Вінницький національний технічний університет
Міністерство освіти і науки України

Кваліфікаційна наукова
праця на правах рукопису

ПОЛЬГУЛЬ ТЕТЯНА ДМИТРІВНА

УДК 004.8:044.89

ДИСЕРТАЦІЯ

ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ ВИЯВЛЕННЯ ШАХРАЙСТВА ПРИ ІНСТАЛЮВАННІ МОБІЛЬНИХ ДОДАТКІВ З ВИКОРИСТАННЯМ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ

05.13.06 – інформаційні технології
Технічні науки

Подається на здобуття наукового ступеня кандидата технічних наук

Дисертація містить результати власних досліджень. Використання ідей, результатів і текстів інших авторів мають посилання на відповідне джерело
_____ Т. Д. Польгуль

Науковий керівник Яровий Андрій Анатолійович
доктор технічних наук, професор

Вінниця – 2020

АНОТАЦІЯ

Польгуль Т. Д. Інформаційна технологія виявлення шахрайства при інсталюванні мобільних додатків з використанням інтелектуального аналізу даних. – Кваліфікаційна наукова праця на правах рукопису.

Дисертація на здобуття наукового ступеня кандидата технічних наук за спеціальністю 05.13.06 «Інформаційні технології». – Вінницький національний технічний університет, Вінниця, 2020.

У зв'язку з появою на ринку значної кількості мобільних додатків, якими користуються мільярди користувачів, компанії-розробники мобільних додатків звертаються за послугами маркетингових кампаній з метою залучення інсталювань саме до їхнього додатку. Саме така потреба у маркетингових кампаніях стала однією з причин появи шахраїв та їх шахрайських способів інсталювання мобільних додатків. Шахраї, у свою чергу, приводять до компаній-розробників необхідну кількість фейкових (несправжніх) «користувачів» та отримують за це відповідну грошову винагороду. Проте такі «користувачі» ніколи не повертаються у мобільний додаток, оскільки є фейковими, ми ж їх називатимемо шахрайськими.

У наш час вже існують такі відомі види шахрайства при інсталюванні додатків, як мобільне викрадення (mobile hijacking), кліковий спам (click spamming), ферми дій (action farms), а також методи та системи виявлення шахрайства при інсталюванні мобільних додатків такі, як Fraudlogix та Kraken, Adjust, Kochava та TCM Attribution Analytics, Protect360 від Appsflyer, FraudScore та AppMetrica. Проте необхідно зазначити, що лише останні дві використовують інтелектуальну складову, причому система AppMetrica просто ґрунтується на алгоритмах та API системи FraudScore. Але вказані системи-аналоги виконують рейтингування користувачів на основі не всіх, а вибіркових вхідних даних, тому відбувається упущення шахраїв системами. Інші вказані системи використовують існуючі бази з шахрайськими даними (наприклад, IP-адресами),

що також призводить до упущення шахраїв, які мають інші властивості, шаблони, поведінку.

Очевидно, що причиною вищевказаних недоліків систем є відсутність єдиного підходу до виявлення шахрайства на основі всіх наявних даних. Також, недоліком існуючих систем є те, що вони розпізнають лише відомі види шахрайства і не можуть розпізнавати нові шахрайські шаблони. А в сучасному світі важливою є можливість системи адаптуватись, тому необхідним є створення відповідних інформаційних технологій, що матимуть змогу самонавчатися.

Все вищенаведене є передумовою актуальності створення інформаційної технології виявлення шахрайства при інсталюванні мобільних додатків, яка б відстежувала та визначала шаблони шахраїв, які непомітні людині. Розв'язанню цієї задачі присвячена дана робота.

Метою роботи дисертаційного дослідження є підвищення точності та швидкодії процесів виявлення шахрайства при інсталюванні мобільних додатків.

Для досягнення вказаної мети в роботі розв'язуються такі основні задачі:

– аналіз методів та постановка задачі виявлення шахрайства при інсталюванні мобільних додатків;

– формалізація процесу виявлення шахрайства як аномалії в даних;

– аналіз та класифікація різнорідних даних при інсталюванні мобільних додатків;

– розробка узагальненого методу виявлення шахрайства при інсталюванні мобільних додатків;

– розробка методу подолання різнорідності вхідних даних;

– розробка інформаційної технології виявлення шахрайства при інсталюванні мобільних додатків.

Науковою новизною виконаного дисертаційного дослідження визначено:

1. 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.

2. 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.

3. 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.

Практична цінність отриманих в дисертації результатів 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 доповідались та обговорювались на 9 науково-технічних конференціях: XLV, XLVI, XLVIII науково-технічних конференціях професорсько-викладацького складу, співробітників та студентів ВНТУ (2016, 2017, 2019 рр.); науково-практичній конференції «Сучасні тенденції розвитку системного програмування» (м. Київ, Національний авіаційний університет, 2016 р.); XIV Міжнародній конференції «Контроль і управління в складних системах (КУСС-2018)» (м. Вінниця, ВНТУ, 2018 р.); V Міжнародній науково-технічній конференції студентів, магістрів та аспірантів «Інформатика, управління та штучний інтелект» (м. Харків, Національний технічний університет «Харківський політехнічний інститут», 2018 р.); 5th International Winter School on Big Data BigDat2019 (м. Кембридж, University of Cambridge, Великобританія, 2019 р.); 577th International Conference on Innovative Engineering Technologies (ICIET) (м. Бангкок, Таїланд, 2019 р.); The 10th International Conference on Dependable Systems, Services and Technologies (DESSERT’2019) (м. Лідс, Великобританія, Leeds Beckett University, 2019 р.), де отримано нагороду «Best Paper Award»; The 14th International conference "Computer sciences and Information technologies" (CSIT 2019) (м. Львів, Україна, 2019 р.); The 2019 10th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS), (м. Метц, Франція, 2019 р.).

Результати дисертаційної роботи впроваджені на підприємствах та навчальних закладах: Garuda AI B. V. (м. Схіпгол, Нідерланди) – інформаційна технологія; ТОВ «ВІН ІНТЕРАКТИВ» (м. Вінниця, Україна) – алгоритми подолання різнорідності даних, модель процесу подолання різнорідності даних; ТОВ «4ХайТек» (м. Вінниця, Україна) – модель процесу подолання різнорідності вхідних даних, методика виявлення шахрая при інсталюванні

мобільних додатків; ПП «Літсофт» (м. Київ, Україна) – алгоритм подолання різнорідності даних, узагальнений метод виявлення шахрайства при інсталюванні мобільних додатків, методика створення узагальненого портрету шахрая; у навчальний процес кафедри комп'ютерних наук Вінницького національного технічного університету (ВНТУ) – узагальнений метод виявлення шахрайства при інсталюванні мобільних додатків та у навчальний процес кафедри інформатики, програмної інженерії та економічної кібернетики Херсонського державного університету – інформаційна технологія виявлення шахрайства при інсталюванні мобільних додатків з використанням інтелектуального аналізу даних.

Ключові слова: інформаційна технологія, інтелектуальний аналіз даних, виявлення шахрайства, виявлення аномалій в даних, коефіцієнти подібності, модель класифікації, машинне навчання, глибинні нейронні мережі, метод подолання різнорідності, матриця невідповідності, інсталювання мобільних додатків.

ABSTRACT

Polhul T. D. Information technology for fraud detection during mobile applications installation using data mining. – Qualification research paper, manuscript copyright.

Thesis for the degree of a candidate of technical sciences in specialty 05.13.06 «Information technology». – Vinnytsia National Technical University, Vinnytsia, 2020.

Nowadays there appeared a need to create information technology for fraud detection. This is due to the emergence of a huge number of new competing products among the billions of users in the mobile application market. To attract most users to their apps, mobile apps developers use the marketing campaigns service. Such a need in marketing campaigns has led to the massive appearance of fraudsters and fraudulent types of mobile applications installation, which can bring companies the required number of "users" and receive appropriate money reward for it [1]. However, it should be noted that such "users" never return to the mobile application because they are fake, we will call them fraudulent ones. Therefore, the creation of information technology for fraud detection during mobile applications installation is an important task.

