Виявлення мережевих атак засобами машинного та глибокого навчання на основі набору даних UNSW-NB15

УДК 004.056.5:004.85

Марина Ксеніта[1], Марія Стадник[2], Володимир Данилюк[3]

Тернопільський національний технічний університет імені Івана Пулюя,
[1]ksenita.marina@gmail.com, [2]maria.stadnyk@gmail.com, [3]vdanilyuk06@gmail.com

Зростання кількості кібератак, ускладнення мережевої інфраструктури та активне використання автоматизованих засобів сканування, експлуатації вразливостей і приховування шкідливої активності зумовлюють потребу в ефективних системах виявлення вторгнень. Традиційні сигнатурні механізми залишаються корисними для відомих загроз, однак вони недостатньо гнучкі для виявлення нових або модифікованих атак. Тому актуальним є застосування методів машинного та глибокого навчання, здатних аналізувати багатовимірні характеристики мережевих потоків і класифікувати трафік як нормальний або шкідливий.

Метою дослідження є розроблення та експериментальна перевірка програмного підходу до бінарної класифікації мережевого трафіку на основі набору даних UNSW-NB15 із використанням алгоритму XGBoost та рекурентних нейронних мереж RNN, LSTM і GRU. Завдання дослідження охоплюють попередню обробку даних, кодування категоріальних ознак, нормалізацію числових характеристик, балансування навчальної вибірки, побудову послідовностей фіксованої довжини для нейронних моделей та порівняння результатів.

Вхідними файлами є навчальна та тестова частини UNSW_NB15_training-set.csv і UNSW_NB15_testing-set.csv. На етапі попередньої обробки вилучаються поля id та attack_cat, категоріальні ознаки proto, service і state перетворюються за допомогою LabelEncoder, пропущені значення видаляються, а всі ознаки масштабуються методом MinMaxScaler. Для зменшення впливу дисбалансу класів навчальна вибірка додатково балансується алгоритмом ADASYN.

Першим базовим класифікатором обрано XGBoost - ансамблевий метод градієнтного бустингу дерев рішень, який поєднує високу точність, регуляризацію та ефективну роботу з табличними даними. Для перевірки стабільності моделі застосовано стратифіковану п'ятикратну крос-валідацію, після чого модель навчається на збалансованій вибірці та тестується на відкладеному наборі. Другий блок експерименту формують нейронні архітектури SimpleRNN, LSTM і GRU, які отримують на вхід послідовності з 10 записів і використовують сигмоїдний вихідний шар для бінарної класифікації. Таке подання дає змогу оцінити, наскільки рекурентні моделі здатні враховувати локальні залежності між послідовними мережевими спостереженнями.

Набір UNSW-NB15 створено в Cyber Range Lab UNSW Canberra із використанням IXIA PerfectStorm для формування поєднання сучасної нормальної активності та синтетичних атак; він містить дев'ять типів атак:

Fuzzers, Analysis, Backdoors, DoS, Exploits, Generic, Reconnaissance, Shellcode і Worms. У межах поданого програмного рішення ці категорії агрегуються до бінарної цільової змінної label, що відповідає практичній задачі первинного виявлення факту атаки. Такий підхід є доцільним для початкового рівня системи NIDS, коли найважливішим є швидке відокремлення потенційно шкідливого трафіку від нормального.

Аналіз ROC-кривих показав (рис. 1) високу ефективність усіх досліджуваних підходів. Найкращий результат продемонструвала модель XGBoost із значенням ROC-AUC ≈ 0,98, що свідчить про її високу здатність розрізняти нормальний мережевий трафік та атаки. Рекурентні нейронні мережі також показали високі результати: RNN досягла ROC-AUC ≈ 0,97, а моделі LSTM та GRU – близько 0,96.

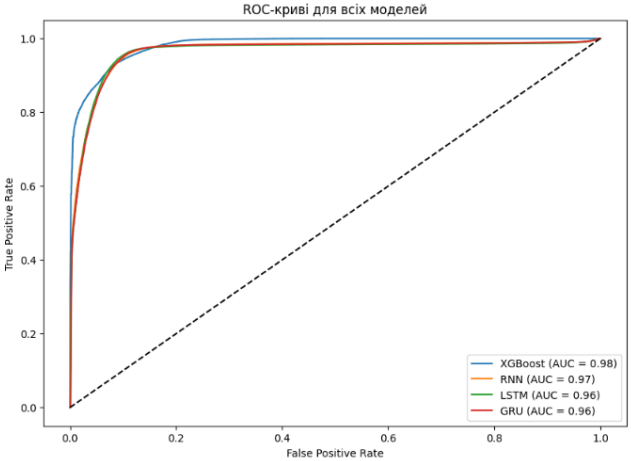

Рис. 1. ROC криві досліджуваних моделей

Отримані значення Precision = 0,9646, Recall = 0,8986, F1-score = 0,9304 та ROC-AUC = 0,9637 підтверджують високу якість класифікації та збалансованість моделі щодо виявлення атак і мінімізації хибних спрацювань. Водночас XGBoost продемонстрував дещо кращі результати порівняно з нейронними мережами при меншій обчислювальній складності.

Проведене дослідження показує, що поєднання класичних ансамблевих алгоритмів і рекурентних нейронних мереж є перспективним напрямом для побудови систем виявлення вторгнень. XGBoost доцільно використовувати як сильний базовий класифікатор для табличних ознак мережевих потоків, тоді як LSTM і GRU можуть бути корисними для моделювання послідовного контексту трафіку. Водночас якість висновків залежить від коректності кодування категоріальних ознак, репрезентативності тестової вибірки, параметрів балансування та кількості епох навчання нейронних мереж. Перспективами подальших досліджень є оптимізація гіперпараметрів, використання attention-механізмів, побудова багатокласової класифікації за типами атак і перевірка моделей на реальному потоковому трафіку.