

МОДЕЛЮВАННЯ ПРОЦЕСУ ВИЯВЛЕННЯ ШАХРАЙСТВА ПРИ ІНСТАЛЮВАННІ МОБІЛЬНИХ ДОДАТКІВ

Вінницький національний технічний університет

Анотація

У даній роботі здійснено аналіз результатів моделювання процесу виявлення шахрайства при інсталюванні мобільних додатків. Для аналізу результатів взято контрольну вибірку з 56962 записами, 56870 з яких – відповідають органічним користувачам, а 92 – користувачам-шахраям. Здійснено перевірку адекватності моделі.

Ключові слова: виявлення шахрайства, виявлення аномалій, матриця невідповідності.

Abstract

A results' analysis of the modeling of mobile application installation fraud detection process was done in this paper. To analyze the results, a control sample of 56962 records was taken, 56870 of which correspond to organic users, and 92 – to fraudsters. The checking of the model's adequacy was done.

Keywords: fraud detection, anomaly detection, confusion matrix.

У наш час компанії-розробники мобільних додатків витрачають великі кошти на маркетингові кампанії, які у свою чергу зобов'язані здійснити вказану кількість інсталяцій додатку. Проте, багато з цих кампаній застосовують шахрайські способи [1 – 3], які показують, що інсталяція відбулася, але в дійсності це не так. Саме через це проблема виявлення шахрайських інсталяцій є актуальною. При моделюванні процесу виявлення шахрайства необхідно вміти оцінити та проаналізувати його результати для вибору найкращого та найефективнішого підходу. Саме тому метою даної роботи є аналіз результатів моделювання розробленого процесу виявлення шахрайства при інсталюванні мобільних додатків.

Обґрунтування та більш детальний опис запропонованого методу виявлення шахрайства при інсталюванні мобільних додатків представлено у роботах [4 – 6]. Розглянемо та проаналізуємо отримані результати. Для доведення адекватності результатів розробленої моделі, необхідно правильно здійснити її оцінку. Для цього визначимо, що насамперед система повинна бути наділена прогностичною здатністю та добре узагальнюватися на нових для неї даних. Також, необхідно зауважити, що для коректного навчання та оцінки системи, а також для уникнення перенавчання системи, необхідно мати три набори даних: тренувальний, перевірочний (валідаційний) та контрольний. Сформуємо графіки втрат (рис. 1а) та точності (рис. 1б) на етапах навчання та перевірки.

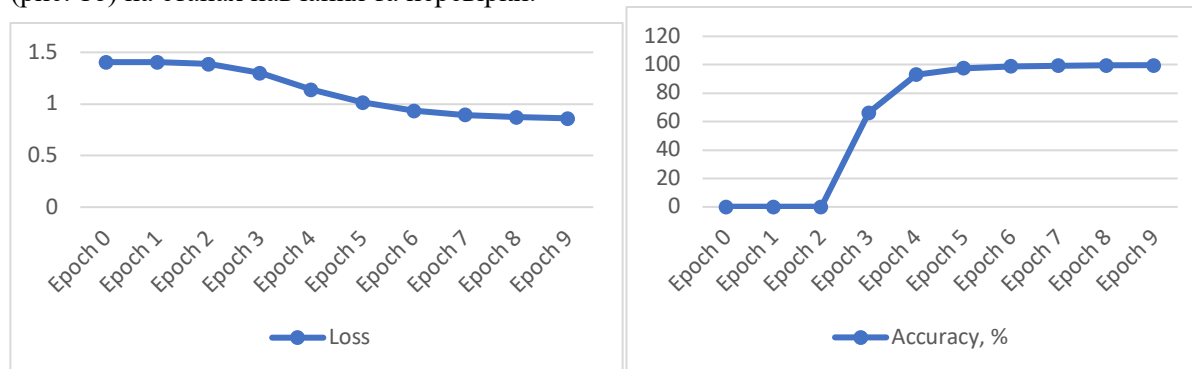


Рис. 1 – Етапи навчання та перевірки: а – втрати, б – точність

Отримавши адекватну модель [7–9], побудуємо нормалізовану матрицю невідповідності (рис. 2) для порівняння отриманих результатів з очікуваними. Для перевірки моделі взято контрольну вибірку з 56962 записами, 56870 з яких – відповідають органічним користувачам, а 92 – користувачам-шахраям. Зазначимо, що представлена на рис. 2 матриця невідповідності побудова з використанням мови програмування Python.

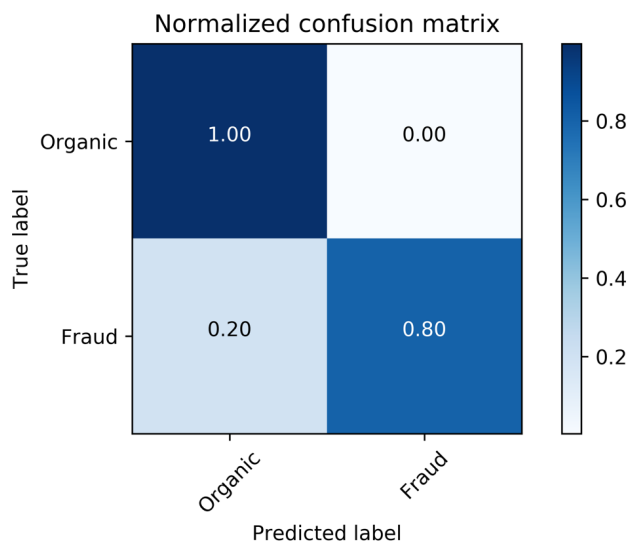


Рис. 2 – Нормалізована матриця невідповідності контрольної вибірки

З рисунку 2 видно, що правильно класифіковано 80,43% користувачів-шахраїв, а 19,57% класифіковано неправильно. В цілому правильно розцінено 99,56% об'єктів (56639 + 74 = 56713 об'єктів з 56962 об'єктів), неправильно класифіковано 0,44% об'єкти.

