ДОСЛІДЖЕННЯ МЕТОДІВ КЛАСИФІКАЦІЇ ДАНИХ ДЛЯ ІДЕНТИФІКАЦІЇ НЕСАНКЦІОНОВАНОГО ДОСТУПУ ДО КОМП'ЮТЕРНИХ МЕРЕЖ

Гавриленко С.Ю., Богданов І.Ю.
Національний технічний університет «Харківський політехнічний інститут», м. Харків.

Одним із напрямків кібербезпеки є виявлення атак та аномалій у трафіку мережі. Для цієї задачі можливо використовувати методи машинного навчання які відрізняються за способом обробки даних і архітектурою моделі.

В даній роботі, з використання хмарного середовище Google Colab Python, досліджено методи класифікації даних: Bagging Classifier, Gradient Boosting Classifier та Random Forest Classifier.

Bagging Classifier поєднує ансамбль базових класифікаторів та визначає результат шляхом простого голосування. У якості базового класифікатора використано класифікатор на основі алгоритму найближчого сусіда KNN. Gradient Boosting Classifier поєднує декілька моделей, таким чином що кожна наступна модель виправляє помилки минулої, поступово покращуючи точність кінцевого результату. Random Forest Classifier – це ансамблевий класифікатор на основі дерев рішень який також визначає кінцевий результат шляхом голосування базових моделей. Крім того, Random Forest Classifier при побудові моделі використовує лише частину ознак об'єкту, що дозволяє запобігти перенавчанню моделі.

У якості вихідних даних використано набір UNSW-NB 15, який містить інформацію про нормальне функціонування мережі та під час вторгнень. Виконано попередню обробку даних

За результатами досліджено отримано, що класифікатор на основі алгоритму Random Forest Classifier має найкращі результати, його точність складає близько 90%.

Література:

1. Kelleher J. , Namee B. and Archi A. Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples and Case Studies ,The MIT Pres, 2015, 642 p.

2. Harshita Singh. Understanding Data Preprocessing, 2020, URL: https://towardsdatascience.com/data-preprocessing-e2b0bed4c7fb

3. Gavrylenko S and Chelak V. Method of computer system state identificationbased on boosting ensemble with special preprocessing procedure, Advanced Information Systems, 2022. vol. 6, No 1, pp. 12–18. DOI:10.20998/2522-9052.2022.1.02