Облікова картка дисертації

I. Загальні відомості

Державний обліковий номер: 0824U001642

Особливі позначки: відкрита

Дата реєстрації: 22-04-2024

Статус: Наказ про видачу диплома

Реквізити наказу МОН / наказу закладу: № НСВС_62_24 від 23.07.2024

II. Відомості про здобувача

Власне Прізвище Ім'я По-батькові:

1. Северін Андрій Іванович
2. Andrii Severin

Кваліфікація:

Ідентифікатор ORCID ID: 0009-0009-1366-8054

Вид дисертації: доктор філософії

Аспірантура/Докторантура: так

Шифр наукової спеціальності: 121

Назва наукової спеціальності: Інженерія програмного забезпечення

Галузь / галузі знань:

Освітньо-наукова програма зі спеціальності: Інженерія програмного забезпечення

Дата захисту: 24-06-2024

Спеціальність за освітою: Інженерія програмного забезпечення

Місце роботи здобувача: Національний технічний університет України "Київський політехнічний інститут імені Ігоря Сікорського"

Код за ЄДРПОУ: 02070921

Місцезнаходження: проспект Берестейський, буд. 37, Київ, 03056, Україна

Форма власності: Державна

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR:

III. Відомості про організацію, де відбувся захист

Шифр спеціалізованої вченої ради (разової спеціалізованої вченої ради): ДФ 26.002.140; ID 5369

Повне найменування юридичної особи: Національний технічний університет України "Київський політехнічний інститут імені Ігоря Сікорського"

Код за ЄДРПОУ: 02070921

Місцезнаходження: проспект Берестейський, буд. 37, Київ, 03056, Україна

Форма власності: Державна

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR:

IV. Відомості про підприємство, установу, організацію, в якій було виконано дисертацію

Повне найменування юридичної особи: Національний технічний університет України "Київський політехнічний інститут імені Ігоря Сікорського"

Код за ЄДРПОУ: 02070921

Місцезнаходження: проспект Берестейський, буд. 37, Київ, 03056, Україна

Форма власності: Державна

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR:

V. Відомості про дисертацію

Мова дисертації: Українська

Коди тематичних рубрик: 20.53.37, 28.23.02, 20.54.07

Тема дисертації:

1. Алгоритмічне та програмне забезпечення захисту приватних наборів даних у задачах класифікації

2. Algorithms and software for protecting private datasets in classification tasks

Реферат:

1. Северін А. І. Алгоритмічне та програмне забезпечення захисту приватних наборів даних у задачах класифікації. – Кваліфікаційна наукова праця на правах рукопису. Дисертація на здобуття наукового ступеня доктора філософії в галузі знань 12 Інформаційні технології за спеціальністю 121 Інженерія програмного забезпечення. – Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Київ, 2024. Впровадження систем аналізу даних і штучного інтелекту набуває все більшого поширення у різних аспектах людського життя. Окрім вже звичних випадків застосування таких систем у електронній комерції (наприклад, підбір рекомендацій користувачеві) та соціальній сфері (виявлення спаму, модерування коментарів), такі інструменти стрімко поширюються й для персонального використання (наприклад, чатботи ChatGPT, Google Bard, Microsoft Copilot, хоча вони з’явились лише впродовж останніх