У ході перевірки адекватності моделі розрахуємо F-міру, що може розглядатися як гармонійне середнє між точністю та чутливістю:

$$F1 = 2 \cdot \frac{PPV \cdot TPR}{PPV + TPR} = \frac{2TP}{2TP + FP + FN}, \quad (1)$$

де PPV – це точність або positive predictive value (2), TPR – це чутливість або ж recall / true positive rate (3).

$$PPV = \frac{TP}{TP + FP} = \frac{56639}{56639 + 18} = 99\%, \quad (2)$$

$$TPR = \frac{TP}{P} = \frac{TP}{TP + FN} = \frac{56639}{56870} = 99\%, \quad (3)$$

де TP – true positive, TN – true negative, FP – false positive, FN – false negative, P – condition positive, тобто дійсна кількість позитивних випадків у даних.

На основі формул (1 – 3) та Fig. 7 – 9 розрахуємо F1 score (4).

$$F1 = 2 \cdot \frac{PPV \cdot TPR}{PPV + TPR} = \frac{2TP}{2TP + FP + FN} = 2 \cdot \frac{0.99 \cdot 0.99}{0.99 + 0.99} = 99\% \quad (4)$$

Отримане значення F1 score на контрольній вибірці є достатньо високим, що обумовлено використанням запропонованого підходу з використанням усіх наявних різномірних вхідних даних та шкалювання даних по кінцевій цілі.

Отже, у даній роботі здійснено аналіз результатів моделювання розробленого процесу виявлення шахрайства при інсталюванні мобільних додатків з використанням мови програмування Python.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Польгуль Т. Д., Яровий А. А. Визначення шахрайських операцій при встановленні мобільних додатків з використанням інтелектуального аналізу даних // Сучасні тенденції розвитку системного програмування. Тези доповідей. Київ, 2016. С. 55–56. URL: http://ccs.nau.edu.ua/wp-content/uploads/2017/12/%D0%A1%D0%A2%D0%A0%D0%A1%D0%9F_2016_07.pdf
2. Польгуль Т. Д., Яровий А. А. Визначення шахрайських операцій при інсталяції мобільних додатків з використанням інтелектуального аналізу даних // Матеріали XLVI науково-технічної конференції підрозділів ВНТУ. Вінниця, 2017. URL: <http://ir.lib.vntu.edu.ua/bitstream/handle/123456789/17200/2158.pdf?sequence=3>
3. Яровий А., Польгуль Т., Крилик Л. Розробка методу виявлення шахрайства при інсталюванні мобільних додатків з використанням інтелектуального аналізу даних // Матеріали конференції «XLVII Науково-технічна конференція підрозділів Вінницького національного технічного університету (2018)». Вінниця, 2018. URL: <http://ir.lib.vntu.edu.ua/bitstream/handle/123456789/22722/079.pdf?sequence=1>
4. Т. Д. Польгуль Метод подолання різномірності даних для виявлення шахрайства при інсталюванні мобільних додатків / Т. Д. Польгуль, А. А. Яровий // Вісник СХУ ім. В. Даля – Сєверодонецьк: СХУ ім. В. Даля. 2018. – № 7 (248), 2018. – С.60-69.
5. Polhul, T., Yarovy, A. (2019). Development of a method for fraud detection in heterogeneous data during installation of mobile applications. Eastern-European Journal of Enterprise Technologies, № 1/2 (97), pp. 65–75 doi: 10.15587/1729-4061.2019.155060
6. Виявлення шахрайства при інсталюванні програмних додатків з використанням інтелектуального аналізу даних / Яровий А. А., Романюк О. Н., Арсенюк І. Р., Польгуль Т. Д. // Наукові праці Донецького національного технічного університету. Серія: “Інформатика, кібернетика та обчислювальна техніка”. 2017. № 2 (25). С. 126–131. URL: http://science.donntu.edu.ua/wp-content/uploads/2018/03/ikvt_2017_2_site-1.pdf
7. Яровий А. А., Польгуль Т. Д. Підвищення продуктивності обчислювальних процесів в паралельно-ієрархічній мережі за допомогою Framework Benchmark Akka // Збірник тез доповіді VII Міжнародної науково-технічної конференції «Фотоніка ОДС-2015». Вінниця, 2015. С. 9.
8. Яровий А. А., Польгуль Т. Д. 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.
9. Яровий А. А., Польгуль Т. Д. 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.

Польгуль Тетяна Дмитрівна – аспірант кафедри комп'ютерних наук ВНТУ, Вінницький національний технічний університет, м. Вінниця, e-mail: tanapolg93@gmail.com

Науковий керівник: Яровий Андрій Анатолійович – д.т.н., професор, завідувач кафедри комп'ютерних наук ВНТУ, Вінницький національний технічний університет, м. Вінниця, e-mail: a.yarovyy@vntu.edu.ua

Tetiana D. Polhul – postgraduate student of Computer Science Department, Vinnytsia National Technical University, Vinnytsia, e-mail: tanapolg93@gmail.com

Scientific Supervisor: Andrii A. Yarovy – Doctor Sc. (Eng), Professor, Head of Computer Science Department, Vinnytsia National Technical University, Vinnytsia, e-mail: a.yarovyy@vntu.edu.ua