РЕЦЕНЗІЯ

доктора технічних наук, професора Поворознюка Анатолія Івановича на дисертаційну роботу **Горносталя Олексія Андрійовича "Ансамблевий метод ідентифікації стану комп'ютерних систем"**, подану на здобуття наукового ступеня доктора філософії з галузі знань 12 – інформаційні технології за спеціальністю 123 – комп'ютерна інженерія

Ступінь актуальності теми дисертаційної роботи. У світі, де технологічні інновації стають все більш важливими та проникають у різні аспекти нашого життя, комп'ютерні системи відіграють ключову роль. Зі зростанням їх складності постійно з'являються різноманітні чинники впливу у їхню роботу, які призводять до вторгнень з дуже значними наслідками: від втрати грошей і важливих даних до прямої шкоди людським життям. Складність та динамічність сучасних інформаційних середовищ не завжди піддається традиційним підходам аналізу, саме тому актуальними є задача вдосконалення та розробки методів ідентифікації стану комп'ютерних систем.

У цьому контексті особливо перспективним є використання ансамблевих методів машинного навчання. Однією з основних їх переваг є здатність поєднувати прогнози кількох базових моделей, що дозволяє зменшити вплив випадкових помилок та підвищити точність передбачень. Завдяки цьому можна стверджувати, що тема дисертаційної роботи Горносталя Олексія Андрійовича, "Ансамблевий метод ідентифікації стану комп'ютерних систем" є актуальною та дозволяє вирішити важливе науково-технічне завдання сучасними методами.

Зв'язок теми дисертаційної роботи з науковими планами, програмами, фундаментальними та прикладними дослідженнями. Дисертаційна робота виконана за важливим як у науково-практичному, так і у суспільному значенні напрямку, який знаходиться у межах планів науково-дослідних робіт кафедри «Комп'ютерна інженерія та програмування» в рамках науково-дослідної теми «Моделі і методи обробки та захисту інформації в

комп'ютерних системах» за замовленням ТОВ «Передові цифрові рішення» (ДР №0122U200526), в якій здобувач був відповідальним виконавцем.

Ступінь обґрунтованості та достовірності наукових положень, висновків та рекомендацій, сформульованих у дисертаційній роботі. У даному дисертаційному дослідженні акцентується увага на використанні ансамблевих беггінг-класифікаторів з метою підвищення ефективності ідентифікації стану комп'ютерних систем. Вибір цього методу для вдосконалення обумовлюється концепцією комбінування кількох базових моделей з метою отримання більш стійких та надійних прогнозів.

Ступінь обґрунтованості та достовірності наукових положень та висновків, сформульованих у дисертаційній роботі, забезпечується шляхом попереднього теоретичного аналізу та ретельного вивчення результатів експериментів, що дозволяє виявити їх переваги та оцінити їхню застосовність у конкретному контексті завдання. Крім того, у роботі приділяється особлива увага методологічним аспектам дослідження, включаючи правильне формулювання гіпотез, вибір відповідних метрик оцінки якості моделей та часових характеристик, а також коректній інтерпретації отриманих результатів.

Сукупність цих факторів дозволяє стверджувати, що дисертаційне дослідження Горносталя Олексія Андрійовича містить обґрунтовані та достовірні положення, висновки та рекомендації.

Наукова новизна положень, висновків та рекомендацій, сформульованих у дисертації. В рамках дисертаційної роботи вдосконалено існуючі та створено нові методи ідентифікації стану комп’ютерних систем з метою вирішення поставленого науково-практичного завдання шляхом використанням ансамблевих методів машинного навчання та за рахунок виконання комплексних теоретичних та експериментальних досліджень. Слід відзначити наступні елементи наукової новизни:

1. Отримав подальший розвиток метод ідентифікації стану комп’ютерної системи на основі дерев рішень та мета-алгоритму беггінг за рахунок вибору оптимальних гіперпараметрів налаштування класифікатора та

використання процедури попередньої обробки даних, яка сфокусована на видаленні аномальних даних та зменшенні статистичної залежності між ознаками.

2. Удосконалено ансамблевий метод ідентифікації стану комп'ютерної системи на основі гомогенного мета-алгоритму беггінг шляхом використання багатошарового перцептрону у якості базової моделі ансамблю та вибору оптимальних гіперпараметрів налаштування класифікатору, а також розробки спеціальної процедури зменшення кількості базових класифікаторів та їх ранжування під час зваженого голосування, що дозволило зменшити час роботи ансамблю та підвищити якість класифікації стану КС.

3. Вперше запропоновано метод ідентифікації стану комп’ютерних систем, що включає в себе розроблений триетапний процес побудови гетерогенного беггінг-ансамблю.