двох років). В основі систем, що використовують методи машинного навчання, лежать дані. Вони є необхідним елементом як для навчання систем аналізу даних і штучного інтелекту, так і для їх тестування. Чим більше різнопланових даних, аналізується, тим точнішою є побудована програмна система. Найчастіше джерелом даних для програмних рішень з використанням машинного навчання є реальний світ. Іноді дані генерують програмним шляхом, намагаючись відтворити певні характеристики даних. Проте, незважаючи на те, що кількість створюваних та оброблюваних даних стрімко зростає, дані досить часто містять щонайменше частину приватної інформації, що обмежує їх використання для систем аналізу даних і штучного інтелекту. Приватні дані – інформація, яка є конфіденційною, чутливою або секретною. Прикладами секретних даних є військові, фінансові та державні дані. Конфіденційні дані – дані, що дозволяють ідентифікувати людину або компанію, їх прикладами є серія та номер паспорту, реєстраційний податковий номер та номер автомобіля. Прикладами чутливих даних є дані, що містять медичні діагнози пацієнтів. Збереження приватності даних є вкрай важливим, адже втрата приватності може призвести до дуже негативних наслідків (передусім різноманітних злочинів та недобросовісної конкуренції). Таким чином, вище описані задачі визначають актуальну науково-технічну задачу вдосконалення алгоритмічного та програмного забезпечення захисту приватних наборів даних у системах з використанням штучного інтелекту, яка вирішується у даній дисертаційній роботі для задачі класифікації. Метою дисертаційної роботи є удосконалення процесу оброблення приватних наборів даних для програмних систем інтелектуального аналізу даних. У першому розділі дисертаційної роботи розглянуто основні етичні аспекти використання систем штучного інтелекту та проблеми до яких може призвести їх ігнорування. Проаналізовано загрози приватності у таких системах, зокрема атаки інверсії, отруєння та логічного висновку. Проведено комплексний порівняльний аналіз методів збереження приватності в машинному навчанні (методи генерації синтетичних даних, анонімізації даних, диференційної приватності, гомоморфного шифрування та федеративного навчання), що дозволило виявити основні проблеми існуючих методів, які потребують досліджень. Розроблено вимоги до програмного забезпечення захисту приватних наборів даних у задачах класифікації. У другому розділі розроблено алгоритмічні методи міжбазисних перетворень елементів скінченних полів. Проаналізовано особливості використання полів Галуа в гомоморфних методах збереження приватності, а також визначено залежність часу виконання операцій над елементами скінченних полів від базису (поліноміального чи нормального), в якому представлені елементи. Запропоновано метод пошуку поліномів, який відрізняється від існуючого використанням простих чисел у десятковому представленні замість поліномів й дозволяє зменшити обчислювальну складність процесу пошуку нормальних многочленів. Розроблено модифікований спосіб для переходу між базисами, який полягає у використанні рекурентної формули, що дозволяє зменшити як кількість пам’яті, що використовується, так і обчислювальну складність. У третьому розділі розроблено алгоритмічно-програмний метод захисту приватних наборів даних. Проаналізовано математичне підґрунтя для побудови алгоритмічно-програмних методів з використанням нейронних мереж. Запропоновано метод функціонального шифрування 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. Запропоновано модифікацію моделі шифрування даних, яка полягає у використанні двовимірних згорткових нейронних мереж і дозволяє застосовувати модель шифрування даних, що представлені набором пікселів, з яких складається зображення. Проаналізовано метрики для оцінки методів захисту наборів даних.

2. Severin A. Algorithms and software for protecting private datasets in classification tasks. – Qualifying scientific work, manuscript. PhD thesis in the field of knowledge 12 Information technologies in specialty 121 Software Engineering. – National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, 2024. The implementation of data analysis and artificial intelligence systems is becoming increasingly widespread in various aspects of human life. In addition to the usual cases of application of such systems in e-commerce (for example, selecting recommendations for the user) and the social sphere (spam detection, comment moderation), such tools are also rapidly spreading for personal use (for example, chatbots ChatGPT, Google Bard, Microsoft Copilot,

