РЕЦЕНЗІЯ

на дисертаційну роботу
Северіна Андрія Івановича
на тему «Алгоритмічне та програмне забезпечення захисту приватних наборів даних у задачах класифікації»,
представлену на здобуття ступеня доктора філософії
в галузі знань 12 Інформаційні технології
за спеціальністю 121 Інженерія програмного забезпечення

Актуальність теми дисертації.

У різних галузях людського життя все частіше застосовуються системи аналізу даних та штучного інтелекту. Підтвердженням цього є поширення систем підбору рекомендацій для користувача у електронній торгівлі, виявлення спаму в сервісах електронної пошти та модерації коментарів користувачів; а також інструменти для особистих задач (наприклад, чатботи ChatGPT, Google Bard, Microsoft Copilot).

Дані є необхідним та одним з ключових елементів для навчання та тестування систем інтелектуального аналізу даних. Аналіз значної кількості різнопланових даних сприяє виявленню закономірностей, а відповідно й побудові програмних систем з високою точністю. З огляду на це, важливим є завдання підбору та підготовки наборів даних, які можна використовувати при побудові таких систем. Однією зі складнощів у цьому завданні є наявність приватної інформації в наборах даних, що обмежує їх використання для систем систем аналізу даних і штучного інтелекту. Така інформація може бути конфіденційною (наприклад, серія та номер паспорту), чутливою (медичні діагнози пацієнтів) або секретною (фінансові, державні та військові дані). Збереження приватності наборів даних є важливим під час розроблення систем інтелектуального аналізу даних.

Таким чином, вдосконалення алгоритмічного та програмного забезпечення захисту приватних наборів даних у системах з використанням штучного інтелекту, яка вирішується у даній дисертаційній роботі для задачі класифікації, є актуальною науково-технічною задачею.

Оцінка обгрунтованості наукових результатів дисертації, їх достовірності та новизни.

Наукова новизна результатів дисертаційного дослідження полягає в наступному:

1. Уперше запропоновано архітектуру програмної системи для вирішення задачі класифікації на основі приватних даних, характерною

особливістю якої є захист приватних наборів даних, шляхом функціонального шифрування, що відбувається на стороні клієнта, і дозволяє збільшити кількість наборів даних для навчання загальнодоступних систем аналізу даних і штучного інтелекту.

2. Уперше запропоновано модифікацію програмної моделі шифрування даних, яка відрізняється від існуючої використанням двовимірних згорткових нейронних мереж, замість одновимірних, і дозволяє застосовувати модель шифрування з використанням нейронних мереж до даних, що представлені набором пікселів, з яких складається зображення.
3. Уперше розроблено алгоритмічно-програмний метод функціонального шифрування наборів даних, особливістю якого є можливість використання приватних наборів даних в загальнодоступних системах аналізу даних та штучного інтелекту шляхом зменшення їх розмірності й функціонального шифрування отриманих даних з використанням приватного ключа.
4. Уперше розроблено алгоритмічно-програмний метод пошуку нормальних поліномів серед незвідних, який відрізняється від існуючого використанням простих чисел у десятковому представленні замість поліномів, що дозволяє зменшити обчислювальні витрати алгоритму пошуку незвідних многочленів з $O(n^3)$ до $O(n \log(\log n))$ і, як наслідок, спростити міжбазисні перетворення у бінарних скінченних полях з метою пришвидшення виконання операцій над елементами поля у методах гомоморфного шифрування даних.
5. Уперше розроблено модифікований спосіб побудови матриці переходу між поліноміальним та нормальним базисами скінченного поля, який полягає у використанні рекурентної формули $\alpha_{i+1} = t^{p^{i+1}} = t^{p^i \cdot p} = (\alpha_i)^p$ замість обчислення остачі від ділення елемента $t^{p^{i+1}}$ на незвідний поліном, що дозволяє зменшити кількість використовуваної пам'яті з n^{p^i} до $n \cdot p$, а також обчислювальну складність з $O\left(m^{p^i}\right)$ до $O\left(m^p\right)$.

Достовірність та обґрунтованість наукових результатів отриманих у дисертаційній роботі забезпечується докладним аналізом джерел за даною проблематикою, коректною постановкою задач дослідження, правильністю застосування математичного апарату, а також даними, які отримані в результаті експериментальних досліджень.

Наукові дослідження були виконані здобувачем на кафедрі програмного забезпечення комп'ютерних систем КПІ ім. Ігоря Сікорського в рамках ініціативної НДР під керівництвом доцента кафедри програмного забезпечення

комп'ютерних систем, кандидата технічних наук, доцента Оная Миколи Володимировича.

Отже, в дисертаційній роботі поставлене наукове завдання вдосконалення алгоритмічного та програмного забезпечення захисту приватних наборів даних у системах з використанням штучного інтелекту для задачі класифікації виконано повністю, здобувач повною мірою оволодів методологією наукової діяльності.

Оцінка змісту дисертації, її завершеність та дотримання принципів академічної доброчесності.

За своїм змістом дисертаційна робота здобувача Северіна А. І. повністю відповідає Стандарту вищої освіти зі спеціальності 121 Інженерія програмного забезпечення та напрямкам досліджень відповідно до освітньої програми Інженерія програмного забезпечення.

Дисертаційна робота є завершеною науковою працею і свідчить про наявність особистого внеску здобувача у науковий напрям інженерії програмного забезпечення.