Наукова та практична цінність одержаних результатів. В рамках дисертаційної роботи чітко сформульовані її основні задачі, які були виконані у повному обсязі шляхом проведення теоретичних та експериментальних досліджень. Здобувач приділив особливу увагу різним підходам до оцінки якості роботи класифікаційних методів. При цьому в рамках дослідження використані сучасні підходи які дозволили перевірити сформовані припущення експериментальним шляхом та сформувати відповідні практичні рекомендації. Завдяки цьому автор виконав поставлену мету, вдосконаливши наявні методи та запропонувавши власний метод ідентифікації стану комп’ютерних систем.

Серед основних практичних результатів роботи слід виділити наступні:

– сформована програмна модель для виконання попередньої обробки даних дозволила збільшити швидкість розпізнавання до 1,62 разів за рахунок оптимізації атрибутів та видалення аномалій;

– розроблена програмна модель ідентифікації стану комп’ютерної системи з використанням запропонованої процедури попередньої обробки даних, формування вхідних послідовностей та вибору оптимальних налаштувань окремих моделей і всього ансамблю дозволив збільшити значення

AUC-ROC класифікатору на тестовій вибірці – на 3%;

– реалізована програмна модель гомогенного беггінг-ансамблю з багатошаровим перцептроном та процедурою вибору оптимальних налаштувань дозволила підвищити значення точності класифікації на 4,67%;

– розроблене програмне забезпечення для комплексного виконання ансамблевої обрізки (з використанням абсолютного показника точності) та зваженого голосування (з використанням функції логарифмічних втрат) дозволило збільшити значення F_1- *Score* – на 2,4%;

– запропонований метод формування гетерогенного ансамблю дозволив отримати різнорідний класифікатор з різними методами машинного навчання у якості базових моделей дозволив збільшити значення метрики F_1-*Score* на 9,5%, якщо порівнювати з гомогенним беггінг-ансамблем на основі дерев рішень зі стандартними налаштуваннями та на 2%, якщо порівнювати з однорідними ансамблями на основі інших методів машинного навчання з найкращим значенням цього показника.

Слід також відзначити, що результати дисертаційного дослідження впровадженні в навчальний процес при викладанні дисциплін за спеціальністю 123 «Комп'ютерна інженерія» в НТУ «ХПІ», а також використовуються в системах моніторингу стану комп'ютерних систем та захисту інформації у ТОВ «Передові цифрові рішення».

Повнота викладення наукових і прикладних результатів дисертації в опублікованих працях. За темою дисертаційного дослідження здобувачем опубліковано 5 наукових публікації у фахових виданнях, 4 з яких виконано у співавторстві з науковим керівником. Водночас з метою їх апробації здобував брав участь у 15 міжнародних симпозіумах та конференціях з публікацією тез доповідей, з них 6 матеріалів входять до наукометричної бази Scopus. Сукупність цих факторів свідчить про повноту представлення та викладення наукових і прикладних результатів у наявних опублікованих працях.

Оцінка змісту дисертації, її завершеності й оформлення. Тема, структура та зміст дисертаційної роботі відповідають усім наявним вимогам та

прийнятим нормам. Вона складається з анотації двома мовами, вступу, 4 розділів, списку використаних джерел з 137 найменуваннями, висновків та 3 додатків. Загальний обсяг роботи складає 170 сторінок. При цьому дисертаційне дослідження відповідає вимогам, що висувається при формуванні роботи на здобуття наукового ступеня доктора філософії.

Вступ містить формулювання актуальності роботи та її загальні характеристики. Зазначені мета, задачі, об'єкт, предмет та методи дослідження. Сформульовані елементи наукової новизни та основні практичні результати. Розглянуті основні публікації за темою дослідження, а також зазначено особистий вклад здобувача в них.

У першому розділі здобувач виконав аналіз наявних статистичних даних, що свідчать про актуальність науково-технічної задачі, що розглядається в рамках дослідження. Виконано огляд основних видів загроз в комп'ютерних системах, а також складових частин систем, що займаються виявленням вторгнень в їх роботу. Розглянуті основні види сигнатурних та евристичних методів виявлення вторгнень, підкреслені недоліки та переваги використання кожного з них. Обґрунтовано вибір ансамблевих методів машинного навчання для подальшого використання та вдосконалення. Виконано постановку науково-технічної задачі дослідження.

У другому розділі здобувачем виконано дослідження впливу різних етапів попередньої обробки даних на ефективність роботи методів машинного навчання, а також особливості вибору різних процедур формування вхідних послідовностей та оптимальних налаштувань при формуванні гомогенного беггінг-ансамблю. Сформовані відповідні рекомендації.

У третьому розділі здобувач розглянув існуючі підходи до покращення процедури ансамблювання. Розглянуто ефективність використання багатошарового перцептрону при побудові гомогенного беггінг-класифікатору з різними налаштуваннями. Виконано експериментальне дослідження ефективності використання ансамблевої обрізки, зваженого голосування, техніки калібрування впевненості, а також адаптації за рахунок мета-ознак та

мета-навчання. Перевірено вплив як окремих підходів, так і їх комбінацій на якість роботи беггінг-ансамблю.

У четвертому розділі здобувачем запропоновано метод формування гетерогенного ансамблю, який складається з трьох етапів та дозволяє оцінити якість роботи різних моделей машинного навчання в рамках однорідних класифікаторів для подальшого їх відбору. Виконано експериментальне дослідження ефективності використання гетерогенних ансамблів в залежності від параметрів сформованого методу.

У висновках сформульовані загальні результати дисертаційного дослідження, які дозволяють стверджувати про виконання усіх поставлених задач, які були запропоновані у першому розділі.

У додатках містяться список публікацій здобувача за темою дослідження, важливі фрагменти програмних компонентів, які були розроблені в рамках дисертаційної роботи, а також акти впровадження отриманих результатів.

Зауваження до дисертаційної роботи. Загальне враження від роботи позитивне, адже вона містить комплексні теоретичні та практичні дослідження та обґрунтовані висновки з рекомендаціями, проте є ряд зауважень:

1. У першому розділі здобувач розглянув основні групи сигнатурних та евристичних методів виявлення вторгнень, а також їх основні переваги та недоліки, проте було б більш наглядно сформувати у схематичному вигляді результати їх порівняння, які б продемонстрували переваги ансамблевих методів машинного навчання.
2. У другому розділі досліджено попередньою обробку даних. На мою думку, було б доцільним виконати порівняльний аналіз ефективності використання різних методів оптимізації кількості атрибутів.
3. У другому розділі досліджено вплив налаштувань дерев рішення та мета-алгоритму беггінг на якість класифікації. За результатами досліджень сформовані практичні рекомендації, проте було б добре також більше уваги приділити теоретичному обґрунтуванню отриманих результатів з метою їх інтерполяції на інші методи машинного навчання.

4. У третьому розділі обґрунтовано використання багатошарового перцептрону у якості базової моделі, а також виконано дослідження ефективності налаштувань на якість моделі. Проте результати дослідження, які наведені у графічному вигляді з використанням індексу класифікатору в упорядкованому списку від найгіршого до найкращого у розрізі певного параметру налаштування сприймаються досить складно, тому було б добре зробити їх більш наглядними та додати більш детальний опис самого експерименту та його результатів..

5. У третьому розділі рис. 3.19-3.22 є занадто громіздкими. Було б доцільно зменшити кількість складових рисунку та сфокусуватися на значущих результатах, або розділити їх на частини.

6. У четвертому розділі запропоновано процедуру формування гетерогенного ансамблю. Разом із тим, доцільно б дослідити вплив налаштувань відібраних базових класифікаторів на якість ансамблю.

7. Висновки до розділів повністю відображають їх зміст, проте є занадто громіздкими, тому варто було б скоротити їх, сфокусувавшись на отриманих результатах, які дозволили сформувати наукову новизну та практичну цінність дисертаційного дослідження.

8. Загальні висновки по роботі сформовані чітко та лаконічно, проте бажано було б додати до них основні практичні рекомендації, які наявні в тексті дисертації.

Відповідність дисертації встановленим вимогам і загальні висновки. За результатами аналізу дисертаційної роботи **не було виявлено фактів порушення академічної доброчесності**. Зазначені зауваження не змінюють загального позитивного враження від роботи, адже не вливають на її наукові та практичні результати і можуть бути враховані при подальших дослідженнях.

За результатами вивчення дисертаційної роботи Горносталя Олексія Андрійовича можна стверджувати, що вона є завершеним науковим дослідженням, адже складається з обґрунтованих теоретичних та практичних результатів, а її тема відповідає спеціальності 123 – «Комп’ютерна інженерія».

За усіма необхідними критеріями дисертаційна робота “Ансамблевий метод ідентифікації стану комп’ютерних систем” відповідає вимогам до оформлення дисертації, затвердженим Наказом МОН України від 12.01.2017 № 40, а також вимогам пунктів 6, 7, 8 і 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 КМУ від 12.01.2022 р. №44, а її автор, Горносталь Олексій Андрійович, заслуговує на присудження наукового ступеня доктора філософії за спеціальністю 123 – комп’ютерна інженерія.

Професор кафедри комп’ютерної інженерії та
програмування Національного технічного університету
«Харківський політехнічний інститут»
доктор технічних наук, професор Анатолій ПОВОРОЗНЮК
“27” травня 2024 р.

Підпис проф. Анатолія Поворознюка
ЗАСВІДЧУЮ:
ВЧЕНИЙ СЕКРЕТАР
НАЦІОНАЛЬНОГО ТЕХНІЧНОГО УНІВЕРСИТЕТУ
“ХАРКІВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ”
“27” 05 2024

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ * Національний технічний університет “Харківський політехнічний інститут” * код 02071180 *