For the time being, there are already known types of mobile applications installation fraud such as mobile hijacking, click spamming, action farms [2-7] and methods and systems for fraud detection during mobile applications installation such as: Fraudlogix and Kraken, Adjust, Kochava and TCM Attribution Analytics, Protect360 by Appsflyer, FraudScore and AppMetrica mentioned in the research paper. However, it should be noted that only the last couple uses the intelligent component, where AppMetrica simply relying on FraudScore and using its algorithms and API. But even the above-mentioned systems perform user rating based on not all but selective input data, so there is an omission of fraudsters by the system. Other specified systems use existing databases with fraudulent data (for example, IP-addresses), which

also leads to the omission of fraudsters having other properties, patterns, behavior, by such systems.

Obviously, the reason for the above mentioned disadvantages of the systems is the lack of a single approach for fraud detection based on available data. Also, the disadvantage of existing systems is that they recognize only known types of fraud and cannot recognize new fraudulent patterns. And the ability of the system to adapt is important in the modern world, therefore, it is necessary to create an intellectual information technology that will be able to learn.

All of the above is a prerequisite to creating an information technology for fraud detection during mobile applications installation that would track and detect fraudulent patterns unnoticed by humans. This research paper is dedicated to solve this problem.

The purpose of the qualification research paper is to improve the accuracy and speed of fraud detection process during mobile applications installation.

In order to achieve the mentioned purpose, the following basic tasks are solved in the work:

– analysis of methods and statement of fraud detection during mobile applications installation problem;

– formalization of the process of detection of fraud as an anomaly in the data;

– analysis and classification of heterogeneous data during mobile applications installation;

– development of a generalized method for fraud detection during mobile applications installation;

– development of a method of overcoming heterogeneity of input data;

– developing information technology for fraud detection during mobile applications installation.

The scientific novelty of the qualification research paper is:

1. For the first time, a method of overcoming the heterogeneity of input data is proposed, which is a set of procedures for feature selecting, dimensionality reduction and data normalization, the difference of which lies in the new model of overcoming

the heterogeneity of data by scaling information, which allows the whole set of heterogeneous user data to be reduced to a vector, reducing the diagnostic value of information.

2. The model of users' classification based on deep neural networks in terms of dimensionality reduction and data normalization has been improved according to the proposed method of overcoming heterogeneity of data, which is the basis for creating of general fraudsters fingerprint in order to simplify their detection processes.

3. For the first time, a generalized method for fraud detection in during installation of mobile applications has been developed, the difference being the use of the proposed users' classification model and the method of overcoming the heterogeneity of the input data, which allows defining users’ classes and increasing the reliability of fraud detection during mobile app installs.

The practical value of the results obtained in the qualification research paper is as follows: classification of heterogeneous data was carried out, which made it possible to simplify the process of analysis of data which is heterogeneous by metrics, dimensions and data templates and to automate it; an algorithm for detecting anomalies in data, algorithms for the process of overcoming the heterogeneity of input data, algorithm for fraud detection during mobile applications installations, an algorithm for generalized fraudster fingerprint formation, and algorithm for minimizing the time of fraud detection based on the parallelization of computing processes, which are the basis of information technology, which has increased the accuracy and speed of detection of fraudsters, were developed; information technology for fraud detection during mobile applications installation has been proposed for the first time, that uses the following: a generalized method for fraud detection during mobile applications installation, a method for overcoming the heterogeneity and classification model of user input data, which, unlike existing Antifraud technologies, allowed to improve accuracy of users’ classification to 99,14 %, in particular, detecting fraudulent users to 82,95 %; “Mobile App Install Fraud Detection System” software for fraud detection during mobile applications installation has been developed.

The results of the qualification research paper were reported and discussed at 9 scientific and technical conferences: XLV, XLVI, XLVIII scientific and technical conferences of teaching staff, staff and students of Vinnitsa National Technical University; scientific-practical conference "Modern tendencies of development of system programming" (Kyiv, National Aviation University, 2016); XIV International Conference "Control and Control in Complex Systems (КУСС-2018)» (Vinnitsa, Vinnitsa National Technical University, 2018); V International Scientific and Technical Conference of Students, Masters and Postgraduate Students "Informatics, Management and Artificial Intelligence" (Kharkiv, National Technical University "Kharkiv Polytechnic Institute", 2018); 5th International Winter School on Big Data BigDat2019 (Cambridge, University of Cambridge, UK, 2019); 577th International Conference on Innovative Engineering Technologies (ICIET) (Bangkok, Thailand, 2019); The 10th International Conference on Dependable Systems, Services and Technologies (DESSERT'2019) (Leeds, UK, Leeds Beckett University, 2019), where the «Best Paper Award» was recieved; The 14th International conference "Computer sciences and Information technologies" (CSIT 2019) (Lviv, Ukraine, September 17-20, 2019); The 2019 10th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS), (Metz, France, September 18-21, 2019).

The results of the qualification research paper were implemented at Garuda AI B.V. (Netherlands) – information technology; LLC «WinInteractive» – algorithms for overcoming data heterogeneity, model for the process of overcoming data heterogeneity; LLC «4HighTech» – a model of the process of overcoming the heterogeneity of the input data, the method of fraud detection when installing mobile applications; PE «Litsoft» – algorithm for overcoming heterogeneity of data, a generalized method of detecting fraud when installing mobile applications, a method for creating a generalized fraudster's fingerprint; to the educational process of the Computer Science Department of Vinnytsia National Technical University – a generalized method for fraud detection during mobile applications installation; and to

the educational process of the Department of Informatics, Software Engineering and Economic Cybernetics of Kherson State University – information technology for fraud detection during mobile applications installation using data mining.

Keywords: information technology, data mining, fraud detection, anomaly detection, similarity coefficients, classification model, machine learning, deep neural networks, method for overcoming heterogeneity of data, confusion matrix, mobile applications installation.

СПИСОК ПУБЛІКАЦІЙ ЗА ТЕМОЮ ДИСЕРТАЦІЇ

[1] T. Polhul, and A. Yarovyi, “Development of a method for fraud detection in heterogeneous data during installation of mobile applications”, *Eastern-European Journal of Enterprise Technologies*, no. 1/2 (97), 2019. doi: 10.15587/1729-4061.2019.155060

[2] A. Yarovyi, and T. Polhul, “Applied Aspects of Implementation of Intelligent Information Technology for Fraud Detection During Mobile Applications Installation”, *Advances in Intelligent Systems and Computing IV. CCSIT 2019: Advances in Intelligent Systems and Computing, Springer*, Cham, Switzerland, vol 1080, pp. 377-386, 2019. doi: https://doi.org/10.1007/978-3-030-33695-0_26

[3] T. Polhul, “Conceptual Model of an Intelligent System for Detecting Fraud During Mobile Applications Installation”, in *Proc. 10th International Conference on Dependable Systems, Services and Technologies (DESSERT)*, Leeds, United Kingdom, pp. 167-174, 2019. doi: 10.1109/DESSERT.2019.8770030. Режим доступу: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8770030&isnumber=8770005.

[4] T.D. Polhul, A.A. Yarovyi, R.Romaniuk, P.Komada, N. Askarova "Method of data anomaly detection in the process of mobile applications installation", *Proc. SPIE 11176, Photonics Applications in Astronomy, Communications, Industry, and High-Energy Physics Experiments 2019*, 111761Y; https://doi.org/10.1117/ 12.2536855

[5] T. Polhul and A. Yarovyi, "Method of Fraudster Fingerprint Formation During Mobile Application Installations", *2019 10th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS)*, Metz, France, 2019, pp. 1099-1103. doi: 10.1109/IDAACS.2019.8924369. Режим доступу: https://ieeexplore.ieee.org/ document/8924369.

[6] A. Yarovyi, and T. Polhul, “Intelligent information technology for fraud detection during mobile applications installation”, in *Proc. 14th International conference "Computer sciences and Information technologies" (CSIT 2019)*, pp. 1-5, Lviv, 2019. doi: 10.1109/STC-CSIT.2019.8929827. Режим доступу: https://ieeexplore.ieee.org/ document/8929827.

