три з яких також входять до міжнародної наукометричної бази IEEE Xplore), 7 робіт опубліковано у збірках матеріалів

конференцій (три з яких міжнародні), отримано 2 свідоцтва про реєстрацію авторського права на твір.

Рівень і кількість публікацій та апробації матеріалів дисертації розкривають основний зміст дисертації та повністю відповідають вимогам МОН України.

Оцінка основного змісту дисертації та її структури. Дисертаційна робота складається зі вступу, чотирьох розділів, висновків, списку використаних джерел і додатків. Основний зміст викладено на 163 сторінках друкованого тексту, містить 63 рисунки, 17 таблиць. Список використаних джерел містить 108 найменувань. Загальний обсяг 245 сторінок.

Оформлення дисертаційної роботи відповідає встановленим вимогам.

Вступ. У вступі до дисертації обґрунтовано актуальність теми, сформульовано мету і завдання досліджень, викладено наукову новизну та практичне значення результатів роботи, особистий внесок здобувача, наведені дані щодо апробації результатів досліджень.

У першому розділі на основі огляду та аналізу сучасної літератури з галузі виявлення шахрайства в інформаційних технологіях, в роботі запропоновано розглядати задачу виявлення шахрайства при інсталюванні мобільних додатків як задачу пошуку аномалій в даних, тому що шахрайство визначено як навмисне породження аномалії в даних сторонньою особою (шахраєм) або механізмом з певною метою.

Розглянуто та проаналізовано методи і здійснено постановку задачі виявлення шахрайства при інсталюванні мобільних додатків. Проведено варіантний аналіз існуючих методів пошуку аномалій в даних, у результаті якого усі розглянуті методи розділено на такі основні групи як: методи класифікації, методи кластеризації, статистичні методи, мікс методів на прикладі глибинного навчання. Здійснено аналіз моделей подібності. Показано, що жоден із зазначених методів не може одночасно здійснювати інтелектуальну обробку повної вхідної інформації та вказувати причину, яка дає змогу позначити дані як аномальні.

Виділено такі відомі шахрайські способи під час інсталювання мобільних додатків як кліковий спам (Click Spamming), мобільне викрадення (Mobile Hijacking), ферми дій (Action Farms). Проаналізовано існуючі системи виявлення шахрайства при інсталюванні мобільних додатків, серед яких: система Fraudlogix, Kraken, Adjust, Kochava, TMC Attribution Analytics, FraudShield, Forensiq, Appsflyer, FraudScore, AppMetrica. Проте зазначено, що більшість з них видають результат у вигляді таблиці рейтингування, а не чіткої відповіді; опираються на бази з відомими шахрайськими даними (наприклад, з IP-адресами) та не в повній мірі використовують усі важливі вхідні дані. У зв’язку з цим, не розпізнаються шахраї, які мають інші властивості, шаблони, поведінку. Також, існує проблема неможливості визначення причини класифікації користувача як шахрайського існуючими системами. Тому й виникла необхідність створення інформаційної технології виявлення шахрайства при інсталюванні мобільних додатків, яка могла б відстежувати та визначати навіть нові шаблони шахраїв і мала змогу самонавчатися.

На основі проведеного аналізу визначено задачі дослідження.

У другому розділі формалізовано процес виявлення шахрайства як аномалії в даних з використанням теорії множин, що дозволило визначити властивості аномальних і неаномальних даних. Це дало змогу визначити властивості даних, які позначають шахраїв, у визначеній предметній області та спростити процес пошуку шахрайства при інсталюванні мобільних додатків.

Завдяки здійсненій у роботі формалізації процесу виявлення шахрайства як аномалії в даних, здійснено процес вибору характеристик даних, які дозволяють визначити клас користувача – шахрай чи органічний (не шахрай). Для визначення таких характеристик у роботі здійснено аналіз та класифікацію даних при інсталюванні мобільних додатків, які показали, що дані в таких системах – різнорідні. Здійснена класифікація різнорідних даних при інсталюванні мобільних додатків на основі процесу вибору ознак дозволила спростити процес аналізу різнорідних за метриками, розмірностями і шаблонами даних та автоматизувати його.