although they appeared only during the last two years). Systems using machine learning methods are based on data. They are a necessary element both for training data analysis and artificial intelligence systems, and for their testing. The more diverse data is analyzed, the more accurate the built software system is. Most often, the data source for software solutions using machine learning is the real world. Sometimes the data is generated by software, trying to reproduce certain characteristics of the data. However, even though the amount of data being created and processed is growing rapidly, the data quite often contains at least some private information, which limits its use for data analysis and artificial intelligence systems. Private data – information that is confidential, sensitive or secret. Examples of secret data are military, financial, and government data. Confidential data – data that allows the identification of a person or company, examples of which are passport series and number, tax registration number and car number. Examples of sensitive data are data containing medical diagnoses of patients. The privacy-preserving of data is extremely important because the loss of privacy can lead to very negative consequences (primarily various crimes and unfair competition). Thus, the presence of these problems determines the actual scientific and technical problem of improving algorithmic and software protection of private datasets in systems using artificial intelligence, which is solved in this dissertation for the classification problem. The purpose of the dissertation is to improve the process of processing private datasets for software systems of intelligent data analysis. In the first section of the thesis, the main ethical aspects of using artificial intelligence systems and the problems that ignoring them can lead to were examined. Threats to privacy in such systems are analyzed, including inversion, poisoning, and logical inference attacks. A comprehensive comparative analysis of privacy preservation methods in machine learning (methods of synthetic data generation, data anonymization, differential privacy, homomorphic encryption, and federated learning) was conducted, which revealed the main problems of existing methods that require research. Software requirements for private datasets protection in classification tasks were developed. In the second section, the algorithmic methods for change-of-basis conversion of finite field elements were developed. The peculiarities of using Galois fields in homomorphic privacy-preserving methods were analyzed, and the dependence of the execution time of operations on the elements of finite fields on the basis (polynomial or normal) in which the elements are represented was determined. A method for finding polynomials was proposed, which differs from the existing one by using simple numbers in decimal representation instead of polynomials and allows to reduce the computational complexity of finding normal polynomials process. A modified way for conversion between bases was developed, which is based on the use of a recurrent formula, which allows to reduce both the amount of memory used and the computational complexity. In the third section, an algorithmic and software method for protecting private datasets were developed. The mathematical basis for building such methods using neural networks was analyzed. A method of functional data encryption was proposed, the feature of which is the possibility of using private datasets in publicly available data analysis and artificial intelligence systems by reducing their size and functionally encrypting the received data using a private key. A modification of the data encryption model was proposed, which is based on the use of two-dimensional convolutional neural networks and allows the encryption model to be applied to data represented by a set of pixels that make up an image. Metrics for evaluating dataset protection methods were analyzed.

Державний реєстраційний номер ДіР:

Пріоритетний напрям розвитку науки і техніки: Інформаційні та комунікаційні технології

Стратегічний пріоритетний напрям інноваційної діяльності: Розвиток сучасних інформаційних, комунікаційних технологій, робототехніки

Підсумки дослідження: Нове вирішення актуального наукового завдання

Публікації:

- 1. Modified Change-of-Basis Conversion Method in GF(2m) / I. A. Dychka, V. P. Legeza, M. V. Onai, A. I. Severin. // Radio Electronics, Computer Science, Control. — 2020. — №2. — С. 117–128 — DOI: 10.15588/1607-3274-2020-2-12

- 2. Method of Performing Operations on the Elements of GF(2m) Using a Sparse Table / I. Dychka, M. Onai, A. Severin, C. Hu. // International Journal of Computer Network and Information Security (IJCNIS). — 2024. — Vol. 16, №1. — pp. 61-72 — DOI: 10.5815/ijcnis.2024.01.05.
- 3. Северін А.І. Методи збереження приватності в машинному навчанні. / М.В. Онай, А.І. Северін // Вісник Хмельницького національного університету Серія: «Технічні науки». — 2023. — №6. — С. 274-280 — DOI: 10.31891/2307-5732-2023-329-6-274-280.
- 4. A. Severin. Architecture of a software system for solving the classification problem based on private data. / M. Onai, A. Severin // Herald of Khmelnytskyi national university. Technical Sciences. — 2024. — №1. — pp. 244-247 — DOI: 10.31891/2307-5732-2024-331-36.
- 5. Северін А.І. Метод захисту набору даних зображень для вирішення задачі класифікації. / М.В. Онай, А.І. Северін // Прикладна математика та комп'ютинг. ПМК-2020 : тринадцята наук. конф. магістрантів та аспірантів, Київ, 18-20 листопада 2020 р. : зб. тез доп. / [ред кол.: Дичка І.А. та ін.] . — К. : Просвіта, 2020. — С. 221-226.
- 6. Северін А.І. Комплексний порівняльний аналіз методів збереження приватності в машинному навчанні. / М.В. Онай, А.І. Северін // Актуальні задачі сучасних технологій : зб. тез доповідей XII міжнар. наук.-практ. конф. Молодих учених та студентів, (Тернопіль, 6-7 грудня 2023) / М-во освіти і науки України, Терн. націон. техн. ун-т ім. І. Пулюя [та ін.]. — Тернопіль: ФОП Паляниця В. А., 2023. — С. 406-407.
- 7. Северін А.І. Модифікований підхід для побудови матриці міжбазисних перетвореннь у GF(pm). / М.В. Онай, А.І. Северін // Матеріали XI науково-технічної конференції «Інформаційні моделі, системи та технології» Тернопільського національного технічного університету імені Івана Пулюя, (Тернопіль, 13-14 грудня 2023 р.). — Тернопіль: Тернопільський національний технічний університет імені Івана Пулюя, 2023 — С. 110.

