

**Гавриленко С.Ю., Зозуля В.Д.**

*Національний технічний університет*

*"Харківський політехнічний інститут", м. Харків*

**Омельченко В.В.,**

*Харківський національний університет радіоелектроніки, м. Харків*

Незбалансовані класифікації створюють проблеми для прогнозного моделювання тому, що більшість класифікаційних алгоритмів машинного навчання були розроблені на основі припущення про кількість прикладів для кожного класу, тобто вони не враховують класового розподілу/пропорції або балансу класів. Це призводить до побудови упереджених моделей із поганою прогностичною ефективністю, особливо для класу меншості. Модель повністю ігнорує клас меншості і маркує всі об'єкти класифікації мітками мажоритарного класу. Особливо це актуально для систем виявлення вторгнень, які, зазвичай, містять невелику кількість прикладів виявлення вторгнень і для яких помилково-негативна класифікація (пропуск цілі) може мати фатальні наслідки.

Один із підходів до вирішення зазначеної проблеми – застосування різних стратегій балансування класів. У даній роботі досліджено методи балансування класів, які базуються на технології Undersampling, Oversampling та їх комбінацій. Для подальшого дослідження обрано такі методи: SMOTEENN, SVM SMOTE, Borderline-SMOTE, ADASYN, SMOTE, KMeansSMOTE. У якості вихідних даних використано набір UNSW-NB 15, який містить інформацію про нормальне функціонування мережі та під час вторгнень. У якості базового класифікатора використано дерево рішень на основі CART алгоритму. За результатами досліджень отримано, що використання методу SMOTEENN надає можливість підвищити якість виявлення вторгнень у функціонування мережі до 30%.

Наукова новизна отриманих результатів полягає у комплексному використанні методів балансування даних та методу класифікації даних на основі дерева рішень з використанням алгоритму CART (Classification And Regression Tree) для виявлення вторгнень у комп'ютерні мережі, що дозволило зменшити кількість помилок II роду.

#### **Література:**

1. Krawczyk, Bartosz. Learning from imbalanced data: open challenges and future directions. Progress in Artificial Intelligence, 2016, V.5, pp.221-232.
2. Abdi L, Sattar H. To combat multi-class imbalanced problems by means of over-sampling techniques. IEEE Trans Knowl Data Eng. 2016, V.28, pp.238-251.