[7] T. Polhul, “Development of an intelligent system for detecting mobile app install fraud”, *International Journal of Advances in Electronics and Computer Science (IJAECS)*, vol. 6, no. 7, pp. 13-17, July, 2019.

[8] А.А. Яровий, О.Н. Романюк, І.Р. Арсенюк, та Т.Д. Польгуль, “Виявлення шахрайства при інсталюванні програмних додатків з використанням інтелектуального аналізу даних”, *Наукові праці Донецького національного технічного університету. Серія “Інформатика, кібернетика та обчислювальна техніка”*, № 2 (25), с. 126-131, 2017. Режим доступу: http://science.donntu.edu.ua/ wp-content/uploads/2018/03/ikvt_2017_2_site-1.pdf

[9] Т. Д. Польгуль, та А. А. Яровий, “Метод подолання різнорідності даних для виявлення шахрайства при інсталюванні мобільних додатків”, *Вісник СНУ ім. В. Даля – Сєвєродонецьк: СНУ ім. В. Даля*, № 7 (248), с.60-69, 2018.

[10] Т.Д. Польгуль, та А.А. Яровий, “Аналіз різнорідних даних в інтелектуальних системах виявлення шахрайства”, *Вісник Вінницького політехнічного інституту*, № 2, с. 78-90, 2019.

[11] Т.Д. Польгуль, “Інформаційна технологія побудови інтелектуальних систем виявлення шахрайства при інсталюванні мобільних додатків”, *Інформаційні технології та комп'ютерна інженерія*, № 1, с. 4-16, 2019.

[12] T. Polhul, “Development of an intelligent system for detecting mobile app install fraud”, in *Proc. IRES 156th International Conference,* Bangkok, Thailand, 2019, pp. 25-29.

[13] А.А. Яровий, та Т.Д. Польгуль, “Комп'ютерна програма “Програмний модуль збору даних інформаційної технології виявлення шахрайства при інсталюванні програмних додатків”, *Свідоцтво про реєстрацію авторського права на твір № 76348*, К.: Міністерство економічного розвитку і торгівлі України, 2018.

[14] А.А. Яровий А. А., та Т.Д. Польгуль, “Комп'ютерна програма “Програмний модуль визначення схожості користувачів інформаційної технології виявлення шахрайства при інсталюванні програмних додатків”, *Свідоцтво про реєстрацію авторського права на твір № 76347*, К.: Міністерство економічного розвитку і торгівлі України, 2018.

[15] Т.Д. Польгуль, та А.А. Яровий, “Визначення шахрайських операцій при встановленні мобільних додатків з використанням інтелектуального аналізу даних”, *Сучасні тенденції розвитку системного програмування. Тези доповідей*, Київ, 2016, с. 55-56. Режим доступу: http://ccs.nau.edu.ua/wp-content/uploads/2017/12/%D0%A1%D0%A2%D0%A0%D0%A1%D0%9F_2016_07.pdf

[16] А. Яровий, Т. Польгуль, та Л. Крилик, “Розробка методу виявлення шахрайства при інсталюванні мобільних додатків з використанням інтелектуального аналізу даних”, *XIV Міжнародна конференція Контроль і управління в складних системах (КУСС-2018). Тези доповідей*, Вінниця, 2018, с. 35.

[17] А.А. Яровий, та Т.Д. Польгуль, “Подолання різнорідності вхідних даних при виявленні шахрайства при інсталюванні мобільних додатків з використанням інтелектуального аналізу даних”, на *П’ятій міжнародній науково-технічній конференції студентів, магістрів, аспірантів «Інформатика, управління та штучний інтелект»*, Національний технічний університет «Харківський політехнічний інститут», Харків, 2018, с. 109.

[18] Т.Д. Польгуль, та А.А. Яровий, “Визначення шахрайських операцій при інсталяції мобільних додатків з використанням інтелектуального аналізу даних”, на *XLVI науково-технічній конференції підрозділів ВНТУ*, Вінниця, 2017. Режим доступу: http://ir.lib.vntu.edu.ua/bitstream/handle/123456789/17200/2158.pdf?sequence=3

[19] Т.Д. Польгуль, “Моделювання процесу виявлення шахрайства при інсталюванні мобільних додатків”, на *XLVIII науково-технічній конференції підрозділів ВНТУ*, Вінниця, 2019. Режим доступу: https://conferences.vntu.edu.ua/index.php/all-fitki/all-fitki-2019/paper/view/6863

[20] Т.Д. Польгуль, “Порівняльний аналіз Apache Spark та Apache Flink для роботи з Big Data”, на *XLV науково-технічній конференції підрозділів ВНТУ*, Вінниця, 2016. Режим доступу: https://ir.lib.vntu.edu.ua/handle/123456789/11619

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ..... 18
ВСТУП 19
РОЗДІЛ 1 АНАЛІЗ МЕТОДІВ ТА ПОСТАНОВКА ЗАДАЧІ ВИЯВЛЕННЯ ШАХРАЙСТВА ПРИ ІНСТАЛЮВАННІ МОБІЛЬНИХ ДОДАТКІВ ....**Ошибка! Закладка не определена.**
1.1 Аналіз об’єкту дослідження ..... **Ошибка! Закладка не определена.**
1.1.1 Визначення поняття шахрайства при інсталюванні мобільних додатків ..... **Ошибка! Закладка не определена.**
1.1.2 Визначення задачі виявлення шахрайства як однієї із задач пошуку аномалій в даних..... **Ошибка! Закладка не определена.**
1.1.3 Аналіз характеристик та властивостей шахрайства як аномалії в даних ..... **Ошибка! Закладка не определена.**
1.2 Варіантний аналіз методів пошуку аномалій в даних ...**Ошибка! Закладка не определена.**
1.2.1 Класифікація методів виявлення аномалій.....**Ошибка! Закладка не определена.**
1.2.2 Методи кластеризації..... **Ошибка! Закладка не определена.**
1.2.3 Методи класифікації ..... **Ошибка! Закладка не определена.**
1.2.4 Статистичні методи..... **Ошибка! Закладка не определена.**
1.2.5 Ансамбль методів на прикладі глибинного навчання.....**Ошибка! Закладка не определена.**
1.3 Аналіз моделей подібності шахрайських шаблонів..**Ошибка! Закладка не определена.**
1.4 Аналіз сучасних систем виявлення шахрайства при інсталюванні мобільних додатків ..... **Ошибка! Закладка не определена.**
1.5 Аналіз проблемних аспектів, що виникають при виявлені шахрайства при інсталюванні мобільних додатків та постановка задач дослідження...**Ошибка! Закладка не определена.**
1.6 Висновки до розділу 1 ..... **Ошибка! Закладка не определена.**

РОЗДІЛ 2 РОЗРОБКА МЕТОДІВ ТА МОДЕЛЕЙ ВИЯВЛЕННЯ ШАХРАЙСТВА ПРИ ІНСТАЛЮВАННІ МОБІЛЬНИХ ДОДАТКІВ З ВИКОРИСТАННЯМ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ ..... **Ошибка! Закладка не определена.**

2.1 Формалізація процесу виявлення шахрайства як аномалії в даних **Ошибка! Закладка не определена.**

2.2 Аналіз та класифікація різнорідних даних при інсталюванні мобільних додатків ..... **Ошибка! Закладка не определена.**

2.3 Розробка узагальненого методу виявлення шахрайства при інсталюванні мобільних додатків ..... **Ошибка! Закладка не определена.**

2.3.1 Розробка методу подолання різнорідності вхідних даних ..... **Ошибка! Закладка не определена.**0

2.3.1.1 Класифікація вхідних даних, їх шкалювання за інформативністю та формалізація..... **Ошибка! Закладка не определена.**

2.3.1.2 Розробка моделей процесу подолання різнорідності вхідних даних ..... **Ошибка! Закладка не определена.**3