Наукова (науково-технічна) продукція: методи, теорії, гіпотези; програмні продукти, програмно-технологічна документація

Соціально-економічна спрямованість: забезпечення промисловості чи населення новим видом інформаційно-комунікаційних послуг

Охоронні документи на ОПІВ:

Впровадження результатів дисертації: Планується до впровадження

Зв'язок з науковими темами: 0121U109925

VI. Відомості про наукового керівника/керівників (консультанта)

Власне Прізвище Ім'я По-батькові:

1. Онай Микола Володимирович

2. Mykola Onai

Кваліфікація: к. т. н., доц., 05.13.05

Ідентифікатор ORCID ID: 0000-0002-4938-8355

Додаткова інформація:

Повне найменування юридичної особи: Національний технічний університет України "Київський політехнічний інститут імені Ігоря Сікорського"

Код за ЄДРПОУ: 02070921

Місцезнаходження: проспект Берестейський, буд. 37, Київ, 03056, Україна

Форма власності: Державна

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR:

VII. Відомості про офіційних опонентів та рецензентів

Офіційні опоненти

Власне Прізвище Ім'я По-батькові:

1. Субботін Сергій Олександрович
2. Sergiy O. Subbotin

Кваліфікація: д.т.н., професор, 05.13.23

Ідентифікатор ORCID ID: 0000-0001-5814-8268

Додаткова інформація:

Повне найменування юридичної особи: Національний університет "Запорізька політехніка"

Код за ЄДРПОУ: 02070849

Місцезнаходження: вул. Жуковського, буд. 64, Запоріжжя, Запорізький р-н., 69063, Україна

Форма власності: Державна

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR:

Власне Прізвище Ім'я По-батькові:

1. Золотухін Олег Вікторович
2. Oleg V. Zolotukhin

Кваліфікація: к. т. н., доцент, 05.13.23

Ідентифікатор ORCID ID: 0000-0002-0152-7600

Додаткова інформація: https://www.webofscience.com/wos/author/record/E-5814-2018;
https://www.scopus.com/authid/detail.uri?authorId=57207774022;
https://scholar.google.com.ua/citations?user=B1sT8KsAAAAJ

Повне найменування юридичної особи: Харківський національний університет радіоелектроніки

Код за ЄДРПОУ: 02071197

Місцезнаходження: проспект Науки, буд. 14, Харків, Харківський р-н., 61166, Україна

Форма власності: Державна

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR:

Рецензенти

Власне Прізвище Ім'я По-батькові:

1. Дорошенко Анатолій Юхимович

2. Anatolii Y. Doroshenko

Кваліфікація: д.ф.-м.н., професор, 01.05.03

Ідентифікатор ORCID ID: 0000-0002-8435-1451

Додаткова інформація:

Повне найменування юридичної особи: Національний технічний університет України "Київський політехнічний інститут імені Ігоря Сікорського"

Код за ЄДРПОУ: 02070921

Місцезнаходження: проспект Берестейський, буд. 37, Київ, 03056, Україна

Форма власності: Державна

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR:

Власне Прізвище Ім'я По-батькові:

1. Клятченко Ярослав Михайлович

2. Yaroslav Klyatchenko

Кваліфікація: к. т. н., доцент, 05.13.05

Ідентифікатор ORCID ID: 0000-0003-4236-4059

Додаткова інформація: https://scholar.google.com.ua/citations?hl=uk&user=LWYsx7QAAAAJ; https://www.scopus.com/authid/detail.uri?authorId=55962506500

Повне найменування юридичної особи: Національний технічний університет України "Київський політехнічний інститут імені Ігоря Сікорського"

Код за ЄДРПОУ: 02070921

Місцезнаходження: проспект Берестейський, буд. 37, Київ, 03056, Україна

Форма власності: Державна

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR:

VIII. Заключні відомості

Власне Прізвище Ім'я По-батькові голови ради Коваль Олександр Васильович

Власне Прізвище Ім'я По-батькові головуючого на засіданні Коваль Олександр Васильович

Відповідальний за підготовку облікових документів Северін Андрій Іванович

Реєстратор

УкрІНТЕІ

Керівник відділу УкрІНТЕІ, що є відповідальним за реєстрацію наукової діяльності

Юрченко Тетяна Анатоліївна