Вперше розроблено узагальнений метод виявлення шахрайства при інсталюванні мобільних додатків, відмінність якого полягає у використанні запропонованої моделі класифікації користувачів та методу подолання різнорідності вхідних даних, що дозволяє визначити класи користувачів та підвищити точність виявлення шахрайства при інсталюванні мобільних додатків.

Вперше запропоновано метод подолання різнорідності вхідних даних, що являє собою сукупність процедур вибору ознак, зниження розмірності та нормалізації даних, відмінність якого полягає у новій моделі процесу подолання різнорідності даних шляхом шкалювання за інформативністю, що дозволяє всю множину різнорідних даних про користувачів звести до вектору уніфікованих ознак без зменшення діагностичної цінності інформації.

Удосконалено модель класифікації користувачів на основі глибинних нейронних мереж у частині зниження розмірності та нормалізації даних згідно запропонованого методу подолання різнорідності даних, яка є основою для створення узагальненого портрету шахрая з метою спрощення процесів їх виявлення.

У третьому розділі розроблено моделі інформаційної технології з використанням технологій системного аналізу та моделювання, а саме: побудована логічна, концептуальна моделі інформаційної технології. Здійснено аналіз та розробку структур даних інформаційної технології, розроблено шаблон шахрая для формування портрету шахрая. Як показали дослідження, усі розроблені моделі є необхідними для ефективного функціонування інформаційної технології виявлення шахрайства на основі запропонованого в роботі узагальненого методу, в основі якого лежать моделі, методи та алгоритми виявлення аномалій в даних.

Розроблено алгоритми процесу подолання різнорідності вхідних даних, алгоритм створення узагальненого портрету шахрая та алгоритм пошуку аномалій в даних, які покладені в основу запропонованої інформаційної технології, що підвищило точність та швидкодію процесу виявлення шахраїв.

Розроблено алгоритм пошуку аномалій в даних як основа функціонування інформаційної технології за допомогою діаграми activity, а також здійснено

аналіз процесів в інформаційній технології виявлення шахрайства за допомогою UML-діаграми послідовності.

Запропоновано інформаційну технологію виявлення шахрайства при інсталюванні мобільних додатків, яка використовує запропоновані метод виявлення шахрайства при інсталюванні мобільних додатків, метод подолання різнорідності вхідних даних, модель класифікації даних, і, на відміну від існуючих систем Antifraud, дозволила підвищити точність класифікації користувачів до 99,14 %, зокрема точність класифікації шахраїв – до 82,76 %.

У четвертому розділі розроблено методику проведення експериментальних досліджень розробленої інформаційної технології.

Розроблено алгоритм мінімізації часу виявлення шахраїв на основі розпаралелення обчислювальних процесів.

Доведено адекватність моделей процесу подолання різнорідності вхідних даних та моделі класифікації вхідних даних на вибірці з розробленого мобільного додатку «MobNsters: Mafia War Strategy» (Orneon Ltd.). Зокрема показано, що точність виявлення класу користувача з використанням розробленої технології було підвищено на 1.26%, а точність виявлення шахрая підвищено на 1.36% у порівнянні з системою-аналогом, що дала найкращий результат.

На основі запропонованої в роботі інформаційної технології розроблено модульне програмне забезпечення “Mobile App Install Fraud Detection System” з використанням алгоритмів, які реалізують розроблені методи, моделі та веб-систему, що впроваджена на підприємстві Garuda AI B.V. (Нідерланди).

У **висновках** наведено основні результати дисертаційної роботи та надано рекомендації щодо практичного застосування теоретичних напрацювань. Загальні висновки по роботі відрізняються чіткістю, лаконічністю, узагальнюють викладені в роботі результати досліджень.

Список використаних джерел є достатнім, охоплює сучасні вітчизняні та зарубіжні публікації, містить 108 найменувань.

У **додатках** представлено документи щодо впровадження результатів роботи; основні лістинги програмного забезпечення; список експертів, які проводили оцінювання; сертифікати на публікації та виступи, а також, список публікацій за темою дисертації.

Відповідність дисертації та автореферату встановленим вимогам.

За своєю структурою, обсягом і оформленням дисертація та автореферат цілком відповідають вимогам, встановленим до кандидатських дисертацій, зокрема пп. 9.11.12 «Порядку присудження наукових ступенів».