2.3.2 Аналіз доцільності використання методів обробки Big Data для виявлення аномалій в даних при інсталяції мобільних додатків ..... **Ошибка! Закладка не определена.**1

2.3.3 Модель класифікації вхідних даних для виявлення аномалій в Big Data ..... **Ошибка! Закладка не определена.**2

2.4 Висновки до розділу 2 ..... **Ошибка! Закладка не определена.**6

РОЗДІЛ 3 РОЗРОБКА ІНФОРМАЦІЙНОЇ ТЕХНОЛОГІЇ ВИЯВЛЕННЯ ШАХРАЙСТВА ПРИ ІНСТАЛЮВАННІ МОБІЛЬНИХ ДОДАТКІВ .... **Ошибка! Закладка не определена.**7

3.1 Концептуальні особливості побудови інформаційної технології виявлення шахрайства на основі методів та моделей..... **Ошибка! Закладка не определена.**7

3.1.1 Розробка концептуальної моделі інформаційної технології виявлення шахрайства ..... **Ошибка! Закладка не определена.**8

3.1.2 Аналіз та розробка структур даних в інформаційній технології виявлення шахрайства ..... **Ошибка! Закладка не определена.**2
3.1.3 Розробка шаблону шахрая..... **Ошибка! Закладка не определена.**7
3.1.4 Розробка нечіткої моделі для формування портрету шахрая ... **Ошибка! Закладка не определена.**8
3.2 Розробка алгоритмів виявлення шахрая при інсталюванні мобільних додатків ..... **Ошибка! Закладка не определена.**0
3.3 Розробка алгоритму створення узагальненого портрету шахрая ... **Ошибка! Закладка не определена.**4
3.4 Розробка алгоритму пошуку аномалій в даних ..... **Ошибка! Закладка не определена.**6
3.5 Аналіз процесів в інформаційній технології виявлення шахрайства ..... **Ошибка! Закладка не определена.**6
3.6 Проектування інформаційної технології виявлення шахрайства ... **Ошибка! Закладка не определена.**0
3.7 Висновки до розділу 3..... **Ошибка! Закладка не определена.**3
РОЗДІЛ 4 ЕКСПЕРИМЕНТАЛЬНІ ДОСЛІДЖЕННЯ ІНФОРМАЦІЙНОЇ ТЕХНОЛОГІЇ ВИЯВЛЕННЯ ШАХРАЙСТВА ПРИ ІНСТАЛЮВАННІ МОБІЛЬНИХ ДОДАТКІВ З ВИКОРИСТАННЯМ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ ..... **Ошибка! Закладка не определена.**5
4.1 Аналіз та підготовка вхідних даних для проведення експериментальних досліджень з використанням інформаційної технології виявлення шахрайства ..... **Ошибка! Закладка не определена.**5
4.2 Розробка методики проведення експериментальних досліджень шахрайства при інсталюванні мобільних додатків ..... **Ошибка! Закладка не определена.**79
4.3 Розробка алгоритму мінімізації часу виявлення шахраїв на основі розпаралелення обчислювальних процесів..... **Ошибка! Закладка не определена.**2

4.4 Дослідження адекватності моделей, точності та швидкодії методу виявлення шахрайства, аналіз результатів тестування...**Ошибка! Закладка не определена.**5

4.5 Аналіз результатів впровадження.....**Ошибка! Закладка не определена.**1

4.5.1 Впровадження інформаційної технології для виявлення шахрайства при інсталюванні мобільних додатків в ТОВ «ВІН ІНТЕРАКТИВ»**Ошибка! Закладка не определена.**1

4.5.2 Впровадження інформаційної технології в Garuda AI B.V.....**Ошибка! Закладка не определена.**2

4.5.3 Впровадження інформаційної технології у ТОВ «4ХайТек»...**Ошибка! Закладка не определена.**3

4.5.4 Впровадження інформаційної технології у ПП «Літсофт».....**Ошибка! Закладка не определена.**4

4.5.5 Впровадження інформаційної технології у навчальний процес .....**Ошибка! Закладка не определена.**4

4.6 Висновки до розділу 4.....**Ошибка! Закладка не определена.**5

ВИСНОВКИ.....**Ошибка! Закладка не определена.**8

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ..... 262

ДОДАТКИ.....**Ошибка! Закладка не определена.**3

Додаток А Документи щодо впровадження результатів роботи .....**Ошибка! Закладка не определена.**4

Додаток Б Основні лістинги програмного забезпечення...**Ошибка! Закладка не определена.**0

Додаток В Список експертів, які проводили оцінювання .**Ошибка! Закладка не определена.**4

Додаток Д Сертифікати на публікації та виступи .....**Ошибка! Закладка не определена.**5

Додаток Е Список публікацій за темою дисертації.....**Ошибка! Закладка не определена.**2

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ

ACID	Atomicity, consistency, isolation, durability
ANN	Artificial neural networks
API	Application programming interface
CNN	Convolutional neural network
DNN	Deep neural network
DQN	Deep Q Learning
EM	Expectation-maximization
GAN	Generative Adversarial Nets
LB	Load Balancer
LOF	Local outlier factor
MTTI	Mean-time-to-install
NN	Neural networks
RNN	Recurrent neural network
SVM	Support Vector Machine
TPR	True positive rate
TTI	Time-to-install Outliers
ІАД	Інтелектуальний аналіз даних
ІТ	Інформаційна технологія
КН	Комп’ютерні науки
ШІ	Штучний інтелект

ВСТУП

Обґрунтування вибору теми дослідження. У зв’язку з появою на ринку значної кількості мобільних додатків, якими користуються мільярди користувачів, компанії-розробники мобільних додатків звертаються за послугами маркетингових кампаній з метою залучення інсталювань саме до їхнього додатку. Саме така потреба у маркетингових кампаніях стала однією з причин появи шахраїв та їх шахрайських способів інсталювання мобільних додатків. Шахраї, у свою чергу, приводять до компаній-розробників необхідну кількість фейкових (несправжніх) «користувачів» та отримують за це відповідну грошову винагороду. Проте такі «користувачі» ніколи не повертаються у мобільний додаток, оскільки є фейковими, ми ж їх називатимемо шахрайськими.

У наш час вже існують такі відомі види шахрайства при інсталюванні додатків, як мобільне викрадення (mobile hijacking), кліковий спам (click spamming), ферми дій (action farms) [1]-[3], а також методи та системи виявлення шахрайства при інсталюванні мобільних додатків такі як Fraudlogix [4] та Kraken [5], Adjust [6], Kochava [7] та TCM Attribution Analytics [8], Protect360 [9] від Appsflyer [10], FraudScore [11] та AppMetrica [12], які згадуються у роботі [13]. Проте необхідно зазначити, що лише останні дві використовують інтелектуальну складову, причому система AppMetrica просто ґрунтується на алгоритмах та API системи FraudScore. Але вказані системи-аналоги виконують рейтингування користувачів на основі не всіх, а вибіркових вхідних даних, тому відбувається упущення шахраїв системами [14]-[16]. Інші вказані системи використовують існуючі бази з шахрайськими даними (наприклад, IP-адресами), що також призводить до упущення шахраїв, які мають інші властивості, шаблони, поведінку.

Очевидно, що причиною вищевказаних недоліків систем є відсутність єдиного підходу до виявлення шахрайства на основі всіх наявних даних. Також, недоліком існуючих систем є те, що вони розпізнають лише відомі види

шахрайства і не можуть розпізнавати нові шахрайські шаблони. А в сучасному світі важливою є можливість системи адаптуватись, тому необхідним є створення відповідних інформаційних технологій, що матимуть змогу самонавчатися.

Все вищенаведене є передумовою актуальності створення інформаційної технології виявлення шахрайства при інсталюванні мобільних додатків, яка б відстежувала та визначала шаблони шахраїв, які непомітні людині [16]. Розв’язанню цієї задачі присвячена дана робота.

Зв’язок роботи з науковими програмами, планами, темами. Дисертаційне дослідження проводилось згідно з планами науково-дослідних робіт кафедри комп'ютерних наук Вінницького національного технічного університету, в тому числі в межах:

- наукового проекту Ф61/199-20150/4711 “Методологія побудови високопродуктивних інтелектуалізованих паралельно-ієрархічних систем на основі сучасних мережевих обчислювальних комплексів з гетерогенною архітектурою” (№ державної реєстрації: 0115U001975, 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 бере участь як виконавець окремих підрозділів.

Мета і завдання дослідження.

Метою роботи є підвищення точності та швидкодії процесів виявлення шахрайства при інсталюванні мобільних додатків.

Для досягнення вказаної мети в роботі розв’язуються такі основні задачі:

– аналіз методів та постановка задачі виявлення шахрайства при інсталюванні мобільних додатків;

– формалізація процесу виявлення шахрайства як аномалії в даних;

– аналіз та класифікація різнорідних даних при інсталюванні мобільних додатків;

– розробка узагальненого методу виявлення шахрайства при інсталюванні мобільних додатків;

– розробка методу подолання різнорідності вхідних даних;

– розробка інформаційної технології виявлення шахрайства при інсталюванні мобільних додатків.

Об'єкт дослідження – процеси виявлення шахрайства як аномалій в даних при інсталюванні мобільних додатків.

Предмет дослідження – моделі, методи та інформаційні технології виявлення шахрайства при інсталюванні мобільних додатків з використанням інтелектуального аналізу даних.

Методи дослідження, що використані в роботі: методи шкалювання під час вирішення задач аналізу та класифікації різнорідних даних при інсталюванні мобільних додатків та розробки методу подолання різнорідності вхідних даних; теорія множин для вирішення задачі формалізації процесу виявлення шахрайства як аномалії в даних, а також методи класифікації, статистичні методи, методи машинного навчання, інтелектуальний аналіз даних, методи кластеризації, нейромережеві методи для вирішення задач розробки узагальненого методу виявлення шахрайства при інсталюванні мобільних додатків та розробки інформаційної технології виявлення шахрайства при інсталюванні мобільних додатків.

Наукова новизна отриманих результатів. В ході розв'язання поставлених задач були отримані наукові результати.

1. 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.

2. 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.

3. 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.

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 підприємствах та навчальних закладах: Garuda AI B. V. (м. Схіпгол, Нідерланди) – інформаційна технологія; ТОВ «ВІН ІНТЕРАКТИВ» (м. Вінниця, Україна) – алгоритми подолання різнорідності даних, модель процесу подолання різнорідності даних; ТОВ «4ХайТек» (м. Вінниця, Україна) – модель процесу подолання різнорідності вхідних даних, методика виявлення шахрая при інсталюванні мобільних додатків; ПП «Літсофт» (м. Київ, Україна) – алгоритм подолання різнорідності даних, узагальнений метод виявлення шахрайства при інсталюванні мобільних додатків, методика створення узагальненого портрету шахрая; у навчальний процес кафедри комп’ютерних наук Вінницького національного технічного університету (ВНТУ) – узагальнений метод виявлення шахрайства при інсталюванні мобільних додатків та у навчальний процес кафедри інформатики, програмної інженерії та економічної кібернетики Херсонського державного університету – інформаційна технологія виявлення шахрайства при інсталюванні мобільних додатків з використанням інтелектуального аналізу даних.

Особистий внесок здобувача. Усі результати, які складають основний зміст дисертації, отримані здобувачем самостійно. У роботах [16], [18], [23], [24], [27], [62] здобувачеві належать усі теоретичні та практичні результати. У роботах, опублікованих у співавторстві, здобувачу належать: [14] – розроблено метод виявлення шахрайства, математичну модель процесу шкалювання, алгоритм шкалювання різнорідних масивів, алгоритм обробки отриманих груп однорідних даних, схему процесу виявлення шахраїв, інтелектуальну систему автоматичного виявлення шахраїв, здійснено класифікацію різнорідних даних при інсталюванні мобільних додатків; [108] – здійснено формалізацію процесу

виявлення шахрайства як аномалії в даних, розроблено метод виявлення аномалій при інсталюванні мобільних додатків; [17], [26] – розроблено метод формування портрету шахрая та алгоритм розробки нечіткої моделі для формування портрету шахрая; [28] – запропоновано інтелектуальну інформаційну технологію виявлення шахрайства при інсталюванні мобільних додатків, удосконалено класифікацію користувачів з використанням глибинних нейронних мереж, створення узагальненого портрету шахрая; [19] – розроблено систему виявлення шахрайства при інсталюванні програмних додатків; [13] – запропоновано метод, моделі та алгоритми подолання різнорідності даних для виявлення шахрайства; [15] – розроблено метод та алгоритм аналізу різнорідних даних, математичну модель процесу аналізу різнорідних даних, схему експериментального дослідження виявлення аномалій в різнорідних даних; [92], [93] – розробка програмного забезпечення для модулю збору даних та для модулю визначення схожості користувачів, розробка алгоритму мінімізації часу виявлення шахраїв, алгоритму пошуку аномалій в даних; [20] – запропоновано модель класифікації користувачів; [22] – розроблено метод виявлення шахрайства при інсталюванні мобільних; [25] – запропоновано метод подолання різнорідності вхідних даних; [21] – запропоновано модель визначення шахрайських способів встановлення мобільних додатків.

Апробація матеріалів дисертації. Результати дисертаційної роботи доповідались та обговорювались на 9 науково-технічних конференціях: XLV, XLVI, XLVIII науково-технічних конференціях професорсько-викладацького складу, співробітників та студентів Вінницького національного технічного університету; науково-практичній конференції «Сучасні тенденції розвитку системного програмування» (м. Київ, Національний авіаційний університет, 2016 р.); XIV Міжнародній конференції «Контроль і управління в складних системах (КУСС-2018)» (м. Вінниця, Вінницький національний технічний університет, 2018 р.); V Міжнародній науково-технічній конференції студентів, магістрів та аспірантів «Інформатика, управління та штучний інтелект» (м. Харків, Національний технічний університет «Харківський політехнічний

інститут», 2018 р.); 5th International Winter School on Big Data BigDat2019 (м. Кембридж, University of Cambridge, Великобританія, 2019 р.); 577th International Conference on Innovative Engineering Technologies (ICIET) (м. Бангкок, Таїланд, 2019 р.); The 10th International Conference on Dependable Systems, Services and Technologies (DESSERT'2019) (м. Лідс, Великобританія, Leeds Beckett University, 2019 р.), де отримано нагороду «Best Paper Award»; The 14th International conference "Computer sciences and Information technologies" (CSIT 2019) (м. Львів, Україна, вересень 17-20, 2019); The 2019 10th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS), (м. Метц, Франція, вересень 18-21, 2019).

Публікації. За темою дисертації опубліковано 20 праць, в тому числі 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.

Структура та обсяг дисертації. 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 сторінки.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

[1] “Our take on mobile fraud detection”, available at: http://geeks.jampp.com/data-science/mobile-fraud/

[2] Vacha Dave, Saikat Guha, and Yin Zhang, “ViceROI: Catching Click-Spam in Search Ad Networks”, available at: http://www.sysnet.ucsd.edu/~vacha/ccs13.pdf

[3] V. Dave, S. Guha, and Y. Zhang, “Measuring and Fingerprinting Click-Spam in Ad Networks”, in *Proc. Annual Conference of the ACM Special Interest Group on Data Communication (SIGCOMM)”,* Helsinki, Finland, 2012, pp. 175-186.

[4] Fraudlogix: Ad Fraud Solutions for Exchanges, Networks, SSPs & DSPs. [Online]. Available: https://www.fraudlogix.com/

[5] Kraken Antibot. [Online]. Available: http://kraken.run/

[6] Adjust . [Online]. Available: https://www.adjust.com/

[7] Kochava Uncovers Global Ad Fraud Scam . [Online]. Available: https://www.kochava.com/

[8] TMC Attribution Analytics. [Online]. Available: https://help.tune.com/marketing-console/attribution-analytics/

[9] Appsflyer: Protect your data from mobile fraud: Protect360. [Online]. Available: https://www.appsflyer.com/product/protect360/

[10] AppsFlyer: Measure In-App To Grow Your Mobile Business. [Online]. Available: https://www.appsflyer.com/

[11] FraudScore: FraudScore fights ad fraud using Machine Learning. [Online]. Available: https://fraudscore.mobi/

[12] AppMetrica. [Online]. Available: https://appmetrica.yandex.ru/

[13] Т. Д. Польгуль, та А. А. Яровий, “Метод подолання різнорідності даних для виявлення шахрайства при інсталюванні мобільних додатків”, *Вісник СНУ ім. В. Даля – Сєвєродонецьк: СНУ ім. В. Даля,* № 7 (248), с.60-69, 2018.

[14] T. Polhul, and A. Yarovyi, “Development of a method for fraud detection in heterogeneous data during installation of mobile applications”, *Eastern-European Journal of Enterprise Technologies*, № 1/2 (97), 2019. doi: 10.15587/1729-

4061.2019.155060

[15] Т.Д. Польгуль, та А.А. Яровий, “Аналіз різнорідних даних в інтелектуальних системах виявлення шахрайства”, *Вісник Вінницького політехнічного інституту*, № 2, с. 78-90, 2019.

[16] Т.Д. Польгуль, “Інформаційна технологія побудови інтелектуальних систем виявлення шахрайства при інсталюванні мобільних додатків”, *Інформаційні технології та комп'ютерна інженерія*, № 1, с. 4-16, 2019.

[17] A. Yarovyi, and T. Polhul, “Applied Aspects of Implementation of Intelligent Information Technology for Fraud Detection During Mobile Applications Installation”, *Advances in Intelligent Systems and Computing IV. CCSIT 2019: Advances in Intelligent Systems and Computing*, Springer, Cham, Switzerland, vol 1080, pp. 377-386, 2019. doi: https://doi.org/10.1007/978-3-030-33695-0_26

[18] T. Polhul, “Conceptual Model of an Intelligent System for Detecting Fraud During Mobile Applications Installation”, in *Proc. 10th International Conference on Dependable Systems, Services and Technologies (DESSERT)*, Leeds, United Kingdom, pp. 167-174, 2019. doi: 10.1109/DESSERT.2019.8770030. Режим доступу: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8770030&isnumber=8770005.

[19] А.А. Яровий, О.Н. Романюк, І.Р. Арсенюк, та Т.Д. Польгуль, “Виявлення шахрайства при інсталюванні програмних додатків з використанням інтелектуального аналізу даних”, *Наукові праці Донецького національного технічного університету. Серія “Інформатика, кібернетика та обчислювальна техніка”*, № 2 (25), с. 126-131, 2017. Режим доступу: http://science.donntu.edu.ua/wp-content/uploads/2018/03/ikvt_2017_2_site-1.pdf

[20] Т.Д. Польгуль, та А.А. Яровий, “Визначення шахрайських операцій при встановленні мобільних додатків з використанням інтелектуального аналізу даних”, *Сучасні тенденції розвитку системного програмування. Тези доповідей*, Київ, 2016, с. 55-56. Режим доступу: http://ccs.nau.edu.ua/wp-content/uploads/2017/12/%D0%A1%D0%A2%D0%A0%D0%A1%D0%9F_2016_07.pdf

[21] Т.Д. Польгуль, та А.А. Яровий, “Визначення шахрайських операцій при інсталяції мобільних додатків з використанням інтелектуального аналізу даних”, на *XLVI науково-технічній конференції підрозділів ВНТУ,* Вінниця, 2017. Режим доступу: http://ir.lib.vntu.edu.ua/bitstream/handle/123456789/17200/2158.pdf?sequence=3

[22] А. Яровий, Т. Польгуль, та Л. Крилик, “Розробка методу виявлення шахрайства при інсталюванні мобільних додатків з використанням інтелектуального аналізу даних”, *XIV Міжнародна конференція Контроль і управління в складних системах (КУСС-2018). Тези доповідей*, Вінниця, 2018, с. 35.

[23] T. Polhul, “Development of an intelligent system for detecting mobile app install fraud”, in *Proc.IRES 156th International Conference*, Bangkok, Thailand, 2019, pp. 25-29.

[24] T. Polhul, “Development of an intelligent system for detecting mobile app install fraud”, *International Journal of Advances in Electronics and Computer Science (IJAECS)*, vol. 6, no. 7, pp. 13–17, July, 2019.

[25] А.А. Яровий, та Т.Д. Польгуль, “Подолання різнорідності вхідних даних при виявленні шахрайства при інсталюванні мобільних додатків з використанням інтелектуального аналізу даних”, на *П’ятій міжнародній науково-технічній конференції студентів, магістрів, аспірантів «Інформатика, управління та штучний інтелект»,* Національний технічний університет «Харківський політехнічний інститут», Харків, 2018, с. 109.

[26] T. Polhul and A. Yarovyi,"Method of Fraudster Fingerprint Formation During Mobile Application Installations", *2019 10th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS)*, Metz, France, 2019, pp. 1099-1103. doi: 10.1109/IDAACS.2019.8924369. Режим доступу: https://ieeexplore.ieee.org/document/8924369.

[27] Т.Д. Польгуль, “Моделювання процесу виявлення шахрайства при інсталюванні мобільних додатків”, на *XLVIII науково-технічної конференції*

підрозділів ВНТУ, Вінниця, 2019. Режим доступу: https://conferences.vntu.edu.ua/index.php/all-fitki/all-fitki-2019/paper/view/6863

[28] A. Yarovyi, and T. Polhul, “Intelligent information technology for fraud detection during mobile applications installation”, in *Proc. 14th International conference "Computer sciences and Information technologies" (CSIT 2019)*, pp. 1-5, Lviv, 2019. doi: 10.1109/STC-CSIT.2019.8929827. Режим доступу: https://ieeexplore.ieee.org/ document/8929827.

[29] Impact: Forensiq by Impact Earns MRC Accreditation for SIVT Detection and Filtration and Viewability Measurement. [Online]. Available: https://impact.com/ad-fraud-detection/

[30] V. Chandola, A. Banerjee, and V. Kumar, *Anomaly detection*. Minnesota, USA: ACM Computing Surveys, 2009, vol. 41, issue 3, pp. 1-58. doi: https://doi.org/10.1145/1541880.1541882

[31] A. Géron, *Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. USA: Aurélien Géron, O’Reilly Media, 2017.

[32] D. Cielen, A. D. B. Meysman, and M. Ali, *Introducing Data Science: Big data, machine learning, and more, using Python tools.* USA: Manning, 2016.

[33] S. Guido, and A. Müller, *Introduction to Machine Learning with Python: A Guide for Data Scientists*. USA: O’Reilly Media, 2016.

[34] F. Chollet, *Deep Learning with Python.* USA: Manning, 2017.

[35] D. Hawkins, *Identification of Outliers*. Chapman and Hall, 1980.

[36] Andrew Ng, *Machine Learning*. Standford, USA. [Online]. Available: https://www.coursera.org/learn/machine-learning?utm_source=gg&utm_medium=sem&utm_content=07-StanfordML-ROW&campaignid=2070742271&adgroupid=80109820241&device=c&keyword=machine%20learning%20mooc&matchtype=b&network=g&devicemodel=&adpostion=1t1&creativeid=369041663186&hide_mobile_promo&gclid=CjwKCAjwpuXpBRAAEiwAyRRPge3ilrZykf2wwoDbEgg4mLES-edSyJsV9n-r8QrtmNuE-WozUpviNBoCW6oQAvD_BwE

[37] X. Song, M. Wu, C. Jermaine, and S. Ranka, “Conditional Anomaly

Detection", *IEEE Transactions on Knowledge and Data Engineering*, vol. 19, iss. 5, pp. 631-645, 2015. doi: https://doi.org/10.1109/tkde.2007.1009

[38] А. В. Гриценко, *Типы аномалий в видеоизображениях*. *Технические науки – от теории к практике: сборник статей по материалам VII международной научно-практической конференции. Часть I*. Новосибирск, Россия: СибАК, 2012. Режим доступа: https://sibac.info/conf/tech/vii/26730

[39] M. A. Prado-Romero, and A. Gago-Alonso, "Detecting contextual collective anomalies at a Glance", in *Proc. 2016 23rd International Conference on Pattern Recognition (ICPR)*, Cancun, Mexico, 2016. doi: https://doi.org/10.1109/ icpr.2016.7900017

[40] R. Agrawal, and R. Srikant, "Mining sequential patterns", in *Proc. Eleventh International Conference on Data Engineering*, Taipei, Taiwan, 1995. doi: https://doi.org/10.1109/icde.1995.380415

[41] D. Agarwal, "An Empirical Bayes Approach to Detect Anomalies in Dynamic Multidimensional Arrays", in *Proc. Fifth IEEE International Conference on Data Mining (ICDM'05)*, Houston, USA, 2005. doi: https://doi.org/10.1109/ icdm.2005.22

[42] C. Siaterlis, and B. Maglaris, "Towards multisensor data fusion for DoS detection", in *Proc. 2004 ACM symposium on Applied computing – SAC '04*, Nicosia, Cyprus, 2004. doi: https://doi.org/10.1145/967900.967992

[43] D. Agarwal, "Detecting anomalies in cross-classified streams: a Bayesian approach*", Knowledge and Information Systems*, vol. 11, iss. 1, pp. 29-44, 2006. doi: https://doi.org/10.1007/s10115-006-0036-4

[44] MachineLearning.ru. Профессиональный информационно-аналитический ресурс, посвященный машинному обучению, распознаванию образов и интеллектуальному анализу даннях. [Электронный ресурс]. Режим доступу: http://www.machinelearning.ru

[45] T. Segaran, *Programming Collective Intelligence: Building Smart Web 2.0 Applications*. Sebastopol, USA: O'Reilly Media, 2008.

[46] D.-Y. Yeung, and C. Chow, "Parzen-window network intrusion detectors*",* in *Proc. Object recognition supported by user interaction for service robots*, Hong Kong,

2002. doi: https://doi.org/10.1109/icpr.2002.1047476

[47] V. J. Hodge, and J. Austin, “A Survey of Outlier Detection Methodologies”, *Artificial Intelligence Review.*, vol. 22, issue 2, pp. 85-126. 2004. doi: https://doi.org/10.1007/s10462-004-4304-y

[48] Agyemang M., Barker K., and Alhajj R., “A comprehensive survey of numeric and symbolic outlier mining techniques*”, Intelligent Data Analysis*, vol. 10, iss. 6, pp. 521-538, 2006. doi: https://doi.org/10.3233/ida-2006-10604

[49] Keogh E., Lin J., and Fu A., “HOT SAX: Efficiently Finding the Most Unusual Time Series Subsequence”, *Fifth IEEE International Conference on Data Mining (ICDM'05)*, Houston, USA, 2005. doi: https://doi.org/10.1109/icdm.2005.79

[50] E. Keogh, J. Lin, S.-H. Lee, and H. V. Herle, “Finding the most unusual time series subsequence: algorithms and applications”, *Knowledge and Information Systems*, vol. 11, iss. 1, pp. 1-27, 2006. doi: https://doi.org/10.1007/s10115-006-0034-6

[51] Donoho S., “Early detection of insider trading in option markets, in *Proc. Tenth ACM SIGKDD international conference on knowledge discovery and data mining – KDD-2004*, USA, 2004. doi: https://doi.org/10.1145/1014052.1014100

[52] A. W. Fu, O. T.-W. Leung, E. Keogh, and J. Lin, “Finding Time Series Discords Based on Haar Transform*”, Lecture Notes in Computer Science*, 2006, pp. 31–41. doi: https://doi.org/10.1007/11811305_3

[53] Baudat G., and Anouar F., “Generalized Discriminant Analysis Using a Kernel Approach”, *Neural Computation*, 2000, vol. 12, iss. 10, pp. 2385–2404. doi: https://doi.org/10.1162/089976600300014980

[54] S. Benndorf, G. Kakulapati, A. Pham and others (2015), “Fighting Mobile Fraud in the Programmatic Era*”, AppLift GmbH*, 2015.

[55] Fraudwatch. [Online]. Available: http://www.fraudshields.com/

[56] Andrii Yarovyi, Raisa Ilchenko, Ihor Arseniuk, Yevhene Shemet, Andrzej Kotyra, and Saule Smailova, "An intelligent system of neural networking recognition of multicolor spot images of laser beam profile", in *Proc. SPIE 10808, Photonics Applications in Astronomy, Communications, Industry, and High-Energy Physics*

Experiments 2018, 108081B, 2018. doi: https://doi.org/10.1117/12.2501691

[57] V. Kozhemyako, L. Timchenko, and A. Yarovyy, "Methodological Principles of Pyramidal and Parallel-Hierarchical Image Processing on the Base of Neural-Like Network Systems", *Advances in Electrical and Computer Engineering*, vol.8, no.2, pp. 54-60, 2008. doi:10.4316/AECE.2008.02010

[58] M. Granik, V. Mesyura, and A. Yarovyi, "Determining Fake Statements Made by Public Figures by Means of Artificial Intelligence", in *Proc. 2018 IEEE 13th International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT),* Lviv, 2018, pp. 424-427. doi: 10.1109/STC-CSIT. 2018.8526631

[59] Анализ клиентских баз данных. Выявление мошенничества (fraud detection) на базе STATISTICA Data Miner. [Online]. Available: http://statsoft.ru/solutions/ExamplesBase/branches/detail.php?ELEMENT_ID=834

[60] А.А. Яровий, та Т.Д. Польгуль, "Підвищення продуктивності обчислювальних процесів в паралельно-ієрархічній мережі за допомогою Framework Benchmark Akka", Збірник тез доповіді на *VII Міжнародній науково-технічній конференції «Фотоніка ОДС-2015»*, Вінниця, 2015, с. 9.

[61] Д.П. Польгуль, та Т.Д. Польгуль, "Концептуальна модель інформаційної безпеки", на *Міжнародній науково-практичній конференції «Тенденції управління фінансовими та інноваційними процесами в умовах ринкових перетворень», секція «Актуальні проблеми розвитку управління сучасним підприємством»,* Вінниця, ВНТУ, 2012, с. 340-342.

[62] Т.Д. Польгуль, "Порівняльний аналіз Apache Spark та Apache Flink для роботи з Big Data", на *XLV науково-технічній конференції підрозділів ВНТУ,* Вінниця, 2016. [Електронний ресурс]. Режим доступу: https://ir.lib.vntu.edu.ua/handle/123456789/11619

[63] А.Г. Кюльян, Т.Д. Польгуль, та М.Б. Хазін, "Математична модель рекомендаційного сервісу на основі методу колаборативної фільтрації", *Комп'ютерні технології та Інтернет в інформаційному суспільстві*, с. 226–227, 2012. Режим доступу: http://ir.lib.vntu.edu.ua/bitstream/handle/

123456789/7911/226227.pdf?sequence=1&isAllowed=y

[64] T. Joachims, “Training linear SVMs in linear time”, in *Proc. 12th ACM SIGKDD international conference on Knowledge discovery and data mining KDD ’06, ACM*, New York, USA, 2006, pp. 217–226.

[65] J. Platt, *Probabilistic outputs for support vector machines and comparison to regularized likelihood methods*. MIT Press, 2000, pp. 61–74.

[66] J. L. Bentley, 1975. “Multidimensional binary search trees used for associative searching”, *Communications of the ACM*, vol. 18, iss. 9, pp. 509–517, 1975.

[67] N. Roussopoulos, S. Kelley, and F. Vincent, “Nearest neighbor queries”, in *Proc. ACM-SIGMOD International Conference on Management of Data*, Місто, 1995.

[68] Метод k-ближайших соседей. [Электронный ресурс]. Режим доступа: http://om.univ.kiev.ua/users_upload/15/upload/file/pr_lecture_03.pdf

[69] К. Воронцов, “Машинное обучение (курс лекций)”. [Электронный ресурс]. Режим доступа: http://www.machinelearning.ru/wiki/index.php?title =Машинное_обучение_(курс_лекций%2C_К.В.Во ронцов)

[70] К.Д. Маннинг, П. Рагхаван, и Х. Шютце, *Введение в информационный поиск*. Москва, Россия: Вильямс, 2011.

[71] J. A. Hartigan, and M. A. Wong, “A k-means clustering algorithm*”, Royal Statistical Society. Series C (Applied Statistics)*, vol. 28, no.1 (1979), pp. 100–108.

[72] E. Eskin, A. Arnold, M. Prerau, L. Portnoy, and S. Stolfo, “A geometric framework for unsupervised anomaly detection”, in *Proc. Applications of Data Mining in Computer Security*. New York, USA: Kluwer Academics, 2002, pp. 78–100.

[73] Jain A. K., *Data Clustering: A Review.* Michigan State University. [Online]. Available: E. Eskin, A. Arnold, M. Prerau, L. Portnoy, and S. Stolfo, 2002. “A geometric framework for unsupervised anomaly detection”, in *Proc. Applications of Data Mining in Computer Security*. Kluwer Academic Publishers, 2002, pp. 77–101.

[74] Jain, A. K. and Dubes, R. C. *Algorithms for Clustering Data*. New Jersey, USA: Prentice Hall, 1988.

[75] Deep Neural Network. [Online]. Available: https://www.techopedia.com/ definition/32902/deep-neural-network

[76] Training Deep Neural Network. [Online]. Available: https://www.techopedia.com/definition/32902/deep-neural-network

[77] Е. Е. Пятикоп, “Исследованине метода коллаборативной фильтрации на основе сходства элементов”. *Наукові праці Донецького національного технічного університету. Серія “Інформатика, кібернетика та обчислювальна техніка”*, № 2 (18), с. 109-114, 2013.

[78] Э. Беккенбах, и Р. Беллман, *Введение в неравенства*. Москва: Мир, 1965.

[79] Р. А. Шмойлова и др., *Теория статистики: Учебник*. Москва, Россия: Финансы и Статистика, 2001.

[80] J. L. Rodgers, and W. A. Nicewander, “Thirteen ways to look at the correlation coefficient”, *The American Statistician*, vol. 42, issue 1, pp. 59-66, 2012.

[81] Ф. Дж. Мак-Вильямс, и Н. Дж. А. Слоэн, *Теория кодов, исправляющих ошибки.* Москва: Связь, 1979.

[82] В. Д. Дмитриенко, и А. Ю. Заковоротный, “Нейронная сеть, использующая расстояние Хемминга, для распознавания изображений на границах нескольких классов”, *Вісник Національного технічного університету “ХПІ”. Інформатика та моделювання*, № 39, с. 57-67, 2013.

[83] Р. Блейхут, *Теория и практика кодов, контролирующих ошибки = Theory and Practice of Error Control Codes*. Москва: Мир, 1986.

[84] *Методичні вказівки до виконання комп’ютерних практикумів з навчальної дисципліни «Інтелектуальний аналіз даних».* Київ, Україна: КПІ ім.Ігоря Сікорського, 2017.

[85] И. И. Елисеева, и В. О. Рукавишников, *Группировка, корреляция, распознавание образов: (статистические методы классификации и измерения связей)*. Москва: Статистика, 1977.

[86] С.І. Альперт, “Основні міри подібності та нові підходи до їх застосування при класифікуванні гіперспектральних космічних зображень”, *Математичні машини і системи*, № 1, 2019.

[87] К.Д. Маннинг, П. Рагхаван, и Х. Шютце, *Введение в информационный поиск*. Москва, Россия: Вильямс, 2011.

[88] Т. О. Говорущенко, О. С. Савенко, С. М. Лисенко, "Забезпечення кібербезпеки комп’ютерних систем в умовах інноваційного розвитку України", *Матеріали Всеукраїнської науково-практичної конференції «Безпека соціально-економічних процесів в кіберпросторі»*, Київ, 27 березня 2019 р., с. 41-42.

[89] Что такое Big data: собрали всё самое важное о больших данных. [Электронный ресурс]. Режим доступа: https://rb.ru/howto/chto-takoe-big-data/

[90] Леоненков А.В., *Самоучитель UML.* Санкт-Петербург, Россия: БХВ-Петербург, 2002. ISBN 5-94157-878-4, 978-5-94157-878-8

[91] A. Dennis, B. Haley, and D. Tegarden, *Systems Analysis and Design with UML* . USA: Wiley, 2012. ISBN 978-1118037423

[92] А. А. Яровий, та Т. Д. Польгуль, Комп'ютерна програма «Програмний модуль збору даних інформаційної технології виявлення шахрайства при інсталюванні програмних додатків». *Свідоцтво про реєстрацію авторського права на твір № 76348.* Київ: Міністерство економічного розвитку і торгівлі України, 2018.

[93] А. А. Яровий, та Т. Д. Польгуль, Комп'ютерна програма «Програмний модуль визначення схожості користувачів інформаційної технології виявлення шахрайства при інсталюванні програмних додатків». *Свідоцтво про реєстрацію авторського права на твір № 76347*. Київ: Міністерство економічного розвитку і торгівлі України, 2018.

[94] T. Polhul, “Fraud detection classifier”, Github. [Online]. Available: https://github.com/tanapolg/fraud-detection-classifier

[95] Kaggle. [Online]. Available: https://www.kaggle.com/

[96] R. Johnson, *Applied Multivariate Statistical Analysis*. New Jersey, USA: Prentice Hall, 1992.

[97] Оценивание и сравнение нейросетевых моделей. [Электронный ресурс]. Режим доступа: https://helpiks.org/6-10672.html

[98] Gamification by University of Pennsilvania. [Online]. Available: https://www.coursera.org/learn/gamification

[99] G. Baudat, and F. Anouar, “Generalized Discriminant Analysis Using a Kernel

Approach”, *Neural Computation*, vol. 12, no. 10, pp. 2385–2404, 2000. doi: https://doi.org/10.1162/089976600300014980

[100] Instagram. URL: https://www.instagram.com/

[101] SocialKit. Популярная программа для Instagram. [Online]. Available: https://socialkit.ru/

[102] В. М. Дубовой, та О. Д. Никитенко, *Оптимізація підсистем збору даних АСУТП в умовах комбінованої невизначеності*. *Монографія.* Вінниця, Україна: УНІВЕРСУМ-Вінниця, 2011. ISBN 978-966-641-434-5

[103] Ш. Эхтер, и Д. Робертс, *Многоядерное программирование*. Санкт-Петербург, Россия: Питер, 2010. ISBN 978-5-388-00091-0

[104] Mark D. Hill, and Michael R. Marty, “Amdahl's Law in the Multicore Era”, *IEEE Computer*, vol. 41, issue 7, pp. 33-38, 2008.

[105] S.L. Blumin, I.A. Shuykova, P.V. Sarajev, and I.V. Cherpakov, *Fuzzy Logic: Algebraic Foundations and Applications*. *Monograph* . Lipetsk, Russia: LESI, 2002.

[106] L. Zade, *The concept of a linguistic variable and its application to making approximate decisions.* Москва: Mir, 1976.

[107] В. І. Месюра, та Л. М. Ваховська, *Методичні вказівки до виконання курсового проекту з дисципліни "Системи прийняття рішень з нечіткою логікою" для студентів спеціальності "Інтелектуальні системи прийняття рішень"*. Вінниця, Україна: ВНТУ, 2005.

[108] T.D. Polhul, A.A. Yarovyi, R.Romaniuk, P.Komada, N. Askarova "Method of data anomaly detection in the process of mobile applications installation", Proc. SPIE 11176, Photonics Applications in Astronomy, Communications, Industry, and High-Energy Physics Experiments 2019, 111761Y; https://doi.org/10.1117/ 12.2536855