ПІДВИЩЕННЯ ЯКОСТІ ТА ОПЕРАТИВНОСТІ МОДЕЛІ ВИЯВЛЕННЯ ВТОРГНЕНЬ

Абдуллін О. Р., Гавриленко С. Ю.
Національний технічний університет «Харківський політехнічний інститут», м. Харків

Забезпечення захисту комп'ютерних систем і мереж від вторгнень, шкідливого програмного забезпечення, різноманітних ризиків і загроз є однією з найважливіших проблем у сфері інформаційних технологій на сьогоднішній день. Це зумовлено постійним збільшенням обсягів шкідливого програмного забезпечення, а також ускладненням способів і типів вразливостей, що використовуються для реалізації загроз і несанкціонованих проникнень.

Виявлення та класифікація різних типів атак мають вирішальне значення для запобігання шкідливій діяльності в мережі. Необхідно мати систему, яка може виявляти та класифікувати ці атаки в режимі реального часу.

Метою даного дослідження є розробка методу підвищення якості та оперативності виявлення вторгнень в комп'ютерні системи та мережі.

У якості вихідних даних використано великий набір даних NF-UQ-NIDS-V2, який містить дані мережевого трафіку для виявлення вторгнень, що згенеровані з різних потоків мережевих налаштувань та включає різні типи атак. Він поєднує чотири набори даних (UNSW-NB15, BoT-IoT, ToN-IoT та CSE-CIC-IDS2018). Набір даних містить загалом 11 994 893 записи, з яких 9 208 048 (76,77%) є доброякісними потоками, а 2 786 845 (23,23%) – атаками, описується 43 ознаками, а його розмір становить 13,73 ГБ.

На етапі попередньої обробки даних виконано оцінку інформативності ознак та для подальшого застосування вилучено вісім найменш інформативних із них.

У якості класифікатора було обрано модель на основі градієнтного бустингу, яка є особливо є ефективною при роботі з незбалансованими даними та виконано її налаштування. Аналіз оперативності моделі показав що час класифікації моделі є значним. Для підвищення оперативності моделі запропоновано алгоритм циклічного перебору ознак, формування пулу ознак та оцінку якості моделі при цьому. Результати досліджень надали можливість сформувати пул із 16 ознак, використання якого зменшило F1-score моделі лише на 1,8%, одночасно підвищуючи оперативність моделі на 19%.

Література:

1. Rana S. Feature Selection and Importance Using XGBoost in Machine Learning. *Medium*. URL: https://medium.com/@mdshohelrana85/feature-selection-and-importance-using-xgboost-in-machine-learning-483c49a3cea9