Автореферат за змістом ідентичний основним положенням, що викладені в дисертації, та не містить інформації, яка не відображена в самій роботі. Стиль викладу матеріалів досліджень, наукових положень і рекомендацій забезпечує їх адекватне і належне сприйняття.

Наукова новизна відповідає паспорту спеціальності 05.13.06 – інформаційні технології, зокрема підрозділ 2.3 «Розробка узагальненого методу виявлення шахрайства при інсталюванні мобільних додатків», що зокрема включає в себе п. 2.3.1 «Розробка методу подолання різнорідності вхідних даних», пп. 2.3.1.2 «Розробка моделей процесу подолання різнорідності вхідних

даних», п. 2.3.3 «Модель класифікації вхідних даних для виявлення аномалій в Big Data»; розділ 3 «Розробка інформаційної технології виявлення шахрайства при інсталюванні мобільних додатків», що включає в себе п. 3.1.1 «Розробка концептуальної моделі інформаційної технології виявлення шахрайства», п. 3.1.4 «Розробка нечіткої моделі для формування портрету шахрая»; а також, підрозділ 4.4 «Дослідження адекватності моделей, точності та швидкодії методу виявлення шахрайства, аналіз результатів тестування».

Недоліки та зауваження до дисертаційної роботи.

До недоліків роботи, на мою думку, слід віднести:

1. У першому розділі не розглядаються основні положення теорії множин, хоча вони використовуються у другому розділі, а також недостатньо проведено аналіз та класифікацію математичних моделей виявлення аномалій.
2. У другому розділі доцільно було б детальніше обґрунтувати обрання 17 коефіцієнтів, які забезпечують переведення різнорідних даних до однорідних, а також навести пояснення кожної неаномальної множини (рисунок 2.8) та яким чином їх сукупність формує аномальні множини.
3. У другому розділі (п. 2.3.1.1) не достатньо обґрунтовано вибір експертів, які беруть участь в опитуванні.
4. У третьому розділі (п. 3.6) проектування інформаційної технології виявлення шахрайства є занадто детальним і його доцільно було б скоротити.
5. У п. 3.1.4 було б бажано навести графіки функції належності для кожного визначеного параметру.
6. У розділі 4 не наведена загальна структура модулів розробленого програмного забезпечення, що утруднює сприйняття розробленої інформаційної технології виявлення шахрайства при інсталюванні мобільних додатків з використанням інтелектуального аналізу даних.
7. При викладенні матеріалу зустрічаються понятійні, стилістичні та термінологічні неточності, а також граматичні та орфографічні помилки, на які вказано автору.

Відмічені зауваження не вплинули на загальну позитивну оцінку дисертаційної роботи та можуть розглядатися як рекомендації до подальших наукових досліджень та впроваджень отриманих результатів.

ВИСНОВКИ.

1. Дисертаційна робота **Польгуль Тетяни Дмитрівни** на тему: «Інформаційна технологія виявлення шахрайства при інсталюванні мобільних додатків з використанням інтелектуального аналізу даних» є завершеною науковою працею, яка розв’язує актуальну наукову задачу підвищення точності та швидкодії процесів виявлення шахрайства при інсталюванні мобільних додатків.
2. Автореферат повністю відповідає змісту дисертації і описує суть одержаних результатів та висновків у дисертаційній роботі і оформлений згідно з вимогами.
3. Дисертаційна робота відповідає спеціальності 05.13.06 – інформаційні технології вимогам ДАК України, зокрема пп. 9.11.12 «Порядку присудження

наукових ступенів», затвердженого постановою кабінету Міністрів від 24 липня 2013 року № 565 (зі змінами затвердженими постановою кабінету Міністрів України від 19 серпня 2015 року № 656), які висуваються до робіт на здобуття наукового ступеня кандидата технічних наук, так як вони містять нові науково обґрунтовані результати проведених досліджень.

4. Автор дисертації **Польгуль Тетяна Дмитрівна** заслуговує на присудження їй наукового ступеня кандидата технічних наук за спеціальністю 05.13.06 – інформаційні технології.

Офіційний опонент,
завідувач кафедри
автоматизованих систем управління,
Національного університету
«Львівська політехніка»
д.т.н., професор **/ Цмоць І.Г. /**

Підпис засвідчую

Вчений секретар **/Р.Б. Брилинський/**