Розглянувши звіт подібності за результатами перевірки дисертаційної роботи на текстові співпадіння, можна зробити висновок, що дисертаційна робота Северіна Андрія Івановича є результатом самостійних досліджень здобувача і не містить елементів фальсифікації, компіляції, фабрикації, плагіату та запозичень. Використані ідеї, результати і тексти інших авторів мають належні посилання на відповідне джерело.

Мова та стиль викладення результатів.

Дисертаційна робота написана українською мовою.

Текст дисертації викладено структуровано, послідовно та логічно. Дисертація оформлена відповідно до чинних норм та є доступною для сприйняття та розуміння. Автор дотримується наукового стилю й використовує загальноприйняту термінологію. Текст роботи в достатній мірі проілюстрований таблицями, блок-схемами та фрагментами програмного коду.

Дисертація складається з вступу, 4 розділів, висновків, списку літератури та додатків. Загальний обсяг дисертації 254 сторінки.

У вступі обґрунтовано актуальність теми дисертації, визначено об'єкт, предмет, мету і задачі дослідження, а також сформульовано наукову новизну й практичне значення одержаних результатів.

У першому розділі дисертації розглянуто основні етичні аспекти використання штучного інтелекту. Проведено аналіз загроз приватності в системах з використанням машинного навчання. Комплексно проаналізовано методи збереження приватності. Сформульовано функціональні й нефункціональні вимоги до програмного забезпечення.

У другому розділі проаналізовано особливості використання полів Галуа для збереження приватності. Докладно розглянуто методи виконання операцій над елементами скінченних полів, залежно від обраного базису. Розроблено метод пошуку нормальних поліномів серед незвідних, який відрізняється від існуючого використанням простих чисел у десятковому представленні замість поліномів. Розроблено модифікований спосіб побудови матриці переходу між поліноміальним та нормальним базисами, який базується на використанні рекурентної формули.

У третьому розділі розроблено алгоритмічно-програмний метод захисту приватних наборів даних. Розглянуто математичне підґрунтя для побудови методу функціонального шифрування з використанням нейронних мереж. Запропоновано модифікацію моделі шифрування даних, яка полягає у використанні двовимірних згорткових нейронних мереж. Розроблено метод функціонального шифрування наборів даних, який надає можливість використання приватних наборів даних в загальнодоступних системах аналізу даних та штучного інтелекту. Проаналізовано й обрано метрики для оцінки розроблених методів.

Четвертий розділ присвячено проєктуванню програмного забезпечення захисту приватних наборів даних та проведенню експериментальних досліджень. Запропоновано архітектуру програмної системи для вирішення задачі класифікації на основі приватних даних. Розроблено програмну систему, яка дозволяє виконувати обчислення над елементами поля $GF(p^m)$, використовуючи поліноміальне й нормальне представлення елементів поля $GF(p^m)$. Проведено експериментальні дослідження розроблених методів міжбазисних перетворень. Розроблено програмну систему вирішення задачі класифікації на приватних наборах даних, що дозволяє вирішувати задачу класифікації на оригінальних та зашифрованих даних. Проведено експериментальні дослідження розробленого методу функціонального шифрування. Проаналізовано шляхи інтеграції розроблених програмних систем.

У висновках підсумовано основні наукові та практичні результати дисертаційного дослідження.

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

Оприлюднення результатів дисертаційної роботи.

Наукові результати дисертації висвітлені у 4 наукових публікаціях здобувача, серед яких: 2 статті у наукових виданнях, включених на дату опублікування до переліку наукових фахових видань України; 2 статті у періодичних наукових виданнях, проіндексованих у базах даних Web of Science

Core Collection та Scopus, з яких 1 статтю у виданні, віднесеному до третього квартилю (Q3) відповідно до класифікації SCImago Journal and Country Rank.

Також результати дисертації були апробовані на 3 наукових фахових конференціях.

Представлені публікації здобувача мають високий науковий рівень й повною мірою відображають головні наукові результати дисертації. Порушень принципів академічної доброчесності в них не виявлено. Особистий внесок здобувача до всіх наукових публікацій, опублікованих із співавторами, є вагомим.

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

Недоліки та зауваження до дисертаційної роботи.

1. Дослідження вирішення задачі класифікації на приватних наборах даних проводиться у роботі з використанням двох методів класифікації (CNN та лінійний класифікатор). Проте, для більш повної оцінки розроблених методів краще було б використати більше методів класифікації.
2. У другому розділі автор вперше використовує термін «нормальний поліном», однак не наводить його формулювання. Попри те, що таке визначення є в списку термінів, скорочень та позначень, краще було б навести його й при першому згадуванні, що спростило б сприйняття матеріалу.
3. У четвертому розділі проведено аналіз шляхів інтеграції розроблених програмних систем. Однак, приклади використання інструментів інтеграції (фрагменти коду) бажано було б навести з урахуванням контексту розроблених систем.

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

Висновок про дисертаційну роботу.

Вважаю, що дисертаційна робота здобувача ступеня доктора філософії Северіна Андрія Івановича на тему «Алгоритмічне та програмне забезпечення захисту приватних наборів даних у задачах класифікації» 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 знань 12 Інформаційні технології. 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 Северін Андрій Іванович заслуговує на присудження ступеня доктора філософії в галузі знань 12 Інформаційні технології за спеціальністю 121 Інженерія програмного забезпечення.

Рецензент:

професор кафедри інформаційних
систем та технологій
КПІ ім. Ігоря Сікорського,
доктор фізико-математичних наук,
професор Анатолій ДОРОШЕНКО

М.П. «24» травня